Learning Individualized Treatment Rules for Multiple-Domain Latent Outcomes

Yuan Chen* Donglin Zeng† Yuanjia Wang*

April 7, 2022

Abstract

For many mental disorders, latent mental status from multiple-domain psychological or clinical symptoms may perform as a better characterization of the underlying disorder status than a simple summary score of the symptoms, and they may also serve as more reliable and representative features to differentiate treatment responses. Therefore, in order to address the complexity and heterogeneity of treatment responses for mental disorders, we provide a new paradigm for learning optimal individualized treatment rules (ITRs) by modeling patients' latent mental status. We first learn the multi-domain latent states at baseline from the observed symptoms under a restricted Boltzmann machine (RBM) model, through which patients' heterogeneous symptoms are represented using an economical number of latent variables and yet remains flexible. We then optimize a value function defined by the latent states after treatment by exploiting a transformation of the observed symptoms based on the RBM without modeling the relationship between the latent mental states before and after treatment. The optimal treatment rules are derived using a weighted large margin classifier. We derive the convergence rate of the proposed estimator under the latent models. Simulation studies are conducted to test the performance of the proposed method. Finally, we apply the developed method to real world studies and we demonstrate the utility and advantage of our method in tailoring treatments for patients with major depression, and identify patient subgroups informative for treatment recommendations.

*Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, NY 10032; †Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC 27516

Key words:

Personalized medicine; Restricted Boltzmann machine (RBM); Machine learning; Clinical trials; Mental disorders

1 INTRODUCTION

For precision medicine of mental disorders, a number of important challenges must be addressed and they include the disease complexity and heterogeneity and highly co-morbid diagnoses that are intrinsic to the nature of mental disorders, as well as substantial diversity in patient's symptomatology within the same diagnostic category. Disease diagnosis of mental disorders for individual patient usually suffers from mismatch between diagnostic categories and the underlying disease pathology, resulting in uninformed understanding of disease mechanism or treatment choices. For example, for major depressive disorder (MDD), the Diagnostic and Statistical Manual of Mental Disorders (DSM-5, APA (2013)) diagnosis is based on nine symptoms and there are 227 unique symptom profiles that lead to the same MDD diagnosis (Fried and Nesse, 2015); there are about 150 known risk genes (Wray et al., 2018); patients' disease course differ substantially (Wakefield and Schmitz, 2014); treatment responses vary across individuals with moderate early phase response rate and inconclusive results after adjunctive treatments. Above all, the overall treatment response is rather poor and less than 50% of MDD patients achieved remission after first-line treatment (Gartlehner et al., 2011).

In particular, one unique challenge to mental disorders in practice, compared to most of symptomatic diseases, is that observed measures are either subscales of disease assessment instruments or questionnaires obtained from several items (an item is a variable on a questionnaire), or simply a summary total score combining all items. For instance, the sum score of 17 items from Hamilton Rating Scale for Depression (HAM-D, Hamilton (1960)) is often used to measure severity of depression. However, there are several limitations with using summary outcomes as measures of treatment responses. First, item scores can be subject to measurement

error and each item may not be an equally discriminating indicator of depression (Cole et al., 2011). Second, the items have overlaps and the scoring system of taking a simple sum is not informed by patient's mental states and usually which threshold should be used to define a responder is subjective. Third and most importantly, all observed diagnostic symptom items are only partial reflections of the patient's underlying disease states or cognitive states; therefore, using a few diagnostic items or their summary can not give a full and accurate assessment of the disorder, let alone to address patient heterogeneity in their true mental states or cognitive states.

Many statistical and machine learning methods have been developed to estimate individualized treatment rules (ITRs) that make treatment recommendations tailored to each patient's characteristics. They include regression-based methods such as Q-learning (Watkins, 1989; Qian and Murphy, 2011), A-learning (Murphy, 2003) and G-computation (Lavori and Dawson, 2004; Moodie et al., 2007), or value-optimization methods such as outcome weighted learning (O-learning, Zhao et al. (2012, 2015)) and its robust version (Zhou et al., 2017; Liu et al., 2018). However, all existing methods on estimating optimal ITRs rely on observed and unambiguous outcomes, so they are not directly applicable to learn ITRs for the latent mental disorders. One recent work by Butler et al. (2018) estimated patients' latent preferences based on item response theory and used them to combine the observed multiple outcomes, but it did not address the true latent disorder outcomes that should be inferred from the diagnostic items.

On the other hand, in psychiatry and psychology, measurement theory has provided theoretical grounding for modeling observed outcomes as arising from multiple domains of underlying (unobserved) constructs (Bollen, 2002). The latent constructs reflect true underlying mental functions, and the observations can be psychological or clinical instruments (i.e., responses on rating scales), biological measurements (e.g., neuroimaging measures), or behavioral tests. Latent constructs change with disease progression or treatment, but the model from latent constructs to observations stay invariant over time according to measurement theories for

reliable instruments (Lohr, 2002). Within the paradigm of latent constructs, a subject at the baseline locates at some point in this latent space, and treatment changes the patient's position in the latent space (e.g., from severely depressed states to mildly depressed states) and thus changes measurements of future symptoms. A challenge in modeling latent constructs is that patients' observed symptoms define an ultra high-dimensional space (e.g., there are roughly 4^{17} unique symptom combinations defined by 17 items in the HAM-D depression rating scale). However, using a set of lower-dimensional latent state variables through measurement models, one can capture the most important variations of the underlying mental states of patients in this multi-domain space.

In this paper, we propose a general framework by embracing and taking advantage of the intrinsic heterogeneity in mental disorders to achieve precision medicine. The central concept is that some of these heterogeneous characteristics are not noise to be eliminated from diagnosis and treatment. Instead, they arise from intrinsic characteristics of mental functions and their interactions with other social and psychological factors, which can be modeled and quantified. Furthermore, in our framework, by using the well-established latent measurement theory, treatment is assumed to change the trajectory of the patient's true underlying latent states over time and subsequently changes the future observations (symptoms). Under this framework, we propose a novel method to learn optimal ITRs for latent outcomes measured from observed symptoms using latent features together with other observed clinical or biological features.

Specifically, we propose to first learn the relationship between the observed symptoms and the underlying latent states through a machine learning model, restricted Bolzmann machine (RBM), which expresses the complex joint distribution of a large number of observed discrete random variables through binary latent variables. Various deep learning architectures use RBM as building blocks (Goodfellow et al., 2016). In this work, RBM is used as a powerful unsupervised learning tool to model high-dimensional discrete psychiatric symptom items or cognitive tests. Unlike factor models which assume continuous underlying constructs, the discrete nature

of RBM automatically partitions patients into subgroups defined by their latent states, which facilitates learning how subgroups relate to treatment responses. Computation-wise, stochastic gradient descent can be used and contrastive divergence algorithm (Carreira-Perpinan and Hinton, 2005) can be adopted to handle the computational burden of a large number of symptoms and latent states. Next, using the derived underlying states from RBM, we propose a novel method to learn optimal treatment rules that are guaranteed to maximize the underlying disease states in a favorable direction. In particular, we derive a transformation of the observed symptom outcomes based on the learned RBM such that the conditional expectation of the transformed outcome given treatment and baseline features remains the same as the conditional expectation of the corresponding latent outcome. Then we apply machine learning methods such as O-learning or its variation to the transformed outcome to obtain optimal ITRs, with the inferred baseline latent states included as tailoring variables. We further study theoretical properties of the proposed method in terms of consistency and convergence rate.

There are several advantages of our multi-domain latent construct approach. Through RBM, we effectively represent patients' heterogeneous observations using an economical number of latent variables and yet remains flexible. The unsupervised learning of feature representation and outcome representation can include baseline data from many studies including those without treatments. Through a transformation of the observed outcomes, our approach learns consistent optimal ITRs even if the underlying true disease status is never observed. Our method does not require correct specification of the model between latent states before and after treatment. When the RBM model captures main sources of heterogeneity, our approach will provide meaningful partitions of patients into subgroups of patients with homogeneous latent states. More importantly, the proposed method directly links these subgroups to biological measures and treatment responses.

The remainder of this article is organized as follows. In Section 2, we formalize the framework in modeling multi-domain latent features and latent outcomes, and provide an estimator

of the optimal ITR that maximizes the underlying latent outcomes. In section 3, we derive the convergence rate of the proposed estimator. In section 4, we evaluate the finite sample performance of the proposed estimator and test its effectiveness compared to the standard method through simulation studies. In section 5, through an application to real world studies, the Establishing Moderators/Biosignatures of Antidepressant Response - Clinical Care (EMBARC, Trivedi et al. (2016)) and Sequenced Treatment Alternatives to Relieve Depression (STAR*D, Rush et al. (2004)), we demonstrate the utility of our method in tailoring treatments (placebo or antidepressant) for patients with major depression, and identify subgroups informative for treatment recommendations. Lastly, we provide discussion of the proposed method and potential extensions in Section 6.

2 METHOD

2.1 Individualized treatment rules with multiple-domain items

Suppose K binary latent (unobserved) variables are to describe a patient's latent mental status (e.g., positive mood versus negative mood), where $Z_{0k} \in \{0,1\}$, $k = 1, ..., K$, denotes the kth latent state variable at the baseline and $Z_{1k}^{(a)} \in \{0,1\}, k = 1, ..., K$, denotes the corresponding potential latent variable after being treated with treatment a. We assume that treatment is dichotomous with $a \in \{-1, 1\}$. We use $\mathbf{X}$ to denote other baseline covariates (e.g., demographics) that are different from $\mathbf{Z}_0 = (Z_{01}, \cdots, Z_{0K})^T$. Our goal is to identify the optimal ITR that prescribes treatment a for different patients so as to maximize an aggregated value of $\mathbf{Z}_1^{(a)} = (Z_{11}^{(a)}, \cdots, Z_{1K}^{(a)})^T$ for $a = -1, 1$.

Since the latent states $\mathbf{Z}_0$ and $\mathbf{Z}_1^{(a)}$ are unobserved, learning the optimal ITR has to rely on measured multiple-domain symptom items (e.g., depression symptoms on the HAM-D rating scale) that are collected in studies. Specifically, let $\mathbf{Y}_0$ denote the symptom item scores from J domains of an instrument measured at the baseline (Y_{0j} for the jth domain takes value in

$0, 1, ..., l_j$), and let $\mathbf{Y}_1^{(a)}$ denote the potential symptom item scores after being treated with treatment a. Here, $\mathbf{Y}_0$ is informative of the latent variables $\mathbf{Z}_0$ and $\mathbf{Y}_1^{(a)}$ is informative of latent variables $\mathbf{Z}_1^{(a)}$. To understand the relationships among these variables, Figure 1 describes their association. Essentially, both baseline mental health status $\mathbf{Z}_0$ and covariates $\mathbf{X}$ affect the potential mental health outcome $\mathbf{Z}_1^{(a)}$, while $\mathbf{Y}_0$ and $\mathbf{Y}_1^{(a)}$ are respectively the reflective indicators (Jarvis et al., 2003) for $\mathbf{Z}_0$ and $\mathbf{Z}_1^{(a)}$. Furthermore, for reliable instruments, the relationship between the item responses and the latent mental states are assumed to be stable over disease phases (Lohr, 2002). Equivalently, we assume

(C.1) $\mathbf{Y}_0$ is independent of all other random variables conditional on $\mathbf{Z}_0$ and similarly, $\mathbf{Y}_1^{(a)}$ is independent of other random variables conditional on $\mathbf{Z}_1^{(a)}$.

(C.2) The conditional distribution of $\mathbf{Y}_0$ given $\mathbf{Z}_0$ is the same as the conditional distribution of $\mathbf{Y}_1^{(a)}$ given $\mathbf{Z}_1^{(a)}$.

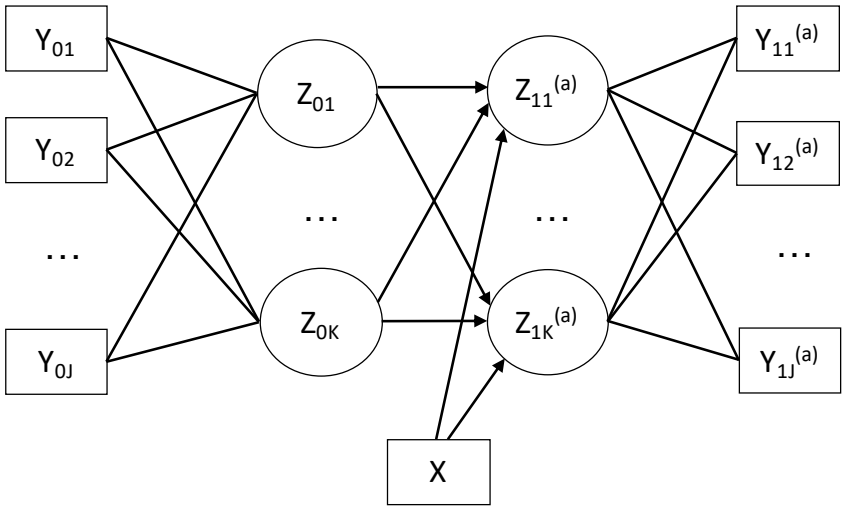

Figure 1: Graphical representation of latent mental states and observed item responses. $Y_{01}, ..., Y_{01}$: baseline observed item responses. $Z_{01}, ..., Z_{0K}$: baseline latent mental states. $Z_{01}^{(a)}, ..., Z_{0K}^{(a)}$: potential latent mental states after being treated with treatment a. $Y_{01}^{(a)}, ..., Y_{01}^{(a)}$: potential item responses after being treated with treatment a. $\mathbf{X}$: a vector of baseline covariates. Note that $Z_{01}, ..., Z_{0K}$ can be correlated with $\mathbf{X}$.

