Outcome-Weighted Learning for Personalized Medicine with Multiple Treatment Options

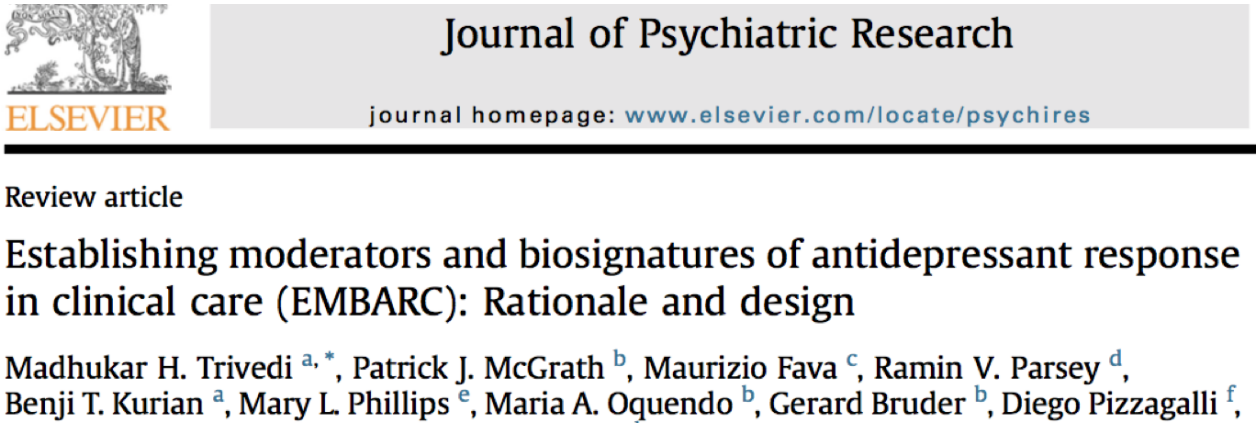
Donglin Zeng

Department of Biostatistics
University of North Carolina

Introduction

Personalized Medicine

- Substantial treatment response heterogeneity observed (anti-depressant remission rate 30-40%; Trivedi et al. 2016)
- Personalized Medicine: **"the tailoring of medical treatment to the individual characteristics of each patient".**
- Need clinical, biological, or behavioral markers to distinguish in advance which treatment will benefit a patient.



ELSEVIER

Journal of Psychiatric Research

journal homepage: www.elsevier.com/locate/psychires

Review article

Establishing moderators and biosignatures of antidepressant response in clinical care (EMBARC): Rationale and design

Madhukar H. Trivedi [a,*], Patrick J. McGrath [b], Maurizio Fava [c], Ramin V. Parsey [d], Benji T. Kurian [a], Mary L. Phillips [e], Maria A. Oquendo [b], Gerard Bruder [b], Diego Pizzagalli [f],

Individualized Treatment Rules

Individualized treatment rules (ITRs): decision rules mapping patient's pre-treatment variables into the space of possible treatment decisions (e.g., switch to an alternative treatment).

Example: Healing Emotion After Loss (HEAL) trial

Original Investigation

Optimizing Treatment of Complicated Grief
A Randomized Clinical Trial

M. Katherine Shear, MD; Charles F. Reynolds III, MD; Naomi M. Simon, MD, MSc; Sidney Zisook, MD; Yuanjia Wang, PhD; Christine Mauro, PhD; Naihua Duan, PhD; Barry Lebowitz, PhD; Natalia Skritskaya, PhD

JAMA Psychiatry, 2016;73(7):685-694

ITR: Administer clinical management as the initial treatment; if a patient does not respond within 8 weeks then offer an anti-depressant (Citalopram).

