# Learning Explainable Treatment Policies with Clinician-Informed Representations: A Practical Approach

**Johannes O. Ferstad**
*OpenAI (work done while at Stanford University)*

**Emily B. Fox**
*Stanford University & CZ Biohub*

**David Scheinker**
**Ramesh Johari**                                                              RJOHARI@STANFORD.EDU
*Stanford University*

## Abstract

Digital health interventions (DHIs) and remote patient monitoring (RPM) have shown great potential in improving chronic disease management through personalized care. However, barriers like limited efficacy and workload concerns hinder adoption of existing DHIs, and limited sample sizes and lack of interpretability limit the effectiveness and adoption of purely black-box algorithmic DHIs. In this paper, we address these challenges by developing a pipeline for learning explainable treatment policies for RPM-enabled DHIs.

We apply our approach in the real-world setting of RPM using a DHI to improve glycemic control of youth with type 1 diabetes. Our main contribution is to reveal the importance of *clinical domain knowledge* in developing state and action representations for effective, efficient, and interpretable targeting policies. We observe that policies learned from clinician-informed representations are significantly more efficacious and efficient than policies learned from black-box representations. This work emphasizes the importance of collaboration between ML researchers and clinicians for developing effective DHIs in the real world.

**Keywords:** digital health interventions; policy optimization; remote patient monitoring

**Data and Code Availability**   Our data is from three clinical trials with $N = 281$ patients with type 1 diabetes (Scheinker et al., 2022; Prahalad et al., 2022, 2024). These data are not yet publicly available. The code used to generate our results is available at http://github.com/jferstad/ml4h-explainable-policies.

**Institutional Review Board (IRB)**   Our research was approved by the Stanford University Institutional Review Board, under protocol 52812.

## 1. Introduction

Digital health interventions (DHIs) and remote patient monitoring (RPM) have the potential to revolutionize patient care with treatment strategies dynamically personalized to each patient's characteristics and context. DHIs and RPM have been associated with improved management of many of chronic conditions including heart disease, diabetes, and mental health (Prahalad et al., 2024; Whitelaw et al., 2021; Liverpool et al., 2020). Relative to standards of care that rely on fixed-cadence clinic visits, RPM-enabled DHIs promise more timely, personalized, and frequent patient support (Scheinker et al., 2022). These technologies could help move population-level outcomes towards those typically seen in environments with more healthcare resources, particularly for underserved communities (Tikkanen, 2017; Anderson et al., 2003; Rodríguez and Campbell, 2017; Prahalad et al., 2024).

We consider RPM-enabled DHIs that involve the following typical workflow. On a regular cadence (e.g., weekly), the RPM platform takes as input rich, high-dimensional patient data, including granular data from sensors such as continuous glucose monitors (CGMs) and activity trackers. These data are used to form a representation of *patient states*, based on which patients are prioritized to receive a DHI (or *action*). Actions may include messaging the patient, recommending activity or treatment adjustments (e.g., a dose change). A *targeting policy* determines how to rank patients for interventions based

on patient state. The results of this ranking inform a whole-population care model in which clinicians determine what actions to take (e.g., which patients to message and what to say in the message).

While the majority of clinicians plan to use RPM-enabled DHIs, their adoption has been limited by significant, well-documented challenges (Stevens et al., 2022; Peterson, 2024; Vivalink, 2023; Lawrence et al., 2023; Sasangohar et al., 2018; Cresswell and Sheikh, 2013; Borges do Nascimento et al., 2023; Borghouts et al., 2021). *First* is uncertain efficacy, i.e., the difficulty of learning effective policies. A major hurdle here is the difficulty of developing representations of patient states from high-dimensional patient-level data from relatively few patients (e.g., at most a few hundred). *Second* is workload concerns: because clinical teams have limited capacity, the targeting policy must respect resource constraints by directing attention to patients who will benefit most from intervention. *Third* is a difficulty understanding or interpreting the technology. This lack of interpretability leads to a reluctance of care teams to adopt solutions that rely on black-box models.

In this paper we develop a pipeline to support the optimization of a care model that addresses the preceding concerns. Our work is carried out in the real-world context of an RPM-enabled DHI for individuals with type 1 diabetes (T1D). In the setting we consider, as in Prahalad et al. (2024), patients wear a continuous glucose monitor (CGM) that measures glucose every 5 minutes, generating a high-dimensional time series. The real-world deployment in Prahalad et al. (2024) generates patient states using clinically-informed single-dimensional summaries of this CGM data, such as the percentage of time glucose levels are in a desired target range (70-180 mg/dL). Clinicians provide guidance on how to improve management through telehealth interactions using natural language messages sent to the patient (e.g., "Increase your pre-dinner insulin dose") based on the dashboard presentation of recent patient CGM data. The objective of technology-based RPM for T1D is to improve patients' glucose management on an ongoing basis, through personalized, timely interventions.

Our main contribution is to reveal the importance of *clinical domain knowledge* in developing state and action representations for effective, efficient, and interpretable targeting policies. Because real-world settings have limited sample sizes, the inductive bias from clinical domain knowledge provides significant benefits to real-world policy performance. In par-

ticular, in the preceding T1D RPM context, we evaluate several approaches to low-dimensional state and action representations: from black-box machine learning methods, to clinician-informed learned representations. *We observe that policies derived from clinician-informed representations significantly outperform policies learned from black-box-learned representations in terms of efficacy and efficiency.* In fact, our evaluation reveals that learned policies outperform random targeting only when the state and action representations are clinically informed – amplifying the importance of clinical inductive bias in practice. Further, the use of clinical domain knowledge also yields policies that are more interpretable than black-box policies with state and action representations that maintain clinically relevant features and interventions.

To carry out our evaluation, we develop an end-to-end pipeline for learning targeting policies: we (1) learn low-dimensional state and action representations; (2) construct targeting policies by ranking patients based on estimated conditional average treatment effects (CATEs); and (3) evaluate the policies in the presence of capacity constraints. While each component of this pipeline has been studied in practice, our paper presents a coherent implementation of these steps together to carry out the evaluation described above. This pipeline may be of independent interest to digital health researchers carrying out similar optimization and evaluation of targeting policies in other real-world settings.

Our approach is broadly applicable to the evaluation and optimization of digital health interventions. Notably, our work suggests that interaction between machine learning researchers and healthcare domain experts is essential for developing practical, effective, and interpretable data-driven treatment policies that can be adopted in clinical practice.

## 2. Related work

Many studies have focused on the offline evaluation of conditional average treatment effect (CATE) estimators and of learned treatment policies, which are foundational to our approach (Mahajan et al., 2023; Dwivedi et al., 2020; Tang and Wiens, 2021; Yadlowsky et al., 2021; Sverdrup et al., 2023; Bouneffouf et al., 2020; Feuerriegel et al., 2024; Imai and Li, 2023, 2024). Our work combines methods from these works into a novel pipeline. Like us, they seek to facilitate the development of better treatment policies. But unlike their work, we do not treat our state and action repre-

sentations as fixed. Instead, we learn and evaluate targeting policies across many different representations, including interpretable representations defined by clinicians.

Our focus on explainable and interpretable causal inference and machine learning methods in healthcare aligns with the growing recognition of the importance of interpretability in this domain. Rasheed et al. (2022) provide a survey of explainable machine learning methods in healthcare, highlighting the need for transparent and understandable models. Mello and Rose (2024) discuss the practical challenges associated with non-interpretable recommendations generated by machine learning models broadly used by insurance companies. Our approach complements this line of research by demonstrating the value of clinician-informed representations for learning interpretable treatment policies in digital health interventions.

## 3. Data and context: States, actions, rewards

We use data from three IRB-approved clinical trials of remote monitoring of $N = 281$ patients with type 1 diabetes (T1D) (Scheinker et al., 2022; Prahalad et al., 2022, 2024). Participants in the trials wear CGMs that regularly transmit glucose measurements to TIDE, a remote patient monitoring platform (Ferstad et al., 2021; Kim et al., 2024). At regular intervals, e.g., weekly, clinicians use TIDE to review patient CGM data and decide whether to send treatment recommendations through asynchronous secure messaging (Figure 1). Due to constraints on provider time, at each review interval TIDE presents data for a subset of patients prioritized based on simple metrics from the consensus guidelines established by the American Diabetes Association, e.g., patients with a relatively high percentage of very low CGM readings.

Our data consist of variable numbers of days of data for each patient, depending on when the patient started the study; we let $T_i$ denote the number of days of data for patient $i$. We let $X_{it}^d$ denote patient demographics for patient $i$ on day $t$; $X_{it}^d$ is a combination of time-invariant patient demographics like sex and race/ethnicity, and time-variant demographics like age and insulin pump use. In addition, clinicians and the TIDE dashboard consider the previous two weeks of CGM readings—taken every 5 minutes—in determining patient prioritization for intervention; we let $X_{it}^g \in \Re^{4032}$ to be the (high-dimensional) vector of CGM recordings for patient $i$ over the two pre-

vious weeks prior to day $t$ (which may include missing values due to, e.g., the patient not wearing their CGM). As such, the individual $X_{it}^g$ are defined using a day-by-day rolling window on the raw CGM trace for patient $i$. Taken together, we call $X_{it} = (X_{it}^d, X_{it}^g)$ the *high-dimensional patient state* for patient $i$ at time $t$; we let $\mathcal{X}$ denote the state space.

Given our clinical context, we focus on messages as the action taken by clinicians. In particular, we let $M_{it}$ denote the raw text of treatment messages sent to patient $i$ on day $t$. If no message is sent, then $M_{it} = 0$. We refer to $M_{it}$ as the *high-dimensional action* taken on patient $i$ at time $t$; we let $\mathcal{M}$ denote the action space. Note that in our data, patients are rarely messaged more than once per week.

At a high level, the goal of any targeting policy is to direct clinicians' limited resources to take actions (i.e., send appropriate messages) to those patients who are most in need of intervention (given their patient state). Such a policy is considered effective if it leads to improvements in patients' glucose management. In particular, we define the *reward* $r_{it}$ to be the improvement in the *time-in-range* (TIR) of patient $i$ at day $t$ in the subsequent week relative to the prior week. TIR is the fraction of a patient's glucose readings between 70-180 mg/dL, one of the most commonly used outcome metrics in T1D care (Battelino et al., 2019).

## 4. Methods: A pipeline for policy learning and evaluation

In this section, we outline the three step approach to learning targeting policies for remote patient monitoring of T1D: (1) learning low-dimensional state and action representations; (2) constructing targeting policies by ranking patients based on estimated conditional average treatment effects (CATEs); and (3) evaluating the policies in the presence of capacity constraints (Figure 2). Our approach is applicable to other domains with similar characteristics; where possible we describe each step using general notation and specialize as appropriate to our specific clinical context.