The conditional independence assumption (C.1) is often assumed in factor analysis models and item response theory (Harman, 1976). However, unlike item response theory which uses a single normally distributed latent variable to describe correlation among items $\mathbf{Y}_0$ or $\mathbf{Y}_1^{(a)}$,

we use multiple latent variables $\mathbf{Z}_0$ or $\mathbf{Z}_1^{(a)}$. (C.2) is an invariance assumption in measurement theory that is expected to hold for reliable psychological or clinical instruments. We will use a flexible machine learning model (RBM, details in Section 2.2) to represent the joint distribution of observed discrete symptom items $\mathbf{Y}_0$ through a set of binary latent variables $\mathbf{Z}_0$. Corresponding to Figure 1, each patient has a unique representation of the latent mental states, and the symptom items are conditionally independent given this set of latent states. Using $\mathbf{Z}_0$, the proposed RBM clusters heterogeneous presentations of patient symptoms into homogeneous subgroups while capturing correlations among domains in $\mathbf{Y}_0$. Subjects with different states of $\mathbf{Z}_0$ form distinct subgroups. Patients' baseline mental states $\mathbf{Z}_0$ can be correlated with other baseline characteristics $\mathbf{X}$. Through the interaction between $\mathbf{Z}_0$, $\mathbf{X}$ and treatment a on $\mathbf{Z}_1^{(a)}$, subgroups and feature variables are associated with the potential outcomes under treatment a.

Although an ideal ITR could use most of patient's baseline characteristics including $\mathbf{X}$ and $\mathbf{Y}_0$, there are disadvantages to arbitrarily incorporate $\mathbf{Y}_0$ to construct the ITR. First, in real world applications, there may be some items that do not predict patient's mental states after treatments and thus not useful for prescribing treatments. Additionally, many items may be designed to collect similar information regarding patient's mental status so are redundant. Thus, a treatment rule depending on some specific items may not be clinically useful due to measurement errors, non-differentiating items, and redundancy within a domain. Instead of directly including $\mathbf{Y}_0$ to build ITRs, a more interpretable and useful ITR should be based on some inferred values for the latent states $\mathbf{Z}_0$, denoted as $\mathbf{W}$. For example, in this paper, the inferred $\mathbf{Z}_0$ is given as the predicted mean of $\mathbf{Z}_0$ based on $\mathbf{Y}_0$, i.e., $\mathbf{W} = E[\mathbf{Z}_0|\mathbf{Y}_0]$. This model will downweight unrelated symptom measures through computing the posterior means $\mathbf{W}$, which also reflect uncertainties in each of the items.

Furthermore, instead of learning an ITR to maximize a sum score based on observed $\mathbf{Y}_1$, we learn an ITR to maximize a known aggregate function of the de-noised and more representative underlying mental state variables $\mathbf{Z}_1^{(a)}$, denoted as $g(\mathbf{Z}_1^{(a)})$. This aggregation function can be

chosen such that $g(\mathbf{Z}_1^{(a)})$ reflects the patient's overall mental wellness. For example, $g(\mathbf{Z}_1^{(a)})$ can be the sum of all components in $\mathbf{Z}_1^{(a)}$ if all mental domains are equally important and a higher value in each domain is more desirable. In practice, total score of the observed item responses are usually used to measure patient's disease symptom severity. For example, in the STAR*D study (Rush et al., 2004), the sum of the 16-item Quick Inventory of Depressive Symptomatology (QIDS) scores was used to define the clinical status of each patient (i.e., remission, partial response, and nonresponse). Similarly, taking $g()$ as the simple summation function can be used for latent states $\mathbf{Z}_1$. The optimal ITR for the multiple-domain latent outcomes, denoted by $d^* : (\mathbf{X}, \mathbf{W}) \to \{-1, 1\}$, is the one such that the expected value based on $g(\mathbf{Z}_1^{(a)})$ is maximized when a is the same as $d^*(\mathbf{X}, \mathbf{W})$. Under this definition, the optimal ITR is given by

$$d^*(\mathbf{X}, \mathbf{W}) = \text{argmax}_{a\in\{-1,1\}} E[g(\mathbf{Z}_1^{(a)})|\mathbf{X}, \mathbf{W}] = \text{sign}\{f^*(\mathbf{X}, \mathbf{W})\} \quad (1)$$

with $f^*(\mathbf{X}, \mathbf{W}) = E[g(\mathbf{Z}_1^{(1)})|\mathbf{X}, \mathbf{W}] - E[g(\mathbf{Z}_1^{(-1)})|\mathbf{X}, \mathbf{W}]$.

2.2 Learning optimal ITRs

We assume that data are collected from a randomized trial and consist of the following observations

$$(\mathbf{Y}_{0i}, \mathbf{X}_i, A_i, \mathbf{Y}_{1i}), \quad i = 1, ..., n,$$

from n independent patients. Here, A denotes the treatment assignment and we allow the randomization probability $A = a$, denoted as $\pi(a, \mathbf{Y}_0, \mathbf{X})$, to depend on $\mathbf{X}$ and $\mathbf{Y}_0$ and we assume it is uniformly bounded away from 0 and 1. Because of randomization, A is independent of all potential outcomes given $\mathbf{Y}_0$ and $\mathbf{X}$. Furthermore, we assume the standard stable unit treatment value assumption (SUTVA) holds, i.e., $\mathbf{Y}_1 = \mathbf{Y}_1^{(a)}$ and $\mathbf{Z}_1 = \mathbf{Z}_1^{(a)}$ when $A = a$.

If $\mathbf{W}_i$ and $\mathbf{Z}_{1i}^{(a)}$ with $A_i = a$ were observed, there would be many existing methods including both regression-based methods (Watkins, 1989; Qian and Murphy, 2011; Murphy, 2003; Lavori and Dawson, 2004; Moodie et al., 2007) and value-based methods (Zhao et al., 2012, 2015; Liu

et al., 2018) to estimate the optimal ITR using data $(\mathbf{W}_i, \mathbf{X}_i, A_i, \mathbf{Z}_{1i}), i = 1, ..., n$. Since $\mathbf{W}_i$ and $\mathbf{Z}_{1i}$ are not observed, ITR learning is confronted by two main challenges. First, we need to infer the latent mental status at the baseline, $\mathbf{W}_i$, using the observed items $\mathbf{Y}_{0i}$. More importantly, we need to find a proper proxy for $g(\mathbf{Z}_{1i}^{(a)})$ based on the observed post-treatment items $\mathbf{Y}_{1i}$, where the proxy would yield the same optimal ITR as if $\mathbf{Z}_{1i}^{(a)}$ with $A_i = a$ were observed.

To tackle both challenges, we adopt an RBM to characterize the relationship between the latent states and the observed symptom items. RBMs have wide applications in dimension reduction (Hinton and Salakhutdinov, 2006), feature learning (Coates et al., 2011) and deep learning (Goodfellow et al., 2016). Specifically, this model assumes the joint distribution of $(\mathbf{Y}_0, \mathbf{Z}_0)$ to have form

$$P(\mathbf{Y}_0, \mathbf{Z}_0) = \mathcal{C}^{-1} \exp\left\{ \sum_j \phi_1(Y_{0j}) + \sum_k \phi_2(Z_{0k}) + \sum_j \sum_k \phi_3(Y_{0j}, Z_{0k}) \right\}, \tag{2}$$

where $\mathcal{C}$ is the normalizing constant. Furthermore, we parameterize the exponent part for each node and the pairwise association between categorical Y_{0j} and binary Z_{0k} as

$$\phi_1(Y_{0j}) = \sum_{p=0}^{l_j} \alpha_{jp} I(Y_{0j} = p), \quad \phi_2(Z_{0k}) = \beta_k Z_{0k}, \quad \phi_3(Y_{0j}, Z_{0k}) = \omega_{jk} Y_{0j} Z_{0k},$$

where $\alpha_{j0} = 0$ for identifiability, $j = 1, ..., J$. This parameterization essentially mimics a log-linear model and the RBM implies the following conditional distributions

$$P(Z_{0k} = 1 | \mathbf{Y}_0) = \frac{\exp(\beta_k + \sum_j \omega_{jk} Y_{0j})}{1 + \exp(\beta_k + \sum_j \omega_{jk} Y_{0j})}, \quad k = 1, ..., K, \tag{3}$$

$$P(Y_{0j} = m | \mathbf{Z}_0) = \frac{\exp(\alpha_{jm} + m \sum_k \omega_{jk} Z_{0k})}{\sum_{p=0}^{l_j} \exp(\alpha_{jp} + p \sum_k \omega_{jk} Z_{0k})}, \quad m = 0, ..., l_j, j = 1, ..., J. \tag{4}$$

Based on our assumptions, $P(\mathbf{Y}_1^{(a)} | \mathbf{Z}_1^{(a)}) = P(\mathbf{Y}_0 | \mathbf{Z}_0)$ so follows the same equations in (4). Furthermore, for identifiablity, we assume that some $w_{jk} > 0$ for each k.

Under the RBM, if all the model parameters are known, it is easy to infer $\mathbf{W} = E[\mathbf{Z}_0 | \mathbf{Y}_0]$ using equation (3). Next, we need to find a proxy for $g(\mathbf{Z}_1^{(a)})$ based on $\mathbf{Y}_1$ such that the proxy can be used to learn the same optimal ITR as if $\mathbf{Z}_1^{(a)}$ were observed. To this end, we first notice

$$P(\mathbf{Y}_1^{(a)} = \mathbf{y}_1 | \mathbf{X}, \mathbf{Y}_0) = \sum_{\mathbf{z}_1} P(\mathbf{Y}_1^{(a)} = \mathbf{y}_1 | \mathbf{Z}_1^{(a)} = \mathbf{z}_1, \mathbf{X}, \mathbf{Y}_0) P(\mathbf{Z}_1^{(a)} = \mathbf{z}_1 | \mathbf{X}, \mathbf{Y}_0)$$

$$= \sum_{\mathbf{z}_1} P(\mathbf{Y}_1^{(a)} = \mathbf{y}_1 | \mathbf{Z}_1^{(a)} = \mathbf{z}_1) P(\mathbf{Z}_1^{(a)} = \mathbf{z}_1 | \mathbf{X}, \mathbf{Y}_0).$$

Let $\mathbf{\Sigma}$ be a matrix of dimension $\Pi_j(l_j + 1) \times 2^K$ with elements $P(\mathbf{Y}_1^{(a)} = \mathbf{y}_1 | \mathbf{Z}_1^{(a)} = \mathbf{z}_1)$ for all possible values of $\mathbf{y}_1$ and $\mathbf{z}_1$. For a function $h : \mathbb{R}^J \to \mathbb{R}$, let $\Big[h(\mathbf{y}_1)\Big]_{\mathbf{y}_1}$ be a vector of length $\Pi_j(l_j + 1)$ where $\mathbf{y}_1$ takes all possible values. For example, if $\mathbf{Y}_1^{(a)}$ is a two-dimensional binary vector, then $\Big[h(\mathbf{y}_1)\Big]_{\mathbf{y}_1} = \Big(h\big((0,0)^T\big), h\big((1,0)^T\big), h\big((0,1)^T\big), h\big((1,1)^T\big)\Big)^T$. Similarly, for a function $h : \mathbb{R}^K \to \mathbb{R}$, let $\Big[h(\mathbf{z}_1)\Big]_{\mathbf{z}_1}$ be a vector of length 2^K where $\mathbf{z}_1$ takes all possible values. Then we obtain

$$\Big[P(\mathbf{Y}_1^{(a)} = \mathbf{y}_1 | \mathbf{X}, \mathbf{Y}_0)\Big]_{\mathbf{y}_1} = \mathbf{\Sigma}\Big[P(\mathbf{Z}_1^{(a)} = \mathbf{z}_1 | \mathbf{X}, \mathbf{Y}_0)\Big]_{\mathbf{z}_1}.$$

When $\mathbf{\Sigma}$ is of full column rank, we can construct a left-inverse matrix for $\mathbf{\Sigma}$ as $\mathbf{\Sigma}_L^- = (\mathbf{\Sigma}^T\mathbf{\Sigma})^{-1}\mathbf{\Sigma}^T$ such that $\mathbf{\Sigma}_L^-\mathbf{\Sigma}$ is an identity matrix. Since the number of items in $\mathbf{Y}_0$ is often much larger than the number of latent variables in $\mathbf{Z}_0$, the full column rank condition usually holds. We note that the left-inverse matrix for $\mathbf{\Sigma}$ is not unique and here we use this particular form for practical convenience. Then we obtain

$$\Big[P(\mathbf{Z}_1^{(a)} = \mathbf{z}_1 | \mathbf{X}, \mathbf{Y}_0)\Big]_{\mathbf{z}_1} = \mathbf{\Sigma}_L^-\Big[P(\mathbf{Y}_1^{(a)} = \mathbf{y}_1 | \mathbf{X}, \mathbf{Y}_0)\Big]_{\mathbf{y}_1}.$$

Thus,

$$E[g(\mathbf{Z}_1^{(a)}) | \mathbf{X}, \mathbf{Y}_0] = \mathbf{G}^T\mathbf{\Sigma}_L^-\Big[P(\mathbf{Y}_1^{(a)} = \mathbf{y}_1 | \mathbf{X}, \mathbf{Y}_0)\Big]_{\mathbf{y}_1},$$

where $\mathbf{G}$ is the vector of $\Big[g(\mathbf{z}_1)\Big]_{\mathbf{z}_1}$.