Existing Methods for Optimizing ITR

Recent machine learning methods:

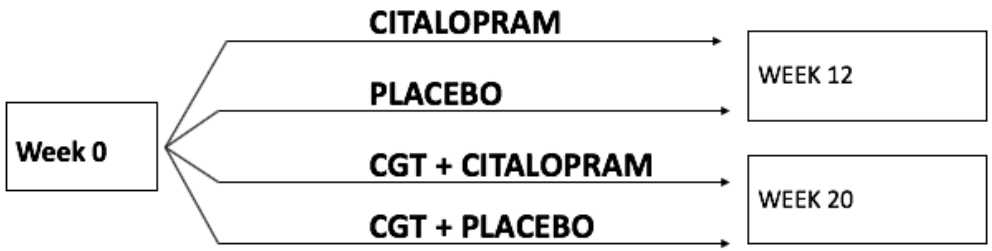
- Q-Learning (*Qian and Murphy 2011*)
- Outcome-Weighted Learning; O-learning (*Zhao et al. 2012; Liu et al. 2018; Qiu et al. 2018*)
- Interaction Trees (*Su et al. 2009*)
- Qualitative interaction trees (QUINT, *Dusseldorp and Van Mechelen 2014*)

Most of the existing work aims at randomized controlled trials (RCTs) with binary treatment decisions.

Multiple Treatments

- In many applications, more than two treatment options are considered (multiple pharmacotherapies/psychotherapies and their combinations).

Figure: HEAL Study Compared Four Treatments



- Our goal: develop O-learning to estimate optimal ITR with multiple treatment options.

A Brief Review of O-learning

- R: Clinical outcome; A: treatment $\{-1, 1\}$; X: feature variables (pre-treatment covariates)
- ITR $\mathcal{D}(X)$: a function mapping X to the domain of A.
- Associated with each $\mathcal{D}$, O-learning maximizes a value function defined as Note

$$\mathcal{V}(\mathcal{D}) = E_{\mathcal{D}}(R) = E\left(\frac{RI(A = \mathcal{D}(X))}{P(A|X)}\right).$$

$\max_{\mathcal{D}} \mathcal{V}(D)$ equivalent to $\min_{\mathcal{D}} E\left(\frac{RI(A \neq \mathcal{D}(X))}{P(A|X)}\right)$.

A Brief Review of O-learning

- R: Clinical outcome; A: treatment $\{-1, 1\}$; X: feature variables (pre-treatment covariates)
- ITR $\mathcal{D}(X)$: a function mapping X to the domain of A.
- Associated with each $\mathcal{D}$, O-learning maximizes a value function defined as Note

$$\mathcal{V}(\mathcal{D}) = E_{\mathcal{D}}(R) = E\left(\frac{RI(A = \mathcal{D}(X))}{P(A|X)}\right).$$

$\max_{\mathcal{D}} \mathcal{V}(D)$ equivalent to $\min_{\mathcal{D}} E\left(\frac{RI(A \neq \mathcal{D}(X))}{P(A|X)}\right)$.

- O-learning is a weighted classification (e.g., SVM) with labels A and weights $R/P(A|X)$:
  - Subjects with high outcomes $\rightarrow$ encourages $\mathcal{D}(X) = A$.
  - Subjects with low outcomes $\rightarrow$ encourages $\mathcal{D}(X) = -A$.

Challenges When Generalizing to More Than Two Treatments

- Improvements to O-learning for binary treatment decisions (*Liu et al.; 2018*):
  - take residuals, weights $|R - m(X)|$, relabel A by $A\text{sign}(R - m(X))$
  - weight augmentation to improve efficiency (especially in small samples)
- Existing multicategory learning (e.g., one-versus-one OVO; or one-versus-all, OVA) is no longer feasible in the presence of negative weights and augmentation.

The Goal of This Work

What we will achieve:

- a new sequential algorithm is developed to extend O-learning to learn optimal treatment rules with more than 2 treatments;
- each stage is an O-learning for two treatment groups;
- the algorithm allows negative rewards or takes residuals as weights;
- the method is shown to be Fisher consistent and possesses the same convergence rate as standard binary O-learning.