### 4.1. State and action representations

In our setting both $X_{it}$ and $M_{it}$ are high-dimensional relative to the number of patients. For this reason, we require dimension-reduced representations of both states and actions. We let $\mathcal{S}$ (resp., $\mathcal{A}$) denote the space of low-dimensional states (resp., actions). Formally, we let $\gamma : \mathcal{X} \to \mathcal{S}$ be a function that maps each
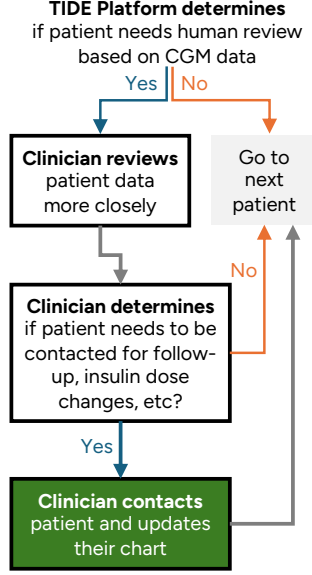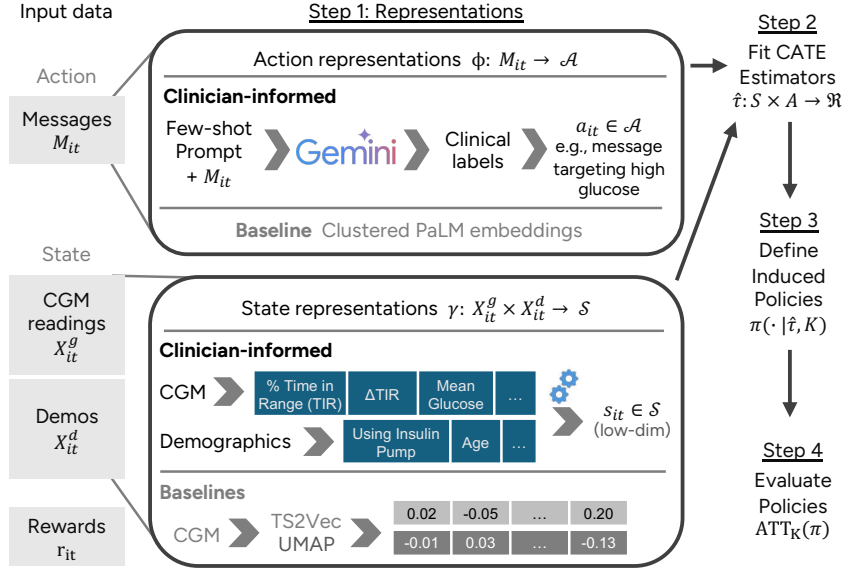
Figure 1: Remote patient monitoring workflow.



Figure 2: Diagram of pipeline for learning targeting policies.

patient's high-dimensional state to a low-dimensional state representation $s_{it} \in \mathcal{S}$, i.e., $s_{it} = \gamma(X_{it})$. Similarly, we let $\phi : \mathcal{M} \to \mathcal{A}$ be a function that maps the high-dimensional action $M_{it}$ to an action representation $a_{it} \in \mathcal{A}$, i.e., $a_{it} = \phi(M_{it})$. We assume that $0 \in \mathcal{A}$, and that $\phi(0) = 0$ uniquely (i.e., the control action maps to itself, and is the only action to do so). For simplicity, we also assume the set $\mathcal{A}$ is *finite* in our development.

We simplify by assuming that patient states are drawn from a stationary *superpopulation* distribution $\mathbb{P}$. Of course, in practice, there may be complex temporal dynamics in the patient state; in general, we presume that any such dynamics can be captured through the high-dimensional underlying state representation $X_{it}$ (e.g., the rich CGM sensor data capturing the trajectory of the patient). With this superpopulation view, we let $R(x, m)$ denote the reward obtained by a patient in state $x$ if they receive action $m$; we can view $R(x, m)$ as the potential outcomes for a patient in state $x$, if we vary the action $m$ (Robins et al., 1994). Note that then $r_{it} = R(X_{it}, M_{it})$, and that given an action $m$, the quantity $\mathbb{E}[R(x, m) | \gamma(x) = s]$ is the expectation of reward (over the superpopulation distribution of patient states) with respect to the dimension-reduced representation. Throughout the paper, we make the following *conditional consistency* assumption; this assumption is similar in nature to the consistency as-

sumption on outcomes in causal inference (Hernán, 2016), except that we have adapted it to apply to the action representation selected.

**Assumption 1** Conditional consistency of action representation. *An action representation $\phi$ is conditionally consistent if for all $s \in \mathcal{S}$ and $m, m' \in \mathcal{M}$ such that $\phi(m) = \phi(m')$, there holds $\mathbb{E}[R(x, m) | \gamma(x) = s] = \mathbb{E}[R(x, m') | \gamma(x) = s]$.*

If $\phi$ satisfies Assumption 1, then for any action $a \in \mathcal{A}$, the quantity $\mathbb{E}[R(x, m) | \gamma(x) = s]$ is the same for any $m$ such that $\phi(m) = a$; thus we can define the reward in terms of the dimension-reduced representations as $\rho(s, a) = \mathbb{E}[R(x, m) | \gamma(x) = s]$ where $m$ satisfies $\phi(m) = a$.

We use two approaches to constructing representations: (1) algorithmic black-box approaches, and (2) clinician-informed representations.

*Black-box baselines: Low-dimensional representations directly from raw data.* For action representation, we get features from the raw message text by generating 728-dimensional embeddings using PaLM (Pathways Language Model) (Anil et al., 2023), a large-scale autoregressive language model. Then we cluster the embeddings into discrete message types using K-means to define discrete actions. For state representations, we consider two methods for learning low-dimensional state representations directly from

the raw CGM traces: TS2Vec (Yue et al., 2021), a universal representation learning framework for time series that applies hierarchical contrastive learning; and UMAP (Uniform Manifold Approximation and Projection), a non-linear dimensionality reduction technique that preserves the local and global structure of the data (McInnes et al., 2018).

*Clinician-informed representations.* In most clinical contexts with high-dimensional states or actions (e.g., sensor time series, imaging, text, etc.), clinicians already have a lower-dimensional set of features they extract for clinical decision-making. Rather than starting from the raw data representation, our approach learns a low-dimensional representation starting from this "medium"-dimensional feature representation. The inductive bias provided by such domain knowledge will prove crucial to learning performant, interpretable policies that are clinically grounded.

For action representations, we extract interpretable clinical features using few-shot labeling with Gemini Pro (Gemini, 2024) to generate labels from each message (e.g., was a dose change recommended, did the dose change focus on high or low glucose). We use the features most predictive of rewards to anchor a set of discrete actions; we then group each of the remaining features with its closest anchor action, or to an "other message type" category, based on the similarity of their clinical meanings. See Appendix C for details.

For state representations, we start with a "medium"-dimensional set of pre-defined CGM clinical features commonly used in practice: time-in-range (TIR; 70-180 mg/dl), mean glucose, time below 70 mg/dl, time below 55 mg/dl, etc. (Battelino et al., 2019). We also include demographics (e.g., age, language preference, insurance type, insulin pump use, etc.). See Appendix D for a full list of included clinician-defined state features. We evaluate representations that are different subsets of these features: the full set, a learned subset predictive of rewards, and a subset defined most relevant by clinicians.

### 4.2. Targeting policies: Ranking via estimated CATEs

A *targeting policy* chooses which patients to prioritize for treatment, and what actions to choose for them, given a capacity constraint. We consider targeting policies that rank patients according to CATEs estimated given the dimension-reduced state and action representations.

A CATE function estimates the effect of an action conditional on the patient's state. The *true* CATE function, denoted $\tau : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$, is $\tau(s,a) = \rho(s,a) - \rho(s,0)$. (Note $\tau(s,0) = 0$ for the control action.) Given a dataset $\mathcal{D} = \{(X_{it}, M_{it}, r_{it})\}_{i,t}$, an *estimated CATE function* $\hat\tau : S \times A \to \mathbb{R}$ is a learned estimate of the expected treatment effect for a given state and action representation (defined by $\gamma$ and $\phi$, respectively).

We estimate CATE functions $\hat\tau$ using several estimators. The S-Learner (Single Learner) is a simple approach that trains a single model to predict the outcome using both the action and state representations as input features. The T-Learner (Two Learners) trains separate models for each treatment action and for a pre-defined control action (e.g., no message), and then estimates the CATE as the difference between their predictions. The X-Learner (X-Learner) (Künzel et al., 2017) is a meta-learner that estimates the CATE by training separate models for each action, and then training a final model on the imputed treatment effects relative to the control action. The Causal Forest (Wager and Athey, 2015) is an extension of random forests that estimates the CATE by recursively partitioning the data based on the covariates and treatment assignment. The DR Forest (Doubly Robust Forest) (Athey et al., 2019) is a variant of the Causal Forest that combines the estimates from a propensity and outcome model to achieve double robustness. We also create an ensemble estimator by combining all of the previous estimators, where the final CATE estimate is obtained by taking a weighted average of the predictions of the individual estimators. The weights for each model in the ensemble are learned using the validation dataset (Mahajan et al., 2023). We fit the CATE estimators using EconML (Syrgkanis et al., 2021) and predictions from nuisance models (prediction propensities and outcomes) trained with AutoML in FLAML (Wang et al., 2021).

For an estimated CATE function $\hat\tau$ and a capacity constraint $K \leq N$ on the number of patients that can receive a treatment $a \neq 0$ (i.e., other than the control treatment), we define a targeting policy that prioritizes treating the patients with the highest estimated rewards under the optimal actions.

**Definition 1 (Induced targeting policy)** *Fix an estimated CATE function $\hat\tau$, capacity constraint $K \leq N$, and a patient state vector $s = (s_1, \ldots, s_N)$.*

The targeting policy $\pi = \pi(\cdot|\hat{\tau}, K)$ induced by $\hat{\tau}$ and $K$ chooses actions for each patient as follows:

1. For each patient $i$, let $a_i^* = \arg\max_{\tilde{a} \in \mathcal{A}} \hat{\tau}(s_i, \tilde{a})$.
2. Rank the patients in descending order of $\hat{\tau}(s_i, a_i^*)$.
3. For the top $K$ ranked patients, set $\pi_i(s) = a_i^*$; for the remaining patients set $\pi_i(s) = 0$.

## 4.3. Evaluating targeting policies

Our goal is to find combinations of state representation $\gamma$, action representation $\phi$, and CATE function $\hat{\tau}$ such that the resulting induced targeting policy $\pi(\cdot|\hat{\tau}, K)$ achieves high quality outcomes, i.e., actually targets patients with the highest treatment effects.

Formally, for any targeting policy $\pi$ with capacity constraint $K$, we define the value function $\mathsf{ATT}_K(\pi)$ as the *average treatment effect on the treated (ATT)*: $\mathsf{ATT}_K(\pi) = \mathbb{E}\left[\frac{1}{K}\sum_{i=1}^{N}\tau(s, \pi_i(s))\right]$. Here the expectation is over the superpopulation, where we assume that patients are sampled i.i.d. from the superpopulation. Note that for any patient that receives the control action under $\pi$, the treatment effect in the sum is zero. As a result, if the policy $\pi$ only provides non-control actions to at most $K$ patients, the right hand side will be the average treatment effect of the treated patients.

Our goal is to learn policies with high ATT, given the capacity constraint $K$. Further, in practice, we will be interested in additional qualitative desiderata, e.g., whether the resulting policy is interpretable or aligns with clinical guidelines. In our empirical evaluation we will test whether these requirements are met, and in particular, whether the use of clinician-informed representations biases selected policies towards being interpretable as well.

Before continuing we comment briefly on the *optimal* policy. In particular, for fixed $K$, let $\pi^* = \pi(\cdot|\tau, K)$; this is the policy that ranks patients according to their true treatment effects. It is easy to check that $\mathsf{ATT}_K(\pi^*) \geq \mathsf{ATT}_K(\pi)$ for any other policy $\pi$ that targets at most $K$ patients; see Proposition 3 in Appendix H for a proof. We show in the following theorem that if we estimate $\tau$ effectively, then the value of the estimated optimal policy converges to the value of the true optimal policy; the result is analogous to existing results for policy learning (Wager and Athey, 2015). See Appendix H for proof details.

