Optimizing the Cost-Benefit of Diagnostic Tests in Clinical Decision-Making

Dong Liu, Doudou Zhou, Tianxi Cai

Abstract

In clinical practice, physicians rely on predictive models to assess disease risk and determine optimal treatments based on patient-specific features, such as demographic data and lab results. While more comprehensive models incorporating a wide range of features tend to offer higher accuracy, the associated costs—whether financial, temporal, or health-related—pose significant constraints. This paper addresses the challenge of optimizing the cost-benefit of diagnostic tests by introducing a novel approach that adapts to the unique characteristics of each patient. We propose a reinforcement learning-based methodology, framed within the Q-learning framework, to optimize the selection and sequencing of diagnostic tests, balancing the need for accurate predictions with the cost of feature collection. Additionally, our algorithm effectively handles informative missing data through a novel importance weighting procedure, ensuring robust performance even when critical predictors are not fully observed. This approach enhances the efficiency and effectiveness of clinical decision-making by minimizing costs while maintaining high predictive accuracy. Through theoretical development, practical applications, and empirical validation, we demonstrate the advantages of this cost-benefit optimization strategy in clinical settings.

1. Introduction

In modern medicine, accurate diagnoses and effective treatment plans often depend on extensive diagnostic data (Jutel, 2024). While comprehensive testing can enhance clinical decision-making, practical constraints such as time and cost make it infeasible to subject every patient to exhaustive evaluations. In practice, most patients require only basic, low-cost tests, whereas severe or complex cases necessitate more expensive and extensive testing (Kent et al., 2023; Martin et al., 2017). Mismatches between diagnostic intensity and patient needs can have serious repercussions: insufficient testing may lead to missed or delayed critical diagnoses, negatively impacting patient outcomes and increasing healthcare costs (Karunathilake and Ganegoda, 2018), while excessive testing can impose unnecessary financial and procedural burdens without proportional benefits (Müskens et al., 2022; Rao and Levin, 2012). Striking a balance between diagnostic cost and accuracy is therefore crucial.

Traditional feature selection methods, such as Lasso and other regularization-based approaches (Tibshirani, 1996; Hastie et al., 2017), optimize predictive features for static models. However, these methods overlook the dynamic nature of clinical decision-making, where test selection evolves in real-time based on the collected patient information and resource constraints. In practice, diagnoses are often achieved through sequential testing, reflecting the conditional and hierarchical relationships among tests (Grobbee et al., 2020). For example, in cardiovascular disease evaluation, initial non-invasive tests (e.g., electrocardiograms or blood tests) may lead to more specialized procedures, such as cardiac catheterization, depending on the results of earlier tests.

To optimize the cost-benefit of diagnostic tests in clinical decision-making, this paper focuses on structured test sequences, where each test provides information that influences subsequent selections. Given a predefined sequence of available tests, after completing the k-th test, the set of subsequent options is restricted to those following it in the sequence, while preceding tests are no longer available. Importantly, intermediate tests may be skipped if deemed unnecessary, guided by the evolving data. The process concludes once sufficient information is obtained for a confident diagnosis. Thus, the resulting sequence is a subsequence of the original tests, maintaining the predefined order. Figure 1 illustrates all possible outcomes when only two tests are available. We term this framework the sequential lab test problem.

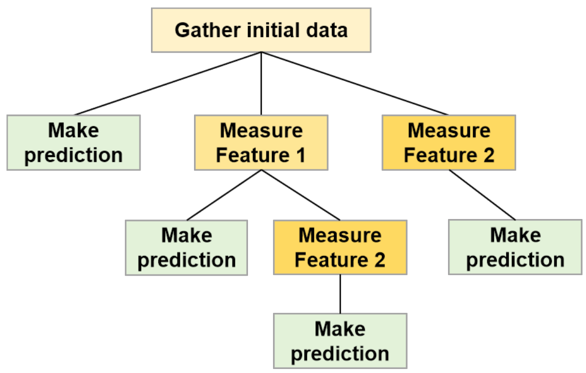

Figure 1: Illustration with two available tests.

A key challenge in this setting is handling missing data, which arises naturally in training datasets due to the sequential nature of testing. These missing values are typically not random but are systematically linked to patient histories and prior test outcomes—a phenomenon known as informative missingness. Ignoring this issue can lead to biased models and poor predictive performance, underscoring the need for specialized methodologies to address it.

While limited research exists on this topic, some related works offer partial solutions. Bryan (2023) proposed a Q-learning framework for dynamic prediction, but their approach assumes that all patients undergo every possible test, limiting its practical applicability. Reinforcement learning-based methods, such as those by Zhu et al. (2015), rely on an initial high-quality prediction model, which is often unavailable in practice. Other frameworks, such as that of Xu et al. (2021), dynamically select features but produce static final models for all individuals. Additionally, methods for optimizing dynamic treatment regimes (DTRs) (Murphy, 2003; Nahum-Shani et al., 2012; Zhao et al., 2015) consider long-term treatment effects, making them unsuitable for the sequential lab test problem.

To address these limitations, we propose a novel algorithm, COST-Q (Cost-Optimized Sequential Testing with Q-learning), specifically designed for datasets with informative missingness. COST-Q innovatively estimates missing data from prior steps using model outputs from subsequent steps and incorporates estimated missing probabilities to mitigate biases caused by informative missingness. By emphasizing practical applications, our approach enhances the efficiency and accuracy of diagnostic processes, reduces unnecessary testing, and improves patient outcomes.

The remainder of this paper is organized as follows: Section 2 formulates the problem, Section 3 introduces our proposed method, Section 4 presents theoretical results, Section 5 provides experimental results, and Section 7 concludes the paper.

2. Problem Setup

We consider a sequential testing problem involving $M+1$ available tests. While we focus on the case $M = 2$ in this paper, the proposed method can be generalized to any M. Let (X_0, X_1, X_2) represent the outcome of all possible features, where $X_0 \in \mathbb{R}^{p_0}$ denotes the basic information that is always available for every patient without cost, and $X_j \in \mathbb{R}^{p_j}, j = 1, 2$ represents the results of the j-th test. The patient's true outcome is denoted by $Y \in \mathbb{R}$. For instance, in our real data example, ...We will denote the space of continuous functions from $\mathbb{R}^d$ to A as $C(\mathbb{R}^d, A)$, where A is the target space.

To address the incomplete nature of the data, we introduce test availability indicators S_1, S_2, where $S_j = 1$ if the j-th test is performed and $S_j = 0$ otherwise. Test availability depends on existing information: S_1 is determined by X_0, while S_2 depends on (X_0, S_1X_1, S_1). Specifically, we model test availability using:

$$\pi((X_0), 1) = \mathbb{P}(S_1 = 1 \mid X_0) \text{ and } \pi((X_0, S_1X_1, S_1), 2) = \mathbb{P}(S_2 = 1 \mid X_0, S_1X_1, S_1). \quad (1)$$

This model naturally leads to the following independence assumption:

Assumption 1 (Independence) *We assume that* $(Y, X_1, X_2) \perp S_1 \mid X_0$ *and* $(Y, X_2) \perp S_2 \mid (X_0, S_1X_1, S_1)$.

Under this setup, the observed data for n samples is:

$$\mathcal{D} = \{(Y^i, X_0^i, S_1^iX_1^i, S_2^iX_2^i, S_1^i, S_2^i)\}_{i=1}^n,$$

where each $(Y^i, X_0^i, S_1^iX_1^i, S_2^iX_2^i, S_1^i, S_2^i)$ is an independent copy of $(Y, X_0, S_1X_1, S_2X_2, S_1, S_2)$.

Beside the above data, we also need a set of predictive models m's with different combinations of predictors as inputs to predict Y, represented by

$$m_0(x_0), m_1(x_0, x_1), m_2(x_0, x_2), m_{12}(x_0, x_1, x_2).$$

For simplicity, these models are assumed to be pre-specified, though we discuss their estimation in Section 3.4.

To evaluate predictive performance, we use a loss function $L(Y, m(x))$ to quantify the error between predicted and actual outcomes. Examples include: $L(Y, m(x)) = \|Y - m(x)\|^2$ (continuous outcome) and $L(Y, m(x)) = Y \log m(x) + (1-Y) \log(1-m(x))$ (binary outcome). In addition to accuracy, we consider the cost of each test, denoted as c_1 and c_2, which are given and assumed to be on a scale comparable to the loss function. There have been many works on evaluating the cost-effectiveness of different diagnostic tests (Detsky and Naglie, 1990; Snowsill, 2023). The combined loss accounts for both prediction error and test costs, enabling a more comprehensive evaluation of test strategies. The combined losses for different test combinations are defined as:

$$\begin{aligned} Err_0(Y, x_0) &= L(Y, m_0(x_0)), \\ Err_1(Y, x_0, x_1) &= L(Y, m_1(x_0, x_1)) + c_1, \\ Err_2(Y, x_0, x_2) &= L(Y, m_2(x_0, x_2)) + c_2, \\ Err_{12}(Y, x_0, x_1, x_2) &= L(Y, m_{12}(x_0, x_1, x_2)) + c_1 + c_2. \end{aligned} \tag{2}$$

For simplicity and to avoid repetition, we adopt shorthand notations such as $Err_0 = Err_0(Y, X_0)$, $Err_0^i = Err_0(Y^i, X_0^i)$, $Err_1 = Err_1(Y, X_0, X_1)$, $Err_1^i = Err_1(Y^i, X_0^i, X_1^i)$, where the arguments are clear from the context.

Our goal is to develop an optimal decision function that guides the selection of diagnostic tests based on the available patient information. Specifically, we construct a sequence of deterministic classifiers, $d = (d_0, d_1)$, defined as follows:

1. $d_0(x_0) \in \{0, 1, 2\}$ determines the first action based on the basic patient information:
   - $d_0(x_0) = 0$: no additional tests are performed; the prediction is made using $m_0(x_0)$.
   - $d_0(x_0) = 1$: the first test X_1 is performed, and d_1 determines the subsequent action.
   - $d_0(x_0) = 2$: the second test X_2 is performed, and the prediction is made using $m_2(x_0, x_2)$.
2. $d_1(x_0, x_1) \in \{0, 2\}$ the next action after the first test X_1:
   - $d_1(x_0, x_1) = 0$: no further tests are performed; the prediction is made using $m_1(x_0, x_1)$;
   - $d_1(x_0, x_1) = 2$: the second test X_2 is performed, and the prediction is made using $m_{12}(x_0, x_1, x_2)$.

