Journal of Data Science 0 (0), 1–23
July 2025

DOI: 10.6339/25-JDS1200
Statistical Data Science

Q-learning with Compound Outcome and Mixed Misclassification and Measurement Error in Covariates

Yasin Khadem Charvadeh[1] and Grace Y. Yi[1,2,*]

[1]Department of Statistical and Actuarial Sciences, University of Western Ontario, London, Ontario, Canada

[2]Department of Computer Science, University of Western Ontario, London, Ontario, Canada

Abstract

Precision medicine is an innovative approach that aims to customize medical treatments and interventions to patients based on their individual characteristics. Several estimation techniques, including Q-learning, have been developed to determine optimal treatment rules. However, the applicability of these methods depends on the availability of precisely measured variables. This study extends the scope of Q-learning to incorporate compound outcomes, deviating from the commonly assumed univariate outcomes, and further accommodates data with mismeasurement in both binary and continuous covariates. Two methods are described to mitigate the impact of mismeasurement. Numerical studies reveal that mismeasurement in covariates leads to notable estimation bias in parameters indexing the optimal treatment, yet the methods addressing the mismeasured effects yield improved results.

Keywords *compound outcome; dynamic treatment regimes; estimating function; misclassification; measurement error; Q-learning; regression calibration; regression models*

1 Introduction

Dynamic treatment regimes (DTRs) refer to individualized treatment strategies that adapt and evolve over time based on a patient's changing characteristics, responses to prior treatments, and other relevant factors. These regimes are designed to guide healthcare providers in making treatment decisions that optimize patient outcomes by accounting for the dynamic nature of diseases and the uniqueness of each patient. Unlike standardized treatment approaches that are often based on average outcomes observed in clinical trials, DTRs recognize that individual responses to treatment can vary due to factors such as genetic differences, co-existing health conditions, lifestyle factors, and personal preferences. A fixed treatment plan may not always provide the best outcome for every patient. By integrating available data and regularly updating treatment decisions based on patient progress, DTRs aim to enhance the precision and effectiveness of medical interventions, ultimately improving patient outcomes and the overall quality of care.

Most methods for determining the optimal DTR primarily focus on univariate outcomes, yet some applications require considering multiple outcomes to better reflect the complexity of the problem. Studies addressing multi-dimensional outcomes include Lizotte et al. (2010) and Wang et al. (2012). In addition to the complexity of outcomes, covariate mismeasurement poses another challenge in determining optimal DTRs. Spicker and Wallace (2020) examined the impact

---

[*] Corresponding author. Email: gyi5@uwo.ca.

Received August 30, 2024; Accepted September 21, 2025

of continuous covariate mismeasurement within the dynamic weighted ordinary least squares (dWOLS) framework, employing regression calibration to improve parameter estimates. Khadem Charvadeh and Yi (2024a) explored the consequences of binary covariate misclassification on Q-learning and proposed methods to mitigate its impact. However, these studies considered only misclassified discrete covariates or mismeasured continuous covariates in isolation, without addressing scenarios in which both types of errors occur simultaneously.

To bridge this gap, our study investigates the combined effects of misclassified discrete covariates and mismeasured continuous covariates on Q-learning. We extend the frameworks of Spicker and Wallace (2020) and Khadem Charvadeh and Yi (2024a) by moving beyond univariate outcomes to incorporate compound outcomes. Furthermore, building upon the estimating equations method proposed by Khadem Charvadeh and Yi (2024a), which was originally designed to mitigate binary covariate misclassification, we generalize it to accommodate mixed discrete and continuous covariate errors. Using the modeling approach of Lizotte et al. (2010), we analyze the impact of these mixed errors on parameter estimation. By addressing both misclassification and measurement error within a unified framework, our study provides a more comprehensive methodology for handling mismeasured covariates in DTR estimation.

This research is partly motivated by the analysis of a publicly available COVID-19 dataset, with a focus on assessing both the health and economic costs of the crisis. Our study provides insights into the optimal balance between health and economic considerations. By examining the interplay between COVID-19 data and various country-specific factors, we seek to identify effective preventive policies that minimize the overall impact of the pandemic. Given the ubiquity of measurement error in applications, we further investigate its potential detrimental effects on determining effective preventive policies.

The remainder of the paper is organized as follows. Section 2 introduces the basic notations and the implementation of Q-learning with bivariate outcomes. Section 3 presents the mismeasurement models and simulation studies to assess the mismeasurement effects. Sections 4 and 5 introduce correction strategies to reduce bias caused by applying a naive Q-learning procedure that disregards the mismeasurement effects. Section 6 assesses the performance of the correction strategies through simulation studies. Section 7 applies the methods described in Sections 2, 4, and 5 to a COVID-19 dataset, and Section 8 concludes the manuscript. Additional details are deferred to the online Supplementary Material.

2 Q-learning with Bivariate Loss/Reward Functions

Q-learning in DTRs focuses on determining optimal treatment strategies for individuals as their health conditions evolve over time. The key idea in Q-learning for DTRs is "backward induction," which involves working backward from the final stage to each of the earlier stages, where we calculate the expected cumulative reward, usually a measure of health outcome, for each possible treatment at each time point. The algorithm identifies the optimal treatment at each stage in a backward manner. While this approach allows for the development of personalized treatment plans that adapt to an individual's changing health conditions and address possible treatment delayed effects, it is often used for settings with a single health outcome. Nevertheless, problems involving bivariate outcomes are also of particular interest, as analyzing each outcome separately with marginal methods does not capture their joint complexity. In what follows, we describe how Q-learning can be applied to bivariate outcomes.

Suppose that the study has K stages. For $k = 1, \ldots, K$, let the first and second outcomes at the end of stage k be denoted by Y_{k1} and Y_{k2}, respectively. Assume that larger cumulative outcomes are more desirable. For $k = 1, \ldots, K$, let A_k denote the binary action taken at the beginning of stage k, taking value 0 or 1; and let X_k and C_k respectively denote error-prone binary and continuous covariates, which are both scalar, where X_k takes on value 0 or 1. Let Z_k denote the vector of precisely measured covariates. Covariates associated with each stage are measured prior to the receipt of the treatment at that stage. For $k = 1, \ldots, K$, let $\overline{X}_k = \Big\{X_1, \ldots, X_k\Big\}$, $\overline{C}_k = \Big\{C_1, \ldots, C_k\Big\}$, $\overline{Z}_k = \Big\{Z_1, \ldots, Z_k\Big\}$, and $\overline{A}_k = \Big\{A_1, \ldots, A_k\Big\}$.

The first and second outcomes at stage k can be represented as a function, say $g_j(\cdot)$ with $j = 1, 2$, of the history of the treatment, $\overline{A}_{k-1}$, together with the current treatment A_k, and the history of the covariates, $\overline{X}_k$, $\overline{C}_k$ and $\overline{Z}_k$, as well as the covariates X_{k+1}, C_{k+1} and Z_{k+1} in the next stage. That is, $Y_{kj} = g_j(\overline{A}_k, \overline{X}_{k+1}, \overline{C}_{k+1}, \overline{Z}_{k+1})$ for $j = 1, 2$ and $k = 1, \ldots, K$, where X_{k+1}, C_{k+1} and Z_{k+1} are null when $k = K$.

2.1 Composite Q-functions

To construct a sequence of optimal decision rules, we first define Q-functions for the first and second outcomes for K stages separately. For $j = 1, 2$, the Q-function for stage K is defined as:

$$Q_K^j(\overline{A}_K, \overline{X}_K, \overline{C}_K, \overline{Z}_K) = E(Y_{Kj} \mid \overline{A}_K, \overline{X}_K, \overline{C}_K, \overline{Z}_K),$$

and for stage k with $k = K - 1, \ldots, 1$, Q-functions are defined as

$$Q_k^j(\overline{A}_k, \overline{X}_k, \overline{C}_k, \overline{Z}_k) = E\Big\{Y_{kj} + \hat{Y}_{(k+1)j} \;\Big|\; \overline{A}_k, \overline{X}_k, \overline{C}_k, \overline{Z}_k\Big\}, \tag{1}$$

with $\hat{Y}_{(k+1)j} = \max_{a_{k+1}} E(Y_{(k+1)j} \mid \overline{A}_k, \overline{X}_{k+1}, \overline{C}_{k+1}, \overline{Z}_{k+1}, a_{k+1})$, called a pseudo-outcome. Here, Q-functions are defined recursively in a backward direction; and in the pseudo-outcome $\hat{Y}_{(k+1)j}$, $\overline{A}_{k+1}$ is written as $\overline{A}_k \cup \{a_{k+1}\}$ so that the future treatment a_{k+1} for stage k can be evaluated to find the optimal one.

To delineate how these conditional expectations are influenced by the history of the treatment and covariates, we employ regression approaches, such as linear regression models. For $j = 1, 2$, consider the regression model

$$Q_k^j(\overline{A}_k, \overline{X}_k, \overline{C}_k, \overline{Z}_k) = \beta_{kj}^{\mathrm{T}} H_{k0} + (\psi_{kj}^{\mathrm{T}} H_{k1}) A_k \quad \text{for} \quad k = K, \ldots, 1, \tag{2}$$

where we rewrite $\overline{A}_{k-1} \cup \overline{X}_k \cup \overline{C}_k \cup \overline{Z}_k$ as $\{H_{k0}, H_{k1}\}$, with H_{k0} representing the covariates that have a predictive effect on the outcome and H_{k1} standing for the covariates that interact with the treatment; H_{k0} and H_{k1} may include a constant, or the intercept, and they may include the same covariates. For $j = 1, 2$ and $k = K, \ldots, 1$, β_{kj} and ψ_{kj} are the regression coefficients, and we write $\theta_{k1} = (\beta_{k1}^{\mathrm{T}}, \psi_{k1}^{\mathrm{T}})^{\mathrm{T}}$ and $\theta_{k2} = (\beta_{k2}^{\mathrm{T}}, \psi_{k2}^{\mathrm{T}})^{\mathrm{T}}$.

With the Q-functions defined for the two outcomes in (1), one approach is to derive optimal decision rules separately by using the Q-functions in (2) for $j = 1, 2$, with $k = K, \ldots, 1$. Specifically, for $j = 1, 2$, consider $d_{jk} = \arg\max_{a_k} Q_k^j(\overline{A}_{k-1}, \overline{X}_k, \overline{C}_k, \overline{Z}_k, a_k)$, which we refer to as marginally optimal decision rules as they ignore the intrinsic association between the two outcomes. Alternatively, one may consider a joint criterion that simultaneously evaluates both Q-functions:

$$d_k = \arg\max_{a_k} \begin{pmatrix} Q_k^1(\overline{A}_{k-1}, \overline{X}_k, \overline{C}_k, \overline{Z}_k, a_k) \\ Q_k^2(\overline{A}_{k-1}, \overline{X}_k, \overline{C}_k, \overline{Z}_k, a_k) \end{pmatrix} \quad \text{for} \quad k = K, \ldots, 1. \tag{3}$$

However, optimal treatment for the first outcome is not necessarily optimal for the second outcome, and vice versa. Thus, joint maximization (3) can be fruitless. To get around this issue, we convert the maximization of the two objective functions in (3) to the maximization of a single objective function, formed by a linear combination of the two objective functions in (3).

To this end, consider a pre-specified weight parameter δ, taking a value between 0 and 1 to show the relative importance of the first and second outcomes, and for $k = K, \ldots, 1$, define the composite Q-function:

$$Q_k(\overline{A}_k, \overline{X}_k, \overline{C}_k, \overline{Z}_k, \delta) = \delta Q_k^1(\overline{A}_k, \overline{X}_k, \overline{C}_k, \overline{Z}_k) + (1-\delta) Q_k^2(\overline{A}_k, \overline{X}_k, \overline{C}_k, \overline{Z}_k), \tag{4}$$

where $Q_k^j(\overline{A}_k, \overline{X}_k, \overline{C}_k, \overline{Z}_k)$ is determined by (1) for $j = 1, 2$ and $k = K, \ldots, 1$.

The weight parameter δ allows us to adjust the relative importance of each outcome in the composite Q-function. To emphasize the first outcome, we assign δ a value closer to 1; to put more weight on the second outcome, we take δ to be closer to 0. By combining the two Q-functions in (3) using a weight parameter δ, we provide a way to balance the two objectives in order to choose an optimal decision rule that performs reasonably well with respect to both $Q_{k+1}^1(\cdot)$ and $Q_{k+1}^2(\cdot)$. In other words, by adjusting the value of δ, we can prioritize one objective over the other, or strike a trade-off between the two.

Consequently, for any given δ, combined optimal decision rules, denoted d_k, are determined by

$$d_k = \arg\max_{a_k} \; Q_k(\overline{A}_{k-1}, \overline{X}_k, \overline{C}_k, \overline{Z}_k, a_k, \delta) \qquad \text{for} \quad k = K, \ldots, 1. \tag{5}$$

The implementation of (5) hinges on the knowledge of the Q-functions for the first and second outcomes, $Q_k^j(\overline{A}_k, \overline{X}_k, \overline{C}_k, \overline{Z}_k)$ with $j = 1, 2$ and $k = K, \ldots, 1$, which can be typically modeled by regression models, such as (2), i.e., the composite Q-function (4) for $k = K, \ldots, 1$, can be expressed as

$$Q_k(\overline{A}_k, \overline{X}_k, \overline{C}_k, \overline{Z}_k, \delta) = \delta \left\{ \beta_{k1}^{\mathrm{T}} H_{k0} + (\psi_{k1}^{\mathrm{T}} H_{k1}) A_k \right\} + (1-\delta) \left\{ \beta_{k2}^{\mathrm{T}} H_{k0} + (\psi_{k2}^{\mathrm{T}} H_{k1}) A_k \right\}. \tag{6}$$

This form will be used in the following development.