Consequently, we obtain a proxy variable based on symptom items defined as

$$\widetilde{g}(\mathbf{Y}_1^{(a)}) = \mathbf{G}^T\mathbf{\Sigma}_L^-\Big[I(\mathbf{Y}_1^{(a)} = \mathbf{y}_1)\Big]_{\mathbf{y}_1}, \tag{5}$$

such that

$$E[g(\mathbf{Z}_1^{(a)}) | \mathbf{X}, \mathbf{Y}_0] = E[\widetilde{g}(\mathbf{Y}_1^{(a)}) | \mathbf{X}, \mathbf{Y}_0].$$

It follows that

$$E[g(\mathbf{Z}_1^{(a)}) | \mathbf{X}, \mathbf{W}] = E[\widetilde{g}(\mathbf{Y}_1^{(a)}) | \mathbf{X}, \mathbf{W}].$$

Therefore, estimating the optimal ITR in (1) is equivalent to solving

$$\text{argmax}_a E[\widetilde{g}(\mathbf{Y}_1^{(a)})|\mathbf{X}, \mathbf{W}].$$

In other words, in order to learn the optimal ITR that maximizes the latent value $g(\mathbf{Z}_1^{(a)})$, it is equivalent to learn the optimal ITR that maximizes the proxy value $\widetilde{g}(\mathbf{Y}_1^{(a)})$. This surrogate outcome $\widetilde{g}(\mathbf{Y}_1^{(a)})$ can be constructed based on (5) for each subject and we provide a simple example in the Supplementary Material. When data are collected from a randomized trial, since A is independent of the potential outcome $\mathbf{Y}_1^{(a)}$ conditional on $\mathbf{X}$ and $\mathbf{Y}_0$, the optimal ITR maximizes

$$E\left[\frac{I(A = \mathcal{D}(\mathbf{X}, \mathbf{W}))}{\pi(A, \mathbf{Y}_0, \mathbf{X})}\widetilde{g}(\mathbf{Y}_1)\right]. \tag{6}$$

Therefore, many machine learning methods can be used to maximize the empirical counterpart of (6) to estimate the optimal ITR. The detailed algorithm will be given in the next section.

Remark 1. Our model and method have the following advantages. First, we aim to learn optimal treatment rules for latent mental states $\mathbf{Z}_1$ inferred from observed symptoms $\mathbf{Y}_1$, and a benefit of using $\mathbf{Z}_1$ instead of $\mathbf{Y}_1$ is that the derived ITR is not subject to a large amount of measurement errors or random noise in $\mathbf{Y}_1$ because the dimension of the latent outcomes is usually much smaller than the number of the items. Second, using machine learning method, we avoid explicitly modeling potentially complex and inestimable relationship between $\mathbf{Z}_1$ and $\mathbf{Z}_0$ as well as pre-treatment covariates including $\mathbf{W}$. Finally, our method naturally incorporates invariant relation structures between $\mathbf{Y}$'s and $\mathbf{Z}$'s in both pre- and post-treatment phases.

2.3 Details of the algorithm

The computational algorithm can be divided into the following steps. In the first step, we estimate the parameters under the RBM by maximizing the observed data likelihood based on the pre-treatment observations $\mathbf{Y}_{0i}, i = 1, ..., n$, which is given by

$$\prod_{i=1}^{n} P(\mathbf{Y}_{0i}) \;=\; \prod_{i=1}^{n}\left[\sum_{\mathbf{z}_0} P(\mathbf{Y}_{0i}, \mathbf{Z}_0 = \mathbf{z}_0)\right]$$

$$= \prod_{i=1}^{n}\left[\sum_{\mathbf{z}_0}\left\{\mathcal{C}^{-1}\exp\Big(\sum_j\sum_{p=0}^{l_j}\alpha_{jp}I(Y_{0ij}=p)+\boldsymbol{\beta}^T\mathbf{z}_0+\mathbf{Y}_{0i}^T\boldsymbol{\Omega}\mathbf{z}_0\Big)\right\}\right], \quad (7)$$

where $\mathcal{C}=\sum_{\mathbf{y}_0}\sum_{\mathbf{z}_0}\exp(\sum_j\sum_{p=0}^{l_j}\alpha_{jp}I(y_{0j}=p)+\boldsymbol{\beta}^T\mathbf{z}_0+\mathbf{y}_0^T\boldsymbol{\Omega}\mathbf{z}_0)$, and $\boldsymbol{\Omega}$ is a matrix with the (j,k) element being ω_{jk}. The number of latent variables, K, can be determined based on prior research. Without practical guidance from the literature, one can use data-driven method to choose the number of latent variables based on empirical model selection statistics such as the Akaike information criterion (AIC) or the Bayesian information criterion (BIC) for the observed data likelihood.

We adopt stochastic gradient ascent with momentum algorithm (Qian, 1999) to maximize the observed data likelihood in (7) where the gradients are explicitly computed and given in the Appendix A. With the parameters in the RBM estimated, we estimate $\mathbf{W}_i$ as $\widehat{\mathbf{W}}_i$ for each subject i using the posterior mean: the k-th component of $\widehat{\mathbf{W}}_i$ is

$$\widehat{W}_{ik}=E[Z_{0ik}|\mathbf{Y}_{0i};\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\Omega}}]=\frac{\exp(\widehat{\beta}_k+\sum_j\widehat{\omega}_{jk}Y_{0ij})}{1+\exp(\widehat{\beta}_k+\sum_j\widehat{\omega}_{jk}Y_{0ij})},\ k=1,..,K.$$

Next, we estimate $\boldsymbol{\Sigma}$ matrix based on (4) by replacing the parameters with $(\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\Omega}})$. We denote the estimated matrix as $\widehat{\boldsymbol{\Sigma}}$, and then compute its left-inverse matrix $\widehat{\boldsymbol{\Sigma}}_L^-$ as $(\widehat{\boldsymbol{\Sigma}}^T\widehat{\boldsymbol{\Sigma}})^{-1}\widehat{\boldsymbol{\Sigma}}^T$. The proxy post-treatment outcome for subject i is obtained as

$$\widehat{R}_i=\mathbf{G}^T\widehat{\boldsymbol{\Sigma}}_L^-\Big[I(\mathbf{Y}_{1i}=\mathbf{y}_1)\Big]_{\mathbf{y}_1},\ i=1,...,n. \quad (8)$$

Finally, using the derived data, $(\widehat{\mathbf{W}}_i,\mathbf{X}_i,A_i,\widehat{R}_i), i=1,...,n$, we modify the robust augmented outcome-weighted learning (AOL) algorithm in Liu et al. (2018) to learn the optimal ITR. More speifically, the decision function $f(\mathbf{X},\widehat{\mathbf{W}})$ that yields the rule $\mathcal{D}(\mathbf{X},\widehat{\mathbf{W}})=\text{sign}(f(\mathbf{X},\widehat{\mathbf{W}}))$, is obtained by solving

$$\text{argmin}_{f\in\mathcal{H}}\left\{n^{-1}\sum_{i=1}^{n}\frac{|\widehat{R}_i-\widehat{s}(\mathbf{X}_i,\widehat{\mathbf{W}}_i)|}{\pi(A_i,\mathbf{Y}_{0i},\mathbf{X}_i)}\phi\Big(A_i\,\text{sign}\Big(\widehat{R}_i-\widehat{s}(\mathbf{X}_i,\widehat{\mathbf{W}}_i)\Big)f(\mathbf{X}_i,\widehat{\mathbf{W}}_i)\Big)+\lambda_n||f||_{\mathcal{H}}^2\right\}, \quad (9)$$

where $\phi(\cdot)$ is a given large-margin loss, and $\mathcal{H}$ is a reproducing kernel Hilbert space (RKHS) of functions generated by the kernels in constructing f. Here, $\widehat{s}$ is an estimated function to remove

the main effect of $(\mathbf{X}_i, \widehat{\mathbf{W}}_i)$ from $\widehat{R}_i$, which is shown to reduce the variability of the estimated ITR by reducing the second moment of the outcome weights (Liu et al., 2018), and it is usually estimated by a least squares regression with $\widehat{R}_i$ being the outcome and $(\mathbf{X}_i, \widehat{\mathbf{W}}_i)$ being the covariates. Note that $\phi(x) = (1-x)_+$ gives the usual support vector machine (SVM) and $\phi(x) = \log(1+e^{-2x})$ gives the logistic loss, and thus the optimization problem in (9) can be solved by a weighted SVM or a weighted logistic regression, where $|\widehat{R}_i - \widehat{s}(\mathbf{X}_i, \widehat{\mathbf{W}}_i)|/\pi(A_i, \mathbf{Y}_{0i}, \mathbf{X}_i)$ is the subject-specific weight and $A_i \operatorname{sign}\big(\widehat{R}_i - \widehat{s}(\mathbf{X}_i, \widehat{\mathbf{W}}_i)\big)$ is the class label for subject i, $i = 1, ..., n$. We denote $\widehat{f}$ as the final estimator for f, then the estimated optimal treatment rule $\widehat{\mathcal{D}}(\mathbf{X}, \widehat{\mathbf{W}})$ is given by $\operatorname{sign}\big(\widehat{f}(\mathbf{X}, \widehat{\mathbf{W}})\big)$.

3 THEORETICAL RESULTS

In this section, we provide theoretical results on the convergence rate of the value function if the estimated ITR is adopted in the population. For notation convenience, we use $\mathbf{H}$ to denote $(\mathbf{W}, \mathbf{X})$ and $\widehat{\mathbf{H}}$ to denote $(\widehat{\mathbf{W}}, \mathbf{X})$, and assume $\mathbf{H}$ is of d-dimension. The value function associated with the true optimal ITR, $\mathcal{D}^*(\mathbf{H}) = \operatorname{sign}(f^*(\mathbf{H}))$, is the expected aggregated latent outcome when the treatment is determined by $\mathcal{D}^*(\mathbf{H})$. Equivalently, this value function is given as

$$\mathcal{V}(f^*) = E[g(\mathbf{Z}_1^{(a)}) I(a f^*(\mathbf{H}) > 0)].$$

Similarly, the value function associated with the estimated ITR, $\widehat{\mathcal{D}}(\widehat{\mathbf{H}}) = \operatorname{sign}(\widehat{f}(\widehat{\mathbf{H}}))$, is the expected aggregated latent outcome when the treatment is given by $\widehat{\mathcal{D}}(\widehat{\mathbf{H}})$, i.e.,

$$\mathcal{V}(\widehat{f}) = E[g(\mathbf{Z}_1^{(a)}) I(a \widehat{f}(\widehat{\mathbf{H}}) > 0)].$$

To present the theoretical results, we assume that the RKHS $\mathcal{H}$ is the space of functions generated by a Gaussian kenel, $k(\mathbf{H}, \mathbf{H}') = \exp(-\sigma_n^2 ||\mathbf{H} - \mathbf{H}'||^2), \mathbf{H}, \mathbf{H}' \in \mathcal{O}$, and σ_n is a positive parameter varying with n. Moreover, we assume a geometric noise condition regarding the distribution behavior near the Bayesian boundary, which has been used in Steinwart et al.

(2007), Zhao et al. (2012), and Liu et al. (2018). We let

$$\eta(\mathbf{H}) = \frac{E(g(\mathbf{Z}_1^{(1)})|\mathbf{H})}{E(g(\mathbf{Z}_1^{(1)})|\mathbf{H}) + E(g(\mathbf{Z}_1^{(-1)})|\mathbf{H})},$$

so $2\eta(\mathbf{H}) - 1$ is the decision boundary for the true optimal ITR. We define

$$\mathcal{O}^+ = \mathbf{H} \in \text{Supp}(\mathbf{H}) : 2\eta(\mathbf{H}) - 1 > 0,\ \mathcal{O}^- = \mathbf{H} \in \text{Supp}(\mathbf{H}) : 2\eta(\mathbf{H}) - 1 < 0,$$

and let $\Delta(\mathbf{H}) = dist(\mathbf{H}, \mathcal{O}^+)$ if $\mathbf{H} \in \mathcal{O}^+$, and $\Delta(\mathbf{H}) = \text{dist}(\mathbf{H}, \mathcal{O}^-)$ if $\mathbf{H} \in \mathcal{O}^-$, where $dist(\mathbf{H}, \mathcal{O})$ denotes the distance of $\mathbf{H}$ from a set $\mathcal{O}$. Then $\eta(\mathbf{H}) \in (0, 1)$ is assumed to satisfy geometric noise exponent (GNE) condition with a geometric noise exponent of $0 < q < \infty$, if there exists a constant $C > 0$ such that for any $t > 0$,

$$E\Big[\exp\big(-\frac{\Delta(\mathbf{H})^2}{t}\big)|2\eta(\mathbf{H}) - 1|\Big] \leq Ct^{qd/2}.$$

Note that in the case when there is a complete separation, i.e., there is zero probability that subjects will belong to a neighborhood of $2\eta(\mathbf{H}) - 1$, the geometric noise exponent q can be as large as possible.

Theorem 1 *In addition to (C.1) and (C.2), we assume that the RBM model holds with the true parameter value $\boldsymbol{\alpha}_0$, $\boldsymbol{\beta}_0$ and $\boldsymbol{\Omega}_0$. Moreover, we assume*

(A1) η satisfies the GNE condition with geometric noise exponent q;

(A2) $E[\mathbf{Z}_1^{(a)}|\mathbf{H} = \mathbf{h}]$ is continuously differentiable with respect to $\mathbf{h}$ in the support of $\mathbf{H}$, which is assumed to be a compact set.

(A3) $|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0| + |\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0| + |\widehat{\boldsymbol{\Omega}}_0 - \boldsymbol{\Omega}_0| = O_p(n^{-\alpha_1})$, and under $(\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \boldsymbol{\Omega}_0)$, $(\boldsymbol{\Sigma}^T\boldsymbol{\Sigma})^{-1}$ exists and is bounded. Also there exists some function $s^*(\mathbf{H})$ with bounded derivative such that $E\Big|s^*(\mathbf{H}) - \widehat{s}(\mathbf{H})\Big| = O_p(n^{-\alpha_2})$ for some $\alpha_1, \alpha_2 > 0$, and $||\widehat{s}||_\infty$ is bounded in probability.

(A4) The bandwidth $\sigma_n = \lambda_n^{-1/(q+1)d}$ and λ_n satisfies $\lambda_n \to 0, n\lambda_n \to \infty$.

(A5) The large margin loss $\phi(\cdot)$ is either the hinge-loss or logistic loss.