**Theorem 2** *For each $N$, let $\hat{\tau}_N$ be a CATE estimator such that $\sup_{s \in \mathcal{S}, a \in \mathcal{A}} |\tau(s, a) - \hat{\tau}_N(s, a)| \to 0$ in*

distribution. *Suppose also that the treatment effects are bounded: $\sup_{s \in \mathcal{S}, a \in \mathcal{A}} |\tau(s, a)| < \infty$. Consider a sequence $K_N$ such that $K_N/N \to c$ as $N \to \infty$, with $0 \leq c \leq 1$. Let $\pi_N = \pi(\cdot|\hat{\tau}_N, K_N)$ be the associated sequence of targeting policies, and let $\pi_N^*$ be the associated sequence of optimal targeting policies. Then $\mathsf{ATT}_{K_N}(\pi_N) - \mathsf{ATT}_{K_N}(\pi_N^*) \to 0$ as $N \to \infty$.*

In general, for a policy $\pi$, to estimate $\mathsf{ATT}_K(\pi)$ for a given $K$, we require an estimate of the treatment effect of each action $a$, for each patient-day $(i, t)$ pair in our evaluation data. A challenge here is that in our data, there is confounding between the actions and rewards. For example, clinicians are more likely to contact a patient with a recent drop in glucose control, and that patient is also more likely to have improved glucose control in the following week even if they are not contacted by a clinician (regression to the mean). If we fail to account for this confounding, we would overestimate the impact of interventions on the reward.

To account for confounding, we use a doubly robust approach. We describe our approach to confounding in detail in Appendix E. In short, to adjust for confounding, we let $X_{it}^c = \gamma^c(X_{it}^g, X_{it}^d)$ denote the representation of the patient state that clinicians see when reviewing patients. We let $\hat{e}(X^c, a)$ denote a learned propensity score model, and let $\hat{r}(X^c, a)$ denote a learned reward model. Now suppose that a patient $i$ is observed at day $t$ in an evaluation dataset $\mathcal{E}$. We define the following doubly robust score for each action $a$; this is an estimate of the treatment effect of action $a$ for patient $i$ at day $t$:

$$\hat{\Gamma}_{it}(a) = (\hat{r}(X_{it}^c, a) - \hat{r}(X_{it}^c, 0))$$
$$+ (r_{it} - \hat{r}(X_{it}^c, a_{it}))\left(\frac{I_{a_{it}=a}}{\hat{e}(X_{it}^c, a)} - \frac{I_{a_{it}=0}}{\hat{e}(X_{it}^c, 0)}\right)$$

We estimate $\mathsf{ATT}_K(\pi)$ on an evaluation dataset $\mathcal{E}$ with $N$ patients and $T$ time periods as:

$$\widehat{\mathsf{ATT}}_K(\pi) = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{K}\sum_{i=1}^{N}\hat{\Gamma}_{it}(\pi_i(s_t)),$$

where $s_t = (s_{1t}, \ldots, s_{Nt})$ is the patient state vector on day $t$. To estimate representative performance in a real clinic, we specifically report $\widehat{\mathsf{ATT}}_K(\pi)$ for $K/N = 0.25$; we refer to this metric as *ATT@25%*. This metric estimates the performance of the policy when treating a proportion of the population that is of similar size to the capacity of our clinic.

**Data splitting.** We split the data into three parts of approximately equal size by randomly dividing the *patients* into three groups and putting the data from each group of patients into separate datasets: train, validation, and test. The training data are used for pre-training state and action representations, CATE estimation, and to train nuisance models (if needed to adjust for confounding). The validation data are used to evaluate the performance of candidate representations and targeting policies induced by the CATE estimators. The test data are used for valid estimation of the performance of the final chosen policy. By creating these splits across patients, we ensure that we learn a treatment policy that will generalize across patients from a similar population.

## 5. Results

**Interpretability and clinical relevance of representations.** We expect to see correlation between states and actions if the representations capture how clinicians make decisions. In managing T1D, clinicians are concerned with *highs* (glucose levels above 180 mg/dL) and *lows* (glucose levels below 70 mg/dL). When these events occur, we expect to see clinicians send messages targeting those events.

Figure 3 shows the correlations between continuous state variables and binary action indicators with clinician-informed (left) or black-box-learned (right) state and action representations. We see that clinician-informed state variables are correlated with clinician-informed actions: messages that target highs (resp., lows) are sent when states representing highs (resp., lows) are observed. TS2Vec-learned state variables are less correlated with embedding-based actions. In other words, the black-box-learned representations are not learning relationships that are relevant to clinicians' actions, while (as expected) the clinician-informed state and action representations are.

**Policy performance.** Figure 4 compares the ATT@25% of different combinations of state and action representations across CATE estimators. We find that the estimated policy performance is equivalent to random targeting (ATE) for policies learned from most of the baseline representations we tested. Notably, policies derived from clinician-informed representations significantly outperform policies learned from black-box algorithmic baseline representations; not only are they more interpretable and clinically grounded, they also have higher efficacy.

A closer look at the state and action representations provides additional insight. For state representations, we see increasing performance with increasing levels of domain knowledge, moving from the "medium"-dimensional representation (all clinician-informed features), to the learned subset, to the fully clinician-informed subset (TIDE features). In this clinical setting, the TIDE-only features represent strong clinical domain knowledge of CGM features relevant to patient care, and have been developed over many years.

By contrast, our clinician-informed action representations require a learning procedure, since we started with high-dimensional (unlabeled) text messages as our raw actions. We again see the benefits of clinical inductive bias: the clinically-informed action set significantly outperforms a black-box-learned clustering. See Appendix F for additional policy evaluation results (including TOC curves).

After identifying the best-performing policy on the validation set, we evaluate it on the held-out test set to check for potential selection bias inflating our results on the validation data. The ATT@25% for the policy (clinician-informed action representation, TIDE state representations, T-Learner) is 6.6 [95% CI: 5.6-7.6] on the test set, which is similar to the validation set result of 6.7 [5.7-7.7].

**Policy interpretation and clinical alignment.** By inspecting how the CATE predictions vary across clinician-defined features we can assess if they align with clinical knowledge. We want to recommend only those policies that align with clinical best practices (Battelino et al., 2019), since clinicians are unlikely to adopt the policy otherwise.

Although management of T1D requires careful attention to both highs and lows, highs are much more consequential events for TIR than lows (Addala et al., 2021). Lows are often emergent events requiring acute intervention, and also are coupled to other interventions (e.g., alarms on CGMs). By contrast, highs are more persistent, significant, and longer-term in their impacts on TIR. Since TIR is our reward, we expect that clinically aligned policies should focus primarily on reducing highs (Addala et al., 2021). Additionally, we expect that patients with larger drops in TIR week-over-week, and patients with higher mean glucose, are more likely to see a TIR benefit from clinician intervention. Finally, it is clinically well-established that patients using insulin pumps tend to be better able to manage their TIR, and thus patients

| | **Clinician-informed Action** | | | | | **Embedding-based Actions** | | | |
| **Clinician-Informed State Variable** | Message Treating Highs & Lows | Message Treating Highs only | Message Treating Lows only | Other Message | **TS2Vec State Variable** | Embedding Cluster 1 | Embedding Cluster 2 | Embedding Cluster 3 | Embedding Cluster 4 |
|---|---|---|---|---|---|---|---|---|---|
| % Time Very Low | 0.021 | -0.032 | 0.072 | 0.006 | Dim 1 | -0.013 | -0.006 | 0.004 | -0.023 |
| % Time Low | 0.019 | -0.042 | 0.096 | 0.011 | Dim 2 | 0.015 | 0.008 | 0.003 | 0.016 |
| % Time High | 0.020 | 0.044 | -0.066 | 0.004 | Dim 3 | -0.005 | 0.004 | 0.000 | -0.006 |
| % Time Very High | -0.004 | 0.010 | -0.050 | -0.021 | Dim 4 | 0.019 | 0.005 | 0.000 | 0.014 |

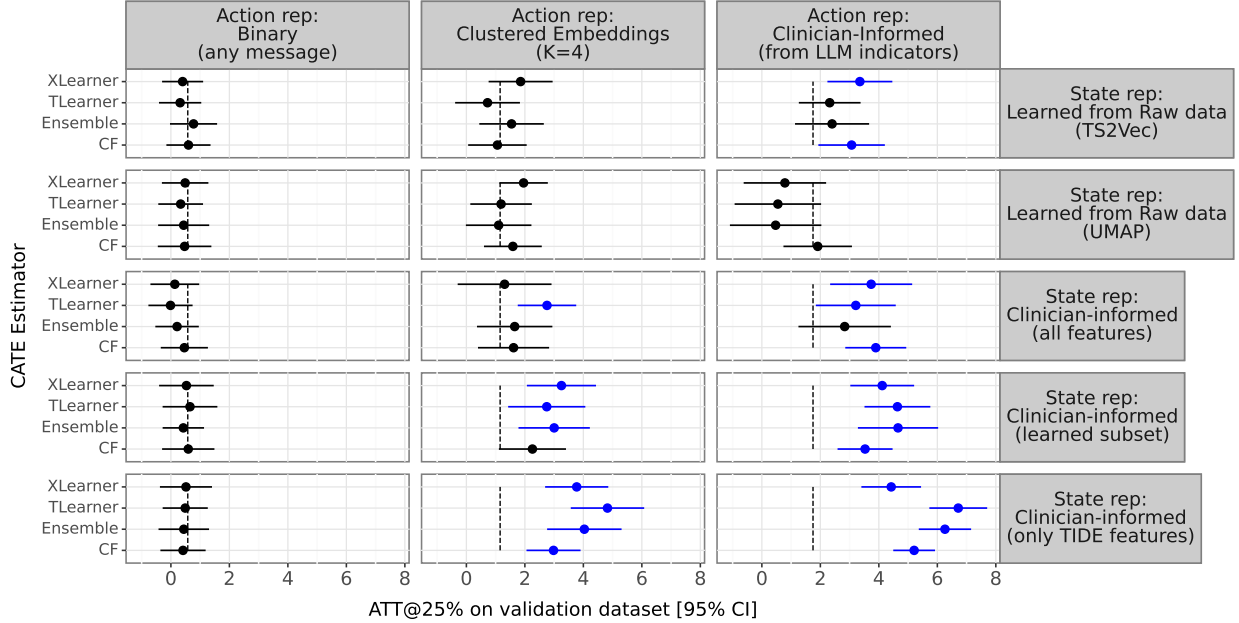Figure 3: Pearson correlations between continuous state variables and binary action indicators.



Figure 4: Estimated ATT@25% with 95% CIs on validation data across representations and CATE estimators. ATEs of the action with the highest predicted ATE shown as vertical dashed lines (expected ATT under random targeting). See Appendix F for more results.

not using insulin pumps are more likely to see a TIR benefit from clinician intervention (Berget et al., 2019).

In Figure 5, we show the values of the CATE predictions for the optimal action, as we vary patient features across panels. For each factor, the policy using clinician-informed representations matches exactly the clinical guidelines described in the previous paragraph: patients with lower TIR, a larger drop in TIR week-over-week, a higher mean glucose, or not using a pump will be prioritized for contact. Conversely, the policies using black-box representations lack this interpretability and are badly misaligned with clinical guidance: notably they prioritize patients with higher TIR, lower drops in TIR, and lower mean glucose.

## 6. Discussion

We introduced an end-to-end pipeline for learning and evaluating policies induced by different CATE estimators across both black-box and low-dimensional, clinically-grounded state and action representations. Using data from clinical trials with remote patient monitoring for type 1 diabetes care, we learned clinician-informed policies that outperformed black-box-learned policies.