2.1.1 Parameter Determination

To determine optimal DTRs, we first estimate the model parameters in the composite Q-functions in (6), which can be achieved by separately estimating parameters associated with (2) using a dataset consisting of n independently and identically distributed (i.i.d.) trajectories, each of the form $\mathcal{D}_i \triangleq \left\{ \{A_{ki}, X_{ki}, C_{ki}, Z_{ki}, Y_{k1i}, Y_{k2i}\} : k = K, \ldots, 1 \right\}$ for $i = 1, \ldots, n$. Let $\mathcal{D} \triangleq \cup_{i=1}^n \mathcal{D}_i$.

To be specific, for $\delta = 1$ and $\delta = 0$, we respectively solve

$$\hat{\theta}_{k1} = \arg\min_{\theta_{k1}} \frac{1}{n} \sum_{i=1}^n \left[\hat{Y}_{k1i} - Q_k^1(\overline{A}_{ki}, \overline{X}_{ki}, \overline{C}_{ki}, \overline{Z}_{ki}; \theta_{k1}) \right]^2 \tag{7}$$

and

$$\hat{\theta}_{k2} = \arg\min_{\theta_{k2}} \frac{1}{n} \sum_{i=1}^n \left[\hat{Y}_{k2i} - Q_k^2(\overline{A}_{ki}, \overline{X}_{ki}, \overline{C}_{ki}, \overline{Z}_{ki}; \theta_{k2}) \right]^2, \tag{8}$$

where for stage K, $\hat{Y}_{K1i} = Y_{K1i}$, $\hat{Y}_{K2i} = Y_{K2i}$; and for $k = K-1, \ldots, 1$ and $j = 1, 2$,

$$\hat{Y}_{kji} = Y_{kji} + Q^{j}_{k+1}(\overline{A}_{ki}, \overline{X}_{(k+1)i}, \overline{C}_{(k+1)i}, \overline{Z}_{(k+1)i}, \hat{d}_{(k+1)i}; \hat{\theta}_{(k+1)j}), \tag{9}$$

representing stage k pseudo-outcomes for subject i. Consequently, we obtain the optimal decision rule

$$\begin{aligned}\hat{d}_{k+1} = \arg\max_{a_{k+1}} \big\{ & \delta Q^{1}_{k+1}(\overline{A}_{k}, \overline{X}_{k+1}, \overline{C}_{k+1}, \overline{Z}_{k+1}, a_{k+1}; \hat{\theta}_{(k+1)1}) \\ & + (1-\delta) Q^{2}_{k+1}(\overline{A}_{k}, \overline{X}_{k+1}, \overline{C}_{k+1}, \overline{Z}_{k+1}, a_{k+1}; \hat{\theta}_{(k+1)2}) \big\} \quad \text{for} \quad k = K-1, \ldots, 0.\end{aligned}$$

For stage K, (7) and (8) are basically derived from the least squares methods minimizing the squared difference between the observed outcome Y_{Kj} and its conditional expectation in (1), with $j = 1$ or 2. For stage k with $k = K-1, \ldots, 1$, (7) and (8) are applied to conceptually minimize the squared difference between the pseudo-outcome $\hat{Y}_{kj}$ in (9) and the corresponding Q-function.

2.2 Estimation Equation Method

Estimation of the model parameters in the Q-functions using (7) and (8) can be alternatively formulated from solving estimating equations. In what follows, we first discuss the estimation for stage K, utilizing the actually observed outcome measurements $\{Y_{Kji} : j = 1, 2; i = 1, \ldots, n\}$, and then using the conceptual pseudo-outcomes (9) for stage $k = K-1, \ldots, 1$.

For stage K and $j = 1, 2$, set $\ell_{Kji} = \{Y_{Kji} - Q^{j}_{K}(\overline{A}_{Ki}, \overline{X}_{Ki}, \overline{C}_{Ki}, \overline{Z}_{Ki}; \theta_{Kj})\}^2$. Define

$$\begin{aligned}& S_{Kj}(\theta_{Kj}; Y_{Kji}, \overline{A}_{Ki}, \overline{X}_{Ki}, \overline{C}_{Ki}, \overline{Z}_{Ki}) = \left(\frac{\partial \ell_{Kji}}{\partial \beta^{\mathrm{T}}_{Kj}}, \frac{\partial \ell_{Kji}}{\partial \psi^{\mathrm{T}}_{Kj}} \right)^{\mathrm{T}} \\ & \triangleq \left(S^{\mathrm{T}}_{K\beta_{Kj}}(\theta_{Kj}; Y_{Kji}, \overline{A}_{Ki}, \overline{X}_{Ki}, \overline{C}_{Ki}, \overline{Z}_{Ki}), S^{\mathrm{T}}_{K\psi_{Kj}}(\theta_{Kj}; Y_{Kji}, \overline{A}_{Ki}, \overline{X}_{Ki}, \overline{C}_{Ki}, \overline{Z}_{Ki}) \right)^{\mathrm{T}},\end{aligned} \tag{10}$$

where

$$\begin{aligned}& S_{K\beta_{Kj}}(\theta_{Kj}; Y_{Kji}, \overline{A}_{Ki}, \overline{X}_{Ki}, \overline{C}_{Ki}, \overline{Z}_{Ki}) \\ & = \{Y_{Kji} - Q^{j}_{K}(\overline{A}_{Ki}, \overline{X}_{Ki}, \overline{C}_{Ki}, \overline{Z}_{Ki}; \theta_{Kj})\} \frac{\partial Q^{j}_{K}(\overline{A}_{Ki}, \overline{X}_{Ki}, \overline{C}_{Ki}, \overline{Z}_{Ki}; \theta_{Kj})}{\partial \beta_{Kj}}\end{aligned} \tag{11}$$

and

$$\begin{aligned}& S_{K\psi_{Kj}}(\theta_{Kj}; Y_{Kji}, \overline{A}_{Ki}, \overline{X}_{Ki}, \overline{C}_{Ki}, \overline{Z}_{Ki}) \\ & \quad = \{Y_{Kji} - Q^{j}_{K}(\overline{A}_{Ki}, \overline{X}_{Ki}, \overline{C}_{Ki}, \overline{Z}_{Ki}; \theta_{Kj})\} \frac{\partial Q^{j}_{K}(\overline{A}_{Ki}, \overline{X}_{Ki}, \overline{C}_{Ki}, \overline{Z}_{Ki}; \theta_{Kj})}{\partial \psi_{Kj}}.\end{aligned} \tag{12}$$

With (2) employed, (11) and (12) are simplified as

$$S_{K\beta_{Kj}}(\theta_{Kj}; Y_{Kji}, \overline{A}_{Ki}, \overline{X}_{Ki}, \overline{C}_{Ki}, \overline{Z}_{Ki}) = \left[Y_{Kji} - \left\{ \beta^{\mathrm{T}}_{Kj} H_{K0} + (\psi^{\mathrm{T}}_{Kj} H_{K1}) A_{K} \right\} \right] H_{K0} \tag{13}$$

and

$$S_{K\psi_{Kj}}(\theta_{Kj}; Y_{Kji}, \overline{A}_{Ki}, \overline{X}_{Ki}, \overline{C}_{Ki}, \overline{Z}_{Ki}) = \left[Y_{Kji} - \left\{\beta_{Kj}^{\mathrm{T}} H_{K0} + (\psi_{Kj}^{\mathrm{T}} H_{K1}) A_K\right\}\right] H_{K1} A_K. \tag{14}$$

Both $S_{K\beta_{Kj}}(\theta_{Kj}; Y_{Kji}, \overline{A}_{Ki}, \overline{X}_{Ki}, \overline{C}_{Ki}, \overline{Z}_{Ki})$ and $S_{K\psi_{Kj}}(\theta_{Kj}; Y_{Kji}, \overline{A}_{Ki}, \overline{X}_{Ki}, \overline{C}_{Ki}, \overline{Z}_{Ki})$ are unbiased estimating functions, i.e., they have zero mean, and thus, solving the estimating equations

$$\sum_{i=1}^{n} S_{Kj}(\theta_{Kj}; Y_{Ki}, \overline{A}_{Ki}, \overline{X}_{Ki}, \overline{C}_{Ki}, \overline{Z}_{Ki}) = 0 \tag{15}$$

for θ_{Kj} yields a consistent estimator of θ_{Kj}, provided regularity conditions (Yi, 2017, Section 2.5).

Similarly, for stage k with $k = K-1, \ldots, 1$, and $j = 1, 2$, let $\hat{\ell}_{kji} = \left\{\hat{Y}_{kji} - Q_k^j(\overline{A}_{ki}, \overline{X}_{ki}, \overline{C}_{ki}, \overline{Z}_{ki}; \theta_{kj})\right\}^2$ and define $\hat{S}_{kj}(\theta_{kj}; \hat{Y}_{kji}, \overline{A}_{ki}, \overline{X}_{ki}, \overline{C}_{ki}, \overline{Z}_{ki}) = \left(\frac{\partial \hat{\ell}_{kji}}{\partial \beta_{kj}^{\mathrm{T}}}, \frac{\partial \hat{\ell}_{kji}}{\partial \psi_{kj}^{\mathrm{T}}}\right)^{\mathrm{T}}$. Then, an estimator of θ_{kj}, denoted $\hat{\theta}_{kj}$, is obtained by solving

$$\sum_{i=1}^{n} \hat{S}_{kj}(\theta_{kj}; \hat{Y}_{kji}, \overline{A}_{ki}, \overline{X}_{ki}, \overline{C}_{ki}, \overline{Z}_{ki}) = 0. \tag{16}$$

We conclude this section with comments on inference about the model parameters using the estimators derived from (7) and (8). When $k = K$, the resulting estimator $\hat{\theta}_{Kj}$ is the least squares estimator of θ_{Kj}, and thus inference about θ_{Kj} can be carried out in a usual way. For instance, coverage rates (CRs) of confidence intervals (CIs) for the Kth stage parameters in (7) and (8) can be calculated using either Wald-type (W-type) CIs or percentile bootstrap (PB) CIs. However, caution should be exercised when obtaining CIs for the estimators in stage k with $k < K$, as those estimators $\hat{\theta}_{kj}$ with $k = K-1, \ldots, 1$ are not the least squares estimators of the θ_{kj} due to the unobservable feature of pseudo-outcomes $\hat{Y}_{kji}$. For $k = K-1, \ldots, 1$, the pseudo-outcome, $\hat{Y}_{kji}$, may be a non-smooth function of $\hat{\theta}_{(k+1)j}$, which creates the non-regularity or weak non-regularity issue of parameters (Robins, 2004, pp.189–326). Non-regularity or weak non-regularity often distorts usual inferential procedures that are rooted in the asymptotic normal distribution. One approach to address non-regularity or weak non-regularity issue is to report double bootstrap (DB) CIs (Chakraborty and Moodie, 2013, Chapter 8).

3 Mismeasurement and Naive Analysis

As discussed by Khadem Charvadeh and Yi (2024a), the validity of the Q-learning procedure relies on the requirement of precisely measured variables and certain assumptions, such as the *stable unit treatment value assumption* (SUTVA) and the *no unmeasured confounders* (NUC) assumption. The SUTVA says that an individual's outcome is unaffected by the treatment allocation for other individuals, and the NUC assumption says that for $k = 1, \ldots, K$, conditional on the observed history $\left\{\overline{A}_{k-1}, \overline{X}_k, \overline{C}_k, \overline{Z}_k\right\}$, the treatment A_k is independent of any future covariates or outcomes $\left\{X_{k+1}, \ldots, X_K; C_{k+1}, \ldots, C_K; Z_{k+1}, \ldots, Z_K; Y_k\right\}$ (Chakraborty and Moodie, 2013). As noted by Khadem Charvadeh and Yi (2024a), the SUTVA may remain unaffected by the presence of error-prone covariates, but the NUC assumption does not necessarily hold for the observed surrogate measurements.

In this section, we consider the case where X_k and C_k are subject to mismeasurement for $k = K, \ldots, 1$. We examine numerically the impact of naively implementing the Q-learning procedure in Section 2, with the mismeasurement effects ignored.