In order to define the optimal decision function, we need to introduce the Q-function first. Given a decision policy $d = \{d_0, d_1\}$ where $d_0 : \mathbb{R}^{p_0} \to \{0, 1, 2\}$ and $d_1 : \mathbb{R}^{p_1} \to \{0, 2\}$, the Q-function is the conditional expected loss under the policy d, which is defined as

$$\begin{aligned} &Q_1^{d_1}(X_0, X_1) = \mathbb{E}\left[Err_1 \mathbb{I}(d_1(X_0, X_1) = 0) + Err_{12} \mathbb{I}(d_1(X_0, X_1) = 2) \mid X_0, X_1\right], \\ &Q_0^{d}(X_0) = \mathbb{E}\left[Err_0 \mathbb{I}(d_0(X_0) = 0) + Q_1^{d_1} \mathbb{I}(d_0(X_0) = 1) + Err_2 \mathbb{I}(d_0(X_0) = 2) \mid X_0\right], \end{aligned} \tag{3}$$

where $\mathbb{I}(\cdot)$ is the indicator function. Then, the optimal decision function minimizes the Q-function at each decision stage:

$$
\begin{aligned}
d_1^*(X_0, X_1) &= \underset{d_1:\mathbb{R}^{p_1}\to\{0,2\}}{\arg\min}\ Q_1^{d_1}(X_0, X_1) \\
&= \underset{d_1:\mathbb{R}^{p_1}\to\{0,2\}}{\arg\min}\ \mathbb{E}\left[Err_1\mathbb{I}(d_1(X_0, X_1) = 0) + Err_{12}\mathbb{I}(d_1(X_0, X_1) = 2) \mid X_0, X_1\right] \\
&= 2\mathbb{I}\left(\mathbb{E}[Err_1 - Err_{12} \mid X_0, X_1] > 0\right), \\
d_0^*(X_0) &= \underset{d_0:\mathbb{R}^{p_0}\to\{0,1,2\}}{\arg\min}\ Q_0^d(X_0) \\
&= \underset{d_0:\mathbb{R}^{p_0}\to\{0,1,2\}}{\arg\min}\ \mathbb{E}\left[Err_0\mathbb{I}(d_0(X_0) = 0) + Q_1^*\mathbb{I}(d_0(X_0) = 1) + Err_2\mathbb{I}(d_0(X_0) = 2) \mid X_0\right] \\
&= \begin{cases} 0, & \text{if } \mathbb{E}[Err_0(Y, X_0) \mid X_0] \le \min\{\mathbb{E}[Q_1^*(X_0, X_1) \mid X_0], \mathbb{E}[Err_2(X_0, X_2) \mid X_0]\}, \\ 1, & \text{if } \mathbb{E}[Q_1^*(X_0, X_1) \mid X_0] < \min\{\mathbb{E}[Err_0(Y, X_0) \mid X_0], \mathbb{E}[Err_2(X_0, X_2) \mid X_0]\}, \\ 2, & \text{if } \mathbb{E}[Err_2(X_0, X_2) \mid X_0] < \min\{\mathbb{E}[Err_0(Y, X_0) \mid X_0], \mathbb{E}[Q_1^*(X_0, X_1) \mid X_0]\}, \end{cases}
\end{aligned}
$$

where

$$
Q_1^* = Q_1^*(X_0, X_1) = \mathbb{E}\left[Err_1\mathbb{I}(d_1^*(X_0, X_1) = 0) + Err_{12}\mathbb{I}(d_1^*(X_0, X_1) = 2) \mid X_0, X_1\right].
$$

We will also use $d^* = (d_0^*, d_1^*)$ to denote the optimal decision function, which sequentially minimizes the total expected loss, considering both test costs and prediction errors.

3. Method

We now introduce the COST-Q algorithm, designed to learn the optimal decision function d^*. Since the goal is to minimize cumulative loss, the optimal decision d_0^* depends on subsequent decisions, particularly d_1^*. From the definition of Q_0^d and $Q_1^{d_1}$, it is evident that Q_0^d depends on both d_0 and d_1, while $Q_1^{d_1}$ only depends on d_1. These dependencies motivate a backward Q-learning approach, where we first estimate d_1^* and subsequently d_0^*.

3.1 Step 1: Estimating d_1^*

Let us initially disregard the issue of informative missing data. In this simplified scenario, a common approach to finding d_1^* is to directly optimize the empirical version of $Q_1^{d_1}$:

$$
\widehat{d}_1 = \underset{d_1:\mathbb{R}^{p_1}\to\{0,2\}}{\arg\min} \sum_{i=1}^{n}[Err_1^i\mathbb{I}(d_1(X_0^i, X_1^i) = 0) + Err_{12}^i\mathbb{I}(d_1(X_0^i, X_1^i) = 2)].
$$

However, due to the nonconvexity of the loss function, directly optimizing this expression is computationally challenging. While relaxation techniques (e.g., surrogate loss functions) are commonly used in literature, recent studies suggest that such relaxations can introduce biases in sequential testing settings (Laha et al., 2024). To address this, we propose an alternative approach to estimate d_1^*. Specifically, the optimal decision rule d_1^* can be expressed in terms of the difference between Err_{12} and Err_1:

$$
d_1^*(X_0, X_1) = 2\mathbb{I}(\Delta_{12|1}^*(X_0, X_1) > 0), \tag{4}
$$

where

$$\Delta^*_{12|1}(X_0, X_1) = \mathbb{E}[Err_{12} - Err_1 \mid X_0, X_1].$$

Now we only need to estimate $\Delta^*_{12|1}$, and we shall take informative missing back into our consideration. The following proposition give us the idea on how to estimate $\Delta^*_{12|1}$ on informative missing data.

Proposition 1 *For samples with $S_1 = 1$, $\pi((X_0, S_1X_1, S_1), 2)^{-1}$ serves to adjust for the missingness in S_2. Specifically:*

$$\Delta^*_{12|1}(X_0, X_1) = \mathbb{E}\left[\frac{Err_{12}\mathbb{I}(S_2 = 1)}{\pi((X_0, S_1X_1, S_1), 2)} - Err_1 \mid X_0, S_1X_1, S_1 = 1\right].$$

The proof of all of the Propositions is given in Supplement A.1. According to Proposition 1, we immediately derive that

$$\Delta^*_{12|1}(X_0, X_1) = \underset{\Delta:\mathbb{R}^{p_0+p_1}\to\mathbb{R}}{\arg\min}\ \mathbb{E}\left[\left(\frac{Err_{12}\mathbb{I}(S_2 = 1)}{\pi((X_0, S_1X_1, S_1), 2)} - Err_1 - \Delta(X_0, X_1)\right)^2 \mid S_1 = 1\right].$$

Thus we can estimate $\Delta^*_{12|1}(X_0, X_1)$ by

$$\widehat{\Delta}_{12|1}(X_0, X_1) = \underset{\Delta\in\mathscr{H}^d_{12|1}}{\arg\min} \sum_{i=1}^{n}\left[Err^i_{12}\cdot\frac{\mathbb{I}(S^i_2 = 1)}{\widehat{\pi}((X_0, X_1, 1), 2)} - Err^i_1 - \Delta(X^i_0, X^i_1)\right]^2 \cdot \mathbb{I}(S^i_1 = 1), \tag{5}$$

where $\widehat{\pi}((X_0, S_1X_1, S_1), 2)$ is the estimated $\pi((X_0, S_1X_1, S_1), 2)$, whose estimation will be discussed in Section 3.3. $\mathscr{H}^d_{12|1} \subset C(\mathbb{R}^{p_0+p_1}, \mathbb{R})$ is the model space for $\widehat{\Delta}_{12|1}$, which can be chosen either parametric or non-parametric based on the specific model chosen. Consequently, a variety of methods, such as Random Forest or kernel regression, can be employed to solve this problem concretely. In this paper, we focus on Neural Networks, which have demonstrated effectiveness in numerous prediction tasks. Now we can define the corresponding estimator of d^*_1 as

$$\widehat{d}_1(X_0, X_1) = 2\mathbb{I}(\widehat{\Delta}_{12|1}(X_0, X_1) > 0). \tag{6}$$

3.2 Step 2: Estimating d^*_0

Next, we estimate d^*_0. Similar to d^*_1, d^*_0 can be expressed in terms of the differences:

$$\Delta^*_{1|0}(X_0) = \mathbb{E}[Q^*_1 - Err_0 \mid X_0], \quad \Delta^*_{2|0}(X_0) = \mathbb{E}[Err_2 - Err_0 \mid X_0].$$

The optimal d^*_0 is then given by

$$d^*_0(X_0) = \begin{cases} 0, & \text{if } 0 \le \min\{\Delta^*_{1|0}(X_0), \Delta^*_{2|0}(X_0)\}, \\ 1, & \text{if } \Delta^*_{1|0}(X_0) < \min\{0, \Delta^*_{2|0}(X_0)\}, \\ 2, & \text{if } \Delta^*_{2|0}(X_0) < \min\{0, \Delta^*_{1|0}(X_0)\}. \end{cases} \tag{7}$$

For informative missingness, Proposition 2 provides the required adjustments.

Proposition 2 *The importance weights $\pi((X_0),1)^{-1}$ and $\pi((X_0,S_1X_1,S_1),2)^{-1}$ adjust for missingness when estimating $\Delta^*_{1|0}(X_0)$ and $\Delta^*_{2|0}(X_0)$, respectively:*

$$\Delta^*_{1|0}(X_0) = \mathbb{E}\left[\frac{Q_1^*\mathbb{I}(S_1=1)}{\pi((X_0),1)} - Err_0 \mid X_0\right],$$
$$\Delta^*_{2|0}(X_0) = \mathbb{E}\left[\frac{Err_2\mathbb{I}(S_2=1)}{\pi((X_0,S_1X_1,S_1),2)} - Err_0 \mid X_0\right].$$

However, a challenge arises in estimating $\Delta^*_{1|0}(X_0)$ because $Q_1^{\widehat{d}_1}$ is not directly observable, even for samples with $S_1 = 1$. To address this issue, we define an estimated version of Err_{12} as follows:

$$\widehat{Err}^i_{12} = \begin{cases} Err^i_{12} & \text{if } S^i_2 = 1, \\ Err^i_1 + \widehat{\Delta}_{12|1}(X^i_0, X^i_1) & \text{if } S^i_2 = 0. \end{cases}$$

Replacing Err_{12} with $\widehat{Err}^i_{12}$ in the definition of Q_1, we can define the "observed" cumulative loss of each sample as

$$\widehat{Q}_1^{i,\widehat{d}_1} = Err^i_1\mathbb{I}(\widehat{d}_1(X^i_0,X^i_1)=0) + \widehat{Err}^i_{12}\mathbb{I}(\widehat{d}_1(X^i_0,X^i_1)=2).$$

Finally, we can estimate $\Delta^*_{1|0}(X_0)$ and $\Delta^*_{2|0}(X_0)$ by

$$\widehat{\Delta}_{1|0}(X_0) = \underset{\Delta\in\mathscr{H}^d_{1|0}}{\arg\min} \sum_{i=1}^n \left[\widehat{Q}_1^{i,\widehat{d}_1}\cdot\frac{\mathbb{I}(S^i_1=1)}{\widehat{\pi}((X^i_0),1)} - Err^i_0 - \Delta(X^i_0)\right]^2,$$
$$\widehat{\Delta}_{2|0}(X_0) = \underset{\Delta\in\mathscr{H}^d_{2|0}}{\arg\min} \sum_{i=1}^n \left[Err^i_2\cdot\frac{\mathbb{I}(S^i_2=1)}{\widehat{\pi}((X^i_0,S^i_1X^i_1,S^i_1),2)} - Err^i_0 - \Delta(X^i_0)\right]^2,$$

where $\widehat{\pi}((X_0),1)$ is the estimated probability of $\mathbb{P}(S_1=1\mid X_0)$, $\widehat{\pi}((X_0,S_1X_1,S_1),2)$ is the estimated probability of $\mathbb{P}(S_2=1\mid X_0,S_1X_1,S_1)$. $\mathscr{H}^d_{1|0}\subset C(\mathbb{R}^{p_0},\mathbb{R})$ and $\mathscr{H}^d_{2|0}\subset C(\mathbb{R}^{p_0},\mathbb{R})$ are the model spaces for $\widehat{\Delta}_{1|0}$ and $\widehat{\Delta}_{2|0}$, respectively.