Iterating on the state and action representations could further improve policy performance. We expect that clustering-embedding-based action representations might produce more effective action representations with increased data, though interpretabil-
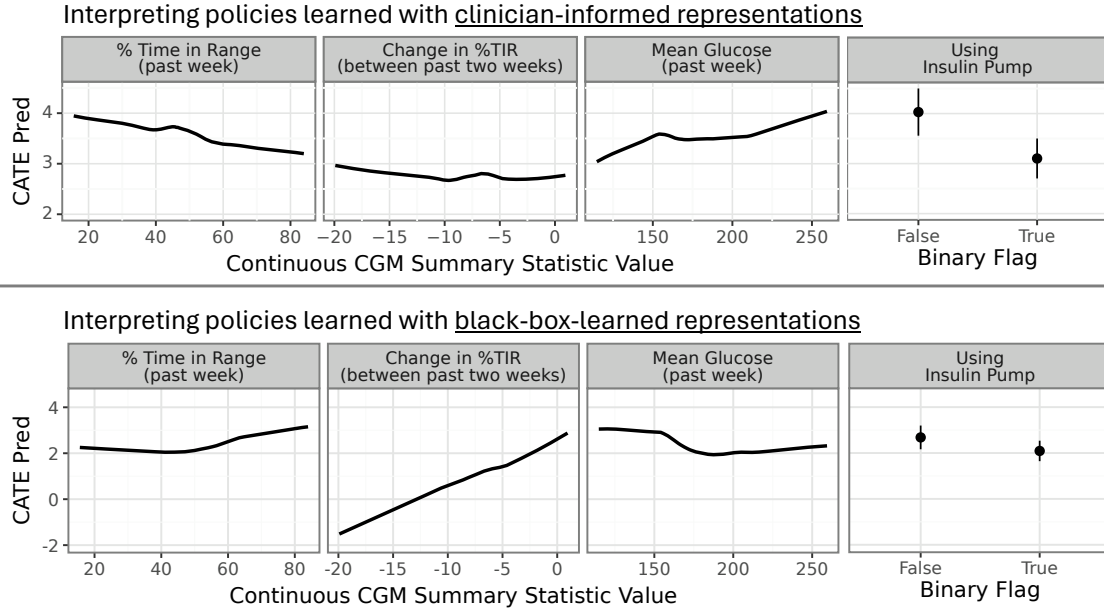
Figure 5: CATE predictions for the optimal action for policies learned using both approaches. Policies learned from clinician-informed representations (top) are interpretable and align with clinical guidelines, while those learned from black-box-learned representations (bottom) do not.

ity would remain a challenge. Any algorithm would likely need a very large amount of data to learn meaningful state representations of the high-dimensional CGM traces that predict where actions fall in the high-dimensional message text space. With low-dimensional clinician-informed representations, it is much easier to learn the relationship between state and action representations (e.g. patients with more low CGM readings are more likely to get a message addressing low glucose). Future work could leverage larger datasets or synthetic data to understand how much data is necessary to learn useful embeddings for policy learning from high-dimensional clinical data.

Despite adjusting for all available information when estimating treatment effects and learning policies, unmeasured confounding may persist, which could bias our results. The only way to guarantee no unmeasured confounding would be to collect data where actions are taken randomly with known propensities, which might be infeasible in most healthcare settings. Future work could include sensitivity analyses to unmeasured confounding.

In our setting, the care team has capacity to take actions on K patients in each time period. Future work could examine a more general setting in which different actions have different capacity costs. Our

analysis focuses on short-term outcomes; evaluating long-term effects requires additional assumptions or a randomized controlled trial of the learned policies over a longer period (Ferstad et al., 2024; Collins et al., 2007).

Our approach can improve digital health interventions with large state and action spaces when clinical domain knowledge is available. For instance, it could enhance interventions based on wearable sensor data (e.g., smart watches measuring pulse and activity). Exactly how the approach is applied will depend on the setting and which clinician-informed state and action representations are available. Our approach enables evaluating policies learned from different candidate representations. Successful deployment of successful targeting policies could boost intervention efficacy and patient outcomes, promoting digital health adoption. However, it is crucial to ensure that patients *not* selected for treatment by a learned policy receive alternative or complementary interventions. Future work requires identifying and evaluating such interventions to ensure all patients receive appropriate care, as well as ensuring equitable access to such interventions (Prahalad et al., 2024).

## Acknowledgments

## References

Ananta Addala, Dessi P Zaharieva, Angela J Gu, Priya Prahalad, David Scheinker, Bruce Buckingham, Korey K Hood, and David M Maahs. Clinically serious hypoglycemia is rare and not associated with time-in-range in youth with new-onset type 1 diabetes. *The Journal of Clinical Endocrinology & Metabolism*, 106(11):3239–3247, 2021.

Gerard F Anderson, Uwe E Reinhardt, Peter S Hussey, and Varduhi Petrosyan. It's the prices, stupid: why the United States is so different from other countries. *Health Affairs*, 22(3):89–105, 2003. Publisher: Project HOPE-The People-to-People Health Foundation, Inc.

Rohan Anil, Andrew M. Dai, Orhan Firat, Melvin Johnson, Dmitry Lepikhin, Alexandre Passos, Siamak Shakeri, Emanuel Taropa, Paige Bailey, Zhifeng Chen, Eric Chu, Jonathan H. Clark, Laurent El Shafey, Yanping Huang, Kathy Meier-Hellstern, Gaurav Mishra, Erica Moreira, Mark Omernick, Kevin Robinson, Sebastian Ruder, Yi Tay, Kefan Xiao, Yuanzhong Xu, Yujing Zhang, Gustavo Hernandez Abrego, Junwhan Ahn, Jacob Austin, Paul Barham, Jan Botha, James Bradbury, Siddhartha Brahma, Kevin Brooks, Michele Catasta, Yong Cheng, Colin Cherry, Christopher A. Choquette-Choo, Aakanksha Chowdhery, Clément Crepy, Shachi Dave, Mostafa Dehghani, Sunipa Dev, Jacob Devlin, Mark Díaz, Nan Du, Ethan Dyer, Vlad Feinberg, Fangxiaoyu Feng, Vlad Fienber, Markus Freitag, Xavier Garcia, Sebastian Gehrmann, Lucas Gonzalez, Guy Gur-Ari, Steven Hand, Hadi Hashemi, Le Hou, Joshua Howland, Andrea Hu, Jeffrey Hui, Jeremy Hurwitz, Michael Isard, Abe Ittycheriah, Matthew Jagielski, Wenhao Jia, Kathleen Kenealy, Maxim Krikun, Sneha Kudugunta, Chang Lan, Katherine Lee, Benjamin Lee, Eric Li, Music Li, Wei Li, YaGuang Li, Jian Li, Hyeontaek Lim, Hanzhao Lin, Zhongtao Liu, Frederick Liu, Marcello Maggioni, Aroma Mahendru, Joshua Maynez, Vedant Misra, Maysam Moussalem, Zachary Nado, John Nham, Eric Ni, Andrew Nystrom, Alicia Parrish, Marie Pellat, Martin Polacek, Alex Polozov, Reiner Pope, Siyuan Qiao, Emily Reif, Bryan Richter, Parker Riley, Alex Castro Ros, Aurko Roy, Brennan Saeta, Rajkumar Samuel, Renee Shelby, Ambrose Slone, Daniel Smilkov, David R. So, Daniel Sohn, Simon Tokumine, Dasha Valter, Vijay Vasudevan, Kiran Vodrahalli, Xuezhi Wang, Pidong Wang, Zirui Wang, Tao Wang, John Wieting, Yuhuai Wu, Kelvin Xu, Yunhan Xu, Linting Xue, Pengcheng Yin, Jiahui Yu, Qiao Zhang, Steven Zheng, Ce Zheng, Weikang Zhou, Denny Zhou, Slav Petrov, and Yonghui Wu. PaLM 2 Technical Report, May 2023. URL https://arxiv.org/abs/2305.10403v3.

Susan Athey, Julie Tibshirani, and Stefan Wager. Generalized random forests. *The Annals of Statistics*, 47(2):1148–1178, April 2019. ISSN 0090-5364, 2168-8966. doi: 10.1214/18-AOS1709. URL https://projecteuclid.org/journals/annals-of-statistics/volume-47/issue-2/Generalized-random-forests/10.1214/18-AOS1709.full. Publisher: Institute of Mathematical Statistics.

Tadej Battelino, Thomas Danne, Richard M. Bergenstal, Stephanie A. Amiel, Roy Beck, Torben

Biester, Emanuele Bosi, Bruce A. Buckingham, William T. Cefalu, Kelly L. Close, Claudio Cobelli, Eyal Dassau, J. Hans DeVries, Kim C. Donaghue, Klemen Dovc, Francis J. Doyle, Satish Garg, George Grunberger, Simon Heller, Lutz Heinemann, Irl B. Hirsch, Roman Hovorka, Weiping Jia, Olga Kordonouri, Boris Kovatchev, Aaron Kowalski, Lori Laffel, Brian Levine, Alexander Mayorov, Chantal Mathieu, Helen R. Murphy, Revital Nimri, Kirsten Nørgaard, Christopher G. Parkin, Eric Renard, David Rodbard, Banshi Saboo, Desmond Schatz, Keaton Stoner, Tatsuiko Urakami, Stuart A. Weinzimer, and Moshe Phillip. Clinical targets for continuous glucose monitoring data interpretation: Recommendations from the international consensus on time in range. *Diabetes Care*, 42(8):1593–1603, August 2019. ISSN 19355548. doi: 10.2337/dci19-0028. URL https://pubmed.ncbi.nlm.nih.gov/31177185/. Publisher: American Diabetes Association Inc.

Cari Berget, Laurel H Messer, and Gregory P Forlenza. A clinical overview of insulin pump therapy for the management of diabetes: past, present, and future of intensive therapy. *Diabetes spectrum: a publication of the American Diabetes Association*, 32(3):194, 2019.

Israel Júnior Borges do Nascimento, Hebatullah Abdulazeem, Lenny Thinagaran Vasanthan, Edson Zangiacomi Martinez, Miriane Lucindo Zucoloto, Lasse Østengaard, Natasha Azzopardi-Muscat, Tomas Zapata, and David Novillo-Ortiz. Barriers and facilitators to utilizing digital health technologies by healthcare professionals. *npj Digital Medicine*, 6(1):1–28, September 2023. ISSN 2398-6352. doi: 10.1038/s41746-023-00899-4. URL https://www.nature.com/articles/s41746-023-00899-4. Publisher: Nature Publishing Group.

Judith Borghouts, Elizabeth Eikey, Gloria Mark, Cinthia De Leon, Stephen M. Schueller, Margaret Schneider, Nicole Stadnick, Kai Zheng, Dana Mukamel, and Dara H. Sorkin. Barriers to and Facilitators of User Engagement With Digital Mental Health Interventions: Systematic Review. *Journal of Medical Internet Research*, 23(3):e24387, March 2021. doi: 10.2196/24387. URL https://www.jmir.org/2021/3/e24387. Company: Journal of Medical Internet Research Distributor: Journal of

Medical Internet Research Institution: Journal of Medical Internet Research Label: Journal of Medical Internet Research Publisher: JMIR Publications Inc., Toronto, Canada.

Djallel Bouneffouf, Irina Rish, and Charu Aggarwal. Survey on Applications of Multi-Armed and Contextual Bandits. In *2020 IEEE Congress on Evolutionary Computation (CEC)*, pages 1–8, July 2020. doi: 10.1109/CEC48606.2020.9185782. URL https://ieeexplore.ieee.org/abstract/document/9185782.

Linda M. Collins, Susan A. Murphy, and Victor Strecher. The Multiphase Optimization Strategy (MOST) and the Sequential Multiple Assignment Randomized Trial (SMART): New Methods for More Potent eHealth Interventions. *American journal of preventive medicine*, 32(5 Suppl):S112–S118, May 2007. ISSN 0749-3797. doi: 10.1016/j.amepre.2007.01.022. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2062525/.

Kathrin Cresswell and Aziz Sheikh. Organizational issues in the implementation and adoption of health information technology innovations: An interpretative review. *International Journal of Medical Informatics*, 82(5):e73–e86, May 2013. ISSN 1386-5056. doi: 10.1016/j.ijmedinf.2012.10.007. URL https://www.sciencedirect.com/science/article/pii/S1386505612001992.

Raaz Dwivedi, Yan Shuo Tan, Briton Park, Mian Wei, Kevin Horgan, David Madigan, and Bin Yu. Stable Discovery of Interpretable Subgroups via Calibration in Causal Studies. *International Statistical Review*, 88(S1):S135–S178, 2020. ISSN 1751-5823. doi: 10.1111/insr.12427. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/insr.12427.