Method

Data

- We assume n i.i.d observations from a randomized trial:

$$(A_i, X_i, R_i), \quad i = 1, ..., n$$

  where A_i denotes treatment assignment from $\{1, 2, ..., K\}$.
- $\pi_a(X)$ denotes $P(A = a|X)$ and is assumed to be positive.
- We aim to learn a prediction rule

$$\mathcal{D} : X \to \{1, 2, ..., K\}$$

  to minimize the value function $\mathcal{V}(\mathcal{D})$:

$$\mathcal{V}(\mathcal{D}) = E\left[\frac{RI(A = \mathcal{D}(X))}{\pi_A(X)}\right].$$

Revisit 2-Treatment O-learning

- Suppose Z to denote a binary treatment. Then O-learning minimizes the following hinge-loss in an asymptotic sense:

$$E\left[R\max(1-Zf(X),0)/\pi_Z(X)\right].$$

- If R is negative or R is replaced by $R-m(X)$, O-learning minimizes

$$E\left[|R|\max(1-Z\text{sign}(R)f(X),0)/\pi_Z(X)\right].$$

- The key result is that the minimizer has the same sign as the rule:

$$\text{sign}(f^*(X)) = \text{sign}(E[R|Z=1,X] - E[R|Z=-1,X]).$$

That is, $f^*(x)$ gives the optimal treatment assignment that gives the higher reward.

Modified O-learning: comparing one treatment versus multiple treatments

- Now, let's consider comparing treatment $A = j_0$ vs $A = j_1, ..., j_m$.
- This is a binary decision problem so $Z = 1$ indicates $A = j_0$ and $Z = -1$ indicates the competing treatment group.
- Since it is unfair to compare one treatment vs more than one treatments, intuitively, we need to give weights m to $Z = 1$.
- This motivates the following modification of the O-learning:

$$\min E\left[\frac{w_Z R \max(1 - Zf(X), 0)}{\pi_A(X)}\right],$$

where $w_1 = m$ and $w_{-1} = 1$.

Key result from modified O-learning

- What is the optimal decision rule:

$$\text{sign}(f^*(X)) = \text{sign}\left\{ E[R|A = j_0, X] - \frac{1}{m}\sum_{l=1}^{m} E[R|A = j_l, X] \right\}.$$

- We compare the reward from treatment j_0 with the average reward from the group of treatments!

Sequential O-Learning

Consider an ordered treatment sequence, $\{1, 2, 3, ..., K\}$.

- Step 1. Perform a binary treatment O-learning to compare treatment 1 vs $\{2, 3, ..., K\}$ and assign weight $(K - 1)$ to treatment 1.
- Step 2. Perform a binary treatment O-learning to compare treatment 2 vs remaining $\{3, ..., K\}$:
  - exclude subjects with observed $A = 1$;
  - exclude subjects whose predicted better treatment is 1 ($f_1^*(X) > 0$) based on Step 1;
  - assign weight $(K - 2)$ to treatment 2.
- From Step 3 to Step K, continue the same procedure to compare a treatment with the remaining choices.

Rationale of the Sequential O-learning

Sufficient condition: *In the sequential O-learning, subjects who pass all steps must have the largest conditional outcome when $A = K$.*

- If a subject passes step 1, the optimal treatment is not 1

$$E[R|A = 1, X] < \frac{1}{K-1} \sum_{j=2}^{K} E[R|A = j, X].$$

- If he/she passes steps $l = 2, ..., K-2$, then

$$E[R|A = l, X] < \frac{1}{K-l} \sum_{j=l+1}^{K} E[R|A = j, X].$$

- The last inequality is $E[R|A = K-1, X] < E[R|A = K, X]$.

Permuted Sequences to Identify All Subjects with K as Optimal Treatment

- Not all subjects with $E[R|A = K, X]$ as the maximum outcome will pass all steps for a fixed order.
  - If $E[R|A = 1, X]$ is larger than the average of $E[R|A = j, X]$ for $j \geq 2 \longrightarrow$ does not pass the first step.
- However, consider order $j_1, j_2, ..., j_{K-1}, K$ such that

$$E[R|A = j_1, X] < E[R|A = j_2, X] < \cdots < E[R|A = K, X],$$

then this subject will pass all steps.

Necessary condition: *For a subject whose $E[R|A = K, X]$ is the largest, there always exists a sequential O-learning using one permutation of $\{1, 2, .., K - 1\}$ so that this subject passes all steps.*

Sequential Outcome-Weighted Multicategory (SOM) Learning

SOM forward learning for K as the optimal treatment:

- Consider all permutations of $\{1, 2, ..., K-1\}$ and K being the last treatment group.
- For each permuted sequence, apply the sequential binary O-learning.
- Any subject who passes at least one of the permuted sequences should have optimal treatment as K.

