Preventive Care Now or Pay Later? A Personalized Medicine Approach for Healthcare Management

Tianjian Guo
University of Texas at Austin, tianjian.guo@mccombs.utexas.edu

Indranil Bardhan
UT Austin, indranil.bardhan@mccombs.utexas.edu

Wen Wen
University of Texas at Austin, wen.wen@mccombs.utexas.edu

Preventive Care Now or Pay Later? A Personalized Medicine Approach for Healthcare Management

Completed Research Paper

Tianjian Guo[1], Indranil Bardhan[2], Wen Wen[3]
McCombs School of Business
UT Austin
[1]tianjian.guo@mccombs.utexas.edu,
[2]indranil.bardhan@mccombs.utexas.edu
[3]wen.wen@mccombs.utexas.edu

Abstract

Preventive care, including routine check-ups and screenings, aims to avert severe illnesses and champion health equity. However, existing recommendations often neglect the need for personalization and patient convenience, resulting in significant underutilization. This study proposes a multi-objective reinforcement learning framework tailored for optimizing patient-centric diabetes-related preventive care, balancing patient convenience and treatment cost. Based on the electronic health records from over 500,000 patients, we show that the optimal preventive care rate could be fourfold the current rate. Our framework could cut annual patient costs by 1.1%, with more pronounced savings for groups such as young adults, the elderly, males, and diabetic patients. We further validate this method with the Michigan Model for Diabetes, a well-established diabetes progression simulation software. Our study contributes to the design of healthcare decision support systems, spotlighting the significance of personalization and the pressing need for value-based incentives to enhance preventive care adoption among targeted patient groups.

Keywords: Multi-objective Reinforcement Learning, Personalized Medicine, Precision Medicine, Value-based Care, Preventive Care Management, Chronic Conditions Management

Introduction

The U.S. spends over $4.3 trillion annually on healthcare, surpassing every other country worldwide. However, there are still significant gaps in care, health equity, and population health outcomes (Centers for Medicare & Medicaid Services, 2021). Effective utilization of preventive care could be a possible solution to address these issues (Agency for Healthcare Research and Quality, 2019). Encouraging patients to undergo regular check-ups and screenings can help healthcare providers identify and address potential health issues before they become more severe and require more expensive treatments. This approach can help patients avoid unnecessary suffering and lead to significant cost savings for both patients and healthcare providers. However, many current recommendations for preventive care are made at the population level rather than being personalized to meet the needs of individual patients. For instance, patients above a certain age may be advised to undergo certain screenings or tests, regardless of their personal health history or risk factors (Centers for Medicare & Medicaid Services, 2022). This generic approach might not sufficiently cater to the specific needs of patients, leading to an inefficient allocation of healthcare resources, especially as the industry shifts towards a value-based care model where the value of healthcare services is measured by the outcomes experienced by patients (Porter, 2010). Furthermore, several deterrents, ranging from logistical inconvenience and medical apprehension to financial worries and social determinants of health, might discourage patients from seeking regular check-ups (Kannan, 2015).

A possible approach to addressing such concerns is to develop patient-centric, personalized recommendations for preventive care based on individual patient characteristics. Healthcare professionals have attempted to deliver tailored treatments by considering patients' unique genetic makeup, lifestyle, and environmental factors through an approach called precision or personalized medicine (Chen et al., 2021). While methods such as dynamic treatment regimens and reinforcement learning (RL) algorithms for personalized medicine have been explored within this domain, the extant literature has largely focused on optimizing a single clinical objective function under specific scenarios (Fox et al., 2020). Few studies have developed methods that prioritize the patient's viewpoint and navigate the delicate balance between multiple, often conflicting, objectives that are paramount from a patient's perspective. For example, while suggestions aimed at reducing treatment costs can be economical, they may not always result in the desired improvement in a patient's quality of life.

In this research, we develop a novel, multi-objective reinforcement learning (RL) approach to provide personalized recommendations for preventive care from a patient's perspective. In the context of diabetes, our framework seeks to recommend a personalized treatment plan of preventive care procedures for managing diabetes, with the goal of optimizing both patient convenience (measured by the annual count of scheduled, routine clinical visits) and annual cost of treatment. Both objectives are essential. Improving patient convenience, such as preventing unnecessary or redundant testing, ensures greater patient compliance and treatment efficacy (Ayabakan et al., 2017). Meanwhile, optimizing the total cost of treatment can lead to significant financial savings for patients. These objectives have practical implications for external stakeholders. Reducing a patient's routine clinical visits may decrease healthcare utilization in the short run. However, in the long run, it could lead to higher costs due to a lack of timely treatments, longer disease progression without treatment which may lead to serious health conditions. Similarly, reducing costs is important to other parties in the healthcare system, such as insurance companies and employers, since insured patients usually only pay a portion of the total cost. Further, instead of providing an overall recommendation, our framework accounts for individual-level differences and enables personalization of diabetes-related preventive care for individual patients. Hence, the focus of our research is to balance the tradeoffs between short-term patient convenience versus long-term cost savings by developing an optimal, personalized recommendation policy that takes into account the heterogeneity across patients based on demographic characteristics and disease progression based on data available in their medical records.

We formulate this problem as a Multi-Objective Markov Decision Process (MDP) and devise an innovative multi-objective, offline RL method to solve the MDP. Due to the sensitive nature of healthcare decisions, testing our proposed approach in an online, live setting with actual patients poses significant challenges and is difficult to implement. Hence, we rely on batch or offline RL techniques for effective training based on a unique longitudinal EHR dataset obtained from a regional health information exchange (HIE) in Central Texas. The proposed method employs a novel double-deep Q network (D2QN) to identify the optimal preventive procedures based on personalized, patient-specific characteristics (such as age and gender) and clinical encounters (such as diagnosis and treatment procedures). The D2QN is trained to optimize over the entire range of possible linear preference functions of the two objectives, and thus, generate different recommendation policies for each unique preference. To evaluate the effectiveness of our RL approach, we extend existing single objective-oriented, off-policy evaluation (OPE) methods to multi-objective settings, which enable us to evaluate multiple recommendation policies for both objectives.

Our multi-objective RL framework reveals several useful insights. First, our results confirm the presence of trade-offs between optimizing for patient convenience versus patient treatment costs. For example, recommendations that emphasize minimization of overall treatment costs are associated with more frequent routine clinical visits. Based on real-world patient behavior data, we observe that rather than minimizing overall treatment costs, patients are more likely to optimize for their convenience by reducing their routine clinical visits. Second, the optimal number of preventive care procedures suggested by our personalized recommendation approach is four times higher, on average, than the actual number of preventive care procedures, and such differences are particularly salient for elderly patients (i.e., above the age of 65 years), males, and diabetic patients. This highlights the importance of providing incentives, especially to these patient subgroups, to encourage their utilization of preventive care.

Our multi-objective OPE method suggests that if patients follow our recommendations for preventive care procedures, instead of behaving as observed, they will need to increase their primary care visits by 7.34%, which results in an overall 1.1% decrease in treatment costs. For male (vs. female) patients and diabetic (vs. non-diabetic) patients, they would need to increase their primary care visits by 8.69% (6.30%) and 12.31% (6.99%), while achieving a reduction in treatment costs of 1.34% (0.92%) and 2.47% (1.00%), based on our policy recommendation. The benefits of our model recommendations are strongest for people under the age of 45 years in terms of overall reduction in treatment costs. These patients would need to increase their routine (primary) care visits by 5.5% which, in turn, would reduce treatment costs by 1.23%. This result contrasts with people between the ages of 45 and 65 years, for whom our models recommend an increase in preventive care visits of 6.4% and a reduction in treatment costs of 0.78%. On the other hand, for patients above 65 years, our personalized care model recommends an increase in preventive care visits by 13% and corresponding reduction in treatment costs by 1.14%.

We further extend our proposed recommendation framework on a well-known diabetes progression simulation model, the Michigan Model for Diabetes (Ye et al., 2015). The results from the simulation validate our insights regarding the tradeoffs between patient convenience and treatment costs. They also highlight a comparable trade-off between patient quality of life and convenience. Furthermore, the simulation underscores that a reduction in treatment costs necessitates a surge in patient medication expenses, potentially neutralizing overall savings in care-related costs. These insights offer policymakers a deeper understanding of the intricate interplay among various medical outcomes, emphasizing the significance of regulating medication costs such as recent efforts by the Biden administration to reduce drug prices for Medicare patients.