Johannes O. Ferstad, Jacqueline J. Vallon, Daniel Jun, Angela Gu, Anastasiya Vitko, Dianelys P. Morales, Jeannine Leverenz, Ming Yeh Lee, Brianna Leverenz, Christos Vasilakis, Esli Osmanlliu, Priya Prahalad, David M. Maahs, Ramesh Johari, and David Scheinker. Population-level management of type 1 diabetes via continuous glucose monitoring and algorithm-enabled patient prioritization: Precision health meets population health. *Pediatric Diabetes*, 22(7):982–991,

2021. ISSN 1399-5448. doi: 10.1111/pedi. 13256. URL https://onlinelibrary.wiley. com/doi/abs/10.1111/pedi.13256.

Johannes O. Ferstad, Priya Prahalad, David M. Maahs, Dessi P. Zaharieva, Emily Fox, Manisha Desai, Ramesh Johari, and David Scheinker. Smart Start — Designing Powerful Clinical Trials Using Pilot Study Data. *NEJM Evidence*, 3(2): EVIDoa2300164, January 2024. doi: 10.1056/ EVIDoa2300164. URL https://evidence.nejm. org/doi/abs/10.1056/EVIDoa2300164. Publisher: Massachusetts Medical Society.

Stefan Feuerriegel, Dennis Frauen, Valentyn Melnychuk, Jonas Schweisthal, Konstantin Hess, Alicia Curth, Stefan Bauer, Niki Kilbertus, Isaac S. Kohane, and Mihaela van der Schaar. Causal machine learning for predicting treatment outcomes. *Nature Medicine*, 30(4):958–968, April 2024. ISSN 1546-170X. doi: 10.1038/ s41591-024-02902-1. URL https://www.nature. com/articles/s41591-024-02902-1. Publisher: Nature Publishing Group.

Team Gemini. Gemini: A Family of Highly Capable Multimodal Models. *arXiv.org*, 2024.

Miguel A. Hernán. Does water kill? A call for less casual causal inferences. *Annals of epidemiology*, 26(10):674–680, October 2016. ISSN 1047-2797. doi: 10.1016/j.annepidem.2016.08. 016. URL https://www.ncbi.nlm.nih.gov/pmc/ articles/PMC5207342/.

Kosuke Imai and Michael Lingzhi Li. Experimental Evaluation of Individualized Treatment Rules. *Journal of the American Statistical Association*, 118(541):242–256, January 2023. ISSN 0162-1459. doi: 10.1080/01621459.2021.1923511. URL https: //doi.org/10.1080/01621459.2021.1923511.

Kosuke Imai and Michael Lingzhi Li. Statistical Inference for Heterogeneous Treatment Effects Discovered by Generic Machine Learning in Randomized Experiments. *Journal of Business & Economic Statistics*, 0(0):1–13, 2024. ISSN 0735-0015. doi: 10.1080/07350015.2024.2358909. URL https: //doi.org/10.1080/07350015.2024.2358909.

Guido W. Imbens and Donald B. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction.* Cambridge University Press, Cambridge, 2015. ISBN 978-0-521-88588-1. doi: 10.1017/CBO9781139025751.

Gloria YK Kim, Rea Rostosky, Franziska K Bishop, Kelly Watson, Priya Prahalad, Aishwari Vaidya, Sharon Lee, Alexander Diana, Clint Beacock, Brian Chu, et al. The adaptation of a single institution diabetes care platform into a nationally available turnkey solution. *npj Digital Medicine*, 7 (1):311, 2024.

David C. Klonoff, Jing Wang, David Rodbard, Michael A. Kohn, Chengdong Li, Dorian Liepmann, David Kerr, David Ahn, Anne L. Peters, Guillermo E. Umpierrez, Jane Jeffrie Seley, Nicole Y. Xu, Kevin T. Nguyen, Gregg Simonson, Michael S. D. Agus, Mohammed E. Al-Sofiani, Gustavo Armaiz-Pena, Timothy S. Bailey, Ananda Basu, Tadej Battelino, Sewagegn Yeshiwas Bekele, Pierre-Yves Benhamou, B. Wayne Bequette, Thomas Blevins, Marc D. Breton, Jessica R. Castle, James Geoffrey Chase, Kong Y. Chen, Pratik Choudhary, Mark A. Clements, Kelly L. Close, Curtiss B. Cook, Thomas Danne, Francis J. Doyle, Angela Drincic, Kathleen M. Dungan, Steven V. Edelman, Niels Ejskjaer, Juan C. Espinoza, G. Alexander Fleming, Gregory P. Forlenza, Guido Freckmann, Rodolfo J. Galindo, Ana Maria Gomez, Hanna A. Gutow, Lutz Heinemann, Irl B. Hirsch, Thanh D. Hoang, Roman Hovorka, Johan H. Jendle, Linong Ji, Shashank R. Joshi, Michael Joubert, Suneil K. Koliwad, Rayhan A. Lal, M. Cecilia Lansang, Wei-An Andy Lee, Lalantha Leelarathna, Lawrence A. Leiter, Marcus Lind, Michelle L. Litchman, Julia K. Mader, Katherine M. Mahoney, Boris Mankovsky, Umesh Masharani, Nestoras N. Mathioudakis, Alexander Mayorov, Jordan Messler, Joshua D. Miller, Viswanathan Mohan, James H. Nichols, Kirsten Nørgaard, David N. O'Neal, Francisco J. Pasquel, Athena Philis-Tsimikas, Thomas Pieber, Moshe Phillip, William H. Polonsky, Rodica Pop-Busui, Gerry Rayman, Eun-Jung Rhee, Steven J. Russell, Viral N. Shah, Jennifer L. Sherr, Koji Sode, Elias K. Spanakis, Deborah J. Wake, Kayo Waki, Amisha Wallia, Melissa E. Weinberg, Howard Wolpert, Eugene E. Wright, Mihail Zilbermint, and Boris Kovatchev. A Glycemia Risk

Index (GRI) of Hypoglycemia and Hyperglycemia for Continuous Glucose Monitoring Validated by Clinician Ratings. *Journal of Diabetes Science and Technology*, 17(5):1226–1242, September 2023. ISSN 1932-2968. doi: 10.1177/19322968221085273.

Sören R. Künzel, Jasjeet S. Sekhon, Peter J. Bickel, and Bin Yu. Meta-learners for Estimating Heterogeneous Treatment Effects using Machine Learning, June 2017. URL https://arxiv.org/abs/1706.03461v6.

Katharine Lawrence, Nina Singh, Zoe Jonassen, Lisa L. Groom, Veronica Alfaro Arias, Soumik Mandal, Antoinette Schoenthaler, Devin Mann, Oded Nov, and Graham Dove. Operational Implementation of Remote Patient Monitoring Within a Large Ambulatory Health System: Multimethod Qualitative Case Study. *JMIR Human Factors*, 10(1):e45166, July 2023. doi: 10.2196/45166. URL https://humanfactors.jmir.org/2023/1/e45166. Company: JMIR Human Factors Distributor: JMIR Human Factors Institution: JMIR Human Factors Label: JMIR Human Factors Publisher: JMIR Publications Inc., Toronto, Canada.

Shaun Liverpool, Catarina Pinheiro Mota, Célia M. D. Sales, Anja Čuš, Sara Carletto, Camellia Hancheva, Sónia Sousa, Sonia Conejo Cerón, Patricia Moreno-Peral, Giada Pietrabissa, Bettina Moltrecht, Randi Ulberg, Nuno Ferreira, and Julian Edbrooke-Childs. Engaging Children and Young People in Digital Mental Health Interventions: Systematic Review of Modes of Delivery, Facilitators, and Barriers. *Journal of Medical Internet Research*, 22(6):e16317, June 2020. doi: 10.2196/16317. URL https://www.jmir.org/2020/6/e16317. Company: Journal of Medical Internet Research Distributor: Journal of Medical Internet Research Institution: Journal of Medical Internet Research Label: Journal of Medical Internet Research Publisher: JMIR Publications Inc., Toronto, Canada.

Divyat Mahajan, Ioannis Mitliagkas, Brady Neal, and Vasilis Syrgkanis. Empirical Analysis of Model Selection for Heterogeneous Causal Effect Estimation, June 2023. URL http://arxiv.org/abs/2211.01939. arXiv:2211.01939 [cs, stat].

Leland McInnes, John Healy, and James Melville. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction, February 2018. URL https://arxiv.org/abs/1802.03426v3.

Michelle M. Mello and Sherri Rose. Denial—Artificial Intelligence Tools and Health Insurance Coverage Decisions. *JAMA Health Forum*, 5(3):e240622, March 2024. ISSN 2689-0186. doi: 10.1001/jamahealthforum.2024.0622. URL https://doi.org/10.1001/jamahealthforum.2024.0622.

Health Technology Institute Peterson. Digital Diabetes Management Solutions, 2024. URL https://phti.com/assessment/digital-diabetes-management-tools/.

Priya Prahalad, Victoria Y Ding, Dessi P Zaharieva, Ananta Addala, Ramesh Johari, David Scheinker, Manisha Desai, Korey Hood, and David M Maahs. Teamwork, Targets, Technology, and Tight Control in Newly Diagnosed Type 1 Diabetes: the Pilot 4T Study. *The Journal of Clinical Endocrinology & Metabolism*, 107(4):998–1008, April 2022. ISSN 0021-972X. doi: 10.1210/clinem/dgab859. URL https://doi.org/10.1210/clinem/dgab859.

Priya Prahalad, David Scheinker, Manisha Desai, Victoria Y. Ding, Franziska K. Bishop, Ming Yeh Lee, Johannes Ferstad, Dessi P. Zaharieva, Ananta Addala, Ramesh Johari, Korey Hood, and David M. Maahs. Equitable implementation of a precision digital health program for glucose management in individuals with newly diagnosed type 1 diabetes. *Nature Medicine*, pages 1–9, May 2024. ISSN 1546-170X. doi: 10.1038/s41591-024-02975-y. URL https://www.nature.com/articles/s41591-024-02975-y. Publisher: Nature Publishing Group.

Khansa Rasheed, Adnan Qayyum, Mohammed Ghaly, Ala Al-Fuqaha, Adeel Razi, and Junaid Qadir. Explainable, trustworthy, and ethical machine learning for healthcare: A survey. *Computers in Biology and Medicine*, 149:106043, October 2022. ISSN 0010-4825. doi: 10.1016/j.compbiomed.2022.106043. URL https://www.sciencedirect.com/science/article/pii/S0010482522007569.

James M. Robins, Andrea Rotnitzky, and Lue Ping Zhao. Estimation of Regression Coefficients When Some Regressors are not Always Observed. *Journal of the American Statistical Association*, September 1994. ISSN 0162-1459. URL https://www.tandfonline.com/doi/abs/10.1080/01621459.1994.10476818. Publisher: Taylor & Francis Group.

José E Rodríguez and Kendall M Campbell. Racial and ethnic disparities in prevalence and care of patients with type 2 diabetes. *Clinical Diabetes*, 35 (1):66–70, 2017. Publisher: Am Diabetes Assoc.

Farzan Sasangohar, Elise Davis, Bita A. Kash, and Sohail R. Shah. Remote Patient Monitoring and Telemedicine in Neonatal and Pediatric Settings: Scoping Literature Review. *Journal of Medical Internet Research*, 20(12):e9403, December 2018. doi: 10.2196/jmir.9403. URL https://www.jmir.org/2018/12/e295. Company: Journal of Medical Internet Research Distributor: Journal of Medical Internet Research Institution: Journal of Medical Internet Research Label: Journal of Medical Internet Research Publisher: JMIR Publications Inc., Toronto, Canada.