3.1 Measurement Error and Misclassification Models

For $k = K, \ldots, 1$, let X^*_k and C^*_k, respectively, denote the observed versions of the true covariates X_k and C_k, and let $\overline{X}^*_k = \left\{X^*_1, \ldots, X^*_k\right\}$ and $\overline{C}^*_k = \left\{C^*_1, \ldots, C^*_k\right\}$. For $j = 0, 1$ and $l = 1 - j$, let $\pi_{jl,i} = P(X^*_{ki} = j \mid \overline{A}_{ki}, \overline{X}^*_{(k-1)i}, \overline{X}_{(k-1)i}, X_{ki} = l, \overline{C}^*_{ki}, \overline{C}_{ki}, \overline{Z}_{ki})$ denote the misclassification probabilities that may depend on either the true or mismeasured covariates or both. To model the misclassification probabilities, one may employ regression models for binary data, such as logistic regression models. For the misclassification process, it is convenient to assume that $P(X^*_{ki} = j \mid \overline{A}_{ki}, \overline{X}^*_{(k-1)i}, \overline{X}_{(k-1)i}, X_{ki} = l, \overline{C}^*_{ki}, \overline{C}_{ki}, \overline{Z}_{ki}) = P(X^*_{ki} = j \mid X_{ki} = l)$, enabling us to express misclassification probabilities as

$$\Pi = \begin{pmatrix} 1 - \pi_{10} & \pi_{01} \\ \pi_{10} & 1 - \pi_{01} \end{pmatrix}, \tag{17}$$

which corresponds to homogeneous misclassification across all subjects.

Now, we describe the measurement error process for the continuous covariates. Let $h(C^*_{ki} \mid \overline{A}_{ki}, \overline{C}^*_{(k-1)i}, \overline{C}_{ki}, \overline{X}^*_{ki}, \overline{X}_{ki}, \overline{Z}_{ki})$ denote the conditional probability density function of C^*_{ki}, given $\overline{A}_{ki} \cup \overline{C}^*_{(k-1)i} \cup \overline{C}_{ki} \cup \overline{X}^*_{ki} \cup \overline{X}_{ki} \cup \overline{Z}_{ki}$. Similar to the misclassification case, one may assume that $h(C^*_{ki} \mid \overline{A}_{ki}, \overline{C}^*_{(k-1)i}, \overline{C}_{ki}, \overline{X}^*_{ki}, \overline{X}_{ki}, \overline{Z}_{ki}) = h(C^*_{ki} \mid C_{ki})$ for simplicity. We can then use a parametric model, say $f(C^*_{ki} \mid C_{ki}; \alpha)$ with parameter α, to modulate $h(C^*_{ki} \mid C_{ki})$.

Here, we consider the classical additive measurement error model

$$C^*_{ki} = C_{ki} + e_{ki}, \tag{18}$$

where the error term e_{ki} is independent of all variables other than the C^*_{ki} and has mean 0 and variance σ^2_k. This model is most widely used in the literature of measurement error models (Carroll et al., 2006; Yi, 2017; Yi et al., 2021).

Similar to assumption (9) of Khadem Charvadeh and Yi (2024a), we assume that for $k = 1, \ldots, K$ and $j = 1, 2$,

$$h(Y_{kj} \mid \overline{Y}_{(k-1)j}, \overline{A}_K, \overline{X}_K, \overline{C}_K, \overline{Z}_K, \overline{X}^*_K, \overline{C}^*_K) = h(Y_{kj} \mid \overline{A}_k, \overline{X}_{k+1}, \overline{C}_{k+1}, \overline{Z}_{k+1}). \tag{19}$$

Assumption (19) says that at stage k, given the history of outcomes $\overline{Y}_{(k-1)j}$ and the information $\overline{A}_K \cup \overline{X}_K \cup \overline{C}_K \cup \overline{Z}_K \cup \overline{X}^*_K \cup \overline{C}^*_K$ over the entire course, the conditional distribution of Y_{kj} depends only on the history $\overline{A}_k \cup \overline{X}_k \cup \overline{C}_k \cup \overline{Z}_k$ at stage k as well as the covariates X_{k+1}, C_{k+1} and Z_{k+1} at the next stage. Assumption (19) implies that the surrogates $\overline{X}^*_K$ and $\overline{C}^*_K$ carry no additional information for conducting inference about the response Y_{kj} if the true covariates $\overline{X}_k$ and $\overline{C}_k$ are given. This independence assumption is similar to the nondifferential misclassification mechanism (Yi, 2017, p.50), which is commonly made in the literature of measurement error models. It enables us to draw inferences about the true variables using their surrogate measurements.

We stress that our following development does not require knowledge of the marginal distribution of the unobserved true covariates. Instead, having the additive form in (18) and the variance σ^2_k of the error term e_{ki} is sufficient for constructing the term C^{***}_k to adjust the measurement error effects, as discussed in Section 5.1.1.

3.2 Naive Q-learning and Simulation Studies

To numerically assess the performance of the naive Q-learning procedure, which ignores the presence of mixed misclassification and measurement error, we consider a randomized treatment setting and set $K = 2$ in a way similar to Section 3.3 of Khadem Charvadeh and Yi (2024a).

For $j = 1, 2$, let $\mu_{Y_{2j}} = E(Y_{2j} \mid X_2, A_1, A_2, C_2, Z_2)$ and let $\mu_{Y_{1j}} = E(Y_{1j} \mid X_1, A_1, C_1, Z_1)$. Then, for $j = 1, 2$ consider models $Y_{2j} = \mu_{Y_{2j}} + \epsilon_{2j}$ and $Y_{1j} = \mu_{Y_{1j}} + \epsilon_{1j}$, where

$$\begin{aligned}\mu_{Y_{2j}} &= \eta_{0j} + \eta_{1j} Z_2 + \eta_{2j} X_2 + \eta_{3j} A_1 + \eta_{4j} A_2 + \eta_{5j} X_2 A_2 + \eta_{6j} C_2 A_2,\\ \mu_{Y_{1j}} &= \gamma_{0j} + \gamma_{1j} Z_1 + \gamma_{2j} X_1 + \gamma_{3j} A_1 + \gamma_{4j} X_1 A_1 + \gamma_{5j} C_1 A_1,\end{aligned}$$

and ϵ_{2j} and ϵ_{1j} are the error terms independently generated from $\mathcal{N}(0, 1)$ for $j = 1, 2$.

Two binary treatments A_1 and A_2, taking values 0 and 1, are generated independently from the Bernoulli distribution, Bernoulli(0.5). Error-free covariates Z_1 and Z_2 are independently generated using $Z_1 \sim$ Bernoulli(0.5) and $Z_2 \sim$ Bernoulli(0.5). Error-prone binary covariate X_1 is independently generated from Bernoulli(0.5), and error-prone binary covariate X_2 is generated from the conditional distribution $X_2 \mid A_1 \sim \text{Bernoulli}\left(\frac{\exp(\nu A_1)}{1+\exp(\nu A_1)}\right)$, with ν set as 0.45. Error-prone continuous covariates are independently generated using $C_1 \sim \mathcal{N}(0, 1)$ and $C_2 \sim \mathcal{N}(0, 1)$.

We consider three settings for the model parameters in the Q-functions, including regular, weak non-regular, and non-regular settings. In the regular setting, the parameter values are set as $(\eta_{01}, \eta_{11}, \eta_{21}, \eta_{31}, \eta_{41}, \eta_{51}, \eta_{61})^{\mathrm{T}} = (1.5, 0.25, 0.8, -0.25, -2, 1.5, 1.75)^{\mathrm{T}}$, $(\eta_{02}, \eta_{12}, \eta_{22}, \eta_{32}, \eta_{42}, \eta_{52}, \eta_{62})^{\mathrm{T}} = (0.5, 0.75, 1.5, -0.15, -1.2, 0.95, 1.1)^{\mathrm{T}}$, $(\gamma_{01}, \gamma_{11}, \gamma_{21}, \gamma_{31}, \gamma_{41}, \gamma_{51})^{\mathrm{T}} = (0.5, 0.15, 0.5, -1.5, 1.25, 0.95)^{\mathrm{T}}$, and $(\gamma_{02}, \gamma_{12}, \gamma_{22}, \gamma_{32}, \gamma_{42}, \gamma_{52})^{\mathrm{T}} = (0.75, 0.2, 0.85, -1.85, 0.85, 1.45)^{\mathrm{T}}$. In the weak non-regular setting, we take $(\eta_{01}, \eta_{11}, \eta_{21}, \eta_{31}, \eta_{41}, \eta_{51}, \eta_{61})^{\mathrm{T}} = (1.5, 0.25, 0.8, -0.25, -2, 2.02, 0)^{\mathrm{T}}$, $(\eta_{02}, \eta_{12}, \eta_{22}, \eta_{32}, \eta_{42}, \eta_{52}, \eta_{62})^{\mathrm{T}} = (0.5, 0.75, 1.5, -0.15, -1.2, 1.22, 0)^{\mathrm{T}}$, $(\gamma_{01}, \gamma_{11}, \gamma_{21}, \gamma_{31}, \gamma_{41}, \gamma_{51})^{\mathrm{T}} = (0.5, 0.15, 0.5, -1.5, 1.25, 0.95)^{\mathrm{T}}$, and $(\gamma_{02}, \gamma_{12}, \gamma_{22}, \gamma_{32}, \gamma_{42}, \gamma_{52})^{\mathrm{T}} = (0.75, 0.2, 0.85, -1.85, 0.85, 1.45)^{\mathrm{T}}$. In the non-regular setting, the parameter values are specified as $(\eta_{01}, \eta_{11}, \eta_{21}, \eta_{31}, \eta_{41}, \eta_{51}, \eta_{61})^{\mathrm{T}} = (1.5, 0.25, 0.8, -0.25, -0.2, 0.2, 0)^{\mathrm{T}}$, $(\eta_{02}, \eta_{12}, \eta_{22}, \eta_{32}, \eta_{42}, \eta_{52}, \eta_{62})^{\mathrm{T}} = (0.5, 0.75, 1.5, -0.15, -0.1, 0.1, 0)^{\mathrm{T}}$, $(\gamma_{01}, \gamma_{11}, \gamma_{21}, \gamma_{31}, \gamma_{41}, \gamma_{51})^{\mathrm{T}} = (0.5, 0.15, 0.5, -1.5, 1.25, 0.95)^{\mathrm{T}}$, and $(\gamma_{02}, \gamma_{12}, \gamma_{22}, \gamma_{32}, \gamma_{42}, \gamma_{52})^{\mathrm{T}} = (0.75, 0.2, 0.85, -1.85, 0.85, 1.45)^{\mathrm{T}}$.

The Q-functions for the two stages are specified as

$$Q_2^j(A_1, X_2, C_2, Z_2, A_2) = \beta_{02j} + \beta_{12j} Z_2 + \beta_{22j} X_2 + \beta_{32j} A_1 + (\psi_{02j} + \psi_{12j} X_2 + \psi_{22j} C_2) A_2,$$

and

$$Q_1^j(X_1, C_1, Z_1, A_1) = \beta_{01j} + \beta_{11j} Z_1 + \beta_{21j} X_1 + (\psi_{01j} + \psi_{11j} X_1 + \psi_{21j} C_1) A_1,$$

where for $j = 1, 2$, β_{k2j} (with $k = 0, 1, 2, 3$), β_{k1j} (with $k = 0, 1, 2$), ψ_{k2j} (with $k = 0, 1, 2$), and ψ_{k1j} (with $k = 0, 1, 2$) are regression parameters. Based on the coefficients of A_2 and A_1, the optimal DTR is given by the decision rules:

$$\begin{aligned} d_2 &= \text{sign}\Big\{\delta(\psi_{021} + \psi_{121} X_2 + \psi_{221} C_2) + (1 - \delta)(\psi_{022} + \psi_{122} X_2 + \psi_{222} C_2)\Big\};\\ d_1 &= \text{sign}\Big\{\delta(\psi_{011} + \psi_{111} X_1 + \psi_{211} C_1) + (1 - \delta)(\psi_{012} + \psi_{112} X_1 + \psi_{212} C_1)\Big\},\end{aligned}$$

where $\text{sign}(t) = 1$ if $t > 0$, and 0 otherwise.

With the misclassification matrix (17) and classical additive model (18), we generate surrogate values X_k^* of X_k and C_k^* of C_k with $k = 1, 2$. We consider three different settings for misclassification probabilities and measurement error degree, with $(\pi_{10}, \pi_{01}, \sigma_k^2)^{\mathrm{T}} = (0.1, 0.1, 1)^{\mathrm{T}}$, $(0.2, 0.2, 1.5)^{\mathrm{T}}$, or $(0.3, 0.3, 2)^{\mathrm{T}}$.

To run simulations, we use the preceding models to generate data of size $n = 2000$ and repeat 500 simulations for each parameter configuration. We implement the Q-learning algorithm similar to Section 3.3 of Khadem Charvadeh and Yi (2024a) to the observed data $\{Z_1, X_1^*, C_1^*, A_1, Y_{11}, Y_{12}, Z_2, X_2^*, C_2^*, A_2, Y_{21}, Y_{22}\}$, called the "naive method", as opposed to the Q-learning procedure in Section 2.1 to the true data $\{Z_1, X_1, C_1, A_1, Y_{11}, Y_{12}, Z_2, X_2, C_2, A_2, Y_{21}, Y_{22}\}$, called the "error-free least squares" (EFLS) method.