Our study also contributes to the design of healthcare information technologies (e.g., Agarwal et al., 2010). While a body of the information systems literature has sought to design novel predictive analytics to predict patient outcomes or diagnostics (Bardhan et al., 2015; Ben-Assuli & Padman, 2020), very little research has offered personalized recommendations for treatment decisions based on prescriptive analytic models that optimizes patient-centric objectives. Closest to our research are Chen et al. (2021) and Zhou et al. (2023), both of which aim to offer personalized healthcare treatment or interventions using a prescriptive approach. However, Chen et al. (2021) investigates specific treatment plans for breast cancer, while Zhou et al. (2023) develop personalized recommendations to help patients discover fitness-related interventions. In contrast, our framework seeks to design personalized recommendations for preventive care that optimizes *both* patient convenience (which may improve patient compliance) and cost of treatment. Our proposed framework highlights the importance and significant potential of patient-centered, health IT to enhance the quality of patient care while benefiting healthcare providers and payers. The proposed prescriptive approach can enhance patient engagement with preventive care initiatives by offering personalized recommendations and estimating their impact on overall costs. This allows patients to understand the importance of preventive care procedures tailored to their unique health context, ultimately improving their overall healthcare experience and outcomes. Furthermore, it provides a foundation for the design of health IT, based on mobile health data and Internet of things (IoT) enabled devices, which can measure patient compliance with care treatment plans and recommend changes as needed based on patient health status.

Background

Preventive Care

The U.S. Centers for Medicare & Medicaid Services (Centers for Medicare & Medicaid Services, 2022) defines preventive care as "routine health care that includes screenings, check-ups, and patient counseling to prevent illnesses, disease, or other health problems." Such care procedures are essential to maintaining good health and preventing severe health risk. Currently, preventive care procedures are recommended in the United States based on broader population-level characteristics such as patient age and gender. For instance, the U.S. Preventive Services Task Force (USPSTF) recommends regular colorectal cancer screening for individuals between 50 to 75 years (Bibbins-Domingo et al., 2017).

The benefits of preventive care are clear from the perspective of policy makers, healthcare professionals, and insurance companies. Failure to utilize preventive care procedures and participation in risky health behavior are significant contributors to morbidity, health disparities, medical care costs, and mortality (National Center for Health, 2015). According to the Institute of Medicine, missed prevention opportunities cost U.S. $55 billion yearly, approximately 30 cents of every healthcare dollar (Institute of Medicine & Committee on the Learning Health Care System in America, 2013). In 2014, the Centers for Disease Control and Prevention (CDC) estimated that at least one-third of all deaths in the U.S., that can be attributed to the top five causes, are preventable by reducing the prevalence of known risk factors (Garcia et al., 2016).

However, as of 2015, only 8 percent of U.S. adults aged thirty-five years and older had received all appropriate clinical preventive services recommended for their age (Borsky et al., 2018). While there are many reasons for such a low utilization rate, recent studies examining patient preferences and behavior have shed some light on the issue. For instance, patients and healthcare providers have discordant preferences about different aspects of treatments, which may affect their perceptions of the benefits from healthcare services (Engel et al., 2021). When evaluating healthcare options, patients not only value their physiological and financial well-being, but also their overall experience and convenience (Higgins et al., 2014). For example, the length and frequency of intravenous (IV) iron infusions has decreased their perceived utility among patients needing IV iron infusions (Takeshima et al., 2023). Furthermore, because preventive care recommendations are made based on broad categorizations such as age and gender, it is often difficult for patients to understand specific benefits for their individual well-being. Recent studies have suggested that personalized recommendations of preventive care services can lead to increased utilization of preventive care (Taksler et al., 2021). In other words, if patients are offered personalized recommendations of preventive care procedures based on a range of objectives that patients value, it can significantly increase their utilization of preventive care.

Personalized Medicine and RL in Healthcare

The medical community has long acknowledged that specific characteristics of diseases and responses to treatments are frequently clustered within individuals, families, and specific population groups. However, the prior literature has mostly deployed a standardized approach for diagnosis and treatment of specific diseases. Due to the abundant digital data now accessible through electronic health records (EHRs) and the emergent integration and availability of genomic data, it has become feasible to offer personalized medicine solutions to administer efficient care that is tailored to patients' unique health needs, preferences, and values (Abul-Husn & Kenny, 2019). During the past decade, the advent of artificial intelligence (AI) enabled solutions has made it possible to create intelligent systems capable of acquiring knowledge regarding clinical treatments and discovering novel insights from the vast quantity of healthcare data (Coronato et al., 2020). Specifically, the medical community has shown greater interest in reinforcement learning due to its potential for creating personalized treatments that align with the broader vision of personalized medicine.

Reinforcement learning is one of the three distinct subfields of machine learning that relies on goal-directed learning, unlike supervised and unsupervised learning. This method involves interactions with the surrounding environment and observing changes to facilitate learning. Its objective is to maximize numerical rewards by learning how to better map situations to actions (Sutton & Barto, 2018). In healthcare, RL has been used to optimize the dosage of insulin administered to diabetes patients for blood sugar control (Fox et al., 2020), prevention of sepsis in intensive care units (ICU) (Raghu et al., 2017), or encouraging more physical activities for patients with chronic conditions (Yom-Tov et al., 2017). Additionally, multi-objective RL, which aims to maximize more than one numerical reward function, has also been applied to healthcare in research contexts where multiple objectives are important. Examples of such research include optimizing the dosage of radiotherapy treatments that balance the need to kill cancerous cells with the adverse effects of radiation on normal cells (Shiranthika et al., 2022).

Overall, although RL has emerged as a promising approach for personalized medicine, the current research has mainly focused on helping healthcare professionals to find the optimal action for clinical outcomes in specific scenarios such as sepsis prevention and cancer treatment.

Research Gap

While it is important to provide personalized preventive care to patients to achieve reduction in healthcare costs and improve patient-centric health outcomes, there has been limited research on prescriptive analytics approaches to address this issue. Although a growing body of literature incorporates prescriptive analytics into the design of optimal treatment plans (Chen et al., 2021; Zhou et al., 2023), prior studies do not provide "optimal" prescriptive directives for general health management. Our study seeks to bridge this important gap and proposes a personalized recommendation approach for preventive care that incorporates multiple objectives of interest for patients. While our empirical context focuses on preventive procedures for Type II diabetes care, the design principles of our framework may be applied to patient health management in other chronic disease settings where there is a need to optimize care utilization and quality of life considerations with overall treatment costs.

A Multi-objective RL framework

Framework Overview

Our research introduces a novel framework using multi-objective reinforcement learning to provide personalized recommendations for preventive care associated with Type II diabetes treatment. Our model focuses on optimizing both patient convenience, measured by the annual count of scheduled, routine care visits that patients incur, and total treatment costs. Both objectives are critical for patients, since greater patient convenience ensures treatment efficacy (adherence) and compliance, while minimizing treatment costs leads to significant financial savings (since even insured patients typically incur some portion of the overall costs in terms of co-pays and deductibles). Further, these objectives are crucial for external stakeholders as reducing patient visits can optimize healthcare utilization, while lowering costs benefits payers such as insurance companies and employers.

We construct this problem as a Multi-Objective Markov Decision Process (MOMDP) with an infinite horizon (White, 1982). We adopt a multi-objective RL approach based on the existing literature to generate multiple recommendation policies, each with different preferences for emphasizing convenience versus treatment costs. Due to the sensitive nature of healthcare, we trained the multi-objective RL agent in an offline/batch manner using a historical dataset based on a large, diverse patient population. In order to evaluate the impact of multiple recommendation policies on both objectives simultaneously, we extend existing off-policy evaluation (OPE) methods to the multi-objective setting. In the following section, we discuss our methodology in detail.

Data

We obtained our primary research data from the Integrated Care Collaboration (ICC), a regional health information exchange in Central Texas. It manages a comprehensive dataset of more than six million clinical encounters across 37 healthcare provider institutions which treated approximately 500,000 patients between the years 2015 to 2020. The dataset includes information at the patient and encounter levels. For each patient, the dataset contains patient demographic information such as age, race, and gender. The dataset includes the associated diagnosis, procedures, and payment information for each medical encounter. Patients and medical encounters are linked based on unique identifiers, allowing us to track the utilization of healthcare resources for each patient across six years. After data processing and cleaning to remove clinical encounters with no associated diagnosis, we have 3.5 million observations at the patient-year level. Each patient-year entry includes all inpatient, outpatient, and emergency room visits for a given patient during that year.