Now we can define the corresponding d_0 as

$$\widehat{d}_0(X_0) = \begin{cases} 0 & \text{if } 0 \le \min\left(\widehat{\Delta}_{1|0}(X_0), \widehat{\Delta}_{2|0}(X_0)\right), \\ 1 & \text{if } \widehat{\Delta}_{1|0}(X_0) < \min\{0, \widehat{\Delta}_{2|0}(X_0)\}, \\ 2 & \text{if } \widehat{\Delta}_{2|0}(X_0) < \min\{0, \widehat{\Delta}_{1|0}(X_0)\}. \end{cases}$$

3.3 Estimating π

In practical applications, the value of π is typically unknown. Fortunately, numerous methods have been developed to estimate this probability. In this study, we adopt a model-based approach for estimating π. To be specific, we have

$$\begin{aligned}
\pi((X_0),1) &= \mathbb{E}[S_1 \mid X_0] \\
&= \underset{\pi:\mathbb{R}^{p_0}\to[0,1]}{\arg\min}\ \mathbb{E}\left[L_{CE}(S_1,\pi(X_0)) \mid X_0\right], \\
\pi((X_0,S_1X_1,S_1),2) &= \mathbb{E}[S_2 \mid X_0,S_1X_1,S_1] \\
&= \underset{\pi:\mathbb{R}^{p_0+p_1}\to[0,1]}{\arg\min}\ \mathbb{E}\left[L_{CE}(S_2,\pi(X_0,S_1X_1,S_1)) \mid X_0,S_1X_1,S_1\right],
\end{aligned}$$

where $L_{CE}(S,x) = -S\log(x) - (1-S)\log(1-x)$ is the cross-entropy loss. As a result, we can estimate π by

$$\begin{aligned}
\widehat{\pi}((X_0),1) &= \underset{\pi\in\mathscr{H}_1^\pi}{\arg\min}\ \sum_{i=1}^n L_{CE}(S_1^i,\pi(X_0^i)), \\
\widehat{\pi}((X_0,S_1X_1,S_1),2) &= \underset{\pi\in\mathscr{H}_2^\pi}{\arg\min}\ \sum_{i=1}^n L_{CE}(S_2^i,\pi(X_0^i,S_1^iX_1^i,S_1^i)),
\end{aligned}$$

where $\mathscr{H}_1^\pi \subset C(\mathbb{R}^{p_0},[0,1])$ and $\mathscr{H}_2^\pi \subset C(\mathbb{R}^{p_0+p_1},[0,1])$ are the model spaces for $\pi((X_0),1)$ and $\pi((X_0,S_1X_1,S_1),2)$, respectively.

Regarding the specific training method of the model as a black box, we can estimate π using the following algorithms:

Algorithm 1: Given X_0, to estimate $\pi((X_0),1)$

Input: $Y_{out} = \{S_1^i\}_{i=1}^n$, $X_{in} = \{X_0^i\}_{i=1}^n$.
Training: Train a prediction model for the outcome S_1^i against the features X_0^i using cross-entropy loss.
Output: Estimated $\pi((X_0),1)$, denoted as $\widehat{\pi}((X_0),1)$.

Algorithm 2: Given X_0, S_1X_1, S_1, to estimate $\pi((X_0,S_1X_1,S_1),2)$

Input: $Y_{out} = \{S_2^i\}_{i=1}^n$, $X_{in} = \{X_0^i, S_1^iX_1^i, S_1^i\}_{i=1}^n$.
Training: Train a prediction model for the outcome S_2^i against the features $(X_0^i, S_1^iX_1^i, S_1^i)$ using cross-entropy loss.
Output: Estimated $\pi((X_0,S_1X_1,S_1),2)$, denoted as $\widehat{\pi}((X_0,S_1X_1,S_1),2)$.

3.4 Estimating m

Although sometimes we may know m in advance, it is more common that we need to estimate m from the data we observed. This is actually a separate issue, but we plan to address it together in this article.

Recall that we used a loss function $L(Y,m(x))$ to measure the difference between the predicted outcome and the real outcome. The optimal m is natually defined as the minimizer of the expected loss, that is,

$$
\begin{aligned}
m_0^*(x_0) &= \mathop{\arg\min}_{m_0:\mathbb{R}^{p_0}\to\mathbb{R}} \mathbb{E}[L(Y, m_0(X_0))], \\
m_1^*(x_0, x_1) &= \mathop{\arg\min}_{m_1:\mathbb{R}^{p_0+p_1}\to\mathbb{R}} \mathbb{E}[L(Y, m_1(X_0, X_1))], \\
m_2^*(x_0, x_2) &= \mathop{\arg\min}_{m_2:\mathbb{R}^{p_0+p_2}\to\mathbb{R}} \mathbb{E}[L(Y, m_2(X_0, X_2))], \\
m_{12}^*(x_0, x_1, x_2) &= \mathop{\arg\min}_{m_{12}:\mathbb{R}^{p_0+p_1+p_2}\to\mathbb{R}} \mathbb{E}[L(Y, m_{12}(X_0, X_1, X_2))].
\end{aligned}
$$

Without the nested sequential structure when estimating d, we can estimate each m independently, which significantly simplifies the problem. However, we still need to address the issue of informative missing data. Now we will discuss how to estimate for each m. Firstly, since X_0 are always available, we can simply estimate m_0 by

$$
\widehat{m}_0 = \mathop{\arg\min}_{m_0\in\mathscr{H}_0^m} \sum_{i=1}^{n} L(Y^i, m_0(X_0^i)), \tag{8}
$$

where $\mathscr{H}_0^m \subset C(\mathbb{R}^{p_0}, \mathbb{R})$ is the model space for $\widehat{m}_0$.

When it comes to m_1, m_2 and m_{12}, informative weights are required because of the missing X_1 and X_2. Similar to the estimation of d, the following proposition provides the required adjustments for the informative missingness.

Proposition 3 *We have*

$$
\begin{aligned}
\mathbb{E}[L(Y, m_1(X_0, X_1))] &= \mathbb{E}\left[L(Y, m_1(X_0, X_1)) \cdot \frac{\mathbb{I}(S_1 = 1)}{\pi((X_0), 1)}\right], \\
\mathbb{E}[L(Y, m_2(X_0, X_2))] &= \mathbb{E}\left[L(Y, m_2(X_0, X_2)) \cdot \frac{\mathbb{I}(S_2 = 1)}{\pi((X_0, S_1X_1, S_1), 2)}\right], \\
\mathbb{E}\left[L(Y, m_{12}(X_0, X_1, X_2))\right] &= \mathbb{E}\left[L(Y, m_{12}(X_0, X_1, X_2)) \cdot \frac{\mathbb{I}(S_1 = 1)\mathbb{I}(S_2 = 1)}{\pi((X_0), 1)\pi((X_0, S_1X_1, S_1), 2)}\right].
\end{aligned}
$$

Proposition 3 shows that for the estimation of m_1 and m_2, $\pi((X_0), 1)^{-1}$ and $\pi((X_0, S_1X_1, S_1), 2)^{-1}$ should be used as the importance weights, respectively. Moreover, for the estimation of m_{12}, due to the inherent independence structure of the data, $\pi((X_0), 1)^{-1} \times \pi((X_0, S_1X_1, S_1), 2)^{-1}$ serves as the corresponding importance weight. As a result, we can estimate m_1, m_2 and m_{12} by

$$
\begin{aligned}
\widehat{m}_1 &= \mathop{\arg\min}_{m_1\in\mathscr{H}_1^m} \sum_{i=1}^{n} L(Y^i, m_1(X_0^i, X_1^i)) \cdot \frac{\mathbb{I}(S_1^i = 1)}{\widehat{\pi}((X_0^i), 1)}, \\
\widehat{m}_2 &= \mathop{\arg\min}_{m_2\in\mathscr{H}_2^m} \sum_{i=1}^{n} L(Y^i, m_2(X_0^i, X_2^i)) \cdot \frac{\mathbb{I}(S_2^i = 1)}{\widehat{\pi}((X_0^i, S_1^iX_1^i, S_1^i), 2)}, \\
\widehat{m}_{12} &= \mathop{\arg\min}_{m_{12}\in\mathscr{H}_{12}^m} \sum_{i=1}^{n} L(Y^i, m_{12}(X_0^i, X_1^i, X_2^i)) \cdot \frac{\mathbb{I}(S_1^i = 1)\mathbb{I}(S_2^i = 1)}{\widehat{\pi}((X_0^i), 1)\widehat{\pi}((X_0^i, S_1^iX_1^i, S_1^i), 2)},
\end{aligned}
$$

where $\mathscr{H}_1^m \subset C(\mathbb{R}^{p_0+p_1}, \mathbb{R})$, $\mathscr{H}_2^m \subset C(\mathbb{R}^{p_0+p_2}, \mathbb{R})$ and $\mathscr{H}_{12}^m \subset C(\mathbb{R}^{p_0+p_1+p_2}, \mathbb{R})$ are the model spaces for $\widehat{m}_1, \widehat{m}_2$ and $\widehat{m}_{12}$, respectively.

Notably, the weights utilized for estimating m are identical to those employed in the estimation of d. This alignment not only enhances computational efficiency but also seamlessly integrates the estimation of m within the overarching methodological framework.

3.5 Summary of the method

We summarize COST-Q as the following algorithm:

Algorithm 3: Flow of COST-Q

Input: $\{Y^i, X_0^i, S_1^i X_1^i, S_2^i X_2^i, S_1^i, S_2^i\}_{i=1}^n$, costs c_1, c_2, loss function $L(Y, m(x))$, (optional) predictive models m_0, m_1, m_2, m_{12}.

Step 1: If m_0, m_1, m_2, m_{12} are not given, estimate $\widehat{m}_0, \widehat{m}_1, \widehat{m}_2, \widehat{m}_{12}$ using the method in section 3.4;

Step 2: Estimate $\widehat{\pi}(X_0), \widehat{\pi}(X_0, S_1X_1, S_1)$ using the method in section 3.3;

Step 3: Estimate $\widehat{d}_1(X_0, X_1)$ using the method in 3.1;

Step 4: Estimate $\widehat{d}_0(X_0)$ using the method in 3.2.

4. Theoretical Results

To ensure clarity in the proof, especially in the convergence rate of the estimators, we will add the superscript (n) to denote the sample size of an estimator. For example, $\widehat{\pi}^{(n)}(X_0)((X_0), 1) = \mathbb{P}(S_1 = 1 \mid X_0)$. Similarly, we will also use symbols like $\widehat{\Delta}_{12|1}^{(n)}, \widehat{d}_1^{(n)}, \widehat{m}_1^{(n)}$ to denote the corresponding estimators. Beside the notation already mentioned, we also define

$$\widehat{Q}_1^{(n),d_1} = \widehat{Q}_1^{(n),d_1}(Y, X_0, X_1) = \mathbb{E}\left[Err_1 \mathbb{I}(d_1(X_0, X_1) = 0) + \widehat{Err}_{12}^{(n)} \mathbb{I}(d_1(X_0, X_1) = 2) \mid X_0, X_1 \right]$$

and

$$\widehat{Err}_{12}^{(n)} = \begin{cases} Err_{12}, & \text{if } S_2 = 1, \\ Err_1 + \widehat{\Delta}_{12|1}^{(n)}(X_0, X_1), & \text{if } S_2 = 0. \end{cases}$$

4.1 Convergence of d for incomplete data

In this section, we will discuss the convergence of our main method. It is worth noting that we shall assume our m is optimal in the discussion in this subsection. We will prove the acumulative loss, i.e. the Q-function, converges to the optimal Q-function, which shows the convergence of our decision function $\widehat{d}^{(n)}$ to the optimal decision function d^*.

Below we list the assumptions we need for the convergence of COST-Q. Firstly, we need to assume the boundedness of π and Err and the convergence of $\widehat{\pi}$.

Assumption 2 (Boundness and convergence of $\widehat{\pi}$) *There exists a constant $c_0 > 0$ such that*

$$\min\{\pi((X_0, S_1X_1, S_1), 2), \pi((X_0), 1)\} \geq c_0,$$
$$\min\{\widehat{\pi}^{(n)}((X_0, S_1X_1, S_1), 2), \widehat{\pi}^{(n)}((X_0), 1)\} \geq c_0,$$

for any X_0, X_1, X_2, n. Moreover, we have

$$\sup_{(X_0, S_1X_1, S_1)} \left| \widehat{\pi}^{(n)}((X_0, S_1X_1, S_1), 2) - \pi((X_0, S_1X_1, S_1), 2) \right| = o_p(n^{-1/4}),$$
$$\sup_{X_0} \left| \widehat{\pi}^{(n)}((X_0), 1) - \pi((X_0), 1) \right| = o_p(n^{-1/4}).$$

Assumption 3 (Boundedness of $\widehat{Err}$) *There exists a constant $c_1 > 0$ such that*

$$0 \leq Err_0, Err_1, Err_2, Err_{12}, \widehat{Err}_{12}^{(n)} \leq c_1,$$

for any X_0, X_1, X_2, n.