Then for any $v \in (0, 2)$ and $\delta > 0$,

$$\mathcal{V}(f^*) - \mathcal{V}(\widehat{f}) = O_p\left(n^{-\alpha_1} + n^{-\alpha_2} + (n\lambda_n)^{-1} + \lambda_n^{-\frac{2}{2+v} - \frac{(2-v)(1+\delta)}{(2+v)(1+q)}} n^{-\frac{2}{2+v}} + \lambda_n^{\frac{q}{q+1}}\right).$$

The conditions in the theorem are similar to Liu et al. (2018) except (A3). Under the usual regularity condition for the RBM model, $\widehat{\boldsymbol{\alpha}}$, $\widehat{\boldsymbol{\beta}}$, and $\widehat{\boldsymbol{\Omega}}$ are the maximum likelihood estimators so Condition (A3) holds naturally for $\alpha_1 = 1/2$ and $\alpha_2 = 1/2$ if $\widehat{s}(\mathbf{H})$ is obtained through a parametric regression model. Condition (A5) is the sufficient condition for the large margin loss to ensure the Fisher consistency and it holds for the hinge loss or the logistic loss.

The theorem provides a convergence rate of the value function under the estimated ITR. In particularly, when choosing $\lambda_n = n^{-(2q+1)/(q+1)}$, we can obtain this rate to be in a polynomial order of n where the order depends on the convergence rates of the estimated parameters in the RBM model, the estimator $\widehat{s}$, and the noise level near the optimal boundary in term of the geometric noise exponent. The proof of Theorem 1 is similar to Liu et al. (2018). However, the major challenge for this proof is that both the feature variable and outcome for learning the optimal ITR are estimated quantities, i.e., $\widehat{\mathbf{W}}$ and $\widehat{R}$. The details of the proof are provided in the Appendix B.

4 SIMULATION STUDIES

We conducted simulation studies to assess the performance of the proposed method in estimating optimal ITRs with multi-domain latent outcomes, and compared to the standard method that directly uses observed items as predictors and the outcome. We simulated data based on the setting and procedures in section 4.1.

4.1 Simulation Design

We assumed there were three latent outcome domains, and the observed outcomes consisted of 9 items, with three items taking values in $\{0, 1\}$, three taking values in $\{0, 1, 2\}$, and three taking values in $\{0, 1, 2, 3\}$. The baseline observed and latent outcomes $(\mathbf{Y}_0, \mathbf{Z}_0)$ were assumed to follow the joint distribution of an RBM under $\alpha_{jp}, \beta_k, \omega_{jk}$ specified in Table 1. $(\mathbf{Y}_0, \mathbf{Z}_0)$ were

simulated by Gibbs sampling based on the conditional distributions specified in (3) and (4). We then took one subsample every 25 Gibbs samples after a burn-in of $10,000$, and treated the subsamples as i.i.d. data. For each subject i in the subsamples, baseline characteristics X_{1i} and X_{2i} were simulated from $\mathcal{N}(0,1)$, and the treatment assignment A_i was simulated taking value of 1 or -1 with equal probability. Each latent outcome status after treatment, $Z_{1ik}, k = 1, 2, 3$, was then simulated based on

$$P(Z_{1ik} = 1 | \mathbf{Z}_{0i}, \mathbf{X}_i, A_i) = \frac{\exp(\theta_{0k} + \boldsymbol{\theta}_{1k}^T \mathbf{Z}_{0i} + \boldsymbol{\theta}_{2k}^T \mathbf{X}_i + \theta_{3k} A_i + \boldsymbol{\theta}_{4k}^T A_i \mathbf{Z}_{0i} + \boldsymbol{\theta}_{5k}^T A_i \mathbf{X}_i)}{1 + \exp(\theta_{0k} + \boldsymbol{\theta}_{1k}^T \mathbf{Z}_{0i} + \boldsymbol{\theta}_{2k}^T \mathbf{X}_i + \theta_{3k} A_i + \boldsymbol{\theta}_{4k}^T A_i \mathbf{Z}_{0i} + \boldsymbol{\theta}_{5k}^T A_i \mathbf{X}_i)},$$

with parameters specified in Table 2. Lastly, $\mathbf{Y}_{1i}$, the observed outcome responses after treatment for each subject were simulated based on $\mathbf{Z}_{1i}$ and the conditional distributions in (4). We assumed that the sum of $\mathbf{Z}_1$ represented the overall severity of a mental disorder, and our goal was to minimize it.

Table 1: Parameters used for modeling $(\mathbf{Y}_0, \mathbf{Z}_0)$ in the simulation study.

α_{jp}			β_k			ω_{jk}		
-1.0			-1.0	1.0	-0.5	1.0	0	-1.0
-0.5						1.2	0.5	0
-1.0						0.8	1.0	-0.5
-1.0	-2.0					0.3	0	-0.5
-0.5	-1.0					0.1	0.3	0
-1.0	-1.0					0.2	0	-1.2
0.5	-0.5	-0.5				0.2	-0.1	-1.0
-0.5	-1.0	-2.0				0.5	-0.2	0
-0.5	-1.0	-2.0				0.3	0	-1.5

Table 2: Parameters for simulating $\mathbf{Z}_1$ in the simulation study.

k	θ_{0k}	$\boldsymbol{\theta}_{1k}$			$\boldsymbol{\theta}_{2k}$		θ_{3k}	$\boldsymbol{\theta}_{4k}$			$\boldsymbol{\theta}_{5k}$	
1	-2.0	1.0	0.5	0	-0.5	-0.5	1.0	-1.0	0.5	0	0.5	-0.5
2	-2.5	0.5	1.5	-1.0	1.0	-0.5	0.5	-0.5	-1.0	0	-1.0	0
3	-1.0	0	-1.0	1.0	-1.0	0.5	0.5	-0.2	0	0.5	0	-0.5

On each simulated training data set, we applied our method and minimized the surrogate outcome $\widehat{R}_i$ as constructed in (8) with $(\widehat{\mathbf{W}}_i, \mathbf{X}_i)$ being the predictors. We compared with the

"standard method" that minimized the sum of observed $\mathbf{Y}_{1i}$ with predictors of $(\mathbf{Y}_{0i}, \mathbf{X}_i)$. Additionally, we compared with the "gold standard" approach where outcome R_i and predictor $\mathbf{W}_i$ were both constructed using the true RBM parameters. For each method, we fit one additional model that also included pairwise interactions among $\widehat{\mathbf{W}}_i$, $\mathbf{Y}_{0i}$, or $\mathbf{W}_i$ respectively in the predictors. Logistic loss and SVM with linear kernel were adopted to solve the weighted classification in O-learning for each method. The estimated ITRs were evaluated on an independent test set of sample size $10,000$, and we calculated the empirical value function with respect to the sum of $\mathbf{Z}_1$ (the lower the better) as well as the accuracy rate of the estimated optimal treatments on the test set under the estimated ITRs.

To assess the robustness of our method, in addition to the O-learning algorithm, we also implemented the Q-learning algorithm (Qian and Murphy, 2011) to estimate the optimal ITRs under the specified covariates and outcome in each of the three methods. Furthermore, we considered model misspecification where $(\mathbf{Y}_0, \mathbf{Z}_0)$ was not generated from an RBM by allowing pairwise interactions among latent domains (i.e., adding $Z_{01}Z_{02} - Z_{02}Z_{03}$ to the exponent in (2)). In our method we still fit an RBM to learn the joint distribution of $(\mathbf{Y}_0, \mathbf{Z}_0)$, while the "gold standard" method used the true distribution of $(\mathbf{Y}_0, \mathbf{Z}_0)$.

When fitting the RBMs in our method, we set the starting value of $(\omega_{11}, \omega_{23}, \omega_{39})$ to be $(1, 1, -1)$ to facilitate learning the latent domains in the right direction, which reflects the real-world situation when we had some pre-knowledge about certain items and potential latent domains before fitting the model. The starting values for the other parameters in the RBM were randomly generated from Uniform distribution $(-0.1, 0.1)$. We fit the RBMs using stochastic gradient ascent with momentum of 0.5 under batch size of 80, 160, and 240 for training sample size of 400, 800, and 1200 respectively, and a learning rate of 2 followed by 1 for 20 and 40 iterations on the whole sample.

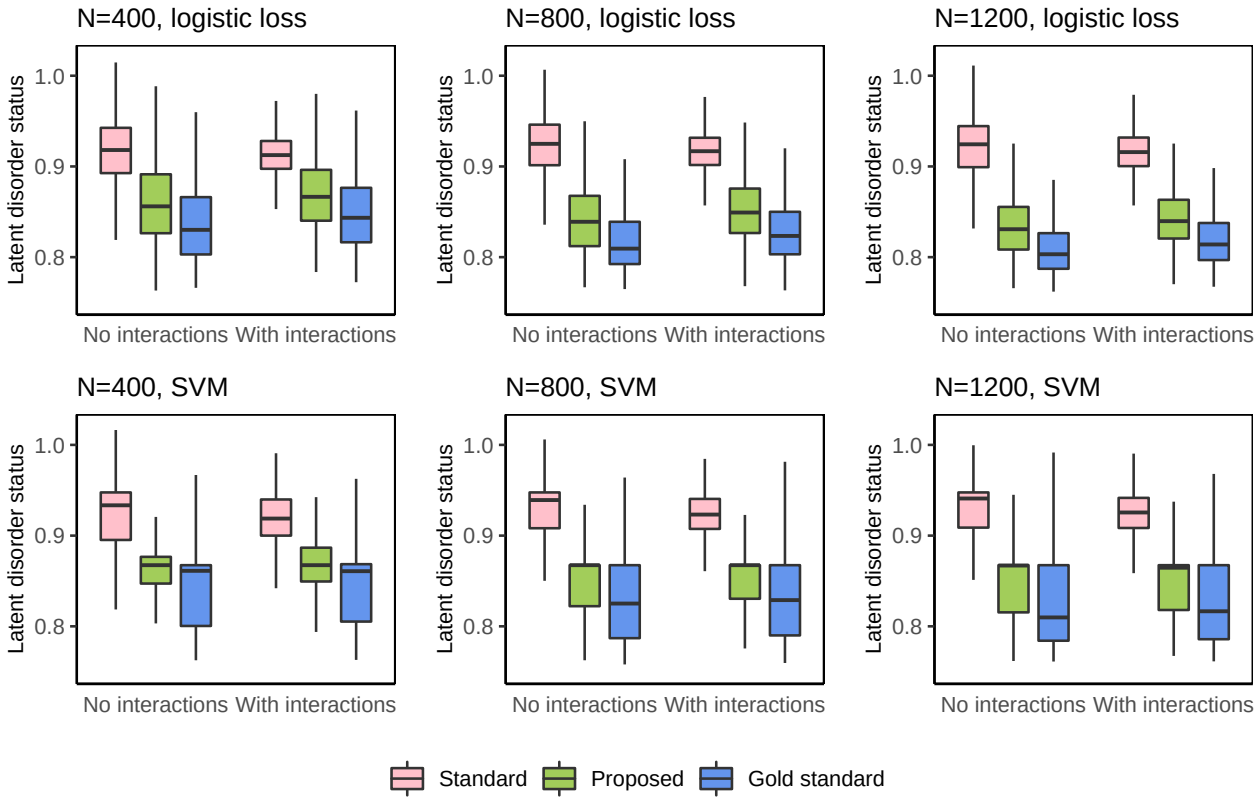

Figure 2: Simulation results: latent disorder status (the lower, the better) evaluated on the test set for the three methods using logistic loss or support vector machine with linear kernel, with or without pairwise interactions among baseline observed items or latent states in the predictors, under training sample size of $N = 400, 800, 1200$. Standard method: observed items $\mathbf{Y}_0$ in the predictors and sum of observed items after treatment $\mathbf{Y}_1$ as the outcome; Proposed method: inferred baseline latent states $\widehat{\mathbf{W}}_i$ in the predictors with the estimated surrogate outcome $\widehat{R}_i$; Gold standard method: baseline latent states $\mathbf{W}_i$ in the predictors and surrogate outcome R_i both constructed with true parameters.

Table 3: Simulation results: latent disorder status (the lower, the better) and accuracy rate of estimated optimal treatments on the test set based on the estimated ITRs from the three methods.

	Standard	Proposed	Gold standard	Standard	Proposed	Gold standard
Latent disorder status (the lower, the better), mean (sd)						
Logistic loss	No interactions			With interactions		
N=400	0.92 (0.04)	0.86 (0.05)	0.84 (0.05)	0.91 (0.02)	0.87 (0.04)	0.85 (0.04)
N=800	0.92 (0.03)	0.84 (0.04)	0.82 (0.04)	0.92 (0.02)	0.85 (0.04)	0.83 (0.04)
N=1200	0.92 (0.03)	0.83 (0.04)	0.81 (0.03)	0.92 (0.02)	0.84 (0.03)	0.82 (0.03)
SVM	No interactions			With interactions		
N=400	0.92 (0.04)	0.87 (0.05)	0.85 (0.06)	0.92 (0.03)	0.87 (0.05)	0.85 (0.06)
N=800	0.93 (0.03)	0.85 (0.04)	0.83 (0.05)	0.92 (0.02)	0.86 (0.04)	0.84 (0.05)
N=1200	0.93 (0.03)	0.85 (0.04)	0.83 (0.05)	0.92 (0.02)	0.85 (0.04)	0.83 (0.05)
Relative latent disorder status with respect to the proposed method, mean (sd)						
Logistic loss	No interactions			With interactions		
N=400	1.07 (0.08)	1.00	0.98 (0.06)	1.05 (0.06)	1.00	0.98 (0.05)
N=800	1.10 (0.07)	1.00	0.98 (0.05)	1.08 (0.06)	1.00	0.98 (0.05)
N=1200	1.11 (0.07)	1.00	0.98 (0.05)	1.09 (0.05)	1.00	0.98 (0.04)
SVM	No interactions			With interactions		
N=400	1.06 (0.08)	1.00	0.98 (0.08)	1.06 (0.07)	1.00	0.98 (0.07)
N=800	1.09 (0.07)	1.00	0.98 (0.07)	1.08 (0.06)	1.00	0.98 (0.07)
N=1200	1.10 (0.07)	1.00	0.98 (0.07)	1.09 (0.06)	1.00	0.98 (0.06)
Accuracy rate, mean (sd)						
Logistic loss	No interactions			With interactions		
N=400	0.67 (0.03)	0.72 (0.04)	0.74 (0.04)	0.68 (0.02)	0.72 (0.03)	0.73 (0.03)
N=800	0.67 (0.03)	0.74 (0.03)	0.75 (0.03)	0.68 (0.02)	0.73 (0.03)	0.74 (0.03)
N=1200	0.67 (0.02)	0.74 (0.03)	0.76 (0.03)	0.68 (0.02)	0.74 (0.02)	0.75 (0.02)
SVM	No interactions			With interactions		
N=400	0.67 (0.03)	0.72 (0.04)	0.73 (0.05)	0.68 (0.02)	0.71 (0.04)	0.73 (0.04)
N=800	0.67 (0.02)	0.73 (0.03)	0.74 (0.04)	0.67 (0.02)	0.73 (0.03)	0.74 (0.04)
N=1200	0.67 (0.02)	0.73 (0.03)	0.75 (0.04)	0.67 (0.02)	0.73 (0.03)	0.75 (0.04)

Standard method: observed items $\mathbf{Y}_0$ in the predictors and sum of observed items after treatment $\mathbf{Y}_1$ as the outcome; Proposed method: inferred baseline latent states $\widehat{\mathbf{W}}_i$ in the predictors with the estimated surrogate outcome $\widehat{R}_i$; Gold standard method: baseline latent states $\mathbf{W}_i$ in the predictors and surrogate outcome R_i both constructed with true parameters.