The ICC dataset does not include billing information. Instead, for each medical encounter, we approximate the cost of treatments, or cost associated with each medical encounter, using the payment received by healthcare institutions, based on data provided by CMS. Every year, for each current procedural terminology (CPT) code, CMS provides a state-wide average of payments received by healthcare institutions for procedures conducted for Medicare patients. We collected the CPT procedure codes associated with each encounter for outpatient and emergency room (ER) encounters that are primarily billed in an itemized manner. Then, for every outpatient and emergency room encounter, we approximated the treatment costs based on the encounter year and all associated procedures.

Inpatient encounters are billed differently, often in a single block based on diagnosis-related groups (DRG). CMS provides state-wide, average payments received by healthcare institutions for inpatient encounters for Medicare patients at the DRG-year level. We used the official DRG classification software published by CMS to determine the DRG for each inpatient encounter in our dataset. Based on the associated DRG and the year in which the encounter occurred, we approximated the treatment costs using the state-wide average total payment received for a given DRG and year. Figure 1 shows the average inpatient, outpatient, and ER treatment costs for patients in our dataset across time. On average, an impatient visit is more expensive than an ER or outpatient visit.

Variable	Mean	Std. Dev.
Selected Patient Variables		
Age (years)	44.66	23.04
Gender: Male	0.368	0.482
Gender: Female	0.631	0.482
Race: African American	0.18	0.384
Race: Asian	0.014	0.118
Race: Caucasian	0.792	0.405
Race: Others	0.013	0.114
Selected Patient-Year Variables		
Count of Outpatient Visits	0.978	2.542
Count of Emergency Room Visits	0.347	1.144
Count of Inpatient Visits	0.076	0.404

Table 1: Descriptive Statistics of Patient Encounter Data and SDOH Variables.

It is important to highlight that our data collection spanned the year 2020, which, as illustrated by Figure 1, exhibited a substantial decrease in clinical encounters compared to prior years due to COVID-19. Our decision to incorporate this year was deliberate, aiming to capture a more extended patient medical history. We anticipate that this inclusion will not markedly influence the generated recommendations. Hence, due to the universal impact of COVID-19, the onset of the pandemic should not bias our recommendation framework.

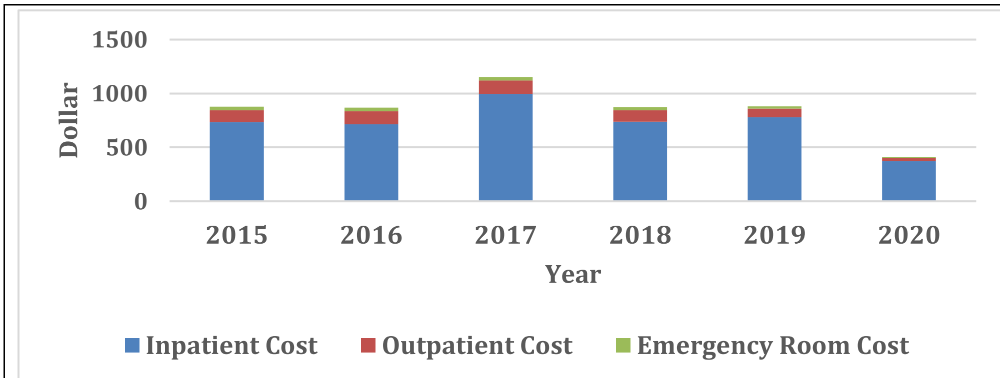


Figure 1. Average Patient Treatment Cost by Year

MOMDP Formulation

We define the problem of making preventive procedure recommendations to patients as an infinite horizon Multi-Objective Markov Decision Process (MOMDP) with linear preference functions represented by the tuple (S, A, P, r, w). In Figure 2, we provide an intuitive representation of the proposed MOMDP. S represents the state space, A represents the action space, P represents an unknown transition distribution $P(s' \mid s, a)$, r represents the reward function $r(s, a, w)$, and W represents the weight space.

State Space

We start by describing the continuous state space in our MOMDP for preventive care recommendations. For a given patient i in year t, we define state $s_{it} \in S$ as a representation of a patient i's health status in year t. Specifically, s_{it} includes two components — one fixed-length component that measures the patient's demographic information such as age, race, and gender; one variable-length component that describes all medical encounters that patient i experienced in year t. This component has variable length because patients may have varying counts of medical encounters in different years. For every encounter, the variable-length component includes information regarding the type of encounter, length of the encounter, diagnosis, and procedures associated with the encounter. Each encounter can be labeled as an outpatient, inpatient, or emergency room visit. The length of the encounter is measured in hours, which may exhibit significant variations for inpatient and emergency room visits. Diagnoses associated with each encounter are recorded in their corresponding ICD-10-CM codes, and medical procedures conducted during each encounter are recorded based on their corresponding ICD-10-PCS or CPT code.

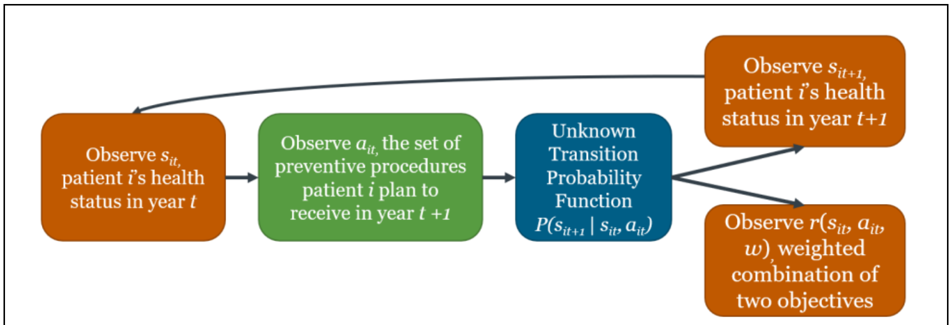


Figure 2. Overview of the Multi-Objective Markov Decision Process

Action Space

The action space includes eight possible combinations of three distinct preventive procedures associated with Type II diabetes care, namely A1C blood sugar tests, lipid panel exams, and urine exams. As such, action a_{it} represents the preventive procedures that patient i plans to receive in year $t+1$ based on her health status in year t, contained in s_{it}. We examined the list of preventive procedures recommended by medical professionals for the screening, prevention, and treatment of Type II diabetes and selected these procedures due to their frequent appearance in the ICC dataset and their prevalence in the literature (Vijan et al., 1997).

Reward and Weight Space

For a given state and action pair (s_{it} and a_{it}) that represents patient i's health status in year t and actions that the patient plans to take in year $t+1$, we construct a reward function as the weighted combination of two distinct objectives, which we are interested in minimizing:

$$r(s_{it}, a_{it}, w) = w_1 * obj_{it+1,\, Convenience} + w_2 * obj_{it+1,\, Cost} = r_{it+1}$$

We define $obj_{it+1,\, Convenience}$, an indicator of the level of inconvenience that patient i experiences in year $t+1$, as the count of scheduled, routine care encounters that patient i experiences in year $t+1$ due to action a_{it} taken in the state s_{it}. We consider all outpatient encounters as scheduled routine encounters, as clinical encounters in outpatient care settings are mostly scheduled for non-emergency purposes. Drawing from the prior literature (Engel et al., 2021), we assume that patients typically have a vested interest in optimizing convenience, which translates to minimizing the count of routine care encounters that consists of outpatient clinic visits. However, the value placed on convenience can differ from one individual to another.

Furthermore, we define $obj_{it+1,\, Cost}$, an indicator of the treatment cost for patient i during year $t+1$, as the total payment, received by healthcare providers for all medical encounters that patient i experiences during year $t+1$ as a result of taking action a_{it} in the state s_{it}. We assume that patients are also interested in minimizing their cost of treatment. Similarly, the degree of this concern may vary across individuals.