Moreover, to insure the convergence of $\widehat{d}_1^{(n)}$, we need to assume the convergence of corresponding nuisance models. We will also need similar assumptions for the convergence of $\widehat{d}_0^{(n)}$.

Assumption 4 *Define*

$$\overline{\Delta}_{12|1}^{(n)}(X_0, X_1) = \mathbb{E}\left[\mathbb{I}(S_2 = 1) \frac{Err_{12}}{\widehat{\pi}^{(n)}((X_0, S_1X_1, S_1), 2)} - Err_1 \mid X_0, X_1, S_1 = 1 \right],$$

and

$$\widetilde{\Delta}_{12|1}^{(n),(n_0)}(X_0, X_1) = \arg\min_{\Delta} \sum_{i=1}^{n} \mathbb{I}(S_1^i = 1) \left[\mathbb{I}(S_2^i = 1) \frac{Err_{12}^i}{\widehat{\pi}^{(n_0)}((X_0^i, X_1^i, 1), 2)} - Err_1^i - \Delta(X_0^i, X_1^i) \right]^2.$$

Notice that we have $\widetilde{\Delta}_{12|1}^{(n),(n)}(X_0, X_1) = \widehat{\Delta}_{12|1}^{(n)}(X_0, X_1)$. Then we have

$$\sup_{X_0, X_1, n_0} \left\{ \widetilde{\Delta}_{12|1}^{(n),(n_0)}(X_0, X_1) - \overline{\Delta}_{12|1}^{(n)}(X_0, X_1) \right\} = o_p(n^{-1/4}),$$

Assumption 4 assumes that the estimated model converge uniformly to its expectation form at the rate of $o_p(n^{-1/4})$. Under this assumption, we can establish the convergence of $\widehat{d}_1^{(n)}$. However, since d_1 is a binary decision function defined over a specific sample space, directly addressing convergence in the function space introduces additional complexities. To simplify the analysis, we instead focus on the convergence of the corresponding Q-function. The following theorem shows the convergence of $Q_1^{\widehat{d}_1^{(n)}}$ to $Q_1^{d_1^*}$.

Theorem 4 (Convergence of $\widehat{d}_1$) *We have* $Q_1^{\widehat{d}_1^{(n)}}(X_0, X_1) \geq Q_1^{d_1^*}(X_0, X_1)$, *and*

$$\sup_{X_0, X_1} \left\{ Q_1^{\widehat{d}_1^{(n)}}(X_0, X_1) - Q_1^{d_1^*}(X_0, X_1) \right\} = o_p(n^{-1/4}).$$

The proof of Theorem 4 is provided in the supplementary material. Similarly, to further ensure the convergence of $\widehat{d}_0^{(n)}$, we also require similar assumptions for the convergence of $\widehat{d}_0^{(n)}$.

Assumption 5 (Convergence of $\widehat{\Delta}_{1|0}^{(n)}, \widehat{\Delta}_{2|0}^{(n)}$) *Define*

$$\begin{aligned}
\overline{\widehat{\Delta}}_{1|0}^{(n)}(X_0) &= \mathbb{E}\left[\mathbb{I}(S_1 = 1) \frac{\widehat{Q}_1^{(n), \widehat{d}_1^{(n)}}}{\widehat{\pi}^{(n)}((X_0), 1)} - Err_0 \mid X_0 \right], \\
\overline{\widehat{\Delta}}_{2|0}^{(n)}(X_0) &= \mathbb{E}\left[\mathbb{I}(S_2 = 1) \frac{Err_2}{\widehat{\pi}^{(n)}((X_0, S_1 X_1, S_1), 2)} - Err_0 \mid X_0 \right].
\end{aligned}$$

Then we have

$$\begin{aligned}
\sup_{X_0} \left\{ \widehat{\Delta}_{1|0}^{(n)}(X_0) - \overline{\widehat{\Delta}}_{1|0}^{(n)}(X_0) \right\} &= o_p(n^{-1/4}), \\
\sup_{X_0} \left\{ \widehat{\Delta}_{2|0}^{(n)}(X_0) - \overline{\widehat{\Delta}}_{2|0}^{(n)}(X_0) \right\} &= o_p(n^{-1/4}).
\end{aligned}$$

Based on Assumption 5 and the convergence of $\widehat{d}_1^{(n)}$, we can finally establish the convergence of $\widehat{d}^{(n)}$ to the optimal decision function d^*, which shows the effectiveness of our main method.

Theorem 5 (Convergence of $\widehat{d}$) *We have* $Q_0^{\widehat{d}^{(n)}}(X_0) \geq Q_0^{d^*}(X_0)$, *and*

$$\sup_{X_0} \left\{ Q_0^{\widehat{d}^{(n)}}(X_0) - Q_0^{d^*}(X_0) \right\} = o_p(n^{-1/4}).$$

4.2 Convergence of m for incomplete data

Firstly, we need the following properties of the loss function.

Assumption 6 (Conditions for Loss function) *$L(y,\hat{y})$ is a bounded loss function. That is, there exists a constant c_2 such that*

$$|L(y,\hat{y})| \le c_2.$$

and $L(y,\hat{y})$ reaches its minimum $L(y,\hat{y})=0$ at $\hat{y}=y$.

Like in the proof of ds, we also require the convergence of corresponding nuisance models to ensure the convergence of $\widehat{m}$.

Assumption 7 (Convergence of m) *Define*

$$\begin{aligned}
\overline{m}_0^{(n)}(X_0) &= m^*(X_0),\\
\overline{m}_1^{(n)}(X_0,X_1) &= \arg\min_m \mathbb{E}\left[\frac{L(Y,m(X_0,X_1))}{\widehat{\pi}^{(n)}((X_0),1)}\mathbb{I}(S_1=1)\mid X_0,X_1\right],\\
\overline{m}_2^{(n)}(X_0,X_2) &= \arg\min_m \mathbb{E}\left[\frac{L(Y,m(X_0,X_2))}{\widehat{\pi}^{(n)}((X_0,S_1X_1,S_1),2)}\mathbb{I}(S_2=1)\mid X_0,X_2\right],\\
\overline{m}_{12}^{(n)}(X_0,X_1,X_2) &= \arg\min_m \mathbb{E}\left[\frac{L(Y,m(X_0,X_1,X_2))}{\widehat{\pi}^{(n)}((X_0,S_1X_1,S_1),2)}\mathbb{I}(S_1=1,S_2=1)\mid X_0,X_1,X_2\right].
\end{aligned}$$

We have

$$\begin{aligned}
\sup_{X_0}\left\{\widehat{m}_0^{(n)}(X_0)-\overline{m}_0^{(n)}(X_0)\right\} &= o_p(n^{-1/4}),\\
\sup_{X_0,X_1}\left\{\widehat{m}_1^{(n)}(X_0,X_1)-\overline{m}_1^{(n)}(X_0,X_1)\right\} &= o_p(n^{-1/4}),\\
\sup_{X_0,X_2}\left\{\widehat{m}_2^{(n)}(X_0,X_2)-\overline{m}_2^{(n)}(X_0,X_2)\right\} &= o_p(n^{-1/4}),\\
\sup_{X_0,X_1,X_2}\left\{\widehat{m}_{12}^{(n)}(X_0,X_1,X_2)-\overline{m}_{12}^{(n)}(X_0,X_1,X_2)\right\} &= o_p(n^{-1/4}).
\end{aligned}$$

Based on the above assumptions, we have the following theorem, which shows the convergence of $\widehat{m}$.

Theorem 6 (Convergence of $\widehat{m}$) *We have*

$$\begin{aligned}
\sup_{X_0}\left\{\widehat{m}_0^{(n)}(X_0)-m_0^*(X_0)\right\} &= o_p(n^{-1/4}),\\
\sup_{X_0,X_1}\left\{\widehat{m}_1^{(n)}(X_0,X_1)-m_1^*(X_0,X_1)\right\} &= o_p(n^{-1/4}),\\
\sup_{X_0,X_2}\left\{\widehat{m}_2^{(n)}(X_0,X_2)-m_2^*(X_0,X_2)\right\} &= o_p(n^{-1/4}),\\
\sup_{X_0,X_1,X_2}\left\{\widehat{m}_{12}^{(n)}(X_0,X_1,X_2)-m_{12}^*(X_0,X_1,X_2)\right\} &= o_p(n^{-1/4}).
\end{aligned}$$

5. Simulation

In this section, we examine the performance of Algorithm 3 from extensive simulation studies for various values of sample size n, data dimension p, also for different data generation mechanisms and missing data patterns. Since COST-Q offers an optional choice of m, we will also examine the performance of COST-Q with or without m given.

5.1 Comparable Methods

We are mainly in examining the performance of the proposed methods, i.e., COST-Q with or without m given. However, to serve as a benchmark, we also include the following methods as competitors.

1. **BOWL**: Kosorok (Zhao et al., 2015) proposed a method for addressing informative missing data in the dynamic treatment regime (DTR) problem. When adapted to the Sequential Lab Test problem, this method can be regarded as a benchmark. Specifically, the BOWL method assigns a weight to each action. For instance, when training d_1, BOWL assigns a weight of $\pi((X_0, S_1, X_1, S_1), 2)$ to samples with $S_2 = 1$, and a weight of $1 - \pi((X_0, S_1, X_1, S_1), 2)$ to all other samples.

2. **Only-complete**: We may only use the complete data to estimate the outcome, thus addressing the problem of missing data. We can then estimate the outcome using the complete data (i.e., samples with $S_1^i = S_2^i = 1$) and use the same method as in COST-Q to estimate the optimal treatment regime. There is no need of estimating $\widehat{\pi}$ and $\widehat{Err}$ in this method.

3. **One-time**: We may directly choose a choice of tests from $\{\{0\}, \{1\}, \{2\}, \{1,2\}\}$ based on X_0. We train regression models $\widehat{\Delta}_1, \widehat{\Delta}_2, \widehat{\Delta}_{12}$ to predict the expectation of $Err_1 - Err_0$, $Err_2 - Err_0$, $Err_{12} - Err_0$, and choose the choice of tests based on the prediction. Moreover, to address the problem of missing data, we may give weights to the samples based on the missing pattern.

5.2 Data Generation Mechanisms and Evaluation Metrics

To assess the performance of the proposed methods, we conduct simulations under a number of scenarios imitating a multi-stage randomized trial. The complete data are generated as follows:

1. **Senario 1:** $M = 2$. The outcome is generated through $Y \sim \text{Ber}(1/10)$. We set $X_2 \sim N(3, 0.5)$ when $Y = 1$, and $X_2 \sim N(0, 0.5)$ when $Y = 0$. The other covariates are generated as $X_1 = X_2/2 + \epsilon_1, X_0 = X_1 + \epsilon_0$, where $\epsilon_1 \sim N(0, 0.175), \epsilon_0 \sim N(0, 0.25)$. $\epsilon_0, \epsilon_1, X_2$ are independent with each other. We set the cost of each test as $c_1 = 0.04, c_2 = 0.3$.

2. **Senario 2:** $M = 2$. The outcome is generated via

$$Y \sim \text{Ber}(p), p = \text{logstic}(X_0 + X_1 + \mathbb{I}(X_1 > 1)X_2),$$

where X_k are numbers, $(X_0, X_1, X_2)^\top \sim N(\mathbf{0}, 0.1\mathbf{1}\mathbf{1}^\top + 0.9I), \varepsilon \sim N(0, 1)$. We set cost of each test as $c_1 = 0.02, c_2 = 0.05$.

3. **Senario 3:** $M = 2$. The outcome is generated via

$$Y \sim \text{Ber}(p), p = \text{logistic}(X_0^\top \mathbf{1} + X_1^\top \mathbf{1} + \langle \mathbb{I}(X_1 > 1), X_2 \rangle),$$

where X_k are 3-dimensional random vectors, $(X_0, X_1, X_2)^\top \sim N(\mathbf{0}, 0.1\mathbf{1}\mathbf{1}^\top + 0.9I), \varepsilon \sim N(0, 1)$. We set cost of each test as $c_1 = 0.02, c_2 = 0.05$.