4.2 Simulation results

Results from 1000 replications are shown in Figure 2, Table 3, and the Supplementary Material. Our proposed method outperforms the alternative standard method in all scenarios with statistical significance, achieving a lower latent disorder status and a higher optimal treatment estimation accuracy. Though there could be outliers, the proposed method yields a better median and mean value than the standard method with non-overlapping 95% confidence intervals. For example, the 95% confidence interval for the mean disorder status is (0.857, 0.863) for the proposed method and (0.918, 0.922) for the standard method under sample size of 400, and is (0.828, 0.832) and (0.918, 0.922) for the two methods respectively under sample size of 1200. With the increase of sample size, our method performs better with a lower mean and smaller variance, while for the standard method the mean value remains about the same. These results demonstrate the desired property of our method that by de-noising the mixed observed outcomes and learning the representative underlying disorder domains, we are able to better optimize the intrinsic mental states of a patient. Furthermore, the performance of our method almost matches the "gold standard", indicating a good estimation efficiency in the RBM parameters.

Including interactions among the predictors ($\mathbf{Y}_0$, $\widehat{\mathbf{W}}$, or $\mathbf{W}$) compared to not including them yields similar mean value and accuracy rate for all methods, but smaller variance especially for the standard method. Using the O-learning algorithm under the logistic loss or SVM and the Q-learning algorithm yielded consistent results for all methods. When $(\mathbf{Y}_0, \mathbf{Z}_0)$ were not from an RBM, our method does not converge to the "gold standard" where $(\mathbf{Y}_0, \mathbf{Z}_0)$ was correctly specified, but the fitted RBMs still perform well in capturing the relationship between $\mathbf{Y}_0$ and $\mathbf{Z}_0$, achieving better latent disorder status than the standard method on the test set (Figure 1.2 in the Supplementary Material).

The RBM model training easily scales up with the sample size due to the use of stochastic gradient descent, with a computing time of 19s, 30s, and 52s under sample size of 400, 800 and

1200 on a 2.7 GHz Intel Core i5 processor with the specified training procedures in section 4.1.

5 REAL DATA APPLICATION

We applied the proposed method to learn optimal ITRs for treating major depressive disorder (MDD) based on a real world study EMBARC (Trivedi et al., 2016). Patients with MDD in the EMBARC study were randomized to receive placebo or sertraline (SER, a selective serotonin reuptake inhibitor [SSRI] antidepressant) for 8 weeks. Hamilton Depression Rating Scale (HAM-D) was recorded for each patient at baseline and after treatment at week 8. Additionally, response status at the end of 8 weeks to the assigned treatment was evaluated by clinicians where "responder" was defined as being at least "much improved" on the Clinical Global Improvement scale (CGI), and we considered the responder status as an external outcome in the analysis. 210 participants in EMBARC were included in the analysis, and the response rate was 0.469 and 0.366 for those who took SER and placebo respectively.

Our aim was to find ITRs that maximize the latent positive mental states after treatment. There were 15 HAM-D items available in EMBARC and we excluded items "loss of insight" and "weight loss" following Chekroud et al. (2017), since there were too few endorsements for "loss of insight" and "weight loss" was not collected in the same way across trials (van Borkulo et al., 2015). We further combined some sparse categories "3" and "4" in the items to increase the learning power. In addition, we considered incorporating the baseline HAM-D scores from 3841 MDD patients from the STAR*D trial (Rush et al., 2004), together with the EMBARC sample, to learn the RBMs more efficiently.

In the literature, 2 to 4 latent factors were detected from depression symptoms (Van Dorn et al., 2016; Van Loo et al., 2012; Sunderland et al., 2013). Hence, we fitted RBMs with 2 to 4 latent mental domains based on the measured HAM-D items, and the RBM with 3 latent domains yielded the lowest BIC. Therefore, we used RBM with 3 latent mental domains to construct the optimal ITRs. To determine the direction of the estimated latent mental domains,

we scored each latent domain by a value from -1 to 1, where the score for the latent domain k was computed as the proportion of negative entries minus proportion of positive entries in $\widehat{\omega}_{.k}$. Since a higher value in HAM-D items indicates more severe symptoms, more negative coefficients in $\widehat{\omega}_{.k}$ for a domain indicates a "good", positive domain. Thus, the scoring system gives positive scores to those "good" mental domains and negative scores to "bad" domains. Therefore, we effectively maximizes the total underlying positive mental states by maximizing the sum of the identified latent domains from the RBM model after scoring. Furthermore, since elevated neuroticism was founded to have greater therapeutic benefit from SSRIs relative to placebo (Tang et al., 2009), we modeled the 12-item Neuroticism subscale from the NEO five-factor inventory-3 (NEO-FFI-3, McCrae and Costa (2010)) at baseline for the EMBARC sample by an RBM with one latent domain. The predicted latent neuroticism variable was included as a predictor in our method.

We compared to the standard method which minimized the sum of observed HAM-D scores after treatment with individual baseline HAM-D items and total neuroticism subscale from NEO-FFI-3 included in the predictor set. Additionally, we considered the method that minimized the sum of observed HAM-D scores but included in the predictors the predicted baseline latent mental states for depression and the predicted latent neuroticism variable based on the fitted RBMs (denoted as "with latent predictors" in Figure 3). A separate model that additionally included in the predictors the pairwise interactions among baseline observed items or predicted latent depression domains was fitted for each method. Other variables in the predictors included age, gender, race, alcohol dependence, and Flanker Interference Accuracy score (higher score indicates reduced cognitive control). Due to a large number of items, single-step contrastive divergence algorithm (Carreira-Perpinan and Hinton, 2005) was adopted to fit the RBMs for faster convergence. SVMs with linear kernel were used for all model fittings in O-learning.

For all methods, we randomly split the EMBARC sample into a training set and a test

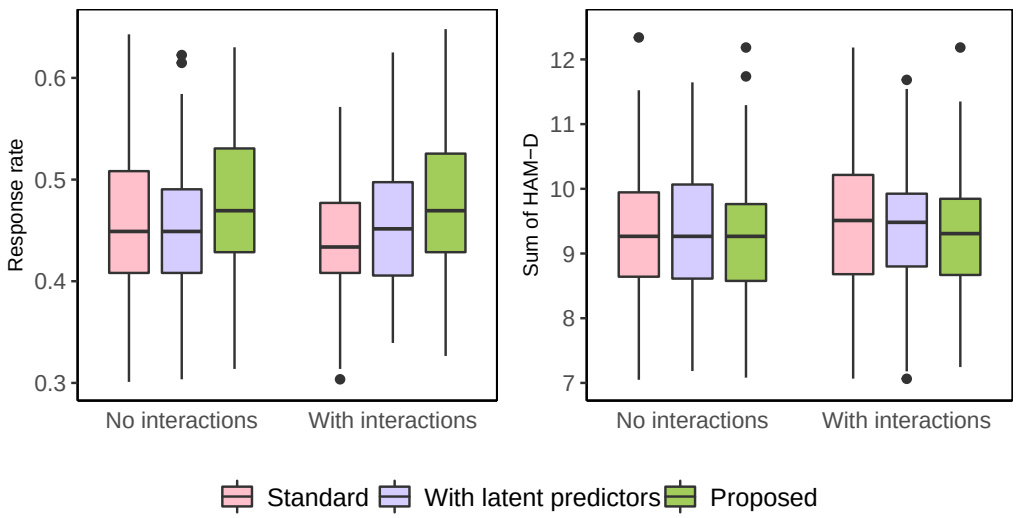

Figure 3: Response rate and total HAM-D depression score (the lower, the better) after treatment evaluated on the test set of the EMBARC study for the three methods from 100 iterations, with or without pairwise interactions among baseline HAM-D items or predicted baseline latent states from HAM-D items included in the predictor set. Standard method: individual HAM-D items and NEO-FFI-3 subscale sum score in the predictors and HAM-D sum score as the outcome; With latent predictors: estimated baseline latent states from HAM-D and estimated latent neuroticism from NEO-FFI-3 subscale in the predictors and HAM-D sum score as the outcome; Proposed method: estimated latent states from HAM-D and estimated latent neuroticism from NEO-FFI-3 subscale in the predictors with the estimated surrogate outcome.

set with a 1:1 ratio and repeated the procedure 100 times. Value function computed from the responder status (the higher the better) on the test set with RBM trained on the pooled EMBARC training sample and STAR*D from the 100 iterations. By incorporating subjects from a broader population, we may learn more representative latent domains and achieve a higher estimation efficiency for parameters in the RBM. We have checked that the RBM learned using pooled STAR*D and EMBARC training sample also had a good generalization properties on the EMBARC test set. The analyses results are presented in Figure 3. Our method performs the best in all settings, and achieves a mean response rate of 0.475 compared to 0.437 for the standard method with interactions. In addition, our proposed ITR performs better than or as well as other standard methods in 81% of random partitions of data into training and testing samples. In all scenarios, our method yields a higher response rate than the universal rule of assigning all to SER, and on average 15% of patients are recommended with placebo instead

of SER under our estimated optimal ITRs. This demonstrates the potential of individualized treatments in improving responses while reducing medical burdens (not recommending SER to all patients) at the same time.

Furthermore, we compared the value function in terms of the total HAM-D score after treatment (the lower the better) in Figure 3. Our method yields the lowest score in all scenarios, with a mean of 9.27 and 9.23 compared to 9.54 and 9.34 from the standard method with or without interactions. That is, our method also minimized the observed depression symptoms even though the algorithm did not directly use total HAM-D score as the outcome variable to minimize, but instead used a denoised and more representative construct $\widetilde{g}(\mathbf{Y}_1)$. On the other hand, compared with the standard method, optimizing the same observed HAM-D total score but using predicted baseline mental states for depression and neuroticism improves the response rate and the total depression score after treatment, indicating that the predicted intrinsic latent variables are more informative to differentiate treatment effects. Lastly, incorporating baseline HAM-D from STAR*D compared to using EMBARC alone in pre-training RBM improved the performance of our estimator. This shows the advantage of our method in that it allows borrowing information from other studies to gain efficiency and potentially achieve higher generalizability in representing the relationship between baseline observed items and the latent states.

We applied our method to the whole EMBARC sample, and achieved a response rate of 0.50 under the estimated ITR with 16% patients recommended to placebo. The heatmap of baseline HAM-D item scores for the pooled EMBARC and STAR*D sample, stratified by subgroups based on the three predicted mental domains identified from the RBM model fitted with HAM-D items is shown in Figure 4. The eight subgroups were defined by median split of each latent domain, where "1" (or "0") denotes higher (or lower) than the median value of that domain. The first latent domain is a positive affect domain (higher value indicates less symptoms) characterized by less insomnia problem; the second is a negative affect domain

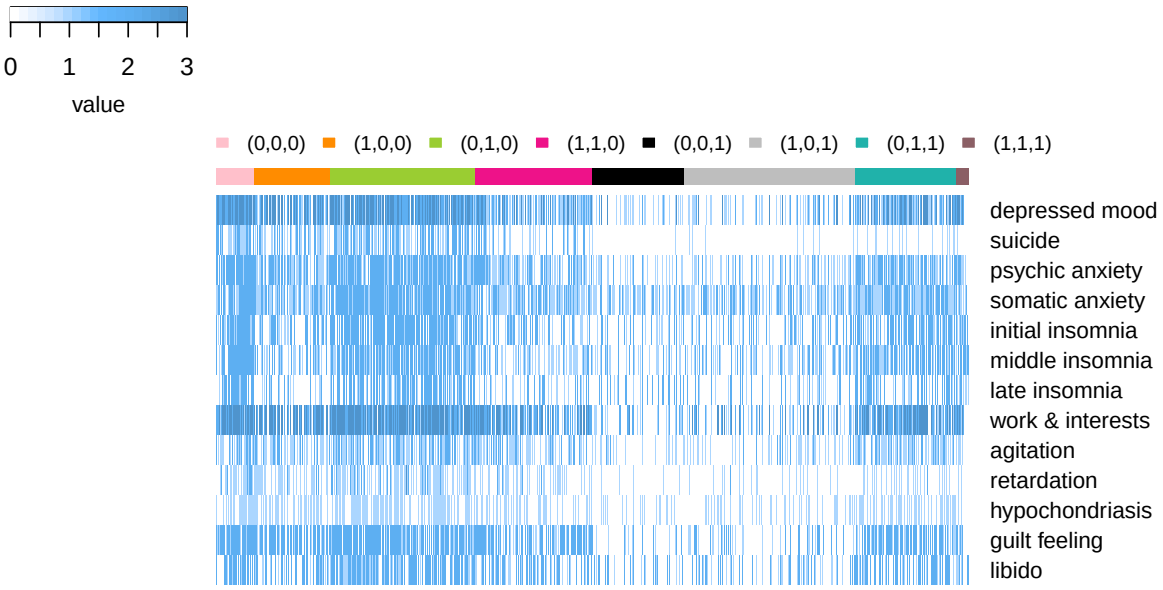

Figure 4: Heatmap of baseline HAM-D item scores by subgroups based on estimated latent mental status for MDD patients in EMBARC and STAR*D. Eight subgroups were identified based on the three latent domains in the RBM fitted with HAM-D items, where "1" (or "0") denotes higher (or lower) than the median value of that domain in the sample.

characterized by higher anxiety and lower work interest; and the last is a positive affect domain characterized by good mood, less suicidal idea and less guilt feeling. These three identified latent domains could reflect positive valence, negative valence and activation (or arousal) of emotion, which are widely identified constructs underlying affective states in psychopathology (Posner et al., 2005). Clear patterns of HAM-D scores by the subgroups can be observed in the figure, which indicates the utility of our proposed method in identifying representative subgroups.