We chose the two objectives based on their significance to both patients and payors, and their association with preventive care procedures. Regarding the first objective (i.e., convenience), since preventive procedures such as A1C tests are often conducted during routine care visits for diabetes, a recommendation for more preventive care procedures will lead to more routine visits. Furthermore, if undergoing more preventive care helps to alert patients and their healthcare providers about potential health issues early on, this could also lead to more routine care visits. With respect to the second objective, a large body of evidence suggests that receiving preventive care procedures is associated with a reduction in the overall cost of treatment, as they may preempt serious complications in the future (National Center for Health, 2015). Hence, we believe that making an optimal recommendation requires balancing the tradeoffs between patient convenience and cost. For example, recommending more preventive care procedures can lead to a reduction in the overall cost of treatment, but patients will have to incur more routine care visits.

Due to data limitations, we are unable to distinguish between treatment costs specific to diabetes and its varied complications versus those related to other conditions. However, we posit that this should not substantially impact the recommendation generation process. The three preventive care procedures to be recommended by our proposed approach are tailored to address diabetes and its complications, thus affecting only the associated treatment costs, but not treatment costs linked to other conditions. In the robustness section, we also validate our proposed RL approach using the Michigan simulation which specifically explores diabetes-related costs and procedures.

We construct the weight space as a two-dimensional space: $w = [w_1, w_2]$, with $0 \leq w_1 \leq 1$ and $w_2 = 1\text{-}w_1$. As the value of w_1 increases, $r(s_{it}, a_{it}, w)$ has more emphasis on optimizing convenience, and as the value of w_1 decreases, $r(s_{it}, a_{it}, w)$ places greater emphasis on optimizing cost.

Optimal Policy

Based on the defined MOMDP, we define the optimal recommendation policy $\pi^*(s_{it}, w)$ as a function of state and weight that, given a specific state and pre-determined weight, generates the optimal action a^*_{it} that minimizes the discounted accumulated reward, with γ as the discount factor. As we defined the MOMDP with an infinite horizon, we take into consideration all future rewards.

$$\pi^*(s_{it}, w) = a^*_{it} \in \text{A s.t. } a^*_{it} = argmin(a_{it}) \quad r_{it+1} + \sum_{n=1}^{\infty} \gamma^n * r^*_{it+1+n}$$

where

$$r_{it+1} = r(s_{it}, a_{it}, w)$$

$$r^*_{it+1+n} = r(s_{it+1+n}, \pi^*(s_{it+1+n}, w), w)$$

Offline Multi-objective Reinforcement Learning

In order to estimate the optimal recommendation policy $\pi^*(s_{it}, w)$, we adopt a Q-learning approach and construct our RL agent with a double Deep Q-Network (D2QN), a variant of Deep Q-Network (DQN) among deep RL algorithms (Van Hasselt et al., 2016). In standard settings with a single objective, DQN models solve for the optimal Q function by minimizing the Bellman error:

$$Q^*(s, a) = argmin(Q)\ |Q(s_t, a_t) - E[r_{t+1} + \gamma * min(a)\ Q(s_{t+1}, a_{t+1})]|$$

In such a scenario, the optimal policy $\pi^*(s)$ is defined as

$$\pi^*(s) = argmin(a)\ Q^*(s, a)$$

In a multi-objective setting, the value of $Q^*(s ,a)$ depends on r, which is a function of state s, action a, and the pre-determined weight w. Hence, changing the weight value w can influence the optimal Q function, thereby influencing the optimal policy. While it is possible to train multiple RL agents, each corresponding to a specific weight, it becomes computationally cumbersome, especially given that the weight space is continuous. Following the extant literature on multi-objective RL, we deploy an envelope Q-learning approach and optimize for $Q^*(s, a, w)$ (Yang et al., 2019).

$$Q^*(s, a, w) = argmin(Q)\ |Q(s_t, a_t, w) - E[r_{t+1} + \gamma * min(a)\ Q(s_{t+1}, a_{t+1}, w)]|$$

In turn, we derive the optimal policy $\pi^*(s, w)$ instead of $\pi^*(s)$:

$$\pi^*(s, w) = argmin(a)\ Q^*(s, a, w)$$

Extending the Q function to take the weight parameter as input allows us to train a single RL agent to provide the optimal recommendation policy for all weight values within the weight space.

In Figure 3, we provide a graphic illustration of the D2QN we constructed. The neural network has two inputs, the state associated with each patient-year and a chosen weight. For each state associated with a given patient-year, the deep neural network first generates a one-dimensional representation of the two components of the state separately, using feed-forward neural layers to process patient demographic characteristics and a long short-term memory (LSTM) layer to process the associated medical encounter information. LSTM is especially suitable for processing medical encounter information as there is a natural temporal order between different medical encounters, and the length of such encounter information varies between patient-years. The one-dimensional representation of the two state components is then combined with the given weight and used as input into another set of feed-forward neural layers, which generates the Q-values associated with each potential action based on the state and weight used as inputs.

Due to the sensitive nature of the healthcare setting, we deployed an offline reinforcement learning approach and trained the RL agent exclusively on the 3.5 million historical observations in our data set. The D2QN was trained for up to 30 million transitions using a mini-batch size of 16, and the final results are selected at the stabilized convergence point with the loss functions. We utilized a discount factor of 0.9 when training the D2DN. This factor determines the present value of future rewards, with a range between 0 and 1. A factor closer to 0 makes the agent more myopic, prioritizing immediate rewards, whereas a factor closer to 1 makes the agent value future rewards more equivalently to immediate ones. We employ a discount factor of 0.9 to strike a balance between the significance of immediate and future rewards, emulating long-term planning. This value consistent with discount factors that have been frequently used in prior research (Cheng et al., 2019). The D2QN was implemented with Pytorch 1.12.0 (Paszke et al., 2019) in Python 3 (Rossum & Drake, 2009).

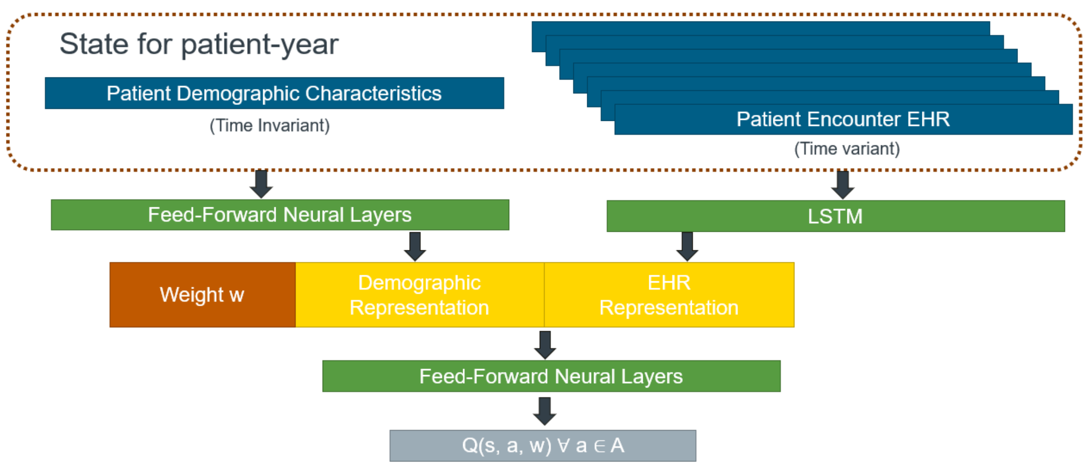


Figure 3. Graphic Illustration of D2QN Implementation

Envelop Off Policy Evaluation

Similar to the limitations faced when training the RL agent, we are restricted to utilizing historical data to evaluate the recommendations generated by the trained RL agent in an offline setting. In such settings, historical data are often generated under a policy that differs from the one that we intend to evaluate. In our scenario, we aim to assess the effectiveness of different policies recommending preventive procedures solely through data collected based on the behavior of patients in real-life, or the behavior policy. This process of evaluating a policy using historical data generated under a different policy, also known as off-policy evaluation (OPE), is more complex than evaluating policies in an online setting. Disparities between the evaluated policy and behavior policy can introduce significant bias into the evaluation. While numerous OPE methods have been proposed for accurately evaluating the performance of generated policies on historical data, most of the extant research focuses on single objective settings, with limited research on OPE methods in a multi-objective setting.