David Scheinker, Priya Prahalad, Ramesh Johari, David M. Maahs, and Rick Majzun. A New Technology-Enabled Care Model for Pediatric Type 1 Diabetes. *NEJM catalyst innovations in care delivery*, 3(5):10.1056/CAT.21.0438, May 2022. ISSN 2642-0007. doi: 10.1056/CAT.21.0438. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9767424/.

Sebastian Stevens, Susan Gallagher, Tim Andrews, Liz Ashall-Payne, Lloyd Humphreys, and Simon Leigh. The effectiveness of digital health technologies for patients with diabetes mellitus: a systematic review. *Frontiers in Clinical Diabetes and Healthcare*, 3, 2022. Publisher: Frontiers Media SA.

Erik Sverdrup, Han Wu, Susan Athey, and Stefan Wager. Qini Curves for Multi-Armed Treatment Rules, September 2023. URL http://arxiv.org/abs/2306.11979. arXiv:2306.11979 [stat].

Vasilis Syrgkanis, Greg Lewis, Miruna Oprescu, Maggie Hei, Keith Battocchi, Eleanor Dillon, Jing Pan, Yifeng Wu, Paul Lo, Huigang Chen, Totte Harinen, and Jeong-Yoon Lee. Causal Inference and Machine Learning in Practice with EconML and CausalML: Industrial Use Cases at Microsoft, TripAdvisor, Uber. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, KDD '21, pages 4072–4073, New York, NY, USA, August 2021. Association for Computing Machinery. ISBN 978-1-4503-8332-5. doi: 10.1145/3447548.3470792. URL https://dl.acm.org/doi/10.1145/3447548.3470792.

Shengpu Tang and Jenna Wiens. Model Selection for Offline Reinforcement Learning: Practical Considerations for Healthcare Settings. *PMLR*, 2021.

Roosa Tikkanen. Multinational Comparisons of Health Systems Data, 2019. *The Commonwealth Fund*, 2017.

Vivalink. Remote Patient Monitoring Adoption Increased over 300 Percent in Two Years, 2023.

Stefan Wager and Susan Athey. Estimation and Inference of Heterogeneous Treatment Effects using Random Forests, October 2015. URL https://arxiv.org/abs/1510.04342v4.

Chi Wang, Qingyun Wu, Markus Weimer, and Erkang Zhu. FLAML: A Fast and Lightweight AutoML Library. *Proceedings of Machine Learning and Systems*, 3:434–447, March 2021. URL https://proceedings.mlsys.org/paper_files/paper/2021/hash/1ccc3bfa05cb37b917068778f3c4523a-Abstract.html.

Sera Whitelaw, Danielle M Pellegrini, Mamas A Mamas, Martin Cowie, and Harriette G C Van Spall. Barriers and facilitators of the uptake of digital health technology in cardiovascular care: a systematic scoping review. *European Heart Journal - Digital Health*, 2(1):62–74, March 2021. ISSN 2634-3916. doi: 10.1093/ehjdh/ztab005. URL https://doi.org/10.1093/ehjdh/ztab005.

Steve Yadlowsky, Scott Fleming, Nigam Shah, Emma Brunskill, and Stefan Wager. Evaluating Treatment Prioritization Rules via Rank-Weighted Average Treatment Effects, November 2021. URL http://arxiv.org/abs/2111.07966.

arXiv:2111.07966 [stat].

Zhihan Yue, Yujing Wang, Juanyong Duan, Tian-
meng Yang, Congrui Huang, Yunhai Tong, and
Bixiong Xu. TS2Vec: Towards Universal Rep-
resentation of Time Series, June 2021. URL
https://arxiv.org/abs/2106.10466v3. arXiv:
2106.10466.

## Appendix A. Data summary

Table 1 presents the summary statistics of the patient population across the three data splits. We have more than a year of data and 10+ messages received (treatments) for most of the patients in our datasets. The data splits have similar summary statistics, as expected with random splitting.

Table 1: Summary statistics by dataset/split

| | Dataset | | |
|---|---|---|---|
| **Summary statistic** | Train | Validation | Test |
| N (# of patients) | 91 | 95 | 95 |
| Pilot study participants | 26 | 24 | 25 |
| 4T study participants | 42 | 42 | 46 |
| TIPS study participants | 24 | 35 | 27 |
| # observations (days); mean (IQR) across patients | 734 (408-1071) | 691 (350-969) | 699 (411-952) |
| # messages received; mean (IQR) across patients | 41 (17-60) | 38 (12-58) | 35 (14-52) |
| Age; mean (IQR) across patients | 13 (10-17) | 12 (9-16) | 13 (10-17) |
| N patients using an insulin pump during study | 66 | 70 | 63 |
| N patients using automated insulin delivery during study | 51 | 34 | 41 |

## Appendix B. Computational resources

To train nuisance models and CATE estimators with EconML and FLAML, we used an instance of Google Compute Engine machine type n2d-standard-224 with 224 vCPUs and 896 GB memory. We also did some training on a n2d-standard-64 instance (64 vCPUs, 256 GB RAM). Generating all of the results in the paper took less than a day with the n2d-standard-224.

To train the TS2Vec encoder, we used an instance of Google Compute Engine machine type n1-highmem-8 with 8 vCPUs, 52 GB memory, and one NVIDIA V100 GPU. Training the encoder took less than a day.

# Appendix C. Action representations

We generate clinical features with a few-shot prompt and language model (LM). Here's an example prompt and output.

### Prompt

```
You are an AI assistant who extracts structured JSON from messages sent by clinicians to patients with
    diabetes.
* First identify whether the message recommends changing insulin (recommends_insulin_dose_change).
* If the message recommends an insulin change, label whether it is a basal / long acting insulin change (
    recommends_changing_basal_or_long_acting_insulin), a reminder to take correction doses (
    recommends_more_correction_doses), whether it adjusts the carb ratio at meal time (
    recommends_changing_carb_ratio), or if it has a reminder to bolus before meals (
    reminds_patient_to_bolus).
* Then identify the targets of the message, which can include nighttime glucose (
    recommends_insulin_change_at_night), daytime/mealtime glucose (
    recommends_insulin_change_during_the_day), high glucose or low time in range (
    recommendations_target_high_glucose_or_low_time_in_range), or low glucose (
    recommendations_target_low_glucose).
* Finally, determine if the message mentiones a recent visit (mentions_recent_visit), or knowledge of the
     patient's schedule (mentions_patient_schedule).

The output should be JSON with this structure: ... # see below

input: Looking at trends this past week compared to the previous week your average increased from 135 to
    245. Has anything changed in your routine? Any insights into your daily routine would help me
    recommend dose changes. For now, I have decreased your bedtime correction target to 150. By giving
    correction insulin before bed we can help keep you in target range overnight. I would like to make
    further recommendation to help bring you down to target range more often , but again can you provide
     insights into your schedule ? How many times per day are you dosing insulin ? Are you giving
    insulin for both carbs and glucose level at mealtime?
output: {{
recommends_insulin_dose_change: True,
recommends_changing_basal_or_long_acting_insulin: False,
recommends_more_correction_doses:  True,
recommends_changing_carb_ratio:  False,
reminds_patient_to_bolus:  False,
recommends_insulin_change_at_night: True,
recommends_insulin_change_during_the_day: False,
recommendations_target_high_glucose_or_low_time_in_range:  True,
recommendations_target_low_glucose:  False,
mentions_recent_visit: False,
mentions_patient_schedule: False
}}
# + 4 more examples

input: X has an average blood glucose of 390 and is in target range (70-150) 1%
output:
```

### Output

```
{
recommends_insulin_dose_change: True,
recommends_changing_basal_or_long_acting_insulin: True,
recommends_more_correction_doses: False,
recommends_changing_carb_ratio: False,
reminds_patient_to_bolus: False,
recommends_insulin_change_at_night: True,
recommends_insulin_change_during_the_day: False,
recommendations_target_high_glucose_or_low_time_in_range: True,
recommendations_target_low_glucose: False,
mentions_recent_visit: False,
mentions_patient_schedule: False
}
```

Table 2: Summary statistics across extracted clinical features

| Clinical feature | Total messages (%) |
|---|---|
| recommends_insulin_dose_change | 4,523 (18%) |
| recommends_changing_basal_or_long_acting_insulin | 2,606 (10%) |
| recommends_more_correction_doses | 1,037 (4%) |
| recommends_changing_carb_ratio | 1,795 (7%) |
| reminds_patient_to_bolus | 1,167 (5%) |
| recommends_insulin_change_at_night | 2,802 (11%) |
| recommends_insulin_change_during_the_day | 2,280 (9%) |
| recommendations_target_high_glucose_or_low_time_in_range | 4,183 (17%) |
| recommendations_target_low_glucose | 3,861 (15%) |
| mentions_recent_visit | 1,510 (6%) |
| mentions_patient_schedule | 1,521 (6%) |

**Defining discrete clinician-informed action representations.** To get discrete representations from the clinical features, we first estimate the average treatment effect on the treated to make sure we include features associated with much greater treatment effects than just receiving any message. Then we group the features into discrete representations with clinically different meanings:

- Message treating highs and lows:
  (recommendations_target_low_glucose) AND
  (recommendations_target_high_glucose_or_low_time_in_range OR
  recommends_more_correction_doses OR reminds_patient_to_bolus)

- Message treating highs only:
  (NOT recommendations_target_low_glucose) AND
  (recommendations_target_high_glucose_or_low_time_in_range
  OR recommends_more_correction_doses OR reminds_patient_to_bolus)

- Message treating lows only:
  (recommendations_target_low_glucose) AND NOT
  (recommendations_target_high_glucose_or_low_time_in_range
  OR recommends_more_correction_doses OR reminds_patient_to_bolus)

- Other Message: Messages falling into none of the categories above.

**Baseline representations.** To extract features from the raw messages, we generate 728-dimensional text embeddings with PaLM 2 (Anil et al., 2023). Then we learn a clustering of those embeddings using K-means on the training data with K=4 to match the cardinality of the clinician-informed action representation. Messages in the other datasets are mapped to the closest cluster centroid.

# Appendix D. State representations

## D.1. Clinical features and representations

Full list of clinical state features:

g_7dr: Mean glucose last 7 days
very_low_7dr: Prop. CGM readings < 54 mg/dL last 7 days
low_7dr: Prop. CGM readings < 70 mg/dL last 7 days
in_range_7dr: Prop. CGM readings 70-180 mg/dL last 7 days
high_7dr: Prop. CGM readings > 180 mg/dL last 7 days
very_high_7dr: Prop. CGM readings > 250 mg/dL last 7 days
gri_7dr: Glycemia Risk Index (Klonoff et al., 2023) last 7 days
g_14dr: Mean glucose last 14 days
very_low_14dr: Prop. CGM readings < 54 mg/dL last 14 days
low_14dr: Prop. CGM readings < 70 mg/dL last 14 days
in_range_14dr: Prop. CGM readings 70-180 mg/dL last 14 days
high_14dr: Prop. CGM readings > 180 mg/dL last 14 days
very_high_14dr: Prop. CGM readings > 250 mg/dL last 14 days
gri_14dr: Glycemia Risk Index (Klonoff et al., 2023) last 14 days
night_very_low_7dr: Prop. CGM readings < 54 mg/dL last 7 days at night time (11pm-5am)
night_low_7dr: Prop. CGM readings < 70 mg/dL last 7 days at night time (11pm-5am)
night_high_7dr: Prop. CGM readings > 180 mg/dL last 7 days at night time (11pm-5am)
night_very_high_7dr: Prop. CGM readings > 250 mg/dL last 7 days at night time (11pm-5am)
day_very_low_7dr: Prop. CGM readings < 54 mg/dL last 7 days at day time (6am-10pm)
day_low_7dr: Prop. CGM readings < 70 mg/dL last 7 days at day time (6am-10pm)
day_high_7dr: Prop. CGM readings > 180 mg/dL last 7 days at day time (6am-10pm)
day_very_high_7dr: Prop. CGM readings > 250 mg/dL last 7 days at day time (6am-10pm)
time_worn_7dr: Prop. of time with CGM readings last 7 days
night_worn_7dr: Prop. of time with CGM readings last 7 days at night time (11pm-5am)
day_worn_7dr: Prop. of time with CGM readings last 7 days at day time (6am-10pm)
gri_7dr_7d_delta: Difference in gri_7d between today and 7 days ago
very_low_7dr_7d_delta: Difference in very_low_7dr between today and 7 days ago
low_7dr_7d_delta: Difference in low_7dr between today and 7 days ago