The heatmap of important tailoring variables (importance measured by the magnitude of their coefficients of the standardized variables on the decision rule) identified from the estimated ITR and the eight baseline subgroups for depression were plotted in Figure 5 stratified by the recommended treatments. Active antidepressant SER is recommended for older patients with lower deficits in cognitive control, which is consistent with prior research findings (Webb et al., 2019). Patients with the race of "black" and who had heavier alcohol dependence also tend to

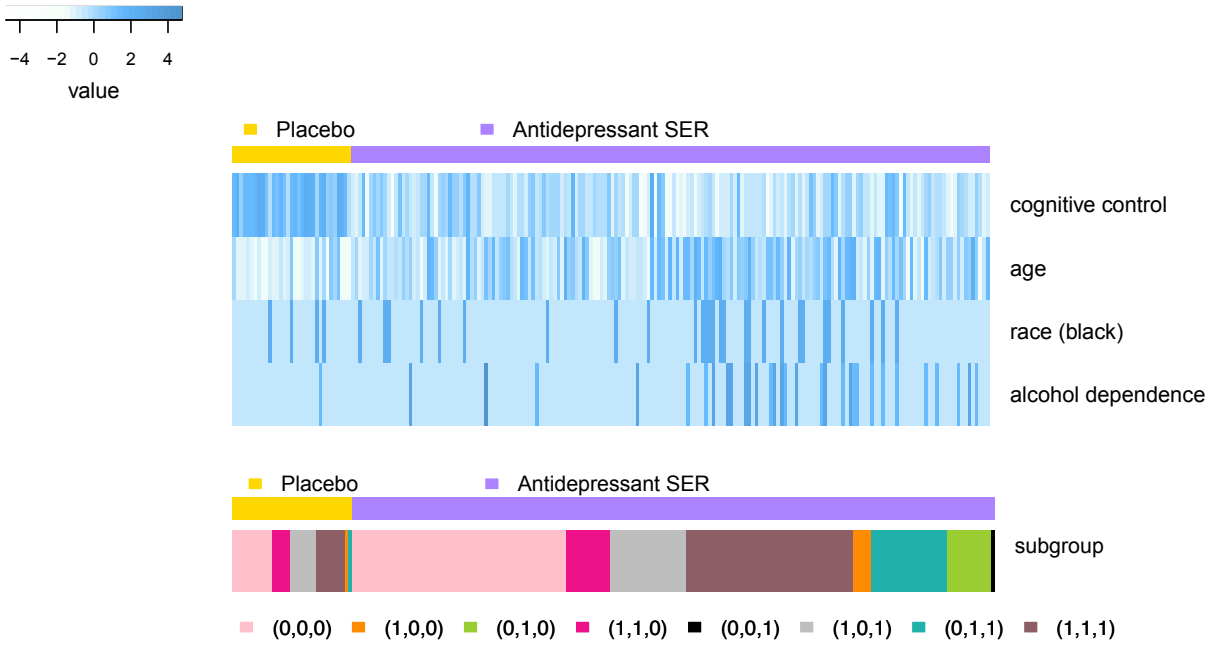

Figure 5: Heatmap of feature variables (standardized) and latent subgroups for EMBARC patients, stratified by recommended treatment under the ITR estimated by the proposed method. Eight subgroups were identified based on the three latent domains in the RBM fitted with HAM-D items from EMBARC and STAR*D, where "1" (or "0") denotes higher (or lower) than the median value of that domain in the sample.

benefit from SER. Additionally, among the eight subgroups, we identified that most patients in subgroup "(1,0,0)" and "(0,1,1)" and all patients in subgroup "(0,1,0)" and "(0,0,1)" are recommended with antidepressant SER. In particular, the subgroup "(0,1,0)" represents the patients with most severe symptoms based on the characterization of the three latent domains, and our result is in line with prior research finding that higher baseline depressive symptom severity predicts advantage of SSRI over placebo (Webb et al., 2019). Beyond that, our subgroup will provide thresholds for selecting most severe patients based on predicted $\mathbf{Z}_{1k}$'s. Another subgroup predicted to benefit from SER is patients with insomnia, more anxious, but more good mood, less guilt and less suicidal ideation (group "(0,1,1)"). Thus, we provide more specific characterization and treatment recommendation for MDD patients who were not among the most severe but seeking effective treatments.

6 DISCUSSION

Under a latent outcome framework, we provide a novel approach for learning optimal ITRs that can address the complexity and heterogeneity that underlie the challenges facing mental disorders. Our approach builds value function in ITR optimization through latent state variables at the baseline and after the treatment. The ITRs connect patients' biological, behavioral and psychological factors to treatment responses defined in terms of their true underlying disease states. By modeling the latent mental status, which are of much lower dimension than the observed psychological or clinical instruments, we effectively denoise and identify subgroups to achieve precision medicine. Specifically, subgroups of patients characterized by homogeneous baseline mental health status can be identified, which facilitates clinical interpretation and are more informative for treatment recommendation than using a simple summary score.

Our method also enjoys the flexibility and potentially higher generalizability in that it can borrow information from other studies, including randomized trials and observational studies, in pre-learning the relationship between observed and latent outcomes through the RBM since only baseline data are required. By incorporating subjects from a wider population, we may learn more representative latent domains while achieving higher estimation efficiency at the same time. Additionally, other latent domain predictors can be learned from separate RBMs to extract meaningful features from high dimensional inputs, such as neuroimaging measures and behavioral tests.

In this paper, we demonstrate our method by optimizing the sum of the latent domains as the objective. However in practice, other meaningful objective functions can be constructed. For instance, domain-specific ITRs can be learned if certain disorder domain is of particular interest. It would also be interesting to compare domain-specific optimal treatment rules and consider clinically meaningful ways of combining them. Lastly, our proposed method can be extended to scenarios when both $\mathbf{Y}$ and $\mathbf{Z}$ are continuous. For example, Welling et al. (2005) proposed an RBM with observed and latent units from exponential families, and it can also be

trained by minimizing contrastive divergence. Furthermore, our method can also be extended to optimizing dynamic treatment rules where at each stage the latent surrogate outcome can be constructed based on the observed items at a given stage and the same conditional distributions learned at the baseline.

A Appendix: Gradient of the RBM likelihood function

The log-likelihood of the observed data based on the pre-treatment observations $\mathbf{Y}_{0i}, i = 1, ..., n$ is

$$\begin{aligned}
& \log(\Pi_i P(\mathbf{Y}_{0i})) \\
= \; & -n \log \mathcal{C} \; + \sum_i \sum_j \sum_{p=0}^{l_j} \alpha_{jp} I(Y_{0ij} = p) + \sum_i \log \left[\sum_{\mathbf{z}_0} \left\{ \exp(\boldsymbol{\beta}^T \mathbf{z}_0 + \mathbf{Y}_{0i}^T \boldsymbol{\Omega} \mathbf{z}_0) \right\} \right] \\
= \; & -n \log \Big[\sum_{\mathbf{y}} \sum_{\mathbf{z}} \exp(\sum_{j=1}^{J} \sum_{p=0}^{l_j} \alpha_{jp} I(y_j = p) + \boldsymbol{\beta}^T \mathbf{z} + \mathbf{y}^T \boldsymbol{\Omega} \mathbf{z}) \Big] + \sum_i \sum_j \sum_{p=0}^{l_j} \alpha_{jp} I(Y_{0ij} = p) \\
& + \sum_i \log \left[\sum_{\mathbf{z}_0} \left\{ \exp(\boldsymbol{\beta}^T \mathbf{z}_0 + \mathbf{Y}_{0i}^T \boldsymbol{\Omega} \mathbf{z}_0) \right\} \right].
\end{aligned}$$

The gradients with respect the parameters are

$$\begin{aligned}
\frac{\partial \log(\Pi_i P(\mathbf{Y}_{0i}))}{\partial \alpha_{jp}} &= -\frac{n}{\mathcal{C}} \sum_{\mathbf{y}} \sum_{\mathbf{z}} \exp \big[\sum_{j=1}^{J} \sum_{p=0}^{l_j} \alpha_{jp} I(y_j = p) + \boldsymbol{\beta}^T \mathbf{z} + \mathbf{y}^T \boldsymbol{\Omega} \mathbf{z} \big] I(y_j = p) \\
&\quad + \sum_i I(Y_{0ij} = p), \\
\frac{\partial \log(\Pi_i P(\mathbf{Y}_{0i}))}{\partial \boldsymbol{\beta}} &= -\frac{n}{\mathcal{C}} \sum_{\mathbf{y}} \sum_{\mathbf{z}} \exp \big[\sum_{j=1}^{J} \sum_{p=0}^{l_j} \alpha_{jp} I(y_j = p) + \boldsymbol{\beta}^T \mathbf{z} + \mathbf{y}^T \boldsymbol{\Omega} \mathbf{z} \big] \mathbf{z} \\
&\quad + \sum_i \left[\frac{\sum_{\mathbf{z}_0} \exp(\boldsymbol{\beta}^T \mathbf{z}_0 + \mathbf{Y}_{0i}^T \boldsymbol{\Omega} \mathbf{z}_0) \mathbf{z}_0}{\sum_{\mathbf{z}_0} \exp(\boldsymbol{\beta}^T \mathbf{z}_0 + \mathbf{Y}_{0i}^T \boldsymbol{\Omega} \mathbf{z}_0)} \right] \\
\frac{\partial \log(\Pi_i P(\mathbf{Y}_{0i}))}{\partial \omega_{jk}} &= -\frac{n}{\mathcal{C}} \sum_{\mathbf{y}} \sum_{\mathbf{z}} \exp \big[\sum_{j=1}^{J} \sum_{p=0}^{l_j} \alpha_{jp} I(y_j = p) + \boldsymbol{\beta}^T \mathbf{z} + \mathbf{y}^T \boldsymbol{\Omega} \mathbf{z} \big] y_j z_k \\
&\quad + \sum_i \left[\frac{\sum_{\mathbf{z}_0} \exp(\boldsymbol{\beta}^T \mathbf{z}_0 + \mathbf{Y}_{0i}^T \boldsymbol{\Omega} \mathbf{z}_0) Y_{0ij} z_{0k}}{\sum_{\mathbf{z}_0} \exp(\boldsymbol{\beta}^T \mathbf{z}_0 + \mathbf{Y}_{0i}^T \boldsymbol{\Omega} \mathbf{z}_0)} \right],
\end{aligned}$$

B Appendix: Proof of Theorem 1

We use $R^{(a)}$ to denote $\widetilde{g}(\mathbf{Y}_1^{(a)})$ and R to denote $\widetilde{g}(\mathbf{Y}_1)$. Then based on Section 2.2, since the conditional mean of $R^{(a)}$ is the same to the conditional mean of $g(\mathbf{Z}_1^{(a)})$ given $\mathbf{X}$ and $\mathbf{Y}_0$, it is clear

$$\mathcal{V}(f^*) = E\left[\frac{I(Af^*(\mathbf{H})>0)R}{\pi(A,\mathbf{Y}_0,\mathbf{X})}\right]$$

and

$$\mathcal{V}(\widehat{f}) = E\left[\frac{I(A\widehat{f}(\widehat{\mathbf{H}})>0)R}{\pi(A,\mathbf{Y}_0,\mathbf{X})}\right].$$

In the subsequent proof, we use c to denote some constant which may be different in each context.