Motivated by the envelop Q learning approach, we extend fitted Q evaluation (FQE), a well-known single objective OPE method, to a multi-objective setting (Le et al., 2019). In single objective settings, it has been shown that FQE is among the best OPE methods currently available (Tang & Wiens, 2021). In a single objective setting, given trained policy $\pi_{\Phi}(s)$, FQE approximates a Q function $Q_{eval}(s, a)$ in an iterative manner:

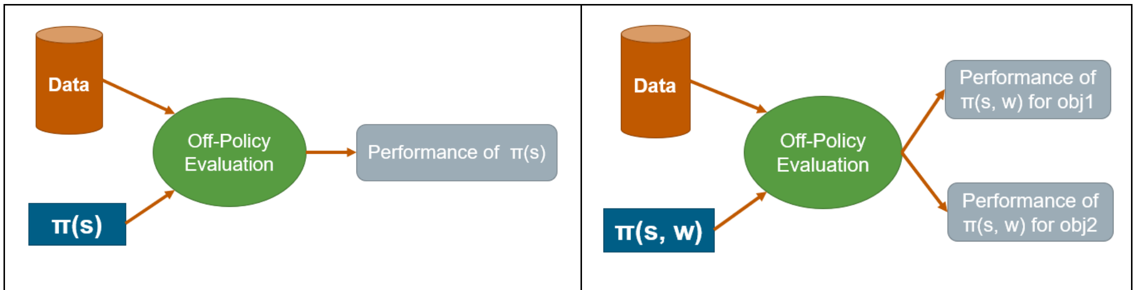
$$Q_{eval}^*(s, a) = argmin(Q)\ |Q_{eval}(s_t, a_t) - E[r_{t+1} + \gamma * min(a)\ Q_{eval}(s_{t+1}, \pi_\Phi(s_{t+1}))]|$$

In order to extend FQE to a multi-objective setting that allows us to evaluate a trained policy $\pi_\Phi(s, w)$ that can produce potentially different outputs based on the chosen weight value, we expand $Q_{eval}(s, a)$ into taking the weight value as an input as well:

$$Q_{eval}^*(s, a, w) = argmin(Q)\ |Q_{eval}(s_t, a_t, w) - E[r_{t+1} + \gamma * min(a)\ Q_{eval}(s_{t+1}, \pi_\Phi(s_{t+1}, w), w)]|$$

Under the modified multi-objective FQE (MO-FQE), we can approximate a single Q function $Q_{eval}^*(s, a, w)$ to evaluate the impact of choosing a specific action given a specific state for all potential weight values, as it internalizes the potential difference in the optimal policy as a result of employing different weight values. Figure 4 provides an intuitive demonstration of the differences between FQE and MO-FQE.

Similar to the training process of the D2QN RL agent described in the previous section, we utilize MO-FQE to evaluate the trained RL agent on our dataset. We constructed the MO-FQE network in the same manner as demonstrated in Figure 3 and trained the MO-FQE network based on the recommendations of the D2QN RL agent for up to 30 million transitions using a mini-batch size of 16. The final results were selected at the stabilized convergence point with the loss functions. We utilized the same discount factor of 0.9.



(a). Original FQE in a single objective setting

(b). MO-FQE in a multi-objective setting

Figure 4. Difference between FQE and MO-FQE

Results

In this section, we report on the characteristics of the recommendation policies generated by the D2QN RL agent and evaluation of such recommendations based on the MO-FQE approach.

Recommendation Policies

Figure 5 demonstrates the distribution of preventive care options recommended by the D2QN RL agent under three selected weight settings:

- $w_{Convenience} = 0.05$, $w_{Cost} = 0.95$, hence referred as the *Cost Focused* weight setting
- $w_{Convenience} = 0.5$, $w_{Cost} = 0.5$, hence referred as the *Balanced* weight setting
- $w_{Convenience} = 0.95$, $w_{Cost} = 0.05$, hence referred as the *Convenience Focused* weight setting

The figure shows a straightforward trade-off between optimizing for convenience versus optimizing for cost. Under the convenience-focused weight setting, which symbolizes recommending preventive procedures with the primary goal of optimizing for convenience, 95% of the patient-year observations are recommended to receive no preventive care procedures, with only 5% of the patient years recommended for some preventive care. At the other extreme, under the cost-focused weight setting, which symbolize recommending preventive procedures with the primary goal of optimizing cost, more than 66% of patient-years are recommended to receive at least one preventive procedure, while less than 33% of patients are recommended with no preventive procedures. These results indicate that the optimal recommendation policy generally depends on the chosen weight settings and varies across different pairs of weight settings.

Furthermore, by comparing the recommended preventive procedures under different weight settings against patients' actual behavior based on our data, we find that patient behavior in the real world is very similar to the recommendations generated under the convenience weight setting. In real life, 94.9% of the patient-years are associated with no preventive care procedures, 2.81% are associated with receiving a single preventive care procedure, 1.88% with receiving two preventive care procedures, and 0.41% with receiving all three preventive care procedures. This suggests that at least during our observation period, patients were likely to decide whether to receive preventive procedures mostly based on convenience. Comparing the optimal recommendation policy that optimizes both convenience and treatment costs to the way that patients behave in real life, we believe that it is possible to introduce significant savings by changing patient behavior.

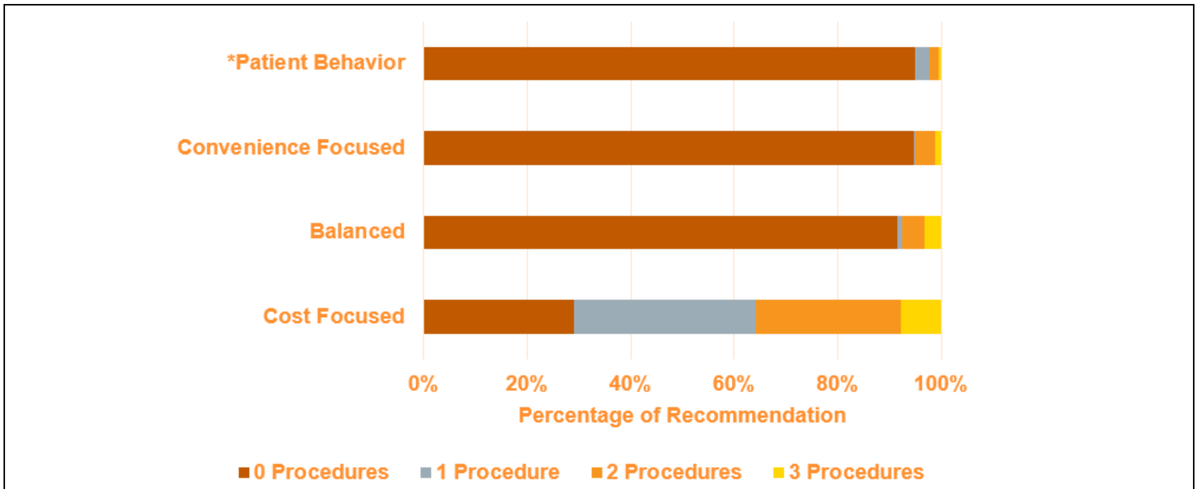


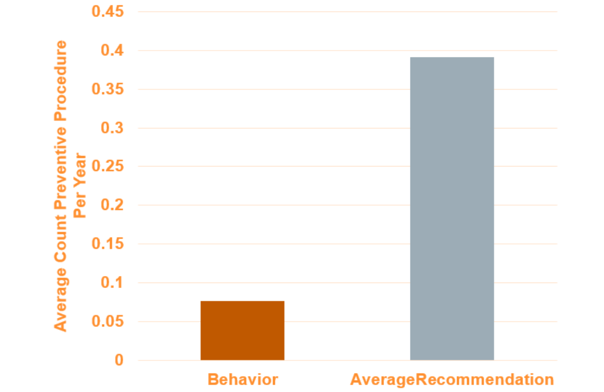
Figure 5. Distribution of Recommendations across Different Pairs of Weight Settings

Next, we study the difference between the count of preventive procedures recommended by the RL agent and how patients behaved. For simplicity, we assume a uniform distribution of weight settings across the entire weight space. As shown in Figure 6a, at the population level, our RL agent recommends that patients receive 0.4 preventive procedures per year or (in other words) one preventive procedure every two to three years. This is four times more than the observed value of 0.08 procedures per year, corresponding to one preventive procedure every 10+ years.