in_range_7dr_7d_delta: Difference in in_range_7dr between today and 7 days ago
very_high_7dr_7d_delta: Difference in very_high_7dr between today and 7 days ago
night_very_low_7dr_7d_delta: Difference in night_very_low_7dr between today and 7 days ago
night_low_7dr_7d_delta: Difference in night_low_7dr between today and 7 days ago
night_high_7dr_7d_delta: Difference in night_high_7dr between today and 7 days ago
sexF: Indicator equal to 1 for female patients, 0 otherwise.
public_insurance: Indicator equal to 1 for publicly insured patients, 0 otherwise
english_primary_language: Indicator equal to 1 for patients with English as their preferred language, 0 otherwise
pop_pilot: Indicator equal to 1 for patients enrolled in the 4T Pilot study (Prahalad et al., 2022), 0 otherwise
pop_4T_1: Indicator equal to 1 for patients enrolled in the 4T Study 1 (Prahalad et al., 2024), 0 otherwise
pop_TIPS: Indicator equal to 1 for patients enrolled in the TIPS Study (Scheinker et al., 2022), 0 otherwise
age: Age of patient
months_since_onset: Months since onset of type 1 diabetes
using_pump: Indicator equal to 1 for patients using an insulin pump, 0 otherwise
using_aid: Indicator equal to 1 for patients using automated insulin delivery (closed loop), 0 otherwise
days_since_msg: Days since the last time the patient received a message
large_tir_drop: Indicator equal to 1 for patients with in_range_7dr_7d_delta<-0.15, 0 otherwise (used by clinicians for risk stratification)
low_tir: Indicator equal to 1 for patients with in_range_7dr<0.65, 0 otherwise (used by clinicians for risk stratification)
lows: Indicator equal to 1 for patients with lows_7dr>0.04, 0 otherwise (used by clinicians for risk stratification
very_lows: Indicator equal to 1 for patients with very_lows_7dr>0.01, 0 otherwise (used by clinicians for risk stratification

We create three different representations using these features:

1. Full: includes all features listed above

2. Subset selected with ML: Top features based on XGBoost variable importance (SHAP) scores when training a model to predicting observed rewards in the training data. Includes large_tir_drop, in_range_7dr_7d_delta, in_range_14dr, in_range_7dr, low_7dr, using_pump, using_aid, time_worn_7dr, day_worn_7dr, day_low_7dr, night_high_7dr_7d_delta, g_7dr, months_since_onset, gri_7dr_7d_delta

3. TIDE subset selected by clinicians: Features shown to clinicians when they review patients in TIDE. Includes very_low_7dr, low_7dr, in_range_7dr, g_7dr, using_pump, in_range_7dr_7d_delta, large_tir_drop, low_tir, lows, very_lows, pop_4T_1, pop_4T_2, pop_TIPS

### D.2. Baseline representations

**TS2Vec.** We generate low-dimensional representations of the CGM traces using TS2Vec (Yue et al., 2021). We train the encoder on the training data, feeding in two weeks of CGM readings (4,032-dimensional vectors). We use the default hyperparameters in the code at https://github.com/zhihanyue/ts2vec/, with n_epochs=1 and output_dims=8. We also tested training for 2 and 3 epochs and found very similar results to n_epochs=1 when using the representations to fit CATE estimators.

**UMAP.** We project two weeks of CGM readings into low-dimensional representations using UMAP as implemented at https://github.com/lmcinnes/umap. We tested projecting down to between 1 and 10 dimensions and then picked the number of dimensions that performed best at predicting observed rewards in the training data. We ended up picking n_components=4 for the projections used for CATE estimation.

## Appendix E. Adjusting for confounding

In our data, there is confounding between the actions and rewards. To account for confounding, we adjust for a set of *control covariates* $X_{it}^c = \left\{ \gamma^c(X_{it}^g), X_i^{d'} \right\}$. This is the representation of the patient state that clinicians see when reviewing patients, which includes a low-dimensional projection of the CGM data $\gamma^c(X_{it}^g)$ and a subset of the demographics $X_{it}^{d'} \subset X_{it}^d$. We make the following (commonly used) assumptions to adjust for confounding and perform doubly robust policy evaluation.

**Assumption 2** Consistency and stable unit treatment value. *The potential outcomes (Imbens and Rubin, 2015) for each patient $i$ at time $t$ under treatment $M_{it} = m$ are the same as the observed outcomes if they actually received treatment $m$, and these potential outcomes depend only on the treatment $M_{it}$ assigned to that patient, not on the treatments assigned to other patients. Formally, $r_{it}(m) = r_{it}$ if $M_{it} = m$.*

**Assumption 3** Conditional ignorability. *Given the control covariates for patient $i$ at time $t$, $X_{it}^c$, the observed actions (treatment messages) $M_{it}$ are independent of the potential reward $r_{it}(m)$ for all possible actions $m \in \mathcal{M}$. Formally, $r_{it}(m) \perp M_{it} \mid X_{it}^c$, $\forall m \in \mathcal{M}$, where $\mathcal{M}$ is the set of all possible actions (messages), and $r_{it}(m)$ denotes the potential reward for patient $i$ at time $t$ under action $m$.*

Under these assumptions, we fit outcome models $\hat{r}(X^c, a)$ predicting the rewards under each action conditional on a vector of control covariates $X^c$, and a model predicting the reward under the control action $\hat{r}(X^c, 0)$. We also fit models $\hat{e}(X^c, a)$, which estimates the propensity scores (probabilities of each action conditional on the control covariates). All of the nuisance models are trained with AutoML in FLAML (Wang et al., 2021).

Now suppose that a patient $i$ is observed at day $t$ in an evaluation dataset $\mathcal{E}$. We define the following doubly robust score for each action $a$; this is an estimate of the treatment effect of action $a$ for patient $i$ at day $t$:

$$
\begin{aligned}
\hat{\Gamma}_{it}(a) = {} & (\hat{r}(X_{it}^c, a) - \hat{r}(X_{it}^c, 0)) \\
& + (r_{it} - \hat{r}(X_{it}^c, a_{it})) \left( \frac{I_{a_{it}=a}}{\hat{e}(X_{it}^c, a)} - \frac{I_{a_{it}=0}}{\hat{e}(X_{it}^c, 0)} \right)
\end{aligned}
$$

The doubly robust score corrects for unmeasured confounding in the following way: if at most one of the reward model $\hat{r}$ or the propensity model $\hat{e}$ is misspecified, the doubly robust score will still be consistent for the true treatment effect. In practice, we cannot completely rule out the possible simultaneous misspecification of both models. However, as noted above, we take advantage of control covariates that capture all information available to clinicians at the time of choosing an action, helping mitigate bias due to confounding.

## Appendix F. Additional policy evaluation results

In plotting our evaluation results for a given policy, we visualize the *targeting operator characteristic (TOC) curve* (Sverdrup et al., 2023; Yadlowsky et al., 2021): this is a plot of $\widehat{\text{ATT}}_K(\pi)$ against $K/N$; when $K = N$, we obtain the average treatment effect (ATE) of the policy. We generate confidence intervals along the TOC by bootstrapping patients in the evaluation data. A common evaluation metric used to evaluate targeting policies is the Area Under the TOC curve (AUTOC), which is the area between the ATT and ATE integrated over K from 0 to 1. See Figure 6 for reference. In this section, we report the ATT@25% and AUTOC values for each state and action representation across CATE estimators.
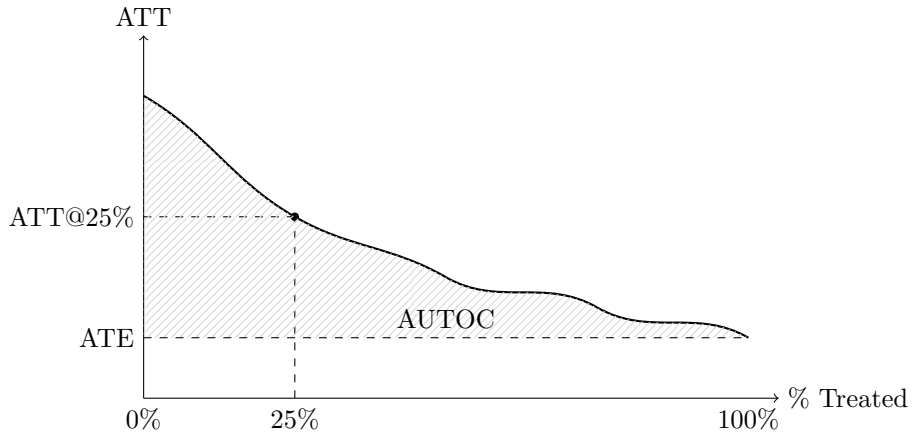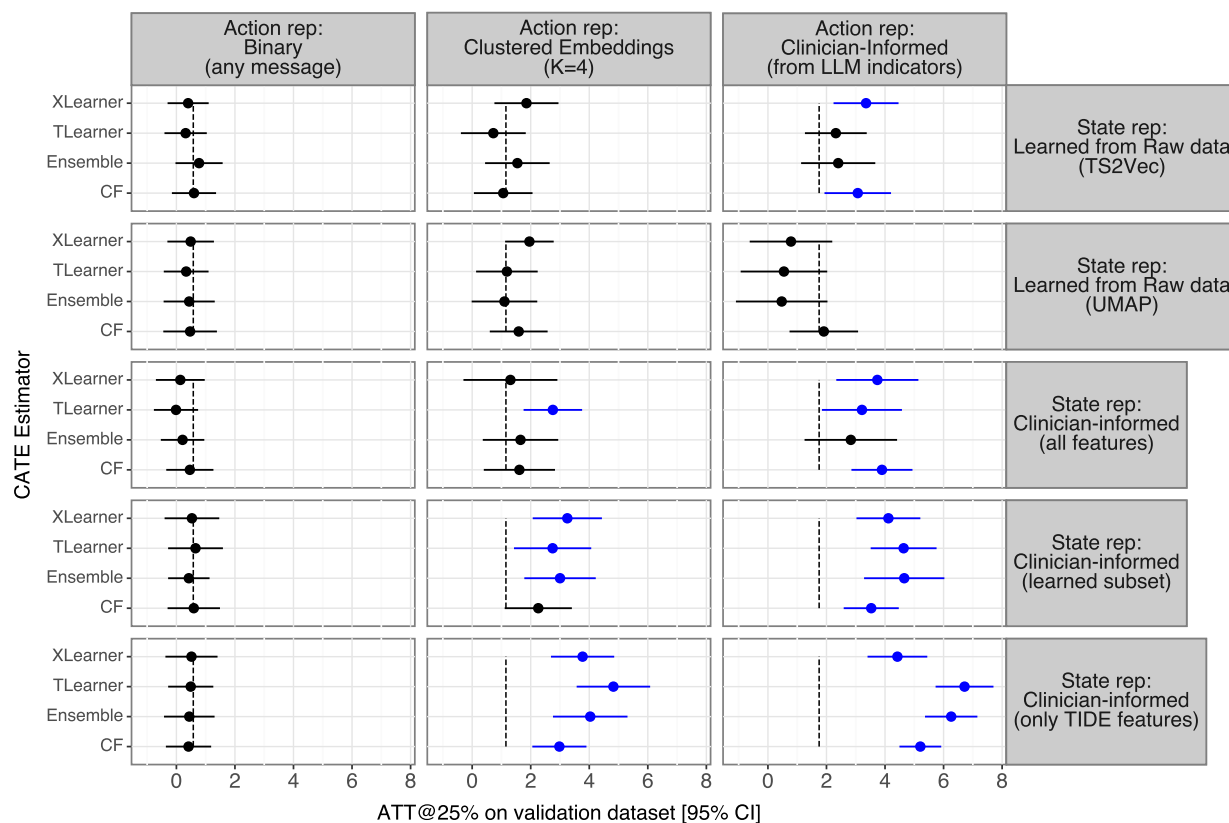


Figure 6: Illustrative TOC curve.