Moreover, the missing policy is as follows:

1. **For Senario 1:**

   (a) **Step 1:** If $X_0 \leq -0.3$, stop with probability 0.8, choose X_1 or X_2 with probability 0.1. If $-0.3 < X_0 \leq 0.3$, choose X_1 with probability 0.8, stop or choose X_2 with probability 0.1. If $X_0 > 0.3$, choose X_2 with probability 0.8, stop or choose X_1 with probability 0.1.

   (b) **Step 2:** If X_2 chosen in step 1, stop the test. If X_1 chosen in step 1, stop or choose X_2 with probability 0.8 and 0.2 if $X_1 \leq -0.5$, with probability 0.5 and 0.5 if $-0.5 < X_1 \leq 0.5$, and with probability 0.2 and 0.8 if $X_1 > 0.5$.

   If X_i is chosen above, we set $S_i = 1$, otherwise $S_i = 0$.

2. **For Senario 2:** We directly set missing probability as $\pi((X_0), 1) = \pi((X_0, 0, 0), 2) = \text{logistic}(X_0)/2$, and $\pi((X_0, X_1, 1), 2) = \text{logistic}(X_0 + X_1)/2$.

3. **For Senario 3:** We directly set missing probability as $\pi((X_0), 1) = \pi((X_0, 0, 0), 2) = \text{logistic}(X_0^\top \mathbf{1})/2$, and $\pi((X_0, X_1, 1), 2) = \text{logistic}(X_0^\top \mathbf{1} + X_1^\top \mathbf{1})/2$.

5.3 Results

We will use COST-Q and other methods in section 3 in our simulation studies. We use Neural Networks to estimate all the models.

In our simulation studies, we employ Neural Networks to estimate all the models. The parameters of the Neural Networks are configured as follows:

We use a neural network with 2 hidden layers. Each hidden layer consists of 64 neurons. The Adam optimizer is employed for training the network, with a learning rate of 0.001. Cross-entropy loss is used as the loss function to measure the performance of the network. A batch size of 10 is used during training. The network is trained for 100 epochs.

6. Application to XX Data

7. Conclusion

References

Bryan. *Inference on the generalization error of machine learning algorithms and the design of hierarchical medical term embeddings.* PhD thesis, University of Cambridge, 2023.

A. S. Detsky and I. G. Naglie. A clinician's guide to cost-effectiveness analysis. *Annals of internal medicine*, 113(2):147–154, 1990.

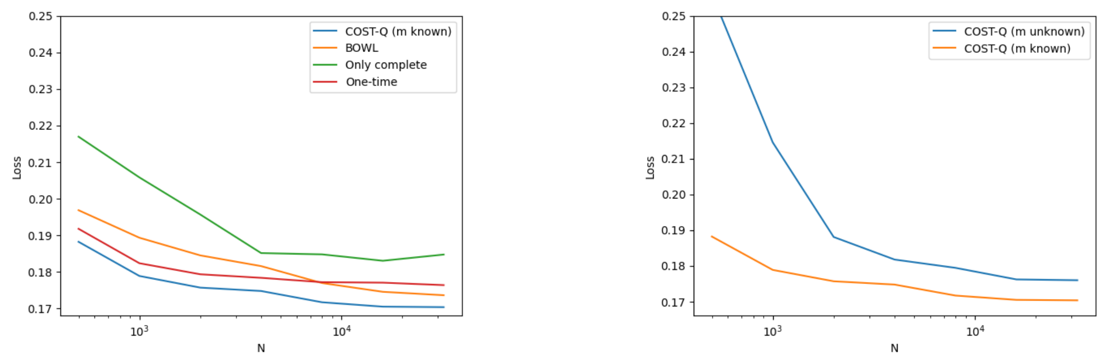

Figure 2: The average loss of COST-Q. We set N ranged from 500 to 32000. Each testing set has 10^6 samples. We repeat the simulation 100 times and report the average loss.

E. J. Grobbee, M. van der Vlugt, A. J. van Vuuren, A. K. Stroobants, R. C. Mallant-Hent, I. Lansdorp-Vogelaar, P. M. Bossuyt, E. J. Kuipers, E. Dekker, and M. C. Spaander. Diagnostic yield of one-time colonoscopy vs one-time flexible sigmoidoscopy vs multiple rounds of mailed fecal immunohistochemical tests in colorectal cancer screening. *Clinical Gastroenterology and Hepatology*, 18(3):667–675, 2020.

T. Hastie, R. Tibshirani, and R. J. Tibshirani. Extended comparisons of best subset selection, forward stepwise selection, and the lasso. *arXiv preprint arXiv:1707.08692*, 2017.

A. Jutel. *Putting a name to it: Diagnosis in contemporary society*. JHU Press, 2024.

S. P. Karunathilake and G. U. Ganegoda. Secondary prevention of cardiovascular diseases and application of technology for early diagnosis. *BioMed research international*, 2018 (1):5767864, 2018.

P. Kent, T. Haines, P. O'Sullivan, A. Smith, A. Campbell, R. Schutze, S. Attwell, J. Caneiro, R. Laird, K. O'Sullivan, et al. Cognitive functional therapy with or without movement sensor biofeedback versus usual care for chronic, disabling low back pain (restore): a randomised, controlled, three-arm, parallel group, phase 3, clinical trial. *The Lancet*, 401 (10391):1866–1877, 2023.

N. Laha, A. Sonabend-W, R. Mukherjee, and T. Cai. Finding the optimal dynamic treatment regimes using smooth fisher consistent surrogate loss. *The Annals of Statistics*, 52 (2):679–707, 2024.

L. Martin, M. Hutchens, C. Hawkins, and A. Radnov. How much do clinical trials cost. *Nat Rev Drug Discov*, 16(6):381–382, 2017.

S. A. Murphy. Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 65(2):331–355, 2003.

J. L. Müskens, R. B. Kool, S. A. van Dulmen, and G. P. Westert. Overuse of diagnostic testing in healthcare: a systematic review. *BMJ quality & safety*, 31(1):54–63, 2022.

I. Nahum-Shani, M. Qian, D. Almirall, W. E. Pelham, B. Gnagy, G. A. Fabiano, J. G. Waxmonsky, J. Yu, and S. A. Murphy. Q-learning: a data analysis method for constructing adaptive interventions. *Psychological methods*, 17(4):478, 2012.

V. M. Rao and D. C. Levin. The overuse of diagnostic imaging and the choosing wisely initiative, 2012.

T. Snowsill. Modelling the cost-effectiveness of diagnostic tests. *PharmacoEconomics*, 41 (4):339–351, 2023.

R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1):267–288, 1996.

R. Xu, M. Li, Z. Yang, L. Yang, K. Qiao, and Z. Shang. Dynamic feature selection algorithm based on q-learning mechanism. *Applied Intelligence*, pages 1–12, 2021.

Y.-Q. Zhao, D. Zeng, E. B. Laber, and M. R. Kosorok. New statistical learning methods for estimating optimal dynamic treatment regimes. *Journal of the American Statistical Association*, 110(510):583–598, 2015.

R. Zhu, D. Zeng, and M. R. Kosorok. Reinforcement learning trees. *Journal of the American Statistical Association*, 110(512):1770–1784, 2015.

A. Supplementary materials

A.1 Proof of Propositions in Section 3

Our proof relies only on Assumption 1, and is constructed through calculations involving conditional expectations.

Proof [Proof of Proposition 1] Firstly, by Assumption 1, we have

$$\mathbb{E}[Err_1 \mid X_0, S_1X_1, S_1 = 1] = \frac{\mathbb{E}[Err_1 \cdot S_1 \mid X_0, X_1]}{\mathbb{P}(S_1 = 1 \mid X_0, X_1)} = \mathbb{E}[Err_1 \mid X_0, X_1].$$

Also, we have

$$\begin{aligned}
&\mathbb{E}\left[Err_{12} \cdot \frac{I(S_2 = 1)}{\mathbb{P}(S_2 = 1 \mid X_0, X_1, S_1 = 1)} \mid X_0, S_1X_1, S_1 = 1\right] \\
=&\mathbb{E}\left[Err_{12} \mid X_0, S_1X_1, S_1 = 1\right] \\
=&\frac{\mathbb{E}[Err_{12} \cdot S_1 \mid X_0, X_1]}{\mathbb{P}(S_1 = 1 \mid X_0, X_1)} \\
=&\mathbb{E}[Err_{12} \mid X_0, X_1].
\end{aligned}$$

Adding the two equations above, we can get the desired result. ■

Proof [Proof of Proposition 2] Directly using Assumption 1, we have

$$\begin{aligned}
\Delta^*_{1|0}(X_0) &= \mathbb{E}\left[Q^*_1 - Err_0 \mid X_0\right] \\
&= \mathbb{E}\left[Q^*_1 \cdot \frac{I(S_1=1)}{\mathbb{P}(S_1=1 \mid X_0)} - Err_0 \mid X_0\right] \\
\Delta^*_{2|0}(X_0) &= \mathbb{E}\left[Err_2 - Err_0 \mid X_0\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[Err_2 - Err_0 \mid X_0, S_1X_1, S_1\right] \mid X_0\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[Err_2 \cdot \frac{I(S_2=1)}{\mathbb{P}(S_2=1 \mid X_0, S_1X_1, S_1)} - Err_0 \mid X_0, S_1X_1, S_1\right] \mid X_0\right] \\
&= \mathbb{E}\left[Err_2 \cdot \frac{I(S_2=1)}{\mathbb{P}(S_2=1 \mid X_0, S_1X_1, S_1)} - Err_0 \mid X_0\right].
\end{aligned}$$

■

Proof [Proof of Proposition 3] The result for m_1 and m_2 is trivial. We only prove the result for m_{12}.

Using Assumption 1, we have

$$\begin{aligned}
&\mathbb{E}\left[L(Y, m_{12}(X_0, X_1, X_2)) \cdot \frac{I(S_1=1)I(S_2=1)}{\mathbb{P}(S_1=1 \mid X_0)\mathbb{P}(S_2=1 \mid X_0, S_1X_1, S_1)}\right] \\
=&\mathbb{E}\left[\mathbb{E}\left[L(Y, m_{12}(X_0, X_1, X_2)) \cdot \frac{I(S_1=1)I(S_2=1)}{\mathbb{P}(S_1=1 \mid X_0)\mathbb{P}(S_2=1 \mid X_0, S_1X_1, S_1)} \mid X_0, S_1X_1, S_1\right]\right] \\
=&\mathbb{E}\left[L(Y, m_{12}(X_0, X_1, X_2)) \cdot \frac{I(S_1=1)}{\mathbb{P}(S_1=1 \mid X_0)} \cdot \mathbb{E}\left[\frac{I(S_2=1)}{\mathbb{P}(S_2=1 \mid X_0, S_1X_1, S_1)} \mid X_0, S_1X_1, S_1\right]\right] \\
=&\mathbb{E}\left[L(Y, m_{12}(X_0, X_1, X_2)) \cdot \frac{I(S_1=1)}{\mathbb{P}(S_1=1 \mid X_0)}\right] \\
=&\mathbb{E}\left[\mathbb{E}\left[L(Y, m_{12}(X_0, X_1, X_2)) \cdot \frac{I(S_1=1)}{\mathbb{P}(S_1=1 \mid X_0)} \mid X_0\right]\right] \\
=&\mathbb{E}\left[L(Y, m_{12}(X_0, X_1, X_2)) \cdot \mathbb{E}\left[\frac{I(S_1=1)}{\mathbb{P}(S_1=1 \mid X_0)} \mid X_0\right]\right] \\
=&\mathbb{E}\left[L(Y, m_{12}(X_0, X_1, X_2))\right].
\end{aligned}$$

■

A.2 Proof of the convergence of $\widehat{d}_1$

The goal of this section is to prove Theorem 4. We first introduce the following lemma.