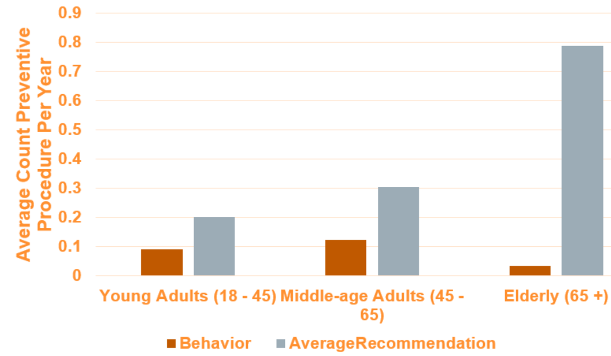
Next, we further stratified our comparison by age groups as shown in Figure 6b. Our results suggest that the count of preventive procedures recommended by our RL model increases with patient age. Our historical data reveals that middle-aged patients between the ages of 45 and 65 years received the most preventive care, followed by young adults between the ages of 18 and 45 years. However, elderly patients (i.e., older than 65 years) received the least amount of preventive care in real life, whereas these patients are most in need of preventive procedures. Our personalized recommendation approach suggests that young adults should receive two times more preventive care procedures and middle-aged adults should receive three times more preventive procedures than the frequency observed in our data, while elderly patients should receive almost 25 times more preventive procedures.

As shown in Figure 6c, male patients received fewer preventive procedures than female patients based on our historical data, whereas they are recommended to receive 10 times more preventive care procedures (compared to female patients who were recommended three times more procedures). Compared to non-diabetic patients, diabetic patients are also recommended to receive more preventive procedures. Interestingly, our RL agent only recommends a three-fold increase in preventive care for diabetic patients compared to a six-fold increase for non-diabetic patients.

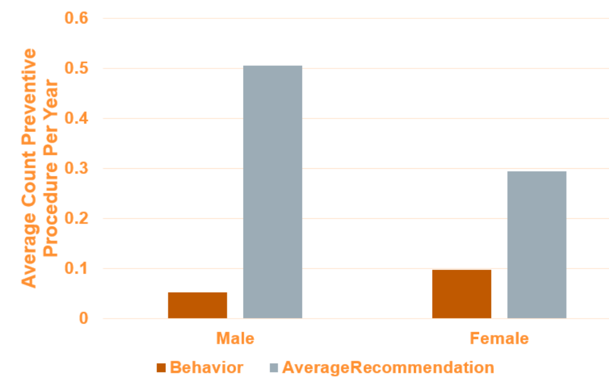
It is worth noting that although there is a relative increase in the recommended preventive care procedures for patients—ranging from twice to twenty-five times—the corresponding yearly increase in routine care visits is more modest in terms of its magnitude. For example, elderly patients receive fewer than 0.05 preventive care visits per year in real-life. Following our recommendations would result in them receiving fewer than one preventive care procedure annually.



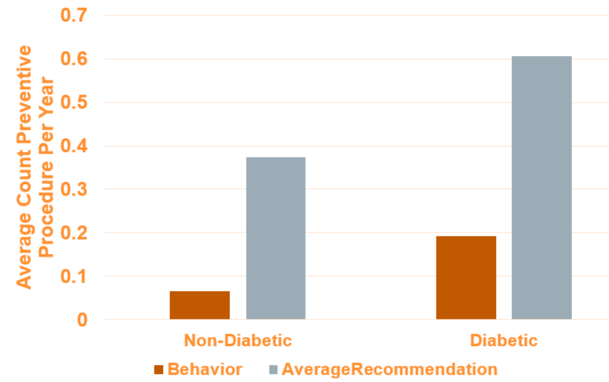
(a) Comparison between observed and recommended count of preventive procedures per year



(b) Comparison between observed and recommended count of preventive procedures, by age



(c) Comparison between observed and recommended count of preventive procedures, stratified by gender



(d) Comparison between observed and recommended count of preventive procedures, stratified by Diabetes status

Figure 6. Comparison of Patient Behavior and Optimal Policy

Evaluations of the Framework

Under the assumption of a uniformly random distribution of weight settings, we evaluate the impact of patients deviating from the optimal policy based on the multi-objective OPE method. In Figure 7a, we describe the average impact on patient convenience, measured by the count of scheduled routine care visits, and cost, measured in thousands of dollars, if patients were to follow the recommendation instead of their actual behavior, for a duration of one year. At the population level, if patients followed our policy recommendations, they would need to increase their routine care visits by 7.34%, to reduce overall treatment costs by 1.1%. It should be noted that the cost associated with the increase in routine care visits is already accounted for in the 7.34% reduction, as it estimates the impact on total treatment cost, including the cost of routine care visits.

If we stratify the impact by age group, we observe that routine care visits increase with patient age, with young adults only experiencing a 5.5% increase in visit count and elderly patients experiencing a 13% increase. The story is different for cost, however, as we observe the most significant reduction in cost for young adults, followed by elderly patients and middle-aged adults. In particular, following our recommendation has the least impact on the number of routine care visits but achieves the most cost savings for young adults. We attribute this phenomenon to the possibility that a few critical preventive procedures in young adults may be able to identify potential chronic conditions early on and reap significant long-term savings. On the other hand, while a similar level of cost reduction is achievable for elderly patients, they need to incur considerably more routine care visits.

As shown in Figures 7c and 7d, if we stratify the impact of our policy recommendations by gender and diabetes status, we observe consistent results as the previous literature. It is possible to achieve greater cost savings for male patients (compared to female patients), and diabetic patients (compared to non-diabetic patients), but more routine care encounters are also needed.

	All Population
Visit Count	↑7.34%
Cost	↓1.10%

(a) Population Level Impact

	Young Adults (18 – 45)	Middle-aged Adults (45-65)	Elderly (65 +)
Visit Count	↑5.52%	↑6.36%	↑12.97%
Cost	↓1.23%	↓0.78%	↓1.14%

(b) Impact by Age Group

	Male	Female
Visit Count	↑8.69%	↑6.30%
Cost	↓1.34%	↓0.92%

(c) Impact by Gender

	Non-Diabetic	Diabetic
Visit Count	↑6.99%	↑12.31%
Cost	↓1.00%	↓2.47%

(d) Impact by Diabetes Status

Figure 7. Impact of Optimal Policy on Patient Convenience and Cost

In summary, we discover that adherence to recommended preventive care procedures is associated with decreased cost of treatment but involves greater frequency of routine care visits. We also observe greater heterogeneity in terms of the impact of our optimal policy on different sub-groups of patients, based on age, gender, and disease progression. Evaluating the impact of our proposed personalized recommendation approach (relative to patient behavior) allows policymakers to pinpoint patient subgroups with the most significant cost-saving potential. Furthermore, our model provides deeper insights into the individualized effects of the recommendation, which in turn, allows policy makers and healthcare professionals to effectively nudge these patients through demonstration of personal impact, thereby enhancing their engagement with preventive care.

Diabetes Progression Simulation Model

While the findings derived from historical ICC data are intriguing and paint a compelling narrative, they have several constraints. As previously mentioned, the ICC dataset includes a spectrum of patient medical encounters, not just those associated with diabetes and its complications. Identifying whether the primary reason for a specific clinical encounter is directly related to diabetes remains a complex and challenging task. Furthermore, the dataset only documents established diagnoses and performed procedures, lacking any results of these procedures or detailed information on patient vital signs. Lastly, in the absence of medication prescription records within the ICC dataset, we cannot gauge or even approximate medication costs, which might significantly influence the overall cost of care.

In order to mitigate these constraints and also validate the findings on an alternative dataset, we test our proposed personalized recommendation framework utilizing the well-established Michigan Model for Diabetes (Ye et al., 2015). The Michigan Model for Diabetes (MMD) is a digital tool designed to simulate the yearly progression of Type II diabetes, its primary complications, and major comorbidities at the individual patient level.[1]

An important feature offered by the MMD is that it provides an estimate of quality adjusted life-year (QALY) — a widely accepted measure of patient quality of life commonly used in health economics literature (Nord, 2014). Given our focus on developing patient-centric, personalized recommendations, in our adaptation of

[1] For a deeper dive into the MMD, we encourage readers to consult the original publication.

the recommendation framework to this setting, we pivoted to optimize for patient convenience and QALY, rather than the cost of treatment. We posit that QALY, as a reflection of a patient's quality of life, resonates more cohesively with patient needs than treatment costs and is more relevant in value-based care settings. Furthermore, there exists an intrinsic link between QALY and the cost of care: severe ailments are typically costly to treat and often deteriorate a patient's QALY.