SOM for the Remaining Treatments

SOM backward elimination for other treatments $1, \cdots, K-1$:

- exclude subjects with observed $A = K$;
- exclude subjects whose predicted better treatment is K ($f_K^*(X) > 0$);
- apply forward learning on the remaining data with only treatments $\{1, 2, ..., K-1\}$ for optimal treatment $(K-1)$;
- repeat this process to learn optimal treatments $(K-2), (K-3)$, in turn.

Remarks

- Each step of SOM is a weighted SVM, so existing binary O-learning (*Liu et al. 2018*) can be applied.
- Because of the sequential data elimination, the size of the input dataset decreases in each step, reducing computational burden.

Asymptotic Results

Theorem 1. *SOM is Fisher consistent: SOM-derived ITR converges to the true optimal ITR such that* $\mathcal{D}^*(X) = a^*$ *if and only if*

$$E(R|X = x, A = a^*) = argmax_{l=1,\ldots,K} E(R|X = x, A = l).$$

Theorem 2. *Under regularity conditions, for any* $\epsilon_0 > 0$, $d/(d+\tau) < p \leq 2$, *there exists a constant* C *such that for any* $\epsilon > 1$, *with probability at least* $1 - e^{-\epsilon}$,

$$\mathcal{V}(\mathcal{D}^*) - \mathcal{V}(\widehat{D})$$

$$\leq C \left\{ \lambda_n^{\frac{\tau}{2+\tau}} \sigma_n^{-\frac{d\tau}{d+\tau}} + \sigma_n^{-\beta} + \epsilon \left(n \lambda_n^p \sigma_n^{\frac{1-p}{1+\epsilon_0 d}} \right)^{-\frac{q+1}{q+2-p}} \right\}^{q/(1+q)}.$$

Simulations and Real World Data Application

Simulation Settings

- We generate 20 feature variables from a multivariate normal distribution, where the first 10 variables $X_1, X_2, ..., X_{10}$ have a pairwise correlation of 0.8, the remaining 10 variables are uncorrelated.
- We consider 3 treatments and they are assigned randomly with equal probabilities.

Reward Models

We consider

- *Setting 1.* $R = X_4 + (X_1 + X_2)I\{A = 2\} + (-X_1 + X_3)I\{A = 3\} + 0.5 \times N(0, 1)$
- *Setting 2.* $R = X_4 + (X_2^2 - X_1^2)I\{A = 2\} + X_3^3 I\{A = 3\} + 0.5 \times N(0, 1)$
- *Setting 3.* $R = (X_1 - 0.2) \times (I\{A = 1\} - I\{X_1 > 0.3\})^2 + (X_2 + 0.3) \times (I\{A = 2\} - I\{X_2 > -0.5\})^2 + (X_3 + 0.5) \times (I\{A = 3\} - I\{X_3 > 0\})^2 + 0.5 \times N(0, 1).$

Numerical Implementation

- Sample size was 300, 600 or 900.
- We compared with the results from one vs all, one vs one and Q-learning method.
- We considered both linear kernel and Gaussian kernel in the estimation.
- For testing data, we generated 3×10^6 independent observations to compute the values.