Lemma 7 (Convergence of $\overline{\Delta}^{(n)}_{12|1}$) *We have*

$$\sup_{X_0, X_1}\left\{\overline{\Delta}^{(n)}_{12|1}(X_0, X_1) - \Delta^*_{12|1}(X_0, X_1)\right\} = o_p(n^{-1/4}).$$

Proof According to Proposition 1, we have

$$\Delta^*_{12|1}(X_0, X_1) = \mathbb{E}\left[\mathbb{I}(S_2 = 1)\frac{Err_{12}}{\pi((X_0, S_1X_1, S_1), 2)} - Err_1 \mid X_0, X_1, S_1 = 1\right].$$

Then, we have

$$\begin{aligned}
&\left|\overline{\Delta}^{(n)}_{12|1}(X_0, X_1) - \Delta^*_{12|1}(X_0, X_1)\right| \\
=&\left|\mathbb{E}\left[\mathbb{I}(S_2 = 1)\frac{Err_{12}}{\widehat{\pi}^{(n)}((X_0, S_1X_1, S_1), 2)} - Err_1 \mid X_0, X_1, S_1 = 1\right]\right. \\
&\left. - \mathbb{E}\left[\mathbb{I}(S_2 = 1)\frac{Err_{12}}{\pi((X_0, S_1X_1, S_1), 2)} - Err_1 \mid X_0, X_1, S_1 = 1\right]\right| \\
\leq&\mathbb{E}\left[\mathbb{I}(S_2 = 1) \cdot Err_{12} \cdot \left|\frac{1}{\widehat{\pi}^{(n)}((X_0, S_1X_1, S_1), 2)} - \frac{1}{\pi((X_0, S_1X_1, S_1), 2)}\right| \mid X_0, X_1, S_1 = 1\right] \\
\leq&\frac{1}{c_0^2}\mathbb{E}\left[\mathbb{I}(S_2 = 1) \cdot Err_{12} \cdot \left|\widehat{\pi}^{(n)}((X_0, S_1X_1, S_1), 2) - \pi((X_0, S_1X_1, S_1), 2)\right| \mid X_0, X_1, S_1 = 1\right] \\
\leq&\frac{c_1}{c_0^2}\left|\widehat{\pi}^{(n)}((X_0, S_1X_1, S_1), 2) - \pi((X_0, S_1X_1, S_1), 2)\right|.
\end{aligned}$$

Taking supremum over X_0, X_1 and using Assumption 2, we can get the desired result. ■

With Lemma 7, we can prove Theorem 4.

Proof [Proof of Theorem 4] Using the definition of d_1^*, we have

$$\begin{aligned}
Q_1^{\widehat{d}_1^{(n)}}(X_0, X_1) &= \mathbb{E}\left[Err_1\mathbb{I}(\widehat{d}_1^{(n)}(X_0, X_1) = 0) + Err_{12}\mathbb{I}(\widehat{d}_1^{(n)}(X_0, X_1) = 2) \mid X_0, X_1\right] \\
&\geq \min_{d_1} \mathbb{E}\left[Err_1\mathbb{I}(d_1(X_0, X_1) = 0) + Err_{12}\mathbb{I}(d_1(X_0, X_1) = 2) \mid X_0, X_1\right] \\
&= Q_1^{d_1^*}(X_0, X_1).
\end{aligned}$$

Then we proceed to estimate the convergence rate. Setting $n_0 = n$ in Assumption 4 and using Lemma 7, we have

$$\sup_{X_0, X_1}\left\{\widehat{\Delta}^{(n)}_{12|1}(X_0, X_1) - \Delta^*_{12|1}(X_0, X_1)\right\} = o_p(n^{-1/4}). \tag{9}$$

It is easy to see that a equivalent form of d_1^* is

$$d_1^*(X_0, X_1) = \begin{cases} 0, & \text{if } 0 = \min\{0, \Delta^*_{12|1}(X_0, X_1)\}, \\ 2, & \end{cases}$$

Then we have

$$
\begin{aligned}
& Q_1^{\widehat{d}_1^{(n)}}(X_0, X_1) - Q_1^{d_1^*}(X_0, X_1) \\
=& \mathbb{E}\left[Err_1 \mathbb{I}(\widehat{d}_1^{(n)}(X_0, X_1) = 0) + Err_{12} \mathbb{I}(\widehat{d}_1^{(n)}(X_0, X_1) = 2) \mid X_0, X_1 \right] \\
& - \mathbb{E}\left[Err_1 \mathbb{I}(d_1^*(X_0, X_1) = 0) + Err_{12} \mathbb{I}(d_1^*(X_0, X_1) = 2) \mid X_0, X_1 \right] \\
=& \mathbb{E}\left[(Err_{12} - Err_1) \mid X_0, X_1 \right] \mathbb{I}(\widehat{d}_1^{(n)}(X_0, X_1) = 2) \\
& - \mathbb{E}\left[(Err_{12} - Err_1) \mid X_0, X_1 \right] \mathbb{I}(d_1^*(X_0, X_1) = 2) \\
=& \Delta^*_{12|1}(X_0, X_1) \mathbb{I}(\widehat{d}_1^{(n)}(X_0, X_1) = 2) - \Delta^*_{12|1}(X_0, X_1) \mathbb{I}(d_1^*(X_0, X_1) = 2) \\
=& \Delta^*_{12|1}(X_0, X_1) \mathbb{I}(\widehat{\Delta}^{(n)}_{12|1}(X_0, X_1) \leq 0) - \Delta^*_{12|1}(X_0, X_1) \mathbb{I}(\Delta^*_{12|1}(X_0, X_1) \leq 0) \\
=& \widehat{\Delta}^{(n)}_{12|1}(X_0, X_1) \mathbb{I}(\widehat{\Delta}^{(n)}_{12|1}(X_0, X_1) \leq 0) - \Delta^*_{12|1}(X_0, X_1) \mathbb{I}(\Delta^*_{12|1}(X_0, X_1) \leq 0) \\
& + (\Delta^*_{12|1}(X_0, X_1) - \Delta^{(n)}_{12|1}(X_0, X_1)) \mathbb{I}(\widehat{\Delta}^{(n)}_{12|1}(X_0, X_1) \leq 0) \\
=& \min\{0, \widehat{\Delta}^{(n)}_{12|1}(X_0, X_1)\} - \min\{0, \Delta^*_{12|1}(X_0, X_1)\} \\
& + (\Delta^*_{12|1}(X_0, X_1) - \widehat{\Delta}^{(n)}_{12|1}(X_0, X_1)) \mathbb{I}(\widehat{\Delta}^{(n)}_{12|1}(X_0, X_1) \leq 0) \\
\leq& 2 \left| \Delta^*_{12|1}(X_0, X_1) - \widehat{\Delta}^{(n)}_{12|1}(X_0, X_1) \right|.
\end{aligned}
$$

Taking supremum over X_0, X_1 and using (9), we can get the result. ■

A.3 Proof of the convergence of $\widehat{d}_0$

The goal of this section is to prove Theorem 5. Different from the proof of Theorem 4, we need to tackle with $\widehat{Q}$, which should be convergent to Q^* in some sense. We first introduce the following lemma.

Lemma 8 (Convergence of $\widehat{Q}_1$) *We have*

$$
\sup_{X_0, n_0} \left| \mathbb{E}\left[\widehat{Q}_1^{(n), \widehat{d}_1^{(n_0)}}(X_0, X_1) - Q_1^{\widehat{d}_1^{(n_0)}}(X_0, X_1) \mid X_0 \right] \right| = o_p(n^{-1/4}).
$$

Proof We only prove the convergence of $\widehat{Q}_1$. By the definition of $\widehat{Q}_1$ and Q_1, we have

$$
\begin{aligned}
&\left|\mathbb{E}\left[\widehat{Q}_1^{(n),\widehat{d}_1^{(n_0)}}(X_0,X_1) - Q_1^{\widehat{d}_1^{(n_0)}}(X_0,X_1) \mid X_0\right]\right| \\
=&\left|\mathbb{E}\left[Err_1 \mathbb{I}(\widehat{d}_1^{(n_0)}(X_0,X_1)=0) + \widehat{Err}_{12}^{(n)} \mathbb{I}(\widehat{d}_1^{(n_0)}(X_0,X_1)=2) \mid X_0\right]\right. \\
&\left. - \mathbb{E}\left[Err_1 \mathbb{I}(\widehat{d}_1^{(n_0)}(X_0,X_1)=0) + Err_{12} \mathbb{I}(\widehat{d}_1^{(n_0)}(X_0,X_1)=2) \mid X_0\right]\right| \\
=&\left|\mathbb{E}\left[\left(\widehat{Err}_{12}^{(n)} - Err_{12}\right) \mathbb{I}(\widehat{d}_1^{(n_0)}(X_0,X_1)=2) \mid X_0\right]\right| \\
=&\left|\mathbb{E}\left[\left(Err_{12} - Err_1 - \widehat{\Delta}_{12|1}^{(n)}(X_0,X_1)\right) I(S_2=0) \mathbb{I}(\widehat{d}_1^{(n_0)}(X_0,X_1)=2) \mid X_0\right]\right| \\
=&\left|\mathbb{E}\left[\mathbb{E}\left[\left(Err_{12} - Err_1 - \widehat{\Delta}_{12|1}^{(n)}(X_0,X_1)\right) I(S_2=0) \mathbb{I}(\widehat{d}_1^{(n_0)}(X_0,X_1)=2) \mid X_0, S_1X_1, S_1\right] \mid X_0\right]\right| \\
=&\left|\mathbb{E}\left[\mathbb{E}\left[Err_{12} - Err_1 - \widehat{\Delta}_{12|1}^{(n)}(X_0,X_1) \mid X_0, S_1X_1, S_1\right] I(S_2=0) \mathbb{I}(\widehat{d}_1^{(n_0)}(X_0,X_1)=2) \mid X_0\right]\right| \\
\leq&\mathbb{E}\left[\left|\mathbb{E}\left[Err_{12} - Err_1 - \widehat{\Delta}_{12|1}^{(n)}(X_0,X_1) \mid X_0, S_1X_1, S_1\right]\right| \mid X_0\right].
\end{aligned}
$$

Since $(Y, X_1, X_2) \perp S_1 \mid X_0$, we have

$$
\begin{aligned}
&\left|\mathbb{E}\left[Err_{12} - Err_1 - \widehat{\Delta}_{12|1}^{(n)}(X_0,X_1) \mid X_0, S_1X_1, S_1\right]\right| \\
=&\left|\mathbb{E}\left[Err_{12} - Err_1 - \widehat{\Delta}_{12|1}^{(n)}(X_0,X_1) \mid X_0, S_1=0\right] I(S_1=0)\right. \\
&\left.+\mathbb{E}\left[Err_{12} - Err_1 - \widehat{\Delta}_{12|1}^{(n)}(X_0,X_1) \mid X_0, X_1, S_1=1\right] I(S_1=1)\right| \\
=&\left|\mathbb{E}\left[Err_{12} - Err_1 - \widehat{\Delta}_{12|1}^{(n)}(X_0,X_1) \mid X_0\right] I(S_1=0)\right. \\
&\left.+\mathbb{E}\left[Err_{12} - Err_1 - \widehat{\Delta}_{12|1}^{(n)}(X_0,X_1) \mid X_0, X_1\right] I(S_1=1)\right| \\
=&\left|\mathbb{E}\left[\Delta_{12|1}^*(X_0,X_1) - \widehat{\Delta}_{12|1}^{(n)}(X_0,X_1) \mid X_0\right] I(S_1=0)\right. \\
&\left.+\left[\Delta_{12|1}^*(X_0,X_1) - \widehat{\Delta}_{12|1}^{(n)}(X_0,X_1)\right] I(S_1=1)\right|.
\end{aligned}
$$

Taking supremum over X_0, n_0 and using Assumption 3, we can get the desired result. ■

Recalling Assumption 5 and the definition of $\overline{\widehat{\Delta}}_{1|0}^{(n)}, \overline{\widehat{\Delta}}_{2|0}^{(n)}$, we want to prove their convergence to models that are closer to $\Delta_{1|0}^*$ and $\Delta_{2|0}^*$. Define that

$$
\begin{aligned}
\overline{\Delta}_{1|0}^{(n)}(X_0) &= \mathbb{E}\left[Q_1^{\widehat{d}_1^{(n)}} - Err_0 \mid X_0\right] = \mathbb{E}\left[\mathbb{I}(S_1=1)\frac{Q_1^{\widehat{d}_1^{(n)}}}{\pi((X_0),1)} - Err_0 \mid X_0\right], \\
\overline{\Delta}_{2|0}^{(n)}(X_0) &= \mathbb{E}\left[Err_2 - Err_0 \mid X_0\right] = \mathbb{E}\left[\mathbb{I}(S_2=1)\frac{Err_2}{\pi((X_0,S_1X_1,S_1),2)} - Err_0 \mid X_0\right].
\end{aligned}
$$

We have the following lemma.