Similar to our baseline model based on the ICC data, we trained the multi-objective RL model with a discount factor of 0.9, using simulated diabetes progression data from MMD of 1,000,000 synthetic patients over 20 years. Subsequently, we assessed the personalized recommendations produced under the same three distinct weight settings as before:

- $w_{Convenience}$ = 0.05, w_{QALY} = 0.95, hence referred as the *QALY Focused* weight setting
- $w_{Convenience}$ = 0.5, w_{QALY} = 0.5, hence referred as the *Balanced* weight setting
- $w_{Convenience}$ = 0.95, w_{QALY} = 0.05, hence referred as the *Convenience Focused* weight setting

The distribution of recommendations across three weight settings are shown in Figure 8. It againhighlights a tradeoff between optimizing for convenience and QALY, similar to our observations in Figure 5. As the emphasis shifts from optimizing QALY to convenience, more patients are recommended to not receive any preventive care procedures.

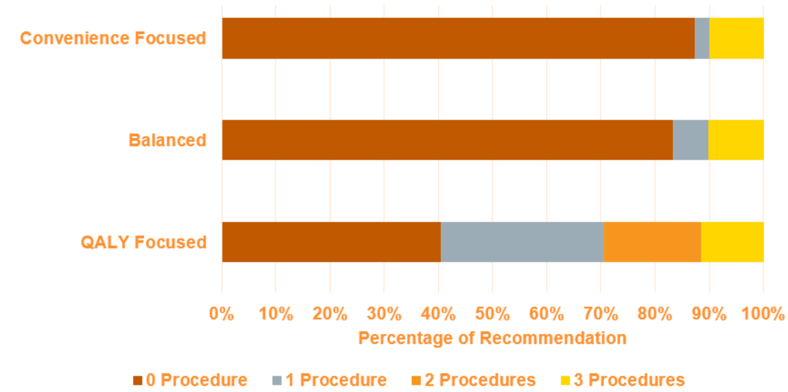


Figure 8. Distribution of Recommendations across Different Weight Settings for Patients Generated by MMD

Table 2 shows the outcomes of following different recommendations. It suggests that following the recommendations crafted under the QALY-focused weight setting allows patients to achieve higher QALY but requires significantly more routine care visits. Over a span of 20 years, on average, patients achieved a 0.15 increase in QALY, while experiencing 28 additional routine care visits when compared to the recommendations generated under the balanced weight setting. In comparison to the convenience-focused setting, they experience an increase of 0.22 in QALY and 30 routine care visits. For reference, blindness in one eye decreases QALY by 0.043 per year, so QALY increases of 0.15 and 0.22, respectively, are quite significant (Coffey et al., 2002). Interestingly, although the cost of treatment associated with hospital stays and ER visits for detrimental health issues is notably reduced for patients if they followed QALY-focused recommendations, their overall care expenses are higher. The improvements in QALY, decreased mortality rate, and reduced treatment cost are associated with a significant increase in the cost of medications.

As outlined previously, our MO-OPE evaluations suggest that, on average, adhering to our proposed recommendation framework for a year can decrease treatment expenses related to hospitalization and ER visits by 1.1%, albeit with a 7.34% increase in routine care visits for real-world patients. Results from the MMD validate the MO-OPE evaluations, while also highlighting the issue that following QALY-focused recommendation would lead to a significant surge in medication costs. Overall, this finding seems to suggest that the net fiscal benefit of personalized preventive care recommendations hinges upon the pricing of the relevant medications. We believe this observation carries profound implications for policymakers, particularly in the context of the burgeoning public discourse advocating for governmental interventions to mitigate the inexorable rise in medication prices (Cubanski et al., 2023). With judicious management of medication-associated expenditures, our findings suggest that the proposed personalized recommendation framework could potentially facilitate an equilibrium that balances both patient satisfaction and fiscal cost savings for patients, insurers, and other stakeholders.

	QALY Focused	*Balanced*	*Convenience Focused*
Average Count of Routine Care Visits Across 20 Years	45.06	17.38	15.51
Average QALY Across 20 Years	11.95	11.80	11.73
Average Mortality Rate Across 20 Years	12.78%	16.98%	18.82%
Average Cost of Treatment Across 20 Years	$ 21,126	$ 33,945	$ 38,634
Average Cost of Medication Across 20 Years	$ 71,827	$ 21,126	$ 17,515
Average Total Cost of Care Across 20 Years	$92,953	$55,071	$56,150

Table 2. Average Health Outcomes for Patients Following Recommendations under Different Weight Settings.

Conclusion

In this study, we develop an innovative, multi-objective, RL-driven framework designed to offer patient-centric personalized recommendations for preventive care related to diabetes. This framework aims to optimize both patient convenience and yearly treatment costs. Such objectives resonate not only with patients but also with external stakeholders such as policymakers, insurers, and employers. We also contribute to the design of healthcare information systems by proposing a novel multi-objective OPE method which can be deployed with IoT-enabled mobile devices for personalized care management. Such a multi-objective OPE method can be applied to other settings where evaluation of policies for multiple objectives is important but restricted to offline settings. We evaluate our personalized recommendation framework using a historical EHR dataset from a HIE in central Texas and juxtapose it with insights from a well-established diabetes simulation tool. Our results underscore the intricate balance between patient convenience and treatment costs, revealing the substantial promise inherent in the personalized recommendation of preventive care. We also validate our proposed recommendation framework based on a simulation model using the MMD. The results from the simulation are consistent, suggesting a tradeoff between patient convenience and quality of life. Moreover, the simulation shows that a reduction in treatment costs is associated with higher medication expenses, potentially neutralizing overall savings in care costs.

To the best of our knowledge, our research is one of the first studies that deploys reinforcement learning to deliver patient-centric, personalized prescriptive solutions for general health management. We believe that such recommendation models can guide policymakers in identifying patient groups with the most promising potential for cost savings. Furthermore, by presenting patients with personalized recommendations, we can illuminate the anticipated benefits specific to their health profiles, thereby enhancing their engagement with preventive care. Lastly, in an environment that increasingly calls for public support for governmental intervention to reduce escalating medication prices, our results suggest that judicious management of these costs in combination with our recommendation framework can harmonize patient content and fiscal responsibility for payors.

Limitations and Future Research

Still, several limitations exist. In the offline setting where we examined historical patient data without direct interactions with the environment, it is challenging to accurately evaluate the impact of the proposed recommendation policies, as there is always a degree of uncertainty with regard to OPE methods. Alternatively, while utilizing the simulation data allows us to accurately evaluate the impact of our recommendation approach, it is difficult to compare the recommendations with existing patient behavior. Furthermore, due to data limitations, we were only able to study primary preventive care procedures recommended for diabetes that are already commonly recommended by physicians, and our approach does not account for the potential impact of other preventive care procedures not included in the action space defined by our model. Nevertheless, we believe that our proposed framework serves as a useful starting point in exploring how to design personalized care recommendations that optimize preventive care utilization and long-term patient health outcomes.

Although we only evaluated the proposed recommendation approach in the context of Type II diabetes management using historical and simulation data, the framework can be applied to other healthcare contexts for patient-centric personalized recommendations. For example, it may be used to balance patient engagement and stress in personalized weight loss programs or to balance the risk of addiction when prescribing pain relief medication. As more personal health data becomes available through mobile health apps and wearable devices, this continuous stream of data can be used to enrich each patient's unique health context. For example, continuous glucose monitoring data from wearable devices could be used to improve diabetes management, or exercise patterns could be used to track weight loss progress. We hope that our work will serve as a foundation for future research on the impact of personalization in healthcare.

References

Abul-Husn, N. S., & Kenny, E. E. (2019). Personalized Medicine and the Power of Electronic Health Records. *Cell*, *177*(1), 58-69.

Agarwal, R., Gao, G. D., DesRoches, C., & Jha, A. K. (2010). The Digital Transformation of Healthcare: Current Status and the Road Ahead. *Information Systems Research*, *21*(4), 796-809.

Agency for Healthcare Research and Quality. (2019). *2019 National Healthcare Quality and Disparities Report*. Retrieved from https://www.ahrq.gov/research/findings/nhqrdr/nhqdr19/index.html

Ayabakan, S., Bardhan, I., Zheng, Z. Q., & Kirksey, K. (2017). The Impact of Health Information Sharing on Duplicate Testing. *MIS Quarterly*, *41*(4), 1083-+.