In Table 1, we report the numerical results for stages 1 and 2 over 500 simulations for the regular, weak non-regular, and non-regular settings, where "Bias" stands for the difference between the true parameter values and the average of their estimates over 500 simulations obtained from the EFLS or naive methods; "SE" represents the average of model-based standard errors (i.e., the standard errors of the least squares parameter estimators over 500 simulations); "ESE" shows the empirical standard error of the estimates, calculated by $\sqrt{\frac{1}{499}\sum_{k=1}^{500}(\mathrm{Est.}_{\cdot k} - \overline{\mathrm{Est.}})^2}$ with $\mathrm{Est.}_{\cdot k}$ representing the estimate for the kth simulation and $\overline{\mathrm{Est.}} = 500^{-1}\sum_{k=1}^{500}\mathrm{Est.}_{\cdot k}$; "MSE" displays the mean squared error of the estimates given by $\mathrm{Bias}^2 + \mathrm{SE}^2$; "WTCR" represents the coverage rate (CR) of 95% W-type CIs with the model-based standard errors used; "PBCR" represents the CR of 95% PB CIs; and "DBCR" represents the CR of 95% DB CIs. The PB CIs are created based on 1000 bootstrap iterations, and the DB CIs are derived from using 1000 first-stage and 100 second-stage bootstrap iterations, with a reduced number of iterations for the second stage to ease the computational burden. We assign the weight parameter δ a value of 0.9. The results demonstrate the favorable performance of the EFLS method in the regular setting, characterized by minimal biases and MSEs, as well as satisfactory CRs of 95% CIs for both stages 1 and 2. In the weak non-regular setting, the CRs of 95% PB CIs are fairly satisfactory, except for $\psi_{0k\delta}$. However, in the non-regular setting, these CRs exhibit under-coverage for both $\psi_{0k\delta}$ and $\psi_{2k\delta}$. We also examined W-type CIs for weak non-regular and non-regular settings in stage 1 (results are not reported here) and observed poor coverage.

The results obtained from the naive method for both stages 1 and 2 incur significant biases, high MSEs, and unacceptably low CRs. The extent of biases due to covariate mismeasurement intensifies as the degrees of mismeasurement increase. These findings underscore the importance of accounting for mismeasurement and implementing suitable correction methods to enhance the accuracy of estimation.

4 Mismeasurement Correction: Regression Calibration

Similar to Khadem Charvadeh and Yi (2024a), here we describe the application of the regression calibration (RC) method (e.g., Yi, 2017; Yi et al., 2021) for the case where internal validation data $\mathcal{D}_V = \Big\{\{A_{ki}, X_{ki}, X_{ki}^*, C_{ki}, C_{ki}^*, Z_{ki}, Y_{k1i}, Y_{k2i}\} : k = 1, \ldots, K;\ i \in \mathcal{V}\Big\}$ are available, in addition to the main study data with surrogate measurements together with measurements for other variables, $\mathcal{D}^* = \Big\{\{A_{ki}, X_{ki}^*, C_{ki}^*, Z_{ki}, Y_{k1i}, Y_{k2i}\} : k = 1, \ldots, K; i \in \mathcal{M}\Big\}$, where $\mathcal{M} \triangleq \{1, \ldots, n\}$ and $\mathcal{V} \subset \mathcal{M}$.

For $i \in \mathcal{M}$ and $k = K, \ldots, 1$, let $X_{ki}^{**} = E(X_{ki} \mid \overline{A}_{(k-1)i}, \overline{X}_{ki}^*, \overline{C}_{ki}^*, \overline{Z}_{ki})$ and $C_{ki}^{**} = E(C_{ki} \mid \overline{A}_{(k-1)i}, \overline{X}_{ki}^*, \overline{C}_{ki}^*, \overline{Z}_{ki})$.

Table 1: Simulation studies for demonstrating biased estimation of the naive method in contrast to the EFLS method: stages 1–2. Entries in bold are obtained from the setting without mismeasurements.

$(\pi_{10}, \pi_{01}, \sigma_k^2)$	Method		regular			weak non-regular			non-regular		
			$\psi_{0k\delta}$	$\psi_{1k\delta}$	$\psi_{2k\delta}$	$\psi_{0k\delta}$	$\psi_{1k\delta}$	$\psi_{2k\delta}$	$\psi_{0k\delta}$	$\psi_{1k\delta}$	$\psi_{2k\delta}$
		Stage 1 ($k = 1$)									
(0,0,0)	EFLS	Bias	**0.001**	**0.001**	**0.002**	**0.003**	**0.001**	**0.003**	**0.006**	**0.004**	**0.001**
		SE	**0.078**	**0.110**	**0.039**	**0.063**	**0.089**	**0.031**	**0.062**	**0.088**	**0.031**
		ESE	**0.092**	**0.113**	**0.040**	**0.081**	**0.091**	**0.033**	**0.079**	**0.094**	**0.034**
		MSE	**0.006**	**0.012**	**0.002**	**0.004**	**0.008**	**0.001**	**0.004**	**0.008**	**0.001**
		PBCR	**0.948**	**0.948**	**0.944**	**0.924**	**0.944**	**0.946**	**0.930**	**0.944**	**0.934**
		DBCR	**0.956**	**0.950**	**0.960**	**0.930**	**0.956**	**0.956**	**0.942**	**0.948**	**0.950**
(0.1,0.1,1)	Naive	Bias	0.119	0.235	0.499	0.159	0.242	0.503	0.124	0.237	0.501
		SE	0.075	0.107	0.027	0.070	0.099	0.025	0.070	0.099	0.025
		ESE	0.096	0.108	0.031	0.095	0.102	0.030	0.088	0.102	0.032
		MSE	0.020	0.067	0.250	0.030	0.068	0.254	0.020	0.066	0.252
		PBCR	0.792	0.392	0.000	0.572	0.296	0.000	0.698	0.300	0.000
		DBCR	0.708	0.258	0.000	0.500	0.208	0.000	0.628	0.238	0.000
(0.2,0.2,1.5)	Naive	Bias	0.253	0.475	0.689	0.312	0.488	0.692	0.258	0.491	0.694
		SE	0.076	0.107	0.021	0.074	0.104	0.020	0.074	0.104	0.020
		ESE	0.108	0.110	0.025	0.094	0.108	0.026	0.084	0.105	0.026
		MSE	0.070	0.237	0.475	0.103	0.249	0.479	0.072	0.252	0.482
		PBCR	0.282	0.000	0.000	0.058	0.000	0.000	0.114	0.000	0.000
		DBCR	0.258	0.000	0.000	0.074	0.002	0.000	0.124	0.000	0.000
(0.3,0.3,2)	Naive	Bias	0.398	0.731	0.801	0.451	0.723	0.800	0.371	0.728	0.800
		SE	0.077	0.108	0.017	0.076	0.108	0.017	0.076	0.108	0.017
		ESE	0.105	0.115	0.021	0.102	0.115	0.022	0.096	0.117	0.021
		MSE	0.164	0.546	0.642	0.209	0.534	0.640	0.143	0.542	0.640
		PBCR	0.020	0.000	0.000	0.002	0.000	0.000	0.002	0.000	0.000
		DBCR	0.036	0.000	0.000	0.004	0.000	0.000	0.016	0.000	0.000
		Stage 2 ($k = 2$)									
(0,0,0)	EFLS	Bias	**0.005**	**0.005**	**0.003**	**0.001**	**0.000**	**0.000**	**0.002**	**0.004**	**0.002**
		SE	**0.061**	**0.082**	**0.029**	**0.061**	**0.082**	**0.029**	**0.061**	**0.082**	**0.029**
		ESE	**0.061**	**0.081**	**0.029**	**0.059**	**0.078**	**0.030**	**0.058**	**0.077**	**0.028**
		MSE	**0.004**	**0.007**	**0.001**	**0.004**	**0.007**	**0.001**	**0.004**	**0.007**	**0.001**
		WTCR	**0.956**	**0.950**	**0.952**	**0.956**	**0.968**	**0.942**	**0.966**	**0.966**	**0.954**
(0.1,0.1,1)	Naive	Bias	0.167	0.283	0.840	0.231	0.391	0.000	0.024	0.040	0.000
		SE	0.086	0.116	0.029	0.071	0.096	0.024	0.062	0.084	0.021
		ESE	0.094	0.121	0.037	0.074	0.097	0.029	0.062	0.082	0.022
		MSE	0.035	0.094	0.706	0.058	0.162	0.001	0.004	0.009	0.000
		WTCR	0.516	0.332	0.000	0.108	0.022	0.892	0.938	0.922	0.950
(0.2,0.2,1.5)	Naive	Bias	0.339	0.578	1.167	0.460	0.780	0.001	0.045	0.081	0.000
		SE	0.094	0.129	0.025	0.077	0.105	0.021	0.063	0.086	0.017
		ESE	0.100	0.133	0.035	0.082	0.109	0.026	0.066	0.088	0.017
		MSE	0.124	0.351	1.363	0.218	0.619	0.000	0.006	0.014	0.000
		WTCR	0.060	0.004	0.000	0.000	0.000	0.888	0.880	0.842	0.952
(0.3,0.3,2)	Naive	Bias	0.505	0.873	1.348	0.672	1.169	0.000	0.065	0.117	0.001
		SE	0.099	0.136	0.022	0.080	0.111	0.018	0.064	0.088	0.014
		ESE	0.104	0.151	0.029	0.082	0.113	0.022	0.066	0.088	0.015
		MSE	0.265	0.781	1.818	0.458	1.379	0.000	0.008	0.021	0.000
		WTCR	0.002	0.000	0.000	0.000	0.000	0.876	0.822	0.728	0.944

Regression modeling techniques can be employed to facilitate the dependence of X_{ki}^{**} and C_{ki}^{**} on the history of the treatment and observed covariate measurements. Because X_{ki} is the binary covariate taking value 0 or 1, X_{ki}^{**} is alternatively expressed as $\pi_i^{(k)} \triangleq P(X_{ki} = 1 \mid \overline{A}_{(k-1)i}, \overline{X}_{ki}^{*}, \overline{C}_{ki}^{*}, \overline{Z}_{ki})$ and then we consider a logistic regression model:

$$\operatorname{logit} \pi_i^{(k)} = m_{X_k}(\overline{A}_{(k-1)i}, \overline{X}_{ki}^{*}, \overline{C}_{ki}^{*}, \overline{Z}_{ki}; \boldsymbol{\zeta}_{X_k}), \tag{20}$$

where $m_{X_k}(\cdot)$ is a specified function, and $\boldsymbol{\zeta}_{X_k}$ is the associated parameter.

To delineate C_{ki}^{**}, we employ a linear regression model

$$C_{ki}^{**} = m_{C_k}(\overline{A}_{(k-1)i}, \overline{X}_{ki}^{*}, \overline{C}_{ki}^{*}, \overline{Z}_{ki}; \boldsymbol{\zeta}_{C_k}), \tag{21}$$

where $m_{C_k}(\cdot)$ is a specified function, and $\boldsymbol{\zeta}_{C_k}$ is the associated parameter.

Next, regression models (20) and (21) are fitted using the validation data $\mathcal{D}_V$, resulting in the estimates of $\boldsymbol{\zeta}_{X_k}$ and $\boldsymbol{\zeta}_{C_k}$, denoted as $\hat{\boldsymbol{\zeta}}_{X_k}$ and $\hat{\boldsymbol{\zeta}}_{C_k}$, respectively. These estimates are then used to estimate X_{ki}^{**} and C_{ki}^{**}, and let $\hat{X}_{ki}^{**}$ and $\hat{C}_{ki}^{**}$ denote them, respectively. The implementation of the calibrated Q-learning algorithm can be modified from stages K to 1 with different treatments of X_{ki} and C_{ki}, where measurements for X_{ki} and C_{ki} are used for $i \in \mathcal{V}$, and $\hat{X}_{ki}^{**}$ and $\hat{C}_{ki}^{**}$ are used to replace X_{ki} and C_{ki} for $i \in \mathcal{M} \setminus \mathcal{V}$.

5 Mismeasurement Correction: Estimating Equation Approach

While regression calibration provides a simple approach to addressing mismeasurement in covariates, it does not always guarantee the consistency of estimators for the model parameters. In this section, we develop an alternative approach by employing estimating function theory, and present criteria for developing unbiased estimating functions. We first describe the unbiased estimating function approach for stage K, and then extend it to other stages to create working estimating functions.

5.1 Corrected Estimation Functions with Known Mismeasurement Degrees

Assume that the misclassification probabilities in (17) and the variance of e_{ki} in (18) are known. In what follows, we first describe how to correct for mismeasurement-induced bias in stage K, and then discuss estimation pertinent to stage k for $k = K-1, \ldots, 1$.

5.1.1 Estimation Related to Stage K

If the true covariates X_{ki} and C_{ki} for $k = 1, \ldots, K$ and $i = 1, \ldots, n$ are not available, but surrogate values X_{ki}^{*} and C_{ki}^{*} are instead collected, then directly using (15) with X_{ki} and C_{ki} replaced by X_{ki}^{*} and C_{ki}^{*} may result in inconsistent estimators, as illustrated in Section 3.2.