Figure 7: ATT@25% for each state (rows) and action (columns) representation across CATE estimators. Results with the S-Learner and DR Forest CATE estimators are excluded because they never outperformed random targeting.
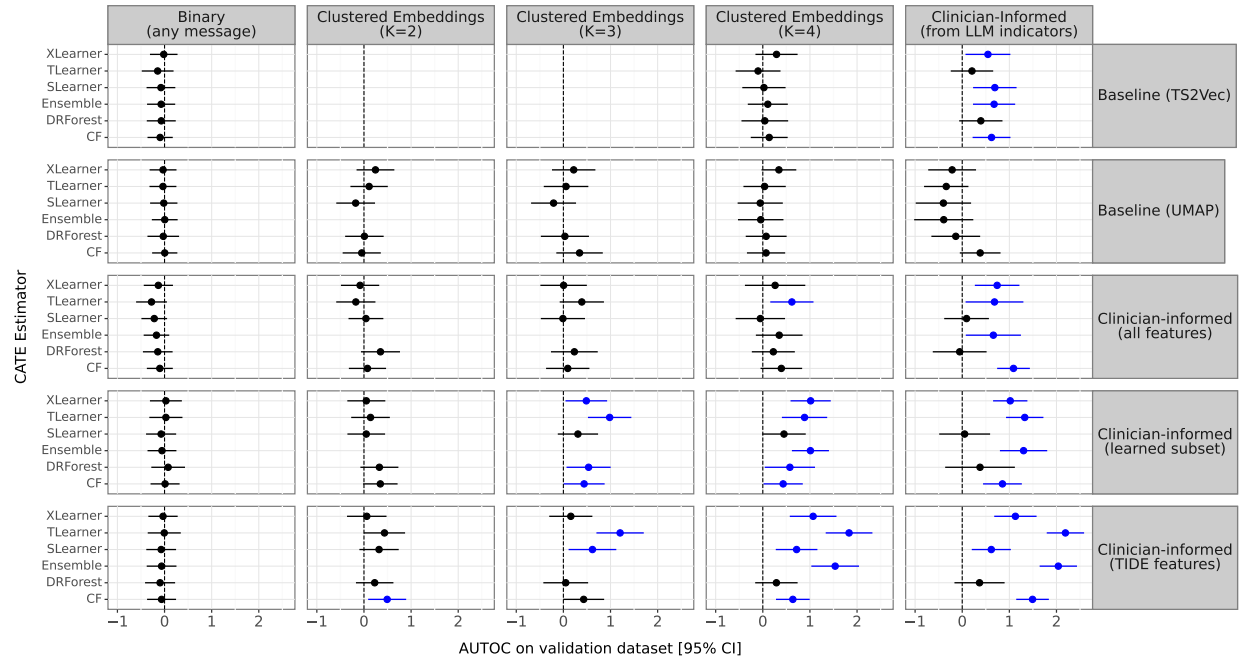
Figure 8: AUTOCs for each state (rows) and action (columns) representation across CATE estimators
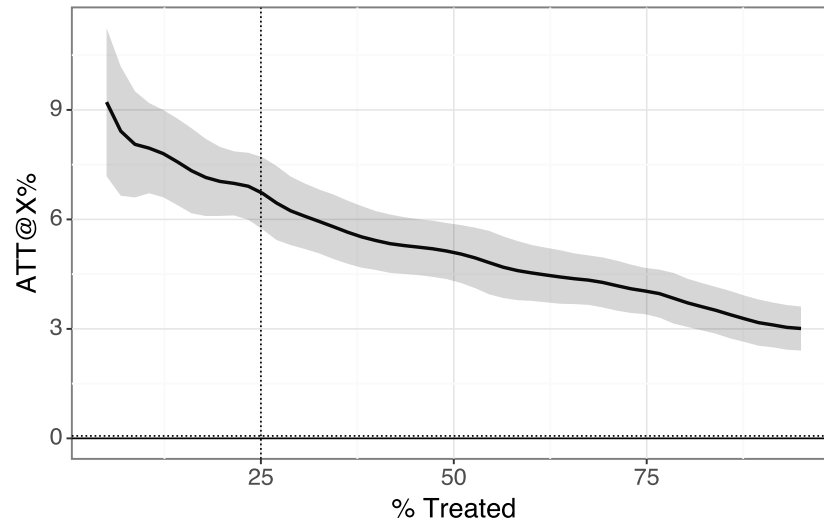


Figure 9: Example TOC curve for the top performing policy on the validation set:
State representation = Clinician-informed (TIDE)
Action representation = Clinician-informed
CATE estimator = TLearner.

## Appendix G. Sensitivity analysis to adding longer patient history in control covariates when evaluating policies

In order to understand if we are getting biased policy evaluation results by only using two weeks of CGM data to construct our control covariates, we also evaluated the best-performing state representation with control covariates that include longer histories (up to 4 weeks), including indicators for messages in previous weeks. The ATT@25% when additional weeks of history are included in control covariates are shown in Figure 10. We see that ATT@25% remains much higher than the ATEs (dashed vertical lines) when we are using clinician-informed action representations (right-most column). There is a slight drop-off in estimated performance as we add history to the control covariates though. This could be noise, but future work might include more sensitivity analyses, and also adding more weeks of history to the state representations.
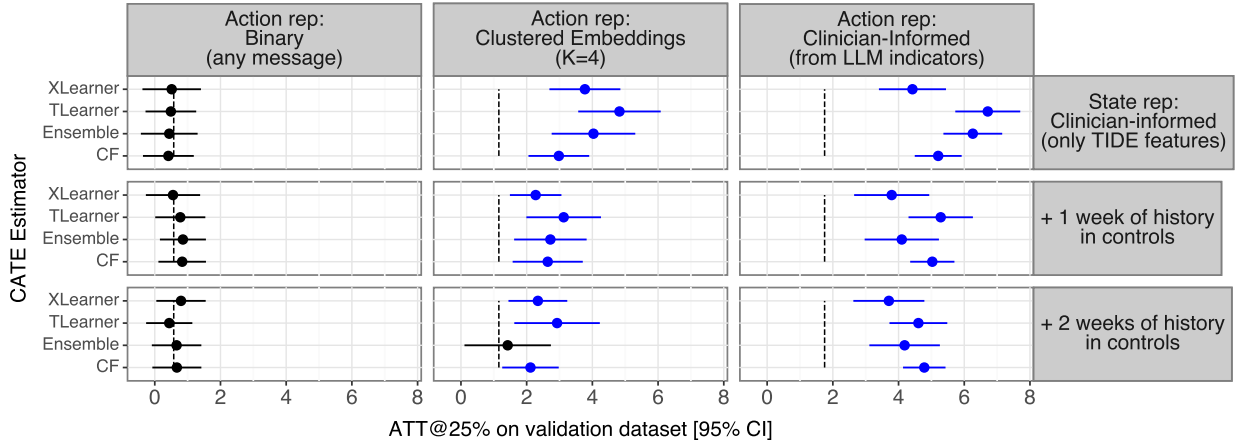


Figure 10: ATT@25% when including more history in the control covariates.

## Appendix H. Additional results and proofs

**Proposition 3** *Fix $K$, and let $\pi^* = \pi(\cdot|\tau, K)$. Let $\pi$ be any other policy that targets at most $K$ patients (i.e., selects at most $K$ patients to receive a non-control action). Then $\mathsf{ATT}_K(\pi^*) \geq \mathsf{ATT}_K(\pi)$.*

*Proof of Proposition 3.* Fix a state vector $(s_1, \ldots, s_N)$ (sampled i.i.d. from the superpopulation). For each patient $i$, let $a_i^* = \arg\max_{\tilde{a} \in \mathcal{A}} \tau(s_i, \tilde{a})$, i.e., the optimal action for that patient. Let $\alpha_i = \tau(s_i, a_i^*)$, and let $(\alpha_{(1)}, \ldots, \alpha_{(N)})$ be the same values sorted in descending order. Observe that since $\pi^*$ selects the $K$ highest ranked patients according to $\alpha_i$, it follows that for any subset $I_K \subset \{1, \ldots, N\}$ of cardinality $K$, there holds:

$$\sum_{i=1}^{N} \tau(s_i, \pi_i^*(s)) = \sum_{i=1}^{K} \alpha_{(i)} \geq \sum_{i \in I_K} \alpha_i.$$

Now let $I_K$ be the set of (at most $K$) patients who receive non-control actions under another feasible policy $\pi$. The preceding allows us to conclude that:

$$\sum_{i=1}^{N} \tau(s_i, \pi_i^*(s)) \geq \sum_{i=1}^{N} \tau(s_i, \pi_i(s)),$$

since $\alpha_i \geq \tau(s_i, \pi_i(s))$ for all $i$. Dividing by $K$ and taking expectations over $s$ concludes the proof. □

*Proof of Theorem 2.* First, we use the Skorohod representation theorem to construct a probability space on which $\sup_{s \in \mathcal{S}, a \in \mathcal{A}} |\tau(s, a) - \hat{\tau}_N(s, a)| \to 0$ almost surely. Next, suppose that $s = (s_1, \ldots, s_N)$ is an i.i.d. sample of $N$ patients from the superpopulation. Note that $\pi_N$ applies optimal treatments to the top $K$ ranked patients according to $\hat{\tau}_N$. Therefore:

$$\frac{1}{K_N} \sum_{i=1}^{N} \hat{\tau}_N(s_i, \pi_{N,i}(s)) \geq \frac{1}{K_N} \sum_{i=1}^{N} \hat{\tau}_N(s_i, \pi_{N,i}^*(s)).$$

For each $i$ and $N$, define $\Delta_{i,N} = \hat{\tau}_N(s_i, \pi_{N,i}(s)) - \tau(s_i, \pi_{N,i}(s))$, and define $\Delta_{i,N}^* = \hat{\tau}_N(s_i, \pi_{N,i}^*(s)) - \tau(s_i, \pi_{N,i}^*(s))$. Then the preceding inequality becomes:

$$\frac{1}{K_N} \sum_{i=1}^{N} (\tau(s_i, \pi_{N,i}(s)) - \tau(s_i, \pi_{N,i}^*(s))) + \frac{1}{K_N} \sum_{i=1}^{N} (\Delta_{i,N} - \Delta_{i,N}^*) \geq 0.$$

Note the second summation on the left hand side converges to zero almost surely. On the other hand, because $\pi_N^*$ is the oracle optimal policy, we also have:

$$\frac{1}{K_N} \sum_{i=1}^{N} (\tau(s_i, \pi_{N,i}(s)) - \tau(s_i, \pi_{N,i}^*(s))) \leq 0.$$

Therefore we conclude:

$$\lim_{N \to \infty} \frac{1}{K_N} \sum_{i=1}^{N} (\tau(s_i, \pi_{N,i}(s)) - \tau(s_i, \pi_{N,i}^*(s))) = 0$$

almost surely.

Since treatment effects are bounded, we can apply the bounded convergence theorem to conclude that $\mathsf{ATT}_{K_N}(\pi_N) - \mathsf{ATT}_{K_N}(\pi_N^*) \to 0$, as required. □