Lemma 9 *We have*

$$\sup_{X_0}\left\{\widehat{\overline{\Delta}}^{(n)}_{1|0}(X_0)-\overline{\Delta}^{(n)}_{1|0}(X_0)\right\}=o_p(n^{-1/4}),$$
$$\sup_{X_0}\left\{\widehat{\overline{\Delta}}^{(n)}_{2|0}(X_0)-\overline{\Delta}^{(n)}_{2|0}(X_0)\right\}=o_p(n^{-1/4}).$$

Proof We only prove the convergence of $\widehat{\overline{\Delta}}^{(n)}_{1|0}$. By the definition of $\widehat{\overline{\Delta}}^{(n)}_{1|0}$ and $\overline{\Delta}^{(n)}_{1|0}$, we have

$$\begin{aligned}
&\left|\widehat{\overline{\Delta}}^{(n)}_{1|0}(X_0)-\overline{\Delta}^{(n)}_{1|0}(X_0)\right|\\
=&\left|\mathbb{E}\left[\mathbb{I}(S_1=1)\left(\frac{\widehat{Q}_1^{(n),\widehat{d}_1^{(n)}}}{\widehat{\pi}^{(n)}((X_0),1)}-\frac{Q_1^{\widehat{d}_1^{(n)}}}{\pi((X_0),1)}\right)\mid X_0\right]\right|\\
\le&\mathbb{E}\left[\left|\frac{\widehat{Q}_1^{(n),\widehat{d}_1^{(n)}}}{\widehat{\pi}^{(n)}((X_0),1)}-\frac{\widehat{Q}_1^{(n),\widehat{d}_1^{(n)}}}{\pi((X_0),1)}\right|+\left|\frac{\widehat{Q}_1^{(n),\widehat{d}_1^{(n)}}}{\pi((X_0),1)}-\frac{Q_1^{\widehat{d}_1^{(n)}}}{\pi((X_0),1)}\right|\mid X_0\right]\\
\le&\mathbb{E}\left[\frac{c_1}{c_0^2}\left|\widehat{\pi}^{(n)}((X_0),1)-\pi((X_0),1)\right|+\frac{1}{c_0}\left|\widehat{Q}_1^{(n),\widehat{d}_1^{(n)}}-Q_1^{\widehat{d}_1^{(n)}}\right|\mid X_0\right]\\
=&\frac{c_1}{c_0^2}\left|\widehat{\pi}^{(n)}((X_0),1)-\pi((X_0),1)\right|+\frac{1}{c_0}\mathbb{E}\left[\left|\widehat{Q}_1^{(n),\widehat{d}_1^{(n)}}-Q_1^{\widehat{d}_1^{(n)}}\right|\mid X_0\right].
\end{aligned}$$

Taking supremum over X_0 on both sides, and using Assumption 2 and Lemma 8, we get

$$\sup_{X_0}\left\{\widehat{\overline{\Delta}}^{(n)}_{1|0}(X_0)-\overline{\Delta}^{(n)}_{1|0}(X_0)\right\}=o_p(n^{-1/4}).$$

■

Now, we convert the convergence of Δ to the convergence of cumulative loss. We have the following lemma.

Lemma 10 *Define*

$$\overline{d}_0^{(n_0)}(X_0)=\begin{cases}0, & \text{if } 0\le\min\{\overline{\Delta}^{(n_0)}_{1|0}(X_0),\overline{\Delta}^{(n_0)}_{2|0}(X_0)\},\\ 1, & \text{if } \overline{\Delta}^{(n_0)}_{1|0}(X_0)<\min\{0,\overline{\Delta}^{(n_0)}_{2|0}(X_0)\},\\ 2, & \text{if } \overline{\Delta}^{(n_0)}_{2|0}(X_0)<\min\{0,\overline{\Delta}^{(n_0)}_{1|0}(X_0)\}.\end{cases}$$

Then we have

$$\begin{aligned}
\sup_{X_0,n_0}&\left\{\mathbb{E}\left[Err_0\mathbb{I}(\widehat{d}_0^{(n)}(X_0)=0)+Q_1^{\widehat{d}_1^{(n_0)}}\mathbb{I}(\widehat{d}_0^{(n)}(X_0)=1)+Err_2\mathbb{I}(\widehat{d}_0^{(n)}(X_0)=2)\mid X_0\right]\right.\\
&\left.-\mathbb{E}\left[Err_0\mathbb{I}(\overline{d}_0^{(n_0)}(X_0)=0)+Q_1^{\widehat{d}_1^{(n_0)}}\mathbb{I}(\overline{d}_0^{(n_0)}(X_0)=1)+Err_2\mathbb{I}(\overline{d}_0^{(n_0)}(X_0)=2)\mid X_0\right]\right\}\\
=&o_p(n^{-1/4}).
\end{aligned}\tag{10}$$

Proof we have

$$
\begin{aligned}
&\mathbb{E}\left[Err_0\mathbb{I}(\widehat{d}_0^{(n)}(X_0)=0)+Q_1^{\widehat{d}_1^{(n_0)}}\mathbb{I}(\widehat{d}_0^{(n)}(X_0)=1)+Err_2\mathbb{I}(\widehat{d}_0^{(n)}(X_0)=2)\mid X_0\right]\\
&\quad-\mathbb{E}\left[Err_0\mathbb{I}(\overline{d}_0^{(n_0)}(X_0)=0)+Q_1^{\widehat{d}_1^{(n_0)}}\mathbb{I}(\overline{d}_0^{(n_0)}(X_0)=1)+Err_2\mathbb{I}(\overline{d}_0^{(n_0)}(X_0)=2)\mid X_0\right]\\
=&\mathbb{E}\left[Q_1^{\widehat{d}_1^{(n_0)}}-Err_0\mid X_0\right]\mathbb{I}(\widehat{d}_0^{(n)}(X_0)=1)+\mathbb{E}\left[Err_2-Err_0\mid X_0\right]\mathbb{I}(\widehat{d}_0^{(n)}(X_0)=2)\\
&\quad-\mathbb{E}\left[Q_1^{\widehat{d}_1^{(n_0)}}-Err_0\mid X_0\right]\mathbb{I}(\overline{d}_0^{(n_0)}(X_0)=1)-\mathbb{E}\left[Err_2-Err_0\mid X_0\right]\mathbb{I}(\overline{d}_0^{(n_0)}(X_0)=2)\\
=&\left[\overline{\Delta}_{1|0}^{(n_0)}(X_0)\mathbb{I}\left(\widehat{d}_0^{(n)}(X_0)=1\right)-\overline{\Delta}_{1|0}^{(n_0)}(X_0)\mathbb{I}\left(\overline{d}_0^{(n_0)}(X_0)=1\right)\right]\\
&\quad+\left[\overline{\Delta}_{2|0}^{(n_0)}(X_0)\mathbb{I}\left(\widehat{d}_0^{(n)}(X_0)=2\right)-\overline{\Delta}_{2|0}^{(n_0)}(X_0)\mathbb{I}\left(\overline{d}_0^{(n_0)}(X_0)=2\right)\right]\\
=&\left[\widehat{\Delta}_{1|0}^{(n)}(X_0)\mathbb{I}\left(\widehat{d}_0^{(n)}(X_0)=1\right)-\overline{\Delta}_{1|0}^{(n_0)}(X_0)\mathbb{I}\left(\overline{d}_0^{(n_0)}(X_0)=1\right)+\left(\overline{\Delta}_{1|0}^{(n_0)}(X_0)-\widehat{\Delta}_{1|0}^{(n)}(X_0)\right)\mathbb{I}\left(\widehat{d}_0^{(n)}(X_0)=1\right)\right]\\
&\quad+\left[\widehat{\Delta}_{2|0}^{(n)}(X_0)\mathbb{I}\left(\widehat{d}_0^{(n)}(X_0)=2\right)-\overline{\Delta}_{2|0}^{(n_0)}(X_0)\mathbb{I}\left(\overline{d}_0^{(n_0)}(X_0)=2\right)+\left(\overline{\Delta}_{2|0}^{(n_0)}(X_0)-\widehat{\Delta}_{2|0}^{(n)}(X_0)\right)\mathbb{I}\left(\widehat{d}_0^{(n)}(X_0)=2\right)\right]\\
=&\min\{0,\widehat{\Delta}_{1|0}^{(n)}(X_0),\widehat{\Delta}_{2|0}^{(n)}(X_0)\}-\min\{0,\overline{\Delta}_{1|0}^{(n_0)}(X_0),\overline{\Delta}_{2|0}^{(n_0)}(X_0)\}\\
&\quad+\left(\overline{\Delta}_{1|0}^{(n_0)}(X_0)-\widehat{\Delta}_{1|0}^{(n)}(X_0)\right)\mathbb{I}\left(\widehat{d}_0^{(n)}(X_0)=1\right)+\left(\overline{\Delta}_{2|0}^{(n_0)}(X_0)-\widehat{\Delta}_{2|0}^{(n)}(X_0)\right)\mathbb{I}\left(\widehat{d}_0^{(n)}(X_0)=2\right)\\
\in&\left[-2\left|\overline{\Delta}_{1|0}^{(n_0)}(X_0)-\widehat{\Delta}_{1|0}^{(n)}(X_0)\right|-2\left|\overline{\Delta}_{2|0}^{(n_0)}(X_0)-\widehat{\Delta}_{2|0}^{(n)}(X_0)\right|,\right.\\
&\quad\left.2\left|\overline{\Delta}_{1|0}^{(n_0)}(X_0)-\widehat{\Delta}_{1|0}^{(n)}(X_0)\right|+2\left|\overline{\Delta}_{2|0}^{(n_0)}(X_0)-\widehat{\Delta}_{2|0}^{(n)}(X_0)\right|\right].
\end{aligned}
$$

Moreover, according to Assumption 5 and Lemma 5, we have

$$
\begin{aligned}
\sup_{X_0}\left\{\widehat{\Delta}_{1|0}^{(n)}(X_0)-\overline{\Delta}_{1|0}^{(n)}(X_0)\right\}=o_p(n^{-1/4}),\\
\sup_{X_0}\left\{\widehat{\Delta}_{2|0}^{(n)}(X_0)-\overline{\Delta}_{2|0}^{(n)}(X_0)\right\}=o_p(n^{-1/4}).
\end{aligned}
$$

Taking supremum over X_0 and n_0, we can get the result. ■

Finally, we are ready to prove Theorem 5.