Our goal here is to construct an estimating function, say $S_{Kj}^{*}(\theta_{Kj}; Y_{Kji}, \overline{A}_{Ki}, \overline{X}_{Ki}^{*}, \overline{C}_{Ki}^{*}, \overline{Z}_{Ki})$, expressed in terms of the observed surrogate measurements together with measurements of the outcomes, treatments, and precisely measured covariates, such that its conditional expectation recovers the unbiased estimating function, $S_{Kj}(\theta_{Kj}; Y_{Kji}, \overline{A}_{Ki}, \overline{X}_{Ki}, \overline{C}_{Ki}, \overline{Z}_{Ki})$ in (10):

$$\begin{aligned} &E\{S_{Kj}^{*}(\theta_{Kj}; Y_{Kji}, \overline{A}_{Ki}, \overline{X}_{Ki}^{*}, \overline{C}_{Ki}^{*}, \overline{Z}_{Ki}) \mid Y_{Kji}, \overline{A}_{Ki}, \overline{X}_{Ki}, \overline{C}_{Ki}, \overline{Z}_{Ki}\} \\ &\qquad = S_{Kj}(\theta_{Kj}; Y_{Kji}, \overline{A}_{Ki}, \overline{X}_{Ki}, \overline{C}_{Ki}, \overline{Z}_{Ki}). \end{aligned} \tag{22}$$

For ease of referral, we call $S_{Kj}(\theta_{Kj}; Y_{Kji}, \overline{A}_{Ki}, \overline{X}_{Ki}, \overline{C}_{Ki}, \overline{Z}_{Ki})$ the "true" estimating function, and $S^*_{Kj}(\theta_{Kj}; Y_{Kji}, \overline{A}_{Ki}, \overline{X}^*_{Ki}, \overline{C}^*_{Ki}, \overline{Z}_{Ki})$ a "corrected" estimating function. With (22), it is immediate that $S^*_{Kj}(\theta_{Kj}; Y_{Kji}, \overline{A}_{Ki}, \overline{X}^*_{Ki}, \overline{C}^*_{Ki}, \overline{Z}_{Ki})$ is an unbiased estimating function due to that

$$E\{S_{Kj}(\theta_{Kj}; Y_{Kji}, \overline{A}_{Ki}, \overline{X}_{Ki}, \overline{C}_{Ki}, \overline{Z}_{Ki})\} = 0.$$

Then, the estimating function theory shows that solving

$$\sum_{i=1}^{n} S^*_{Kj}(\theta_{Kj}; Y_{Kji}, \overline{A}_{Ki}, \overline{X}^*_{Ki}, \overline{C}^*_{Ki}, \overline{Z}_{Ki}) = 0$$

for θ_{Kj} yields a consistent estimator for θ_{Kj}, provided regularity conditions (Yi, 2017, Section 2.5).

Since the dependence of (13) and (14) on X_k and C_k is reflected respectively by $\left\{X_k, X_k X_k^{\mathrm{T}}\right\}$ and $\left\{C_k, C_k^2\right\}$ via linear operations for $k = 1, \ldots, K$, to find $S^*_{Kj}(\theta_{Kj}; Y_{Kji}, \overline{A}_{Ki}, \overline{X}^*_{Ki}, \overline{C}^*_{Ki}, \overline{Z}_{Ki})$ to meet (22), it suffices to find unbiased surrogates for $\left\{X_k, X_k X_k^{\mathrm{T}}\right\}$ and $\left\{C_k, C_k^2\right\}$ in the following sense. That is, we aim to find functions of X_k^*, say X_k^{**} and X_k^{***}, and functions of C_k^*, say C_k^{**} and C_k^{***}, such that

$$E\left\{X_k^{**} \mid Y_{k1i}, Y_{k2i}, \overline{A}_{ki}, \overline{X}_{ki}, \overline{C}_{ki}, \overline{Z}_{ki}\right\} = X_k;\ E\left\{X_k^{***} \mid Y_{k1i}, Y_{k2i}, \overline{A}_{ki}, \overline{X}_{ki}, \overline{C}_{ki}, \overline{Z}_{ki}\right\} = X_k X_k^{\mathrm{T}}; \quad (23)$$

and

$$E\left\{C_k^{**} \mid Y_{k1i}, Y_{k2i}, \overline{A}_{ki}, \overline{X}_{ki}, \overline{C}_{ki}, \overline{Z}_{ki}\right\} = C_k;\ E\left\{C_k^{***} \mid Y_{k1i}, Y_{k2i}, \overline{A}_{ki}, \overline{X}_{ki}, \overline{C}_{ki}, \overline{Z}_{ki}\right\} = C_k^2. \quad (24)$$

To construct X_k^{**} and X_k^{***}, we apply the technique of Akazawa et al. (1998). For $t = 1, 2$, let e_t denote a 2×1 vector with 1 in the t position and zero elsewhere. We now express the two values, 0 and 1, for the binary variable X_k (or X_k^*) as two 2×1 vectors. If $X_k = 1$, then a 2×1 vector is created with the first element set to 0 and the second element set to 1; if $X_k = 0$, then a 2×1 vector is created with the first element set to 1 and the second element set to 0. That is, $X_k = 1$ and $X_k = 0$ can be represented by e_2 and e_1, respectively. Similarly, $X_k^* = 1$ and $X_k^* = 0$ are represented by e_2 and e_1, respectively. Then, setting $X_k^{**} = \Pi^{-1} X_k^*$ and $X_k^{***} = \sum_{t=1}^{2} \left\{X_k^{**\mathrm{T}} e_t\right\} e_t e_t^{\mathrm{T}}$ meets (23), where the misclassification effects for the binary covariate X_k are incorporated in X_k^{**} and X_k^{***} through the adjustment Π^{-1}.

The construction of C_k^{**} and C_k^{***} is straightforward by the classical additive measurement error model (18). Setting $C_k^{**} = C_k^*$ and $C_k^{***} = C_k^{*2} - \sigma_k^2$ makes (24) hold, where the measurement error effects for the continuous covariate C_k are reflected by the subtraction term σ_k^2 in C_k^{***}. Consequently, we define $S^*_{Kj}(\theta_{Kj}; Y_{Kji}, \overline{A}_{Ki}, \overline{X}^*_{Ki}, \overline{C}^*_{Ki}, \overline{Z}_{Ki})$ to be $S_{Kj}(\theta_{Kj}; Y_{Kji}, \overline{A}_{Ki}, \overline{X}_{Ki}, \overline{C}_{Ki}, \overline{Z}_{Ki})$ with $\left(X_{ki}, X_{ki} X_{ki}^{\mathrm{T}}\right)$ and $\left(C_{ki}, C_{ki}^2\right)$ replaced by their unbiased surrogates $\left(X_{ki}^{**}, X_{ki}^{***}\right)$ and $\left(C_{ki}^{**}, C_{ki}^{***}\right)$, respectively. An example of constructing $S^*_{Kj}(\theta_{Kj}; Y_{Kji}, \overline{A}_{Ki}, \overline{X}^*_{Ki}, \overline{C}^*_{Ki}, \overline{Z}_{Ki})$ is given in Section S1 of the Supplementary Material.

Let $\hat{\theta}_{Kjc} = (\hat{\beta}_{Kjc}^{\mathrm{T}}, \hat{\psi}_{Kjc}^{\mathrm{T}})^{\mathrm{T}}$ denote the resultant estimator of θ_{Kj} by solving

$$\sum_{i=1}^{n} S^*_{Kj}(\theta_{Kj}; Y_{Kji}, \overline{A}_{Ki}, \overline{X}^*_{Ki}, \overline{C}^*_{Ki}, \overline{Z}_{Ki}) = 0. \quad (25)$$

Under regularity conditions (Yi, 2017, Section 1.3), $\sqrt{n}\left(\hat{\theta}_{Kjc}-\theta_{Kj}\right) \xrightarrow{d} \mathcal{N}\left(0, \Sigma(\theta_{Kj})\right)$ as $n \to \infty$, where $\Sigma(\theta_{Kj}) = \left\{I(\theta_{Kj})\right\}^{-1} J(\theta_{Kj})\left\{I(\theta_{Kj})\right\}^{-1\mathrm{T}}$, with

$$I(\theta_{Kj}) = E\left\{\frac{\partial S^{*}_{Kj}(\theta_{Kj}; Y_{Kji}, \overline{A}_{Ki}, \overline{X}^{*}_{Ki}, \overline{C}^{*}_{Ki}, \overline{Z}_{Ki})}{\partial \theta_{Kj}}\right\}$$

and

$$J(\theta_{Kj}) = E\left[S^{*}_{Kj}(\theta_{Kj}; Y_{Kji}, \overline{A}_{Ki}, \overline{X}^{*}_{Ki}, \overline{C}^{*}_{Ki}, \overline{Z}_{Ki})\left\{S^{*}_{Kj}(\theta_{Kj}; Y_{Kji}, \overline{A}_{Ki}, \overline{X}^{*}_{Ki}, \overline{C}^{*}_{Ki}, \overline{Z}_{Ki})\right\}^{\mathrm{T}}\right].$$

5.1.2 Estimation Related to Stage $k < K$

In the error-free setting, estimation in stage k can be carried out by solving (16). However, in the case where accurate measurements of X_{ki} and C_{ki} are not available, (16) may produce unreliable results if directly replacing X_{ki} and C_{ki} with their observed surrogate measurements X^{*}_{ki} and C^{*}_{ki}. Similar to the consideration for stage K, we modify $\hat{S}_{kj}(\theta_{kj}; \hat{Y}_{kji}, \overline{A}_{ki}, \overline{X}_{ki}, \overline{C}_{ki}, \overline{Z}_{ki})$ in (16) by replacing $\left(X_{ki}, X_{ki}X^{\mathrm{T}}_{ki}\right)$ and $\left(C_{ki}, C^{2}_{ki}\right)$ with their respective unbiased surrogates, $\left(X^{**}_{k}, X^{***}_{k}\right)$ and $\left(C^{**}_{k}, C^{***}_{k}\right)$; such a replacement is also carried out for X_{ki} and C_{ki} in the pseudo-outcome $\hat{Y}_{kji}$.

Let $\hat{S}^{*}_{kj}(\theta_{kj}; \hat{Y}^{*}_{kji}, \overline{A}_{ki}, \overline{X}^{*}_{ki}, \overline{C}^{*}_{ki}, \overline{Z}_{ki})$ denote the resultant modified estimating function. Then, for $k = K-1, \ldots, 1$, we solve

$$\sum_{i=1}^{n} \hat{S}^{*}_{kj}(\theta_{kj}; \hat{Y}^{*}_{kji}, \overline{A}_{ki}, \overline{X}^{*}_{ki}, \overline{C}^{*}_{ki}, \overline{Z}_{ki}) = 0, \tag{26}$$

for θ_{kj}, and let $\hat{\theta}_{kjc}$ denote the resulting estimator of θ_{kj}.

5.2 Corrected Estimation Functions with Unknown Mismeasurement Degrees

The implementation of the correction procedure described in Section 5.1 relies on having prior knowledge of the misclassification and measurement error degrees. In practice, however, such information is often unavailable, making it necessary to estimate the misclassification probabilities and the extent of measurement error. To address this issue, a validation subsample can be utilized. In this section, along the same line as Khadem Charvadeh and Yi (2024a), we modify the estimation equation method proposed in Section 5.1 to account for the unknown mismeasurement degrees. Our analytical procedure is based on two sets of data: the main study sample $\mathcal{D}^{*}$ and the internal validation subsample $\mathcal{D}_{V}$. The validation sample, which comprises a smaller group of m individuals, is drawn from the main study sample, where $m \leqslant n$.

5.2.1 Estimation of Misclassification Probabilities and Measurement Error Degree

First, we present the process for estimating misclassification probabilities. For $i \in \mathcal{M}$ and $k = 1, \ldots, K$, let

$$\pi_{ki01} = P(X^{*}_{ki} = 0 \mid X_{ki} = 1, \overline{A}_{(k-1)i} \cup \overline{X}_{ki} \cup \overline{C}_{ki} \cup \overline{Z}_{ki} \setminus X_{ki})$$

and

$$\pi_{ki10} = P(X^{*}_{ki} = 1 \mid X_{ki} = 0, \overline{A}_{(k-1)i} \cup \overline{X}_{ki} \cup \overline{C}_{ki} \cup \overline{Z}_{ki} \setminus X_{ki}).$$

To describe how misclassification probabilities are associated with covariates, we employ logistic regression models

$$\begin{aligned} \operatorname{logit} \pi_{ki10} &= \alpha_{k0}^{\mathrm{T}} \mathcal{W}_{ki0}; \\ \operatorname{logit} \pi_{ki01} &= \alpha_{k1}^{\mathrm{T}} \mathcal{W}_{ki1}, \end{aligned} \tag{27}$$

where α_{kl} denotes the vector of regression coefficients and $\mathcal{W}_{kil}$ may include 1 and a subset of covariates $\left\{X_{ki} = l\right\} \cup \overline{A}_{(k-1)i} \cup \overline{X}_{ki} \cup \overline{C}_{ki} \cup \overline{Z}_{ki} \setminus X_{ki}$ that reflects different misclassification mechanisms for $l = 0, 1$. Having 1 in $\mathcal{W}_{kil}$ allows the inclusion of the intercept in (27), and $\mathcal{W}_{kil}$ may contain the entire covariate vector $\{X_{ki} = l\} \cup \overline{A}_{(k-1)i} \cup \overline{X}_{ki} \cup \overline{C}_{ki} \cup \overline{Z}_{ki} \setminus X_{ki}$ or just constant 1 alone, where the latter case corresponds to homogeneous misclassification across all subjects. Let $\alpha_k = (\alpha_{k0}^{\mathrm{T}}, \alpha_{k1}^{\mathrm{T}})^{\mathrm{T}}$ denote the parameter vector for $k = 1, \ldots, K$.