Results for Setting 1

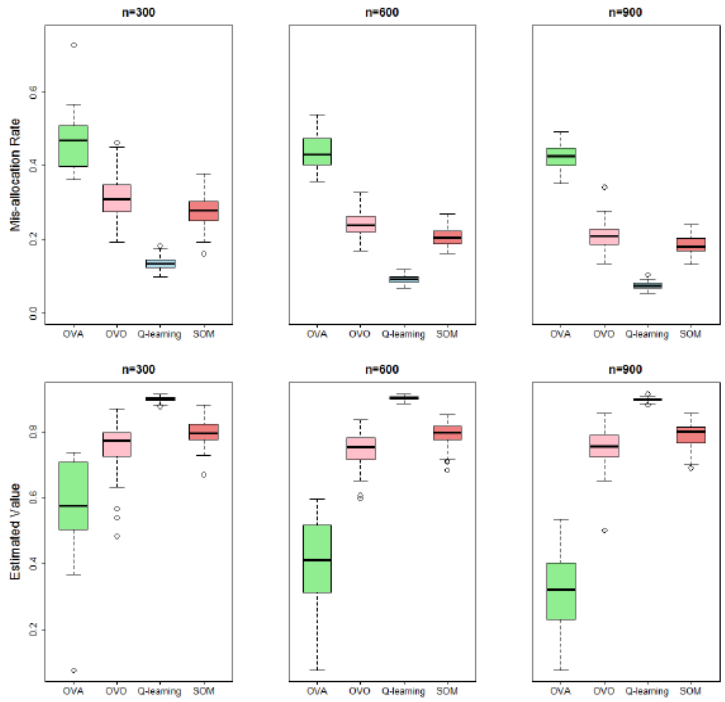


Fig. 1: Box plots of the optimal treatment mis-allocation rates and estimated value functions of SOM, Q-learning, OVA and OVO for setting 1 with sample size of 300, 600 and 900. The optimal value is 0.9245.

Results for Setting 2

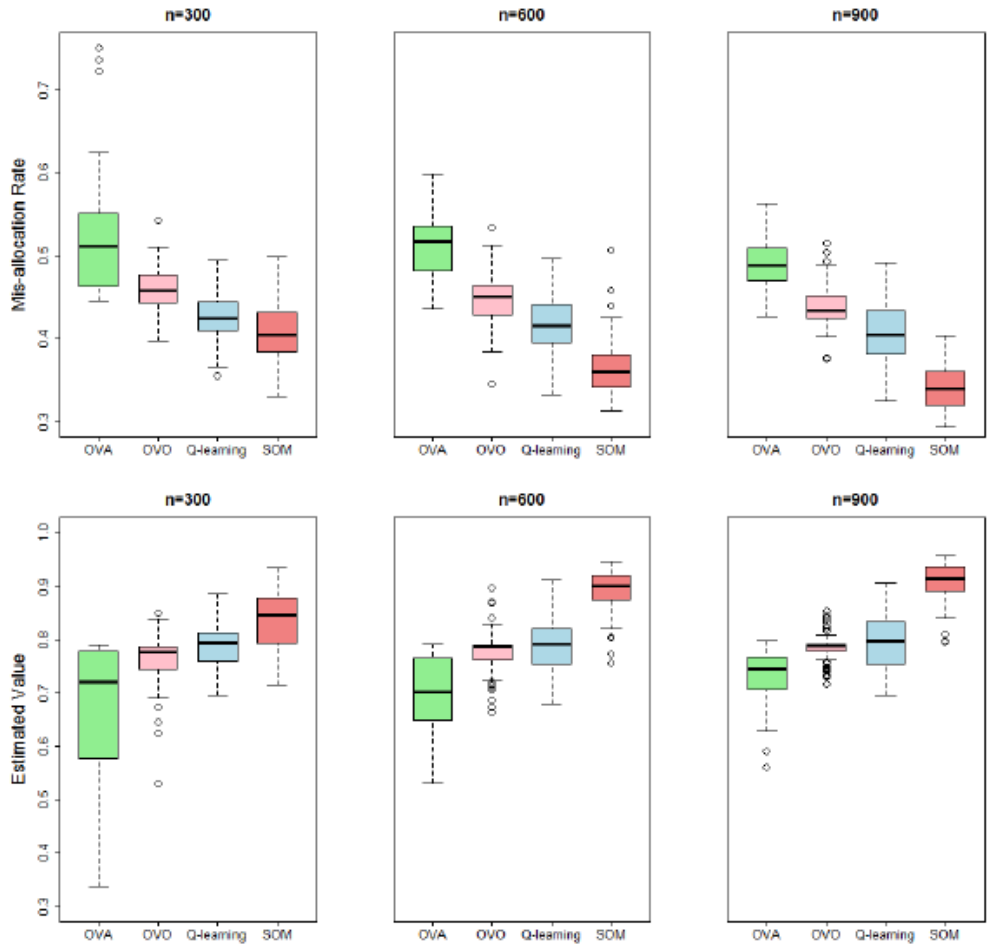


Fig. 2: See Fig. 1. The optimal value of setting 2 is 1.0585.