Bardhan, I., Oh, J. H., Zheng, Z. Q., & Kirksey, K. (2015). Predictive Analytics for Readmission of Patients with Congestive Heart Failure. *Information Systems Research*, *26*(1), 19-39.

Ben-Assuli, O., & Padman, R. (2020). Trajectories of repeated readmissions of chronic disease patients : risk stratification, profiling, and prediction. *MIS Quarterly*, *44*(1), 201-226.

Bibbins-Domingo, K., Grossman, D. C., & Curry, S. J. (2017). US Preventive Services Task Force. Screening for colorectal cancer: US Preventive Services Task Force recommendation statement (vol 315, pg 2564, 2016). *Jama-Journal of the American Medical Association*, *317*(21), 2239-2239.

Borsky, A., Zhan, C. L., Miller, T., Ngo-Metzger, Q., Bierman, A. S., & Meyers, D. (2018). Few Americans Receive All High-Priority, Appropriate Clinical Preventive Services. *Health Affairs*, *37*(6), 925-928.

Centers for Medicare & Medicaid Services. (2021). *National Health Expenditure Data*. Retrieved 4/19 from https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData

Centers for Medicare & Medicaid Services. (2022). *Preventive Services*. Retrieved 4/19 from https://www.cms.gov/Medicare/Prevention/PrevntionGenInfo

Chen, W., Lu, Y. X., Qiu, L. F., & Kumar, S. (2021). Designing Personalized Treatment Plans for Breast Cancer. *Information Systems Research*, *32*(3), 932-949.

Cheng, L. F., Prasad, N., & Engelhardt, B. E. (2019). An Optimal Policy for Patient Laboratory Tests in Intensive Care Units. *Pac Symp Biocomput*, *24*, 320-331.

Coffey, J. T., Brandle, M., Zhou, H., Marriott, D., Burke, R., Tabaei, B. P., . . . Herman, W. H. (2002). Valuing Health-Related Quality of Life in Diabetes. *Diabetes Care*, *25*(12), 2238-2243.

Coronato, A., Naeem, M., De Pietro, G., & Paragliola, G. (2020). Reinforcement learning for intelligent healthcare applications: A survey. *Artificial Intelligence in Medicine*, *109*, Article 101964.

Cubanski, J., Neuman, T., & Freed, M. (2023). *Explaining the Prescription Drug Provisions in the Inflation Reduction Act*. Kaiser Family Foundation. https://www.kff.org/medicare/issue-brief/explaining-the-prescription-drug-provisions-in-the-inflation-reduction-act/

Engel, L., Bryan, S., & Whitehurst, D. G. T. (2021). Conceptualising 'Benefits Beyond Health' in the Context of the Quality-Adjusted Life-Year: A Critical Interpretive Synthesis. *Pharmacoeconomics*, *39*(12), 1383-1395.

Fox, I., Lee, J., Pop-Busui, R., & Wiens, J. (2020). Deep reinforcement learning for closed-loop blood glucose control. Machine Learning for Healthcare Conference,

Garcia, M. C., Bastian, B., Rossen, L. M., Anderson, R., Minino, A., Yoon, P. W., . . . Iademarco, M. F. (2016). Potentially Preventable Deaths Among the Five Leading Causes of Death - United States, 2010 and 2014. *Mmwr-Morbidity and Mortality Weekly Report*, *65*(45), 1245-1255.

Higgins, A., Barnett, J., Meads, C., Singh, J., & Longworth, L. (2014). Does Convenience Matter in Health Care Delivery? A Systematic Review of Convenience-Based Aspects of Process Utility. *Value in Health*, *17*(8), 877-887.

Institute of Medicine, & Committee on the Learning Health Care System in America. (2013). Best Care at Lower Cost: The Path to Continuously Learning Health Care in America. In M. Smith, R. Saunders, L. Stuckhardt, & J. M. McGinnis (Eds.), *Best Care at Lower Cost: The Path to Continuously Learning Health Care in America*. National Academies Press.

Kannan, V. D. (2015). Capsule Commentary on Taber et al., Why Do People Avoid Medical Care? A Qualitative Study Using National Data. *Journal of General Internal Medicine*, *30*(3), 342-342.

Le, H. M., Voloshin, C., & Yue, Y. S. (2019, Jun 09-15). Batch Policy Learning under Constraints.*Proceedings of Machine Learning Research* [International conference on machine learning, vol 97]. 36th International Conference on Machine Learning (ICML), Long Beach, CA.

National Center for Health, S. (2015). Health, United States. In *Health, United States, 2014: With Special Feature on Adults Aged 55–64*. National Center for Health Statistics (US).

Nord, E. (2014). Quality-Adjusted Life-Years. In J. C. Anthony (Ed.), *Encyclopedia of Health Economics* (pp. 231-234). Elsevier.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., . . . Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. *Advances in Neural Information Processing Systems 32 (Nips 2019)*, *32*.

Porter, M. E. (2010). What Is Value in Health Care? *New England Journal of Medicine*, *363*(26), 2477-2481.

Raghu, A., Komorowski, M., Ahmed, I., Celi, L., Szolovits, P., & Ghassemi, M. (2017). Deep reinforcement learning for sepsis treatment. *arXiv preprint arXiv:1711.09602*.

Rossum, G. V., & Drake, F. L. (2009). *Python 3 Reference Manual*. CreateSpace.

Shiranthika, C., Chen, K. W., Wang, C. Y., Yang, C. Y., Sudantha, B. H., & Li, W. F. (2022). Supervised Optimal Chemotherapy Regimen Based on Offline Reinforcement Learning. *Ieee Journal of Biomedical and Health Informatics*, *26*(9), 4763-4772.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction second edition Introduction*.

Takeshima, T., Ha, C., & Iwasaki, K. (2023). Estimation of the utilities of attributes of intravenous iron infusion treatment for patients with iron-deficiency anemia: a conjoint analysis in Japan. *Journal of Medical Economics*, *26*(1), 84-94.

Taksler, G. B., Hu, B., DeGrandis, F., Montori, V. M., Fagerlin, A., Nagykaldi, Z., & Rothberg, M. B. (2021). Effect of Individualized Preventive Care Recommendations vs Usual Care on Patient Interest and Use of Recommendations A Pilot Randomized Clinical Trial. *Jama Network Open*, *4*(11), Article e2131455.

Tang, S., & Wiens, J. (2021). Model selection for offline reinforcement learning: Practical considerations for healthcare settings. Machine Learning for Healthcare Conference,

Van Hasselt, H., Guez, A., & Silver, D. (2016). Deep reinforcement learning with double q-learning. Proceedings of the AAAI conference on artificial intelligence,

Vijan, S., Stevens, D. L., Herman, W. H., Funnell, M. M., & Standiford, C. J. (1997). Screening, prevention, counseling, and treatment for the complications of type II diabetes mellitus - Putting evidence into practice. *Journal of General Internal Medicine*, *12*(9), 567-580.

White, D. J. (1982). Multi-Objective Infinite-Horizon Discounted Markov Decision Processes *Journal of Mathematical Analysis and Applications*, *89*(2), 639-647.

Yang, R. Z., Sun, X. Y., & Narasimhan, K. (2019). A Generalized Algorithm for Multi-Objective Reinforcement Learning and Policy Adaptation. *Advances in Neural Information Processing Systems 32 (Nips 2019)*, *32*.

Ye, W., Brandle, M., Brown, M. B., & Herman, W. H. (2015). The Michigan Model for Coronary Heart Disease in Type 2 Diabetes: Development and Validation. *Diabetes Technology & Therapeutics*, *17*(10), 701-711.

Yom-Tov, E., Feraru, G., Kozdoba, M., Mannor, S., Tennenholtz, M., & Hochberg, I. (2017). Encouraging Physical Activity in Patients With Diabetes: Intervention Using a Reinforcement Learning System. *Journal of Medical Internet Research*, *19*(10), Article e338.

Zhou, T. X., Wang, Y. F., Yan, L., & Tan, Y. (2023). Spoiled for Choice? Personalized Recommendation for Healthcare Decisions: A Multiarmed Bandit Approach. *Information Systems Research*.