Proof [Proof of Theorem 5] Firstly, it is easy to see that

$$
\begin{aligned}
Q_0^{\widehat{d}^{(n)}}(X_0)&=\mathbb{E}\left[Err_0\mathbb{I}(\widehat{d}_0^{(n)}(X_0)=0)+Q_1^{\widehat{d}_1^{(n)}}\mathbb{I}(\widehat{d}_0^{(n)}(X_0)=1)+Err_2\mathbb{I}(\widehat{d}_0^{(n)}(X_0)=2)\mid X_0\right]\\
&\geq\min_{d_0}\ \mathbb{E}\left[Err_0\mathbb{I}(d_0(X_0)=0)+Q_1^{\widehat{d}_1^{(n)}}\mathbb{I}(d_0(X_0)=1)+Err_2\mathbb{I}(d_0(X_0)=2)\mid X_0\right]\\
&\geq\min_{d_0}\ \mathbb{E}\left[Err_0\mathbb{I}(d_0(X_0)=0)+Q_1^{d_1^*}\mathbb{I}(d_0(X_0)=1)+Err_2\mathbb{I}(d_0(X_0)=2)\mid X_0\right]\\
&=Q_0^{d^*}(X_0).
\end{aligned}
$$

Then we proceed to estimate the convergence rate. By Lemma 10 we have

$$\begin{aligned}
\sup_{X_0,n_0}\Bigg\{&\mathbb{E}\left[Err_0\mathbb{I}(\widehat{d}_0^{(n)}(X_0)=0)+Q_1^{\widehat{d}_1^{(n_0)}}\mathbb{I}(\widehat{d}_0^{(n)}(X_0)=1)+Err_2\mathbb{I}(\widehat{d}_0^{(n)}(X_0)=2)\right]\\
&-\mathbb{E}\left[Err_0\mathbb{I}(\overline{d}_0^{(n_0)}(X_0)=0)+Q_1^{\widehat{d}_1^{(n_0)}}\mathbb{I}(\overline{d}_0^{(n_0)}(X_0)=1)+Err_2\mathbb{I}(\overline{d}_0^{(n_0)}(X_0)=2)\right]\Bigg\} \qquad (11)\\
=&o_p(n^{-1/4}).
\end{aligned}$$

We mention that (11) shows that we have a uniform convergence rate, which is independent of X_0 and the choice of $\widehat{d}_1^{(n_0)}$.

Also, we have

$$\begin{aligned}
&\mathbb{E}\left[Err_0\mathbb{I}(\overline{d}_0^{(n_0)}(X_0)=0)+Q_1^{\widehat{d}_1^{(n_0)}}\mathbb{I}(\overline{d}_0^{(n_0)}(X_0)=1)+Err_2\mathbb{I}(\overline{d}_0^{(n_0)}(X_0)=2)\mid X_0\right]\\
&\quad-\mathbb{E}\left[Err_0\mathbb{I}(d_0^*(X_0)=0)+Q_1^*\mathbb{I}(d_0^*(X_0)=1)+Err_2\mathbb{I}(d_0^*(X_0)=2)\mid X_0\right]\\
=&\min_{d_0}\ \mathbb{E}\left[Err_0\mathbb{I}(d_0(X_0)=0)+Q_1^{\widehat{d}_1^{(n_0)}}\mathbb{I}(d_0(X_0)=1)+Err_2\mathbb{I}(d_0(X_0)=2)\mid X_0\right]\\
&\quad-\mathbb{E}\left[Err_0\mathbb{I}(d_0^*(X_0)=0)+Q_1^*\mathbb{I}(d_0^*(X_0)=1)+Err_2\mathbb{I}(d_0^*(X_0)=2)\mid X_0\right]\\
\leq&\mathbb{E}\left[(Q_1^{\widehat{d}_1^{(n_0)}}-Q_1^*)\mathbb{I}(d_0^*(X_0)=1)\mid X_0\right]\\
\leq&\sup_{(X_0,X_1)}\left|Q_1^{\widehat{d}_1^{(n_0)}}-Q_1^*\right|.
\end{aligned}$$

As a result, taking supremum over X_0 and using Theorem 4, we have

$$\begin{aligned}
\sup_{X_0}\Bigg\{&\mathbb{E}\left[Err_0\mathbb{I}(\overline{d}_0^{(n_0)}(X_0)=0)+Q_1^{\widehat{d}_1^{(n_0)}}\mathbb{I}(\overline{d}_0^{(n_0)}(X_0)=1)+Err_2\mathbb{I}(\overline{d}_0^{(n_0)}(X_0)=2)\mid X_0\right]\\
&-\mathbb{E}\left[Err_0\mathbb{I}(d_0^*(X_0)=0)+Q_1^*\mathbb{I}(d_0^*(X_0)=1)+Err_2\mathbb{I}(d_0^*(X_0)=2)\mid X_0\right]\Bigg\}\\
=&o_p(n_0^{-1/4}). \qquad (12)
\end{aligned}$$

Set $n_0=n$ in (12), and add (11) and (12) together. Noticing that the $o_p(n_0^{-1/4})$ term is also independent of X_0, we have

$$\sup_{X_0}\left\{Q_0^{\widehat{d}^{(n)}}(X_0)-Q_0^{d^*}(X_0)\right\}=o_p(n^{-1/4}).$$

■

A.4 Proof of the convergence of m

Proof [Proof of Theorem 6] We only prove the convergence of $\widehat{m}_1, \widehat{m}_{12}$. Firstly, by the definition of $\overline{m}_1$ and m_1^*, we have

$$
\begin{aligned}
&\left|\overline{m}_1^{(n)}(X_0, X_1) - m_1^*(X_0, X_1)\right| \\
=&\left|\mathbb{E}\left[\frac{L(Y, m(X_0, X_1))}{\widehat{\pi}^{(n)}((X_0), 1)}\mathbb{I}(S_1 = 1) \mid X_0, X_1\right] - \mathbb{E}\left[\frac{L(Y, m(X_0, X_1))}{\pi((X_0), 1)}\mathbb{I}(S_1 = 1) \mid X_0, X_1\right]\right| \\
\leq&\mathbb{E}\left[\left|\frac{L(Y, m(X_0, X_1))}{\widehat{\pi}^{(n)}((X_0), 1)} - \frac{L(Y, m(X_0, X_1))}{\pi((X_0), 1)}\right| \mid X_0, X_1\right] \\
\leq&\mathbb{E}\left[\frac{c_2}{c_0^2}\left|\widehat{\pi}^{(n)}((X_0), 1) - \pi((X_0), 1)\right| \mid X_0, X_1\right].
\end{aligned}
$$

Taking supremum over X_0 and using Assumption 2 and Assumption 7, we can get the desired result for $\widehat{m}_1$.

For $\widehat{m}_{12}$, we have

$$
\begin{aligned}
&\left|\overline{m}_{12}^{(n)}(X_0, X_1, X_2) - m_{12}^*(X_0, X_1, X_2)\right| \\
\leq&\mathbb{E}\left[\left|\frac{L(Y, m(X_0, X_1, X_2))}{\widehat{\pi}^{(n)}((X_0^i), 1)\cdot\widehat{\pi}^{(n)}((X_0^i, S_1^i X_1^i, S_1^i), 2)} - \frac{L(Y, m(X_0, X_1, X_2))}{\pi((X_0^i), 1)\cdot\pi((X_0^i, S_1^i X_1^i, S_1^i), 2)}\right| \mid X_0, X_1, X_2\right] \\
\leq&\mathbb{E}\left[\frac{c_2}{c_0^2}\left|\widehat{\pi}^{(n)}((X_0^i), 1)\cdot\widehat{\pi}^{(n)}((X_0^i, S_1^i X_1^i, S_1^i), 2) - \pi((X_0^i), 1)\cdot\pi((X_0^i, S_1^i X_1^i, S_1^i), 2)\right| \mid X_0, X_1, X_2\right] \\
=&\frac{c_2}{c_0^2}\mathbb{E}\left[\left|\widehat{\pi}^{(n)}((X_0^i), 1)\cdot\left(\widehat{\pi}^{(n)}((X_0^i, S_1^i X_1^i, S_1^i), 2) - \pi((X_0^i, S_1^i X_1^i, S_1^i), 2)\right)\right.\right. \\
&\left.\left.+\left(\widehat{\pi}^{(n)}((X_0^i), 1) - \pi((X_0^i), 1)\right)\cdot\pi((X_0^i, S_1^i X_1^i, S_1^i), 2)\right| \mid X_0, X_1, X_2\right] \\
\leq&\frac{c_2}{c_0^2}\mathbb{E}\left[\left|\widehat{\pi}^{(n)}((X_0^i, S_1^i X_1^i, S_1^i), 2) - \pi((X_0^i, S_1^i X_1^i, S_1^i), 2)\right| + \left|\widehat{\pi}^{(n)}((X_0^i), 1) - \pi((X_0^i), 1)\right| \mid X_0, X_1, X_2\right].
\end{aligned}
$$

Taking supremum over X_0, X_1, X_2 and using Assumption 2 and Assumption 7, we can get the desired result for $\widehat{m}_{12}$. ■

A.5 Detailed Comparable Methods

We are mainly in examining the performance of the proposed methods, i.e., COST-Q with or without m given. However, to serve as a benchmark, we also include the following methods as competitors.

1. **BOWL**: Kosorok (Zhao et al., 2015) proposed a method for addressing informative missing data in the dynamic treatment regime (DTR) problem. When adapted to the Sequential Lab Test problem, this method can be regarded as a benchmark. Specifically, the BOWL method assigns a weight to each action. For instance, when training d_1, BOWL assigns a weight of $\pi((X_0, S_1, X_1, S_1), 2)$ to samples with $S_2 = 1$, and a weight of $1 - \pi((X_0, S_1, X_1, S_1), 2)$ to all other samples. Using this approach, the

method defines

$$
\begin{aligned}
&\widehat{\Delta}_{12|1}(X_0, X_1)\\
&= \underset{\Delta:\mathbb{R}^{p_0+p_1}\to\mathbb{R}}{\arg\min} \sum_{i=1}^{n} \left[Err_{12}^i \cdot \frac{I(S_2^i = 1)}{\widehat{\pi}((X_0, X_1, 1), 2)} - Err_1^i \cdot \frac{I(S_2^i = 0)}{1 - \widehat{\pi}((X_0, X_1, 1), 2)} - \Delta(X_0^i, X_1^i) \right]^2 \cdot \mathbb{I}(S_1^i = 1),
\end{aligned}
$$

as the estimator of d_1. d_0 is estimated in a similar way.

2. **Only-complete**: We may only use the complete data to estimate the outcome, thus addressing the problem of missing data. We can then estimate the outcome using the complete data (i.e., samples with $S_1^i = S_2^i = 1$) and use the same method as in COST-Q to estimate the optimal treatment regime. There is no need of estimating $\widehat{\pi}$ and $\widehat{Err}$ in this method.

3. **One-time**: We may directly choose a choice of tests from $\{\{0\}, \{1\}, \{2\}, \{12\}\}$ based on X_0. We train regression models $\widehat{\Delta}_1, \widehat{\Delta}_2, \widehat{\Delta}_{12}$ to predict the expectation of $Err_1 - Err_0$, $Err_2 - Err_0$, $Err_{12} - Err_0$, and choose the choice of tests based on the prediction. Moreover, to address the problem of missing data, we may give weights to the samples based on the missing pattern. $\widehat{\Delta}_1, \widehat{\Delta}_2, \widehat{\Delta}_{12}$ are trained via

$$
\begin{aligned}
\widehat{\Delta}_1(X_0) =& \underset{\Delta:\mathbb{R}^{p_0}\to\mathbb{R}}{\arg\min} \sum_{i=1}^{n} \left[Err_1^i \cdot \frac{I(S_1^i = 1)}{\widehat{\pi}((X_0), 1)} - Err_0^i - \Delta(X_0^i) \right]^2,\\
\widehat{\Delta}_2(X_0) =& \underset{\Delta:\mathbb{R}^{p_0}\to\mathbb{R}}{\arg\min} \sum_{i=1}^{n} \left[Err_2^i \cdot \frac{I(S_2^i = 1)}{\widehat{\pi}((X_0, S_1X_1, S_1), 2)} - Err_0^i - \Delta(X_0^i) \right]^2,\\
\widehat{\Delta}_{12}(X_0) =& \underset{\Delta:\mathbb{R}^{p_0}\to\mathbb{R}}{\arg\min} \sum_{i=1}^{n} \left[Err_{12}^i \cdot \frac{I(S_1^i = 1, S_2^i = 1)}{\widehat{\pi}((X_0), 1)\widehat{\pi}((X_0, S_1X_1, S_1), 2)} - Err_0^i - \Delta(X_0^i) \right]^2.
\end{aligned}
$$