Results for Setting 3

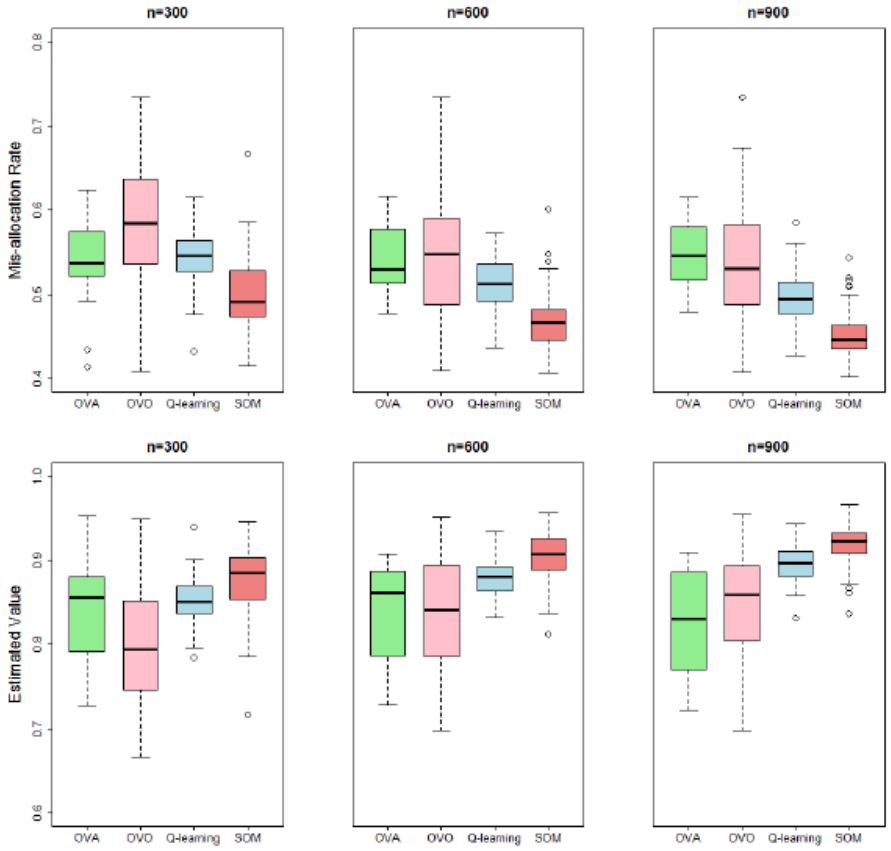


Fig. 3: See Fig. 1. The optimal value of setting 3 is 1.1438.

REVAMP Study

- It is a randomized trial aiming to evaluate the efficacy of adjunctive psychotherapy in treating patients with chronic depression who failed in initial treatment (Phase I) with an antidepressant medication.
- There were 491 patients randomized to
  (1) continued pharmacotherapy and augmentation with brief supportive psychotherapy (MEDS+BSP),
  (2) continued pharmacotherapy and augmentation with cognitive behavioral analysis system of psychotherapy (MEDS+CBASP), or
  (3) continued pharmacotherapy (MEDS) alone, and were followed for 12 weeks.
- The primary outcome was the Hamilton Scale for Depression (HAM-D) scores at the end of 12-week follow-up.

Patient's characteristics

There were 17 baseline feature variables including

- participants' demographics,
- patient's expectation of treatment efficacy,
- social adjustment scale, mood and anxiety symptoms, and depression experience,
- phase I depressive symptom measures such as rate of change in HAM-D score over phase I, HAM-D score at the end of phase I, rate of change of Quick Inventory of Depression Symptoms (QIDS) scores during phase I, and QIDS at the end of phase I.