For $i = 1, \ldots, n$ and $k = 1, \ldots, K$, let

$$L_{ki}(\alpha_k) = P(X_{ki}^* = x_{ki}^* \mid X_{ki} = x_{ki}, \overline{A}_{(k-1)i} \cup \overline{X}_{ki} \cup \overline{C}_{ki} \cup \overline{Z}_{ki} \setminus X_{ki}),$$

which equals $\left\{\pi_{ki10}^{x_{ki}^*}(1 - \pi_{ki10})^{1-x_{ki}^*}\right\}^{1-x_{ki}} \cdot \left\{\pi_{ki01}^{1-x_{ki}^*}(1 - \pi_{ki01})^{x_{ki}^*}\right\}^{x_{ki}}$ for $x_{ki}, x_{ki}^* = 0, 1$. Write $\alpha = (\alpha_1^{\mathrm{T}}, \ldots, \alpha_K^{\mathrm{T}})^{\mathrm{T}}$. Let $S_{ki}(\alpha_k) = \partial \log L_{ki}(\alpha_k)/\partial \alpha_k$ and $S_i(\alpha) = \left(S_{1i}^{\mathrm{T}}(\alpha_1), \ldots, S_{Ki}^{\mathrm{T}}(\alpha_K)\right)^{\mathrm{T}}$.

With internal validation data, solving $\sum_{i \in \mathcal{V}} S_i(\alpha) = 0$ for α yields the maximum likelihood estimate, denoted $\hat{\alpha} = (\hat{\alpha}_1^{\mathrm{T}}, \ldots, \hat{\alpha}_K^{\mathrm{T}})^{\mathrm{T}}$, of α.

Rewriting (18) as $C_{ki}^* = \zeta_{k0} + \zeta_{k1} C_{ki} + e_{ki}$, we estimate the measurement error model parameters by applying the likelihood method to the validation sample $\mathcal{D}_V$. Let $\eta_k = (\zeta_{k0}, \zeta_{k1}, \sigma_k^2)^{\mathrm{T}}$ and $\eta = (\eta_1^{\mathrm{T}}, \ldots, \eta_K^{\mathrm{T}})^{\mathrm{T}}$.

For $i \in \mathcal{V}$ and $k = K, \ldots, 1$, let $L_{ki}(\eta_k) = \left(2\pi\sigma_k^2\right)^{-1/2} \exp\left\{-\frac{(c_{ki}^* - \zeta_{k0} - \zeta_{k1} c_{ki})^2}{2\sigma_k^2}\right\}$ denote the likelihood function contributed from subject i and let $\ell_{ki}(\eta_k) = \partial \log L_{ki}(\eta_k)/\partial \eta_k$. Define $\ell_i(\eta) = \left(\ell_{1i}^{\mathrm{T}}(\eta_1), \ldots, \ell_{Ki}^{\mathrm{T}}(\eta_K)\right)^{\mathrm{T}}$. Then solving

$$\sum_{i \in \mathcal{V}} \ell_i(\eta) = 0 \tag{28}$$

for η yields the maximum likelihood estimator of η; let $\hat{\eta} = \left(\hat{\eta}_1^{\mathrm{T}}, \ldots, \hat{\eta}_K^{\mathrm{T}}\right)^{\mathrm{T}}$ denote the resultant estimator of η.

5.2.2 Estimation for the Parameters of Q-functions

For notation simplicity, let $S_{Kji}^*(\theta_{Kj}, \alpha, \eta)$ represent $S_{Kj}^*\left(\theta_{Kj}, \alpha, \eta; Y_{Kji}, \bar{A}_{Ki}, \bar{X}_{Ki}^*, \bar{C}_{Ki}^*, \bar{Z}_{Ki}\right)$ in (25) with the dependence on α and η spelled out, and let $S_{Kji}(\theta_{Kj})$ represent $S_{Kj}(\theta_{Kj}; Y_{Kji}, \bar{A}_{Ki}, \bar{X}_{Ki}, \bar{C}_{Ki}, \bar{Z}_{Ki})$ in (15). For stage K, we then estimate θ_{Kj} by solving

$$\sum_{i \in \mathcal{M} \setminus \mathcal{V}} S_{Kji}^*(\theta_{Kj}, \hat{\alpha}, \hat{\eta}) + \sum_{i \in \mathcal{V}} S_{Kji}(\theta_{Kj}) = 0$$

for θ_{Kj}, and let $\hat{\theta}_{Kj}$ denote the resultant estimator of θ_{Kj}.

For $j = 1, 2$, let $\vartheta_j = (\beta_{Kj}^{\mathrm{T}}, \psi_{Kj}^{\mathrm{T}}, \alpha^{\mathrm{T}}, \eta^{\mathrm{T}})^{\mathrm{T}}$ and $\hat{\vartheta}_j = (\hat{\theta}_{Kj}^{\mathrm{T}}, \hat{\alpha}^{\mathrm{T}}, \hat{\eta}^{\mathrm{T}})^{\mathrm{T}}$. Under certain regularity conditions and that the ratio m/n approaches a positive constant, say ρ, as $n \to \infty$, $\hat{\vartheta}_j$ is a consistent estimator of ϑ_j, and

$$\sqrt{n}(\hat{\vartheta}_j - \vartheta_j) \xrightarrow{d} \mathcal{N}(0, \Sigma_{Vj}) \qquad \text{as} \quad n \to \infty,$$

where $\Sigma_{Vj} = A_{Vj}^{-1} B_{Vj} A_{Vj}^{-1\mathrm{T}}$, with

$$A_{Vj} = -(1-\rho)\begin{pmatrix} E\left(\frac{\partial S^*_{Kji}(\theta_{Kj},\alpha,\eta)}{\partial \theta_{Kj}}\right) & E\left(\frac{\partial S^*_{Kji}(\theta_{Kj},\alpha,\eta)}{\partial \alpha}\right) & E\left(\frac{\partial S^*_{Kji}(\theta_{Kj},\alpha,\eta)}{\partial \eta}\right) \\ 0 & 0 & 0 \end{pmatrix}$$
$$- \rho \begin{pmatrix} E\left(\frac{\partial S_{Kji}(\theta_{Kj})}{\partial \theta_{Kj}}\right) & 0 & 0 \\ 0 & E\left(\frac{S_i(\alpha)}{\partial \alpha}\right) & 0 \\ 0 & 0 & E\left(\frac{\ell_i(\eta)}{\partial \eta}\right) \end{pmatrix}$$

and

$$B_{Vj} = (1-\rho)\begin{pmatrix} E\left\{S^*_{Kji}(\theta_{Kj}, \alpha, \eta) S^{*\mathrm{T}}_{Kji}(\theta_{Kj}, \alpha, \eta)\right\} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$
$$+ \rho \begin{pmatrix} E\left\{S_{Kji}(\theta_{Kj}) S^{\mathrm{T}}_{Kji}(\theta_{Kj})\right\} & 0 & 0 \\ 0 & E\left\{S_i(\alpha) S_i^{\mathrm{T}}(\alpha)\right\} & 0 \\ 0 & 0 & E\left\{\ell_i(\eta)\ell_i^{\mathrm{T}}(\eta)\right\} \end{pmatrix}.$$

The impact of the misclassification rates (π_{01} and π_{10}) and the measurement error variance σ_k^2 on the asymptotic covariance matrix Σ_{Vj} is reflected by the formulation of $S^*_{Kji}(\theta_{Kj}, \alpha, \eta)$, which is defined before (25). The dependence of Σ_{Vj} on π_{01} and π_{10} is accommodated by X^{**}_{ki} and X^{***}_{ki}, whereas the dependence on σ_k^2 is accounted for by C^{***}_{ki}.

Finally, for $k = K-1, \ldots, 1$, an estimator of θ_{kj} can be obtained by solving (26), where π_{k10} and π_{k01} are determined by (27) with α replaced by $\hat{\alpha}$, and σ_k^2 is replaced with its estimate determined by (28).

6 Simulation Studies

In this section, we conduct simulation studies to evaluate the finite sample performance of the methods described in Sections 4 and 5. We employ the same configuration and parameter settings as in Section 3.2 to generate the main study data $\left\{\left\{A_{ki}, X^*_{ki}, C^*_{ki}, Z_{ki}, Y_{k1i}, Y_{k2i}\right\} : k = 1, \ldots, K; i \in \mathcal{M}\right\}$. Additionally, we create an internal validation subsample by randomly selecting 30% of the study subjects from $\mathcal{M}$ and record their precise measurements of $\{X_{ki}, C_{ki} : k = 1, \ldots, K; i \in \mathcal{V}\}$ to form the validation subsample $\mathcal{D}_V = \left\{\left\{A_{ki}, X_{ki}, X^*_{ki}, C_{ki}, C^*_{ki}, Z_{ki}, Y_{k1i}, Y_{k2i}\right\} : k = 1, \ldots, K;\ i \in \mathcal{V}\right\}$.

We examine the data using three methods. The first method, referred to as "RC", applies the approach outlined in Section 4 to both $\mathcal{D}^*$ and $\mathcal{D}_V$. The second method, termed as "EE-known", applies the method described in Section 5.1 to the data in $\mathcal{D}^*$, assuming that the misclassification probabilities and measurement error degrees are known. The third method, named "EE-estimated", employs the procedure described in Section 5.2 to both $\mathcal{D}^*$ and $\mathcal{D}_V$, where the misclassification probabilities and measurement error degrees are estimated using $\mathcal{D}_V$.

Tables 2–4 present the numerical results pertaining to stages 1 and 2 for the regular, weak non-regular, and non-regular settings, where "Bias", "ESE", "PBCR", and "DBCR" are the same as in Section 3.2, but “SE” and “MSE” may be slightly different, depending on the method used. When the RC method is considered, “SE” represents the average of standard errors obtained from the bootstrap method with 1000 bootstrap samples, and “MSE” represents the mean squared error of the estimates obtained using the bootstrap SE; when the EE-known or EE-estimated method is used, “SE” and “MSE” are the same as in Section 3.2. It is evident that both the RC and EE methods yield satisfactory results in terms of bias, with the RC method exhibiting relatively lower levels of bias compared to the EE methods. The EE methods yield high SEs, consequently leading to higher MSEs, which tend to escalate as the degrees of mismeasurement increase.

As noted in Section 5.2.2, the variability of the estimators depends on the degrees of mismeasurement, characterized by $\pi_{jl} = P(X_k^* = j \mid X_k = l)$ for binary covariates and σ_k^2 for continuous covariates, which is also numerically illustrated in Table 2. For example, when $(\pi_{10}, \pi_{01}, \sigma_k^2) = (0.1, 0.1, 1)$, the SEs of the estimators for the EE-known method at stage 1 are 0.122, 0.191, and 0.090. However, when $(\pi_{10}, \pi_{01}, \sigma_k^2) = (0.3, 0.3, 2)$, these SEs increase to 0.393, 0.723, and 0.350. Similarly, for the EE-estimated method, with $(\pi_{10}, \pi_{01}, \sigma_k^2) = (0.1, 0.1, 1)$, the SEs are 0.145, 0.184, and 0.155, whereas for $(\pi_{10}, \pi_{01}, \sigma_k^2) = (0.3, 0.3, 2)$, the SEs increase to 0.404, 0.717, and 0.425. A similar pattern is observed in stage 2 results from Table 2 under these two mismeasurement scenarios. For the EE-known method, when $(\pi_{10}, \pi_{01}, \sigma_k^2) = (0.1, 0.1, 1)$, the SEs of the estimators are 0.121, 0.173, and 0.105. However, under the higher mismeasurement scenario $(\pi_{10}, \pi_{01}, \sigma_k^2) = (0.3, 0.3, 2)$, these SEs increase to 0.402, 0.661, and 0.455. A similar trend is also observed for stage 2 results of the EE-estimated method, where the SEs under $(\pi_{10}, \pi_{01}, \sigma_k^2) = (0.1, 0.1, 1)$ are 0.131, 0.164, and 0.136. When the mismeasurement parameters increase to $(\pi_{10}, \pi_{01}, \sigma_k^2) = (0.3, 0.3, 2)$, the SEs rise to 0.411, 0.666, and 0.601. These findings highlight how higher misclassification rates and greater measurement error variance may amplify the uncertainty in the estimation.

Unsurprisingly, the performance of the correction methods deteriorates as the degree of mismeasurement increases. Under the regular setting, the RC method maintains PBCR within an acceptable range across all mismeasurement degrees. The EE-known method performs well at low mismeasurement degree but tends to exhibit over-coverage as mismeasurement severity increases. The EE-estimated method generally yields acceptable coverage, though mild under-coverage is observed for the parameter $\psi_{2k\delta}$ when $(\pi_{10}, \pi_{01}, \sigma_k^2) = (0.3, 0.3, 2)$. In the weak non-regular setting, the RC and EE-known methods continue to produce satisfactory PBCR, while the EE-estimated method remains mostly acceptable with occasional under-coverage, again more apparent for $\psi_{2k\delta}$. In the non-regular setting, both RC and EE-known methods maintain reliable coverage performance, but the EE-estimated method shows consistent under-coverage for $\psi_{2k\delta}$, particularly under high mismeasurement degree. Similarly, the DB 95% CIs maintain a generally close coverage to the nominal level across all methods and settings, indicating improved calibration compared to PB intervals. The RC and EE-known methods maintain satisfactory DBCR even under a high mismeasurement degree. However, the EE-estimated method shows slight under-coverage for $\psi_{2k\delta}$ in non-regular settings, reflecting the added uncertainty from estimating mismeasurement degrees. The Wald-type 95% CIs for the RC method have reasonable coverage, while for the EE methods, they may have over-coverage.