We define

$$\widetilde{\mathcal{V}}(f) = E\Big[\frac{\widehat{R}\,I(Af(\widehat{\mathbf{H}})>0)}{\pi(A,\mathbf{Y}_0,\mathbf{X})}\Big]$$

and let $\check{f}$ be the optimal decision function to maximize $\widetilde{\mathcal{V}}(f)$, i.e., the optimal ITR based on the feature $\widehat{\mathbf{H}}$ and outcome $\widehat{R}$. Note that

$$\Big|\mathcal{V}(f^*) - E\Big[\frac{\widehat{R}\,I(Af^*(\mathbf{H})>0)}{\pi(A,\mathbf{Y}_0,\mathbf{X})}\Big]\Big| \le cE[|R-\widehat{R}|]$$

and the latter is bounded by a multiplier of $|\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}|+|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}|+|\widehat{\boldsymbol{\Omega}}-\boldsymbol{\Omega}|$. The similar inequality holds for

$$\Big|\mathcal{V}(\widehat{f}) - E\Big[\frac{\widehat{R}\,I(A\widehat{f}(\widehat{\mathbf{H}})>0)}{\pi(A,\mathbf{Y}_0,\mathbf{X})}\Big]\Big|.$$

Then we obtain

$$\begin{aligned}\mathcal{V}(f^*) - \mathcal{V}(\widehat{f}) \;\le\; & c(|\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0|+|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0|+|\widehat{\boldsymbol{\Omega}}-\boldsymbol{\Omega}_0|)\\ & +E\Big[\frac{\widehat{R}\,I(Af^*(\mathbf{H})>0)}{\pi(A,\mathbf{Y}_0,\mathbf{X})}\Big] - E\Big[\frac{\widehat{R}\,I(A\widehat{f}(\widehat{\mathbf{H}})>0)}{\pi(A,\mathbf{Y}_0,\mathbf{X})}\Big].\end{aligned}$$

On the other hand, according to the Lipschitz continuity in Condition (A2), the same condition holds for the conditional mean of $\mathbf{Y}_1^{(a)}$ given $\mathbf{H}$. Observing that $\widehat{R}$ is a linear combination of $\mathbf{Y}_1^{(a)}$ when $A=a$, it is easy to see

$$E\Big[\frac{\widehat{R}\,I(Af^*(\mathbf{H})>0)}{\pi(A,\mathbf{Y}_0,\mathbf{X})}\Big] - \widetilde{\mathcal{V}}(f^*) \le c(|\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0|+|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0|+|\widehat{\boldsymbol{\Omega}}-\boldsymbol{\Omega}_0|).$$

We thus obtain from Condition (A3)

$$\begin{aligned}\mathcal{V}(f^*) - \mathcal{V}(\widehat{f}) &\leq c\{|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0| + |\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0| + |\widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}_0|\} + \widetilde{\mathcal{V}}(f^*) - \widetilde{\mathcal{V}}(\widehat{f}) \\ &\leq c\{|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0| + |\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0| + |\widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}_0|\} + \widetilde{\mathcal{V}}(\check{f}) - \widetilde{\mathcal{V}}(\widehat{f}) \\ &= O_p(n^{-\alpha_1}) + \widetilde{\mathcal{V}}(\check{f}) - \widetilde{\mathcal{V}}(\widehat{f}).\end{aligned}$$

In other words, the value loss for the estimated ITR, $\widehat{f}$, can be bounded by the value loss when the feature variable is $\widehat{\mathbf{H}}$ and the outcome is $\widehat{R}$, plus the error due to the parameter estimates in the RBM model.

The rest of the proof thus proceeds in a similar way to Liu et al (2018) except that the feature variable is now $\widehat{\mathbf{H}}$ and the outcome is $\widehat{R}$ and that the hinge-loss in Liu et al. (2018) is replaced by a more general margin loss satisfying Condition (A5). First, since $\phi(x)$ is either the hinge-loss or logistic loss, according to Bartlett et al. (2006), we have

$$\begin{aligned}\widetilde{\mathcal{V}}(\check{f}) - \widetilde{\mathcal{V}}(\widehat{f}) &= E\Big[\frac{\widehat{R} - s^*(\widehat{\mathbf{H}})}{\pi(A, \mathbf{Y}_0, \mathbf{X})} I(A\check{f}(\widehat{\mathbf{H}}) > 0)\Big] - E\Big[\frac{R - s^*(\widehat{\mathbf{H}})}{\pi(A, \mathbf{Y}_0, \mathbf{X})} I(A\widehat{f}(\widehat{\mathbf{H}})) > 0)\Big] \\ &\leq \mathcal{R}_\phi\Big(\check{f}, s^*\Big) - \mathcal{R}_\phi\Big(\widehat{f}, s^*\Big),\end{aligned}$$

where the loss function R_ϕ is defined as

$$\mathcal{R}_\phi(f, s) = E\Big[\frac{|\widehat{R} - s(\widehat{\mathbf{H}})|}{\pi(A, \mathbf{Y}_0, \mathbf{X})} \phi(A\,\mathrm{sign}(\widehat{R} - s(\widehat{\mathbf{H}}))f(\widehat{\mathbf{H}}))\Big].$$

Furthermore, we have

$$\begin{aligned}&\mathcal{R}_\phi(\check{f}, s^*) - \mathcal{R}_\phi(\widehat{f}, s^*) \\ =\ &\mathcal{R}_\phi(\check{f}, s^*) - \mathcal{R}_\phi(\check{f}, \widehat{s}) - \mathcal{R}_\phi(\widehat{f}, s^*) + \mathcal{R}_\phi(\widehat{f}, \widehat{s}) \\ &+\mathcal{R}_\phi(\check{f}, \widehat{s}) - \mathcal{R}_\phi(\widehat{f}, \widehat{s}). \qquad \text{(B.1)}\end{aligned}$$

For the first two terms on the right-hand side, using the Lipschitz continuity of ϕ (Condition (A5)), we can bound them by

$$O_p(E[|\widehat{s}(\widehat{\mathbf{H}}) - s^*(\widehat{\mathbf{H}})|]) = O_p(n^{-\alpha_1}) + O_p(n^{-\alpha_2})$$

due to conditions (A3). Finally, for the last term in the right-hand side of (B.1), we note that $\mathcal{R}_\phi(\check{f},\widehat{s})-\mathcal{R}_\phi(\widehat{f},\widehat{s})$ is the excess risk if we treat $\mathcal{R}_\phi(f,\widehat{s})$ as a loss function. Recall that $\widehat{f}$ minimizes the regularized empirical version of $\mathcal{R}_\phi(f,\widehat{s})$:

$$n^{-1}\sum_{i=1}^{n}\Big[\frac{|\widehat{R}_i-\widehat{s}(\widehat{\mathbf{H}}_i)|}{\pi(A_i,\mathbf{Y}_{0i},\mathbf{X}_i)}\big(1-A_i\,\mathrm{sign}(\widehat{R}_i-\widehat{s}(\widehat{\mathbf{H}}_i))f(\widehat{\mathbf{H}}_i)\big)_+\Big]+\lambda_n||f||^2_{\mathcal{H}}.$$

Hence, we can follow the same arguments in the proof of Theorem 3.4 in Zhao et al. (2012), whereas the loss function here is $\mathcal{R}_\phi(f,\widehat{s})$. We note that

$$\begin{aligned}
&\mathcal{R}_\phi(\check{f},\widehat{s})-\mathcal{R}_\phi(\widehat{f},\widehat{s})\le\lambda_n||\widehat{f}||^2_{\mathcal{H}}+\mathcal{R}_\phi(\check{f},\widehat{s})-\mathcal{R}_\phi(\widehat{f},\widehat{s})\\
\le\ &\Big[\lambda_n||\widehat{f}||^2_{\mathcal{H}}+\mathcal{R}_\phi(\widehat{f},\widehat{s})-\inf_{f\in\mathcal{H}}(\lambda_n||f||^2_{\mathcal{H}}+\mathcal{R}_\phi(f,\widehat{s}))\Big]\\
&+\Big[\inf_{f\in\mathcal{H}}(\lambda_n||f||^2_{\mathcal{H}}+\mathcal{R}_\phi(f,\widehat{s}))-\mathcal{R}_\phi(\check{f},\widehat{s})\Big] \qquad \text{(B.2)}
\end{aligned}$$

The second term on the right hand side of (B.2) can be bound using Theorem 2.7 in (Steinwart et al., 2007) by $O(\lambda_n^{\frac{q}{q+1}})$, when we set $\sigma_n=\lambda_n^{-1/(q+1)d}$. Next, we adopt Theorem 5.6 in Steinwart et al. (2007) to obtain a bound for the first term in the right and side of (B.2). First we observe that by the definition of $\widehat{f}$, for any $f\in\mathcal{H}$,

$$P_n\Big(\mathcal{R}_\phi(\widehat{f},\widehat{s})+\lambda_n||\widehat{f}||^2_{\mathcal{H}}\Big)\le P_n\Big(\mathcal{R}_\phi(f,\widehat{s})+\lambda_n||f||^2_{\mathcal{H}}\Big),$$

where P_n is the empirical measure. Take $f=0$, we obtain

$$||\widehat{f}||^2_{\mathcal{H}}\le\frac{1}{\lambda_n}\frac{1}{n}\sum_i\frac{|\widehat{R}_i-\widehat{s}(\widehat{\mathbf{H}}_i)|}{\pi(A_i,\mathbf{Y}_{0i},\mathbf{X}_i)}\le\frac{M}{\lambda_n},$$

where $M=2E|\widehat{R}-\widehat{s}(\widehat{\mathbf{H}})|/c_0$ and c_0 is the lower bound for $\pi(A,\mathbf{Y},\mathbf{X})$. Hence, it suffice to consider the subspace of $\mathcal{H}$, $B_{\mathcal{H}}(\sqrt{M/\lambda_n})$, the ball of $\mathcal{H}$ with radius $\sqrt{M/\lambda_n}$. We define the function class

$$\mathcal{G}_{\lambda_n}=\Big\{R_\phi(f,\widehat{s})+\lambda_n||f||^2_{\mathcal{H}}-R_\phi(f^*_{\lambda_n},\widehat{s})-\lambda_n||f^*_{\lambda_n}||^2_{\mathcal{H}}:f\in B_{\mathcal{H}}(\sqrt{M/\lambda_n})\Big\},$$

where $f^*_{\lambda_n}=\arg\min_{f\in B_{\mathcal{H}}(\sqrt{M/\lambda_n})}(R_\phi(f,\widehat{s})+\lambda_n||f||^2_{\mathcal{H}})$. Similarly as in Zhao et al. (2012) , by the Lipschitz continuity of ϕ, convexity of R_ϕ, and condition (A3), there exist constant B and c such that for all $g\in\mathcal{G}_{\lambda_n}$,

$$|g|\le B,\ \ E(g^2)/E(g)\le c,$$

and $B = O(\sqrt{M}\lambda_n^{-1/2})$ and $c = O(\lambda_n^{-1})$ when n is large. Also we obtain that for all $\sigma_n > 0, 0 < v < 2, \delta > 0, \epsilon > 0$,

$$N(B^{-1}\mathcal{G}_{\lambda_n}, \epsilon, L_2(P_n)) \leq ce^{-\tau}\sigma_n^{(1-v/2)(1+\delta)d}.$$

Additionally, since $\widehat{\mathbf{H}}$ converges to $\mathbf{H}$, the GNE condition holds for $\widehat{\mathbf{H}}$ when n is large. Hence we satisfy all conditions in Theorem 5.6 in Steinwart et al. (2007). Therefore, we obtain that with at least probability $1 - c'e^{-\tau}$,

$$\mathcal{R}_\phi(\check{f}, \widehat{s}) - \mathcal{R}_\phi(\widehat{f}, \widehat{s}) \leq c\Big(\lambda_n^{-\frac{2}{2+v}-\frac{(2-v)(1+\delta)}{(2+v)(1+q)}} n^{-\frac{2}{2+v}} + \frac{\tau}{n\lambda_n} + \lambda_n^{\frac{q}{q+1}}\Big) \tag{B.3}$$

where c is a constant depending on (v, d) and the second moment of $(\widehat{R} - \widehat{s}(\widehat{\mathbf{H}}))$. Therefore, combining all above results, we obtain

$$\mathcal{V}(f^*) - \mathcal{V}(\widehat{f}) = O_p\left(n^{-\alpha_1} + n^{-\alpha_2} + \lambda_n^{-\frac{2}{2+v}-\frac{(2-v)(1+\delta)}{(2+v)(1+q)}} n^{-\frac{2}{2+v}} + (n\lambda_n)^{-1} + \lambda_n^{\frac{q}{q+1}}\right).$$

We obtain the theorem.

Supplementary Material

The online Supplementary Material includes additional simulation results, an example of computing the surrogate outcome $\widetilde{g}(\mathbf{Y}_1^{(a)})$ in the proposed method, and the R code to implement the proposed method.

Funding

This research is supported by U.S. NIH grants NS073671, GM124104.

References

APA (2013). *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition.* American Psychiatric Association, Washington, DC.

Bartlett, P. L., M. I. Jordan, and J. D. McAuliffe (2006). Convexity, classification, and risk bounds. *Journal of the American Statistical Association 101*(473), 138–156.

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology 53*(1), 605–634.

Butler, E. L., E. B. Laber, S. M. Davis, and M. R. Kosorok (2018). Incorporating patient preferences into estimation of optimal individualized treatment rules. *Biometrics 74*(1), 18–26.

Carreira-Perpinan, M. A. and G. E. Hinton (2005). On contrastive divergence learning. In *Artificial Intelligence and Statistics*, Volume 10, pp. 33–40.

Chekroud, A. M., R. Gueorguieva, H. M. Krumholz, M. H. Trivedi, J. H. Krystal, and G. McCarthy (2017). Reevaluating the efficacy and predictability of antidepressant treatments: a symptom clustering approach. *JAMA Psychiatry 74*(4), 370–378.

Coates, A., A. Ng, and H. Lee (2011). An analysis of single-layer networks in unsupervised feature learning. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pp. 215–223.

Cole, D. A., L. Cai, N. C. Martin, R. L. Findling, E. A. Youngstrom, J. Garber, J. F. Curry, J. S. Hyde, M. J. Essex, B. E. Compas, et al. (2011). Structure and measurement of depression in youths: applying item response theory to clinical data. *Psychological Assessment 23*(4), 819.

Fried, E. I. and R. M. Nesse (2015). Depression is not a consistent syndrome: an investigation of unique symptom patterns in the STAR*D study. *Journal of Affective Disorders 172*, 96–102.

Gartlehner, G., R. A. Hansen, L. C. Morgan, K. Thaler, L. Lux, M. Van Noord, U. Mager, P. Thieda, B. N. Gaynes, T. Wilkins, et al. (2011). Comparative benefits and harms of second-

generation antidepressants for treating major depressive disorder: an updated meta-analysis. *Annals of Internal Medicine 155*(11), 772–785.

Goodfellow, I., Y. Bengio, and A. Courville (2016). *Deep learning*. MIT press.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry 23*(1), 56.

Harman, H. H. (1976). *Modern Factor Analysis*. University of Chicago press.

Hinton, G. E. and R. R. Salakhutdinov (2006). Reducing the dimensionality of data with neural networks. *Science 313*(5786), 504–507.

Jarvis, C. B., S. B. MacKenzie, and P. M. Podsakoff (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research 30*(2), 199–218.

Lavori, P. W. and R. Dawson (2004). Dynamic treatment regimes: practical design considerations. *Clinical Trials 1*(1), 9–20.

Liu, Y., Y. Wang, M. Kosorok, Y. Zhao, and D. Zeng (2018). Augmented outcome-weighted learning for estimating optimal dynamic treatment regimens. *Statistics in Medicine 36*(7), 1099–1117.

Lohr, K. N. (2002). Assessing health status and quality-of-life instruments: attributes and review criteria. *Quality of Life Research 11*(3), 193–205.

McCrae, R. R. and P. T. Costa (2010). *NEO Inventories For The NEO Personality Inventory-3 (NEO-PI-3), NEO Five-Factor Inventory-3 (NEO-FFI-3), NEO Personality Inventory-Revised (NEO PI-R): Professional Manual*. PAR.

Moodie, E. E., T. S. Richardson, and D. A. Stephens (2007). Demystifying optimal dynamic treatment regimes. *Biometrics 63*(2), 447–455.

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

Posner, J., J. A. Russell, and B. S. Peterson (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology 17*(3), 715–734.

Qian, M. and S. A. Murphy (2011). Performance guarantees for individualized treatment rules. *Annals of Statistics 39*(2), 1180.

Qian, N. (1999). On the momentum term in gradient descent learning algorithms. *Neural Networks 12*(1), 145–151.

Rush, A. J., M. Fava, S. R. Wisniewski, P. W. Lavori, M. H. Trivedi, H. A. Sackeim, M. E. Thase, A. A. Nierenberg, F. M. Quitkin, T. M. Kashner, et al. (2004). Sequenced treatment alternatives to relieve depression (STAR*D): rationale and design. *Controlled Clinical Trials 25*(1), 119–142.

Steinwart, I., C. Scovel, et al. (2007). Fast rates for support vector machines using gaussian kernels. *The Annals of Statistics 35*(2), 575–607.

Sunderland, M., N. Carragher, N. Wong, and G. Andrews (2013). Factor mixture analysis of dsm-iv symptoms of major depression in a treatment seeking clinical population. *Comprehensive Psychiatry 54*(5), 474–483.

Tang, T. Z., R. J. DeRubeis, S. D. Hollon, J. Amsterdam, R. Shelton, and B. Schalet (2009). Personality change during depression treatment: a placebo-controlled trial. *Archives of General Psychiatry 66*(12), 1322–1330.

Trivedi, M. H., P. J. McGrath, M. Fava, R. V. Parsey, B. T. Kurian, M. L. Phillips, M. A. Oquendo, G. Bruder, D. Pizzagalli, M. Toups, et al. (2016). Establishing moderators and

biosignatures of antidepressant response in clinical care (EMBARC): rationale and design. *Journal of Psychiatric Research 78*, 11–23.

van Borkulo, C., L. Boschloo, D. Borsboom, B. W. Penninx, L. J. Waldorp, and R. A. Schoevers (2015). Association of symptom network structure with the course of depression. *JAMA Psychiatry 72*(12), 1219–1226.

Van Dorn, R. A., S. L. Desmarais, K. J. Grimm, S. J. Tueller, K. L. Johnson, B. G. Sellers, and M. S. Swartz (2016). The latent structure of psychiatric symptoms across mental disorders as measured with the panss and bprs-18. *Psychiatry Research 245*, 83–90.

Van Loo, H. M., P. De Jonge, J.-W. Romeijn, R. C. Kessler, and R. A. Schoevers (2012). Data-driven subtypes of major depressive disorder: a systematic review. *BMC Medicine 10*(1), 156.

Wakefield, J. C. and M. F. Schmitz (2014). Uncomplicated depression is normal sadness, not depressive disorder: further evidence from the nesarc. *World Psychiatry 13*(3), 317–319.

Watkins, C. (1989). *Learning from delayed rewards*. Ph. D. thesis, King's College, Cambridge.

Webb, C. A., M. H. Trivedi, Z. D. Cohen, D. G. Dillon, J. C. Fournier, F. Goer, M. Fava, P. J. McGrath, M. Weissman, R. Parsey, et al. (2019). Personalized prediction of antidepressant v. placebo response: evidence from the embarc study. *Psychological Medicine 49*(7), 1118–1127.

Welling, M., M. Rosen-Zvi, and G. E. Hinton (2005). Exponential family harmoniums with an application to information retrieval. In *Advances in Neural Information Processing Systems*, pp. 1481–1488.

Wray, N. R., S. Ripke, M. Mattheisen, M. Trzaskowski, E. M. Byrne, A. Abdellaoui, M. J. Adams, E. Agerbo, T. M. Air, T. M. Andlauer, et al. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics 50*(5), 668.

Zhao, Y., D. Zeng, A. J. Rush, and M. R. Kosorok (2012). Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association 107*(499), 1106–1118.

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

Zhou, X., N. Mayer-Hamblett, U. Khan, and M. R. Kosorok (2017). Residual weighted learning for estimating individualized treatment rules. *Journal of the American Statistical Association 112*(517), 169–187.

Supplementary Material for Learning Individualized Treatment Rules for Multiple-Domain Latent Outcomes

Yuan Chen* Donglin Zeng† Yuanjia Wang*

Our R codes and examples are available at our Github page (`https://github.com/ychen178/ITRs-for-Latent-Outcomes`). In this supplementary material, we provide additional simulation results and a simple example of computing the surrogate outcome $\widetilde{g}(\mathbf{Y}_1^{(a)})$ in the proposed method.

1 More Simulation Results

Using the Q-learning algorithm for all methods (Figure 11) yielded consistent results with using the O-learning algorithm as presented in the main paper. When $(\mathbf{Y}_0, \mathbf{Z}_0)$ were not from an RBM, our method does not converge to the "gold standard" where $(\mathbf{Y}_0, \mathbf{Z}_0)$ was correctly specified, but the fitted RBMs still perform well in capturing the relationship between $\mathbf{Y}_0$ and $\mathbf{Z}_0$, achieving better latent disorder status than the standard method on the test set (Figure 12).

*Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, NY 10032; †Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC 27516

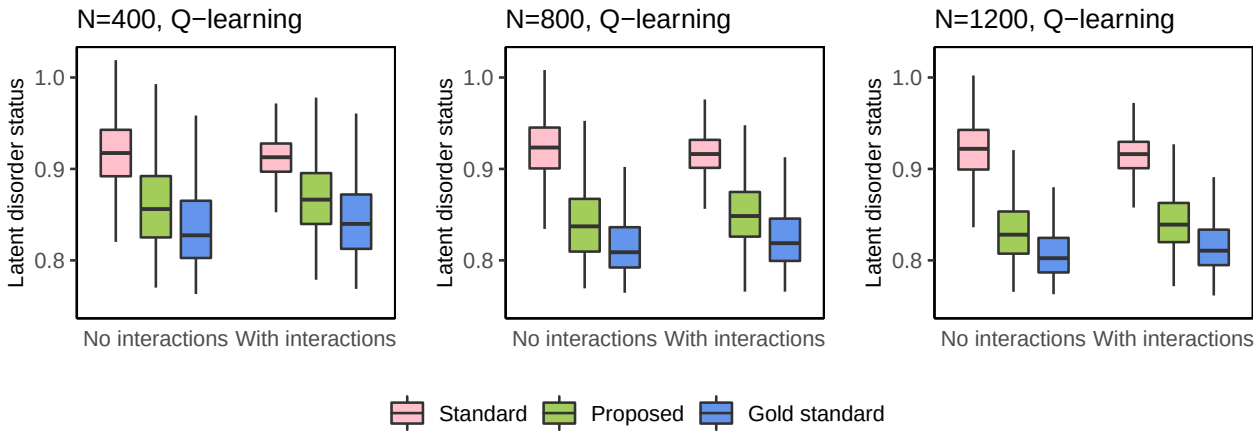

Figure 11: Simulation results: latent disorder status (the lower, the better) evaluated on the test set for the three methods using the Q-learning algorithm, with or without pairwise interactions among baseline observed items or latent domains in the predictors, under training sample size of $N = 400, 800, 1200$. Standard method: observed items $\mathbf{Y}_0$ in the predictors and sum of observed items after treatment $\mathbf{Y}_1$ as the outcome; Proposed method: inferred baseline latent states $\widehat{\mathbf{W}}_i$ in the predictors with the estimated surrogate outcome $\widehat{R}_i$; Gold standard method: baseline latent states $\mathbf{W}_i$ in the predictors and surrogate outcome R_i both constructed with true parameters. Outliers are not plotted in the figure.

2 Simple Example for constructing the surrogate outcome $\widetilde{g}(\mathbf{Y}_1^{(a)})$

Assuming 1-dimensional binary latent $\mathbf{Z}_1^{(a)}$ and 2-dimensional binary $\mathbf{Y}_1^{(a)}$, then the probability matrix $\mathbf{\Sigma}$ is constructed as

$$\mathbf{\Sigma} = \begin{pmatrix} P(\mathbf{Y}_1^{(a)} = (0,0)^T \mid \mathbf{Z}_1^{(a)} = 0) & P(\mathbf{Y}_1^{(a)} = (0,0)^T \mid \mathbf{Z}_1^{(a)} = 1) \\ P(\mathbf{Y}_1^{(a)} = (0,1)^T \mid \mathbf{Z}_1^{(a)} = 0) & P(\mathbf{Y}_1^{(a)} = (0,1)^T \mid \mathbf{Z}_1^{(a)} = 1) \\ P(\mathbf{Y}_1^{(a)} = (1,0)^T \mid \mathbf{Z}_1^{(a)} = 0) & P(\mathbf{Y}_1^{(a)} = (1,0)^T \mid \mathbf{Z}_1^{(a)} = 1) \\ P(\mathbf{Y}_1^{(a)} = (1,1)^T \mid \mathbf{Z}_1^{(a)} = 0) & P(\mathbf{Y}_1^{(a)} = (1,1)^T \mid \mathbf{Z}_1^{(a)} = 1) \end{pmatrix}$$

A left inverse of $\mathbf{\Sigma}$, $\mathbf{\Sigma}_L^- = (\mathbf{\Sigma}^T\mathbf{\Sigma})^{-1}\mathbf{\Sigma}^T$ is a 2 by 4 matrix. For a subject with $\mathbf{Y}_{1i}^{(a)} = (0,1)^T$, the surrogate outcome for this subject is constructed as

$$\begin{aligned} \widetilde{g}(\mathbf{Y}_{1i}^{(a)}) &= \big[g(\mathbf{z}_1)\big]_{\mathbf{z}_1}^T \mathbf{\Sigma}_L^- \Big[I(\mathbf{Y}_{1i}^{(a)} = \mathbf{y}_1)\Big]_{\mathbf{y}_1} \\ &= \begin{pmatrix} g(0) \\ g(1) \end{pmatrix}^T \mathbf{\Sigma}_L^- \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \end{aligned}$$

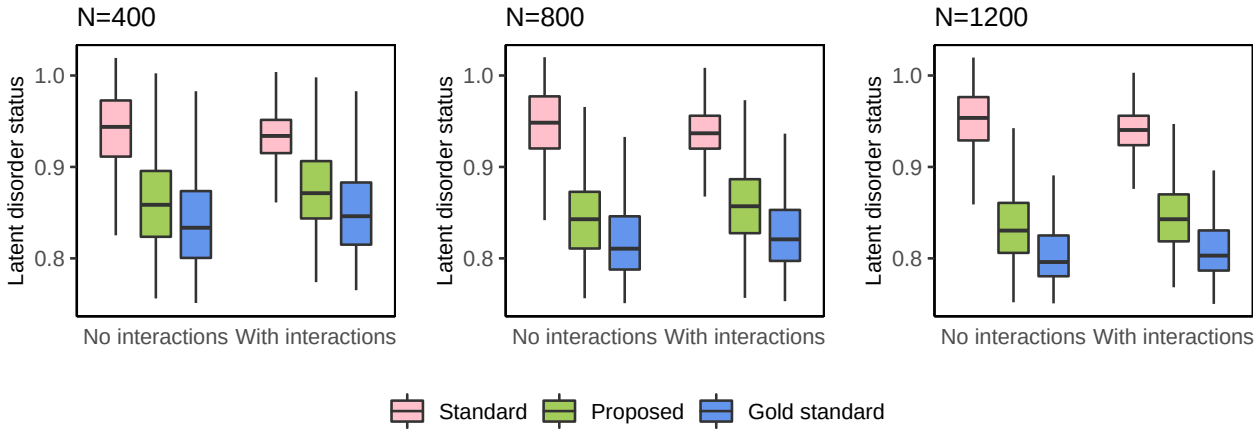

Figure 12: Simulation results: latent disorder status (the lower, the better) evaluated on the test set for the three methods using the O-learning algorithm with logistic loss when $(\mathbf{Y}_0, \mathbf{Z}_0)$ is not from an RBM, with or without pairwise interactions among baseline observed items or latent domains in the predictors, under training sample size of $N = 400, 800, 1200$. Standard method: observed items $\mathbf{Y}_0$ in the predictors and sum of observed items after treatment $\mathbf{Y}_1$ as the outcome; Proposed method: inferred baseline latent states $\widehat{\mathbf{W}}_i$ in the predictors with the estimated surrogate outcome $\widehat{R}_i$; Gold standard method: baseline latent states $\mathbf{W}_i$ in the predictors and surrogate outcome R_i both constructed with true parameters. Outliers are not plotted in the figure.

and for a subject with $\mathbf{Y}_{1i}^{(a)} = (1\ 1)$, the surrogate outcome for this subject is constructed as

$$\widetilde{g}(\mathbf{Y}_{1i}^{(a)}) = \begin{pmatrix} g(0) \\ g(1) \end{pmatrix}^T \mathbf{\Sigma}_L^- \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$