To assess the performance of the proposed methods in determining optimal treatments for future patients, in Table S1 in the Supplementary Material, we report the proportion of optimally treated future patients, denoted POTFP, where in Scenario 1, the true covariate measurements are treated as available; and in Scenario 2, the true error-prone binary and continuous covariate measurements are unavailable, but their surrogate measurements are available. The results

Table 2: Simulation studies for assessing the performance of the RC, EE-known, and EE-estimated methods: stages 1–2 and regular case.

		RC			EE-known			EE-estimated		
$(\pi_{10}, \pi_{01}, \sigma_k^2)$		$\psi_{0k\delta}$	$\psi_{1k\delta}$	$\psi_{2k\delta}$	$\psi_{0k\delta}$	$\psi_{1k\delta}$	$\psi_{2k\delta}$	$\psi_{0k\delta}$	$\psi_{1k\delta}$	$\psi_{2k\delta}$
		Stage 1 $(k = 1)$								
(0.1,0.1,1)	Bias	0.020	0.008	0.002	0.005	0.015	0.010	0.002	0.006	0.012
	SE	0.106	0.134	0.058	0.122	0.191	0.090	0.145	0.184	0.155
	ESE	0.097	0.129	0.055	0.135	0.200	0.099	0.135	0.188	0.095
	MSE	0.012	0.018	0.003	0.015	0.037	0.008	0.021	0.034	0.024
	PBCR	0.936	0.942	0.950	0.948	0.942	0.946	0.954	0.968	0.936
	DBCR	0.942	0.958	0.962	0.954	0.954	0.944	0.966	0.968	0.932
(0.2,0.2,1.5)	Bias	0.019	0.007	0.002	0.008	0.044	0.036	0.000	0.021	0.038
	SE	0.118	0.157	0.066	0.199	0.340	0.173	0.209	0.321	0.216
	ESE	0.125	0.162	0.067	0.223	0.368	0.177	0.204	0.321	0.186
	MSE	0.014	0.025	0.004	0.040	0.118	0.031	0.044	0.103	0.048
	PBCR	0.934	0.938	0.950	0.962	0.956	0.956	0.945	0.949	0.936
	DBCR	0.948	0.952	0.954	0.972	0.972	0.974	0.966	0.953	0.936
(0.3,0.3,2)	Bias	0.012	0.009	0.001	0.011	0.018	0.065	0.014	0.002	0.068
	SE	0.129	0.184	0.071	0.393	0.723	0.350	0.404	0.717	0.425
	ESE	0.133	0.189	0.072	0.399	0.735	0.337	0.371	0.657	0.343
	MSE	0.017	0.034	0.005	0.155	0.523	0.127	0.163	0.514	0.185
	PBCR	0.949	0.945	0.953	0.967	0.965	0.974	0.958	0.950	0.930
	DBCR	0.939	0.939	0.961	0.965	0.959	0.976	0.944	0.944	0.922
		Stage 2 $(k = 2)$								
(0.1,0.1,1)	Bias	0.009	0.008	0.006	0.013	0.023	0.025	0.003	0.001	0.003
	SE	0.098	0.127	0.066	0.121	0.173	0.105	0.131	0.164	0.136
	ESE	0.096	0.124	0.063	0.134	0.184	0.112	0.106	0.153	0.113
	MSE	0.010	0.016	0.004	0.015	0.030	0.012	0.017	0.027	0.019
	WTCR	0.958	0.954	0.964	0.928	0.934	0.926	0.986	0.954	0.988
(0.2,0.2,1.5)	Bias	0.002	0.002	0.002	0.028	0.043	0.045	0.013	0.024	0.066
	SE	0.114	0.152	0.076	0.206	0.318	0.221	0.204	0.299	0.247
	ESE	0.111	0.142	0.078	0.198	0.319	0.240	0.192	0.284	0.265
	MSE	0.013	0.023	0.006	0.043	0.103	0.051	0.042	0.090	0.065
	WTCR	0.952	0.962	0.950	0.970	0.952	0.944	0.964	0.966	0.954
(0.3,0.3,2)	Bias	0.002	0.005	0.005	0.021	0.032	0.110	0.033	0.038	0.104
	SE	0.123	0.166	0.079	0.402	0.661	0.455	0.411	0.666	0.601
	ESE	0.117	0.163	0.080	0.386	0.620	0.420	0.336	0.541	0.497
	MSE	0.015	0.028	0.006	0.162	0.438	0.219	0.170	0.445	0.372
	WTCR	0.964	0.948	0.946	0.982	0.973	0.937	0.977	0.991	0.907

Table 3: Simulation studies for assessing the performance of the RC, EE-known, and EE-estimated methods: stages 1–2 and weak non-regular case.

		RC			EE-known			EE-estimated		
$(\pi_{10}, \pi_{01}, \sigma_k^2)$		$\psi_{0k\delta}$	$\psi_{1k\delta}$	$\psi_{2k\delta}$	$\psi_{0k\delta}$	$\psi_{1k\delta}$	$\psi_{2k\delta}$	$\psi_{0k\delta}$	$\psi_{1k\delta}$	$\psi_{2k\delta}$
		Stage 1 ($k = 1$)								
(0.1,0.1,1)	Bias	0.001	0.003	0.006	0.023	0.003	0.001	0.018	0.005	0.007
	SE	0.094	0.118	0.052	0.094	0.145	0.074	0.114	0.142	0.127
	ESE	0.096	0.114	0.052	0.115	0.150	0.081	0.099	0.140	0.081
	MSE	0.009	0.014	0.003	0.009	0.021	0.005	0.013	0.020	0.016
	PBCR	0.954	0.942	0.958	0.938	0.952	0.954	0.936	0.944	0.936
	DBCR	0.958	0.958	0.962	0.946	0.962	0.966	0.950	0.952	0.918
(0.2,0.2,1.5)	Bias	0.002	0.000	0.001	0.040	0.006	0.013	0.017	0.027	0.019
	SE	0.106	0.140	0.060	0.151	0.255	0.145	0.160	0.243	0.179
	ESE	0.102	0.131	0.058	0.177	0.291	0.155	0.160	0.250	0.171
	MSE	0.011	0.020	0.004	0.024	0.065	0.021	0.026	0.060	0.032
	PBCR	0.964	0.950	0.928	0.950	0.938	0.936	0.953	0.945	0.924
	DBCR	0.964	0.956	0.960	0.954	0.950	0.958	0.951	0.940	0.920
(0.3,0.3,2)	Bias	0.003	0.008	0.002	0.019	0.066	0.039	0.001	0.067	0.071
	SE	0.114	0.157	0.064	0.311	0.567	0.299	0.338	0.599	0.419
	ESE	0.122	0.167	0.067	0.318	0.598	0.295	0.323	0.579	0.339
	MSE	0.013	0.025	0.004	0.097	0.326	0.091	0.114	0.363	0.181
	PBCR	0.956	0.938	0.940	0.962	0.966	0.958	0.962	0.964	0.932
	DBCR	0.948	0.942	0.952	0.964	0.972	0.958	0.950	0.960	0.930
		Stage 2 ($k = 2$)								
(0.1,0.1,1)	Bias	0.001	0.004	0.001	0.008	0.010	0.004	0.010	0.016	0.005
	SE	0.082	0.113	0.043	0.088	0.121	0.060	0.108	0.121	0.112
	ESE	0.078	0.112	0.045	0.087	0.119	0.067	0.089	0.120	0.054
	MSE	0.007	0.013	0.002	0.008	0.015	0.004	0.012	0.015	0.013
	WTCR	0.954	0.950	0.948	0.944	0.942	0.926	0.984	0.954	1.000
(0.2,0.2,1.5)	Bias	0.000	0.000	0.001	0.016	0.030	0.002	0.015	0.025	0.006
	SE	0.097	0.137	0.051	0.131	0.184	0.117	0.141	0.176	0.150
	ESE	0.097	0.140	0.052	0.127	0.181	0.124	0.138	0.186	0.101
	MSE	0.009	0.019	0.003	0.017	0.035	0.014	0.020	0.032	0.023
	WTCR	0.952	0.958	0.942	0.958	0.948	0.958	0.964	0.934	1.000
(0.3,0.3,2)	Bias	0.002	0.002	0.001	0.061	0.098	0.019	0.060	0.101	0.008
	SE	0.105	0.153	0.056	0.231	0.339	0.258	0.246	0.351	0.311
	ESE	0.101	0.143	0.052	0.228	0.320	0.236	0.224	0.299	0.217
	MSE	0.011	0.023	0.003	0.057	0.125	0.067	0.064	0.133	0.097
	WTCR	0.959	0.963	0.970	0.956	0.954	0.993	0.968	0.968	0.995

Table 4: Simulation studies for assessing the performance of the RC, EE-known, and EE-estimated methods: stages 1–2 and non-regular case.

		RC			EE-known			EE-estimated		
$(\pi_{10}, \pi_{01}, \sigma_k^2)$		$\psi_{0k\delta}$	$\psi_{1k\delta}$	$\psi_{2k\delta}$	$\psi_{0k\delta}$	$\psi_{1k\delta}$	$\psi_{2k\delta}$	$\psi_{0k\delta}$	$\psi_{1k\delta}$	$\psi_{2k\delta}$
		Stage 1 $(k = 1)$								
(0.1,0.1,1)	Bias	0.002	0.006	0.001	0.001	0.013	0.007	0.007	0.003	0.011
	SE	0.091	0.117	0.052	0.091	0.141	0.073	0.111	0.138	0.125
	ESE	0.092	0.115	0.053	0.109	0.145	0.084	0.102	0.147	0.086
	MSE	0.008	0.014	0.003	0.008	0.020	0.005	0.012	0.019	0.016
	PBCR	0.948	0.942	0.938	0.960	0.954	0.958	0.942	0.932	0.900
	DBCR	0.956	0.968	0.956	0.960	0.956	0.964	0.938	0.946	0.910
(0.2,0.2,1.5)	Bias	0.017	0.013	0.005	0.016	0.010	0.018	0.009	0.031	0.020
	SE	0.101	0.139	0.060	0.139	0.234	0.140	0.148	0.222	0.172
	ESE	0.100	0.137	0.063	0.149	0.247	0.158	0.147	0.242	0.159
	MSE	0.010	0.019	0.004	0.020	0.055	0.020	0.022	0.050	0.030
	PBCR	0.946	0.968	0.950	0.962	0.956	0.950	0.943	0.937	0.905
	DBCR	0.938	0.952	0.946	0.968	0.964	0.974	0.939	0.943	0.895
(0.3,0.3,2)	Bias	0.000	0.003	0.003	0.005	0.000	0.031	0.021	0.056	0.077
	SE	0.109	0.157	0.064	0.255	0.461	0.267	0.278	0.491	0.404
	ESE	0.111	0.156	0.062	0.264	0.473	0.236	0.249	0.448	0.309
	MSE	0.012	0.025	0.004	0.065	0.213	0.072	0.078	0.244	0.169
	PBCR	0.940	0.924	0.944	0.958	0.970	0.944	0.940	0.950	0.890
	DBCR	0.952	0.952	0.934	0.946	0.966	0.942	0.938	0.964	0.906
		Stage 2 $(k = 2)$								
(0.1,0.1,1)	Bias	0.001	0.002	0.001	0.000	0.000	0.002	0.003	0.004	0.001
	SE	0.069	0.098	0.037	0.074	0.106	0.044	0.099	0.108	0.106
	ESE	0.068	0.095	0.036	0.071	0.100	0.045	0.070	0.097	0.039
	MSE	0.005	0.010	0.001	0.005	0.011	0.002	0.010	0.012	0.011
	WTCR	0.950	0.954	0.958	0.956	0.964	0.952	0.994	0.966	1.000
(0.2,0.2,1.5)	Bias	0.001	0.006	0.003	0.003	0.000	0.002	0.009	0.014	0.002
	SE	0.077	0.114	0.042	0.097	0.148	0.066	0.115	0.144	0.116
	ESE	0.081	0.117	0.043	0.099	0.146	0.063	0.089	0.134	0.061
	MSE	0.006	0.013	0.002	0.009	0.022	0.004	0.013	0.021	0.013
	WTCR	0.950	0.946	0.926	0.952	0.950	0.964	0.990	0.962	1.000
(0.3,0.3,2)	Bias	0.002	0.008	0.001	0.016	0.019	0.006	0.014	0.018	0.003
	SE	0.085	0.131	0.045	0.150	0.238	0.114	0.164	0.237	0.164
	ESE	0.085	0.131	0.047	0.145	0.228	0.113	0.125	0.193	0.109
	MSE	0.007	0.017	0.002	0.023	0.057	0.013	0.027	0.056	0.027
	WTCR	0.944	0.951	0.954	0.940	0.954	0.977	0.998	0.975	1.000

obtained from Scenario 1 reveal that, in stage 1, both the RC and EE methods surpass the naive method. In stage 2, it is evident that both the RC and EE methods outperform the naive method in regular and weak non-regular settings. However, in the non-regular setting, the naive method yields larger POTFP values than the RC and EE methods.

Turning our attention to Scenario 2, the findings demonstrate that the RC method yields superior results for stage 1 estimation compared to the naive method. Furthermore, when the mismeasurement degree is set to $(0.1, 0.1, 1)$, the EE-estimated method outperforms the naive method. However, for other mismeasurement degrees, the naive method consistently yields larger values of POTFP compared to the EE-estimated method. The comparison of the EE-known method and the naive method shows that the naive method consistently yields larger POTFP values, regardless of the mismeasurement degree. For stage 2, in both regular and weak non-regular settings, both the RC and EE-estimated methods outperform the naive method. However, in the non-regular setting, the naive method results in larger POTFP values compared to both the RC and EE-estimated methods. The EE-known method exhibits superior performance to the naive method only for the weak non-regular setting.

To evaluate the performance of correction strategies with different sample sizes, we present results for a reduced sample size ($n = 1000$) in Tables S2-S5 in the Supplementary Material. While the overall trends remain similar, it is observed that the EE-estimated method occasionally yields smaller MSEs than the EE-known method. This phenomenon appears counterintuitive, as EE-estimated relies on estimated misclassification probabilities and measurement error variance, which could introduce additional variability. This pattern may be attributed to numerical instability with small or moderate sample sizes, as well as a known paradox in estimating equation methods. The efficiency of parameter estimates does not always improve and may even decrease when nuisance parameters are known rather than estimated. This phenomenon, observed by Robins et al. (1994) and Ning et al. (2018), among others, is further explained by Henmi and Eguchi (2004) by examining the orthogonality among the estimating functions.

We further evaluate the performance of correction strategies with a reduced validation subsample size. To be specific, we create an internal validation subsample by randomly selecting only 2.5% of the study subjects from $\mathcal{M}$ and recording their precise measurements of $\{X_{ki}, C_{ki} : k = 1, \ldots, K; i \in \mathcal{V}\}$ to form the validation subsample $\mathcal{D}_V = \Big\{\big\{A_{ki}, X_{ki}, X^*_{ki}, C_{ki}, C^*_{ki}, Z_{ki}, Y_{k1i}, Y_{k2i}\big\} : k = 1, \ldots, K;\ i \in \mathcal{V}\Big\}$. The results, presented in Tables S6-S8 in the Supplementary Material, indicate that this substantial reduction in the validation subsample size significantly worsens the performance of both RC and EE-estimated compared to cases with a 30% validation subsample size. This degradation is particularly evident in increased bias and SE. Unsurprisingly, EE-known demonstrates superior performance under small validation subsample sizes, maintaining lower bias and SE. However, its performance occasionally declines for certain parameters when the degree of mismeasurement is severe (e.g., (0.3, 0.3, 2)). These results highlight the impact of the validation subsample size and mismeasurement severity on the method performance.

7 Data Analysis

We now apply the methods in Sections 2, 4, and 5 to analyze a dataset on COVID-19. The data analysis here builds upon Khadem Charvadeh and Yi (2024b), which employed the standard Q-learning algorithm to examine how reducing the COVID-19 case fatality rate (*CFR*) may be influenced by preventive policies and country-specific features. The study here utilizes the extended framework discussed in Section 2 to target not only reducing *CFR* but also minimizing the economic cost associated with the pandemic. The dataset analyzed here differs from that considered by Khadem Charvadeh and Yi (2024b) in terms of the sample size, outcome variables considered, and tailoring factors, though both datasets share some common variables. Such discrepancies reflect a deliberate shift in focus toward examining the intricate interplay between health outcomes and economic repercussions in the context of COVID-19, though the analysis steps in both studies share some similarities.

A detailed description of the dataset, including data sources, pre-processing steps, feature engineering, as well as the complete analysis results and their interpretation, is provided in the Supplementary Material.

8 Discussion

In this paper, we consider the Q-learning procedure with a composite outcome in the presence of mixed misclassification and measurement error in covariates and demonstrate its substantially adverse effects on the estimation of the associated parameters. To mitigate the mismeasurement effects, we describe regression calibration and unbiased estimation equation approaches. Our development is directed to the scenario where an internal validation subsample is available to characterize the mismeasurement degrees. In situations where additional data is not available to quantify mismeasurement degrees or estimate the parameters associated with the mismeasurement models, sensitivity analyses are often employed to assess the impact of mismeasurement on the outcomes of the Q-learning algorithm. This involves selecting a set of representative values for $\boldsymbol{\zeta}_{X_k}$ and $\boldsymbol{\zeta}_{C_k}$, and using models (20) and (21) to estimate X_{ki}^{**} and C_{ki}^{**} for $i = 1, \ldots, n$. The calibrated Q-learning algorithm is then repeated to evaluate how the results may vary with different mismeasurement scenarios.

The Q-learning procedure involves the examination of multiple stages, where the number of stages K is determined by decision points, which basically depend on the study design or prior knowledge of the problem. In clinical trials or longitudinal studies, for example, K corresponds to the number of predefined intervals for treatment decisions. In other contexts, it could be determined by practical considerations such as the follow-up period duration or the expected time frame for observing treatment effects. While the specification of K may be struck to balance the benefits of tailoring treatments to patients with different conditions and the risk of excessive model complexity, exploring systematic approaches to optimize K is useful. This could potentially be achieved through simulation studies or sensitivity analyses to assess its impact across various scenarios.

In formulating the composite Q-function, the parameter δ plays a crucial role in determining the relative importance of each outcome: higher values prioritize the first outcome, while lower values emphasize the second one. Lizotte et al. (2010) explored the role of the trade-off parameter δ using a piecewise linear spline to express the optimal Q-function for each state. However, there is no analytical method to determine the optimal value of δ. The choice of δ is typically guided by domain knowledge or specific trade-offs. For example, selecting a treatment to minimize symptom severity may favor aggressive but side-effect-prone drugs, while prioritizing minimal side effects would lead to milder, less effective options. This situation is analogous to setting a significance level (denoted α) in hypothesis testing, where there is no analytical solution for determining an optimal α. Instead, its value is user-specified based on the tolerance for Type I or Type II errors. While different choices of α may lead to varying conclusions, it is important that the interpretation of the results explicitly considers the chosen significance level rather than being made without reference to it.

Finally, we make a comment on the procedure in Section 2.1. While the composite Q-function is introduced to incorporate bivariate outcomes, one may wonder whether it is equivalent to introducing a weighted average of the outcomes and then applying the standard Q-learning procedure to this new outcome variable. Specifically, for δ in (4) and $k = K, \ldots, 1$, let $Y_k^w = \delta Y_{k1} + (1-\delta) Y_{k2}$ denote the weighted average of the bivariate outcomes at stage k. Then applying the usual Q-learning procedure to Y_k^w by mimicking the formulation in (1), we define the corresponding Q-function for stage K:

$$Q_K^w(\overline{A}_K, \overline{X}_K, \overline{C}_K, \overline{Z}_K) = E(Y_K^w \mid \overline{A}_K, \overline{X}_K, \overline{C}_K, \overline{Z}_K) \tag{29}$$

and the stage k Q-function:

$$Q_k^w(\overline{A}_k, \overline{X}_k, \overline{C}_k, \overline{Z}_k) = E\{Y_k^w + \hat{Y}_{k+1}^w \mid \overline{A}_k, \overline{X}_k, \overline{C}_k, \overline{Z}_k\}, \tag{30}$$

for $k = K-1, \ldots, 1$, with $\hat{Y}_{k+1}^w = \max_{a_{k+1}} E(Y_{k+1}^w \mid \overline{A}_k, \overline{X}_{k+1}, \overline{C}_{k+1}, \overline{Z}_{k+1}, a_{k+1})$.

While $Q_K^w(\overline{A}_K, \overline{X}_K, \overline{C}_K, \overline{Z}_K)$ in (29) for stage K is identical to $Q_K(\overline{A}_K, \overline{X}_K, \overline{C}_K, \overline{Z}_K, \delta)$ in (4), expression (30) differs from (4) for stage k with $k < K$, which is primarily due to the inequality between the pseudo-outcome of the weighted average outcomes $\hat{Y}_{k+1}^w$ and the weighted average of the pseudo-outcomes $\delta \hat{Y}_{(k+1)1} + (1-\delta)\hat{Y}_{(k+1)2}$. That is,

$$\begin{aligned}\max_{a_{k+1}} E(Y_{k+1}^w \mid \overline{A}_k, \overline{X}_{k+1}, \overline{C}_{k+1}, \overline{Z}_{k+1}, a_{k+1}) &= \delta \max_{a_{k+1}} E(Y_{(k+1)1} \mid \overline{A}_k, \overline{X}_{k+1}, \overline{C}_{k+1}, \overline{Z}_{k+1}, a_{k+1}) \\ &\quad + (1-\delta) \max_{a_{k+1}} E(Y_{(k+1)2} \mid \overline{A}_k, \overline{X}_{k+1}, \overline{C}_{k+1}, \overline{Z}_{k+1}, a_{k+1})\end{aligned}$$

does not hold in general.

Moreover, although creating the weighted average outcome enables us to procedurally apply the standard Q-learning method to handle bivariate outcomes, this approach does not offer the same interpretation as that presented in Section 2.1. The weighted average of bivariate outcomes often lacks meaningful interpretability, as Y_{k1} and Y_{k2} represent different variables. For example, if Y_{k1} and Y_{k2} represent the kth month salary and exercise amount, respectively, it makes no sense to construct a weighted average $\hat{Y}_k$ and to maximize its conditional expectation. However, separately evaluating the conditional expectations of Y_{k1} and Y_{k2} retains interpretability, and using a linear combination of them offers us a way to create a scalar objective function.

Supplementary Material

S1. An Example of Constructing $S_{Kj}^*(\theta_{Kj}; Y_{Kji}, \overline{A}_{Ki}, \overline{X}_{Ki}^*, \overline{C}_{Ki}^*, \overline{Z}_{Ki})$
S2. Proportion of Optimally Treated Future Patients
S3. Simulation Results for Correction Strategies with Reduced Sample Size
S4. Simulation Results for Correction Strategies with Reduced Validation Subsample Size
S5. Data Analysis

Acknowledgement

The authors thank the editor, associate editor, and referees for their constructive comments which has led to significant improvement of this paper.

Funding

Yi is the Canada Research Chair in Data Science (Tier 1). Her research is supported by funding from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canada Research Chairs Program.

References

Akazawa K, Kinukawa N, Nakamura T (1998). A note on the corrected score function adjusting for misclassification. *Journal of the Japan Statistical Society*, 28(1): 115–123. https://doi.org/10.14490/jjss1995.28.115

Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*. CRC press.

Chakraborty B, Moodie EE (2013). *Statistical Methods for Dynamic Treatment Regimes*. Springer.

Henmi M, Eguchi S (2004). A paradox concerning nuisance parameters and projected estimating functions. *Biometrika*, 91(4): 929–941. https://doi.org/10.1093/biomet/91.4.929

Khadem Charvadeh Y, Yi GY (2024a). Accommodating misclassification effects on optimizing dynamic treatment regimes with Q-learning. *Statistics in Medicine*, 43(3): 578–605. https://doi.org/10.1002/sim.9973

Khadem Charvadeh Y, Yi GY (2024b). Understanding effective virus control policies for COVID-19 with the Q-learning method. *Statistics in Biosciences*, 16(1): 265–289. https://doi.org/10.1007/s12561-023-09382-w

Lizotte DJ, Bowling MH, Murphy SA (2010). Efficient reinforcement learning with multiple reward functions for randomized controlled trial analysis. In: *Twenty-Seventh International Conference on Machine Learning (ICML)*, 695–702.

Ning Y, Yi GY, Reid N (2018). A class of weighted estimating equations for semiparametric transformation models with missing covariates. *Scandinavian Journal of Statistics*, 45(1): 87–109. https://doi.org/10.1111/sjos.12289

Robins JM (2004). Optimal structural nested models for optimal sequential decisions. In: Lin, DY, Heagerty, PJ (eds.), *Proceedings of the Second Seattle Symposium in Biostatistics*. Lecture Notes in Statistics, vol. 179. Springer. New York, NY. https://doi.org/10.1007/978-1-4419-9076-1_11

Robins JM, Rotnitzky A, Zhao LP (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89(427): 846–866. https://doi.org/10.1080/01621459.1994.10476818

Spicker D, Wallace MP (2020). Measurement error and precision medicine: Error-prone tailoring covariates in dynamic treatment regimes. *Statistics in Medicine*, 39(26): 3732–3755. https://doi.org/10.1002/sim.8690

Wang L, Rotnitzky A, Lin X, Millikan RE, Thall PF (2012). Evaluation of viable dynamic treatment regimes in a sequentially randomized trial of advanced prostate cancer. *Journal of the American Statistical Association*, 107(498): 493–508. https://doi.org/10.1080/01621459.2011.641416

Yi GY (2017). *Statistical Analysis with Measurement Error or Misclassification: Strategy, Method and Application*. Springer Science+Business Media LLC, New York.

Yi GY, Delaigle A, Gustafson P (2021). *Handbook of Measurement Error Models*. CRC Press.