Results: Value Function

Table: Value function (mean, sd) of the HAM-D under universal "one-size-fits-all" rule and ITR (2-fold cross-validation procedure with 500 repetitions)

Treatment†	MEDS+BSP	MEDS+CBASP	MEDS	
Value*	12.90 (0.04)	10.62 (0.04)	12.53 (0.11)	
ITR Method	**SOM learning**	**Q-learning**	**OVA**	**OVO**
Value*	9.95 (2.09)	12.64 (2.01)	11.97 (1.15)	11.15 (1.46)

†: Non-personalized, "one-size-fits-all" assignment rule.

*: Value function is the average HAM-D score at the end of phase II for patients following an estimated optimal treatment (smaller HAM-D indicates a better outcome) in the testing samples.

Results: Feature Variables and Optimal Treatment

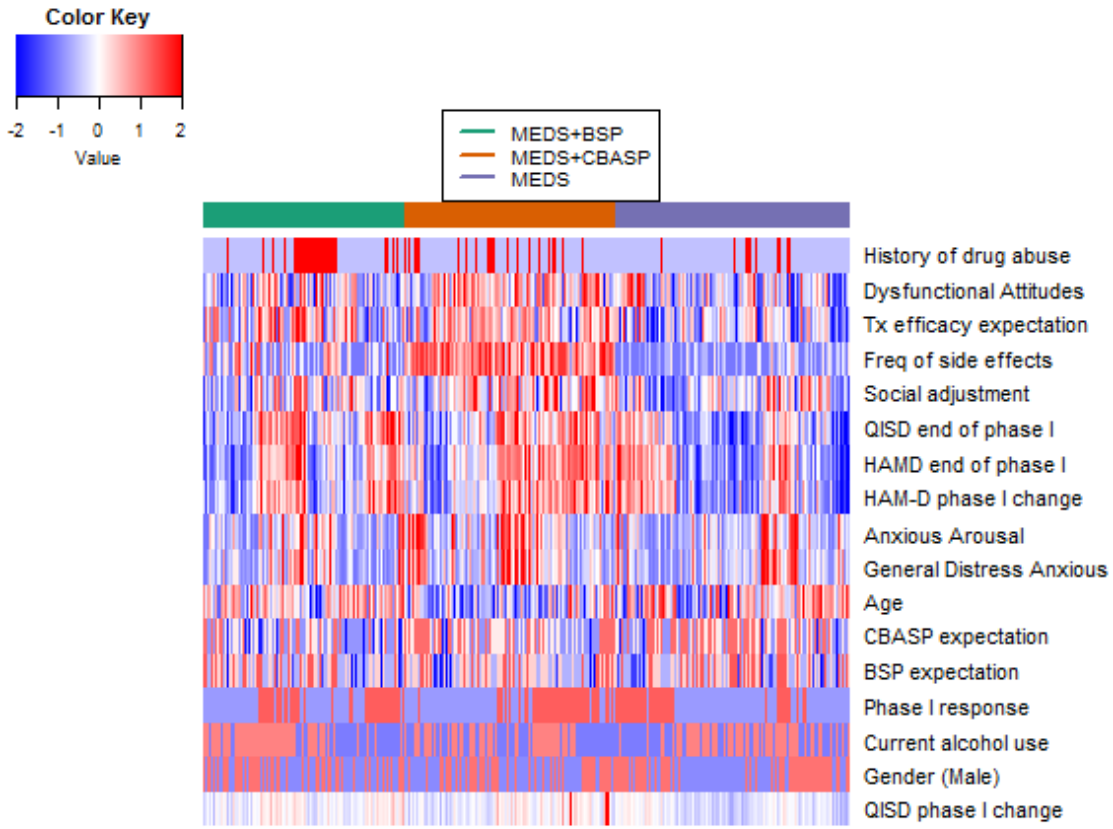


Figure: 17 standardized feature variables on all patients grouped by predicted optimal treatment. Row corresponds to feature variables and column corresponds to patients stratified by predicted optimal treatment.

Conclusion

Concluding Remarks

- SOM extends O-learning for binary treatments to learn optimal rules for K treatments ($K \geq 2$) by solving sequential SVMs.
- SOM is Fisher consistent: When n is infinity, SOM identifies the optimal treatment among K options.
- Limitation: When the number of treatments is large, grouping is necessary.
- Extensions:
  - multi-stage SOM
  - other binary classifiers and weighting scheme in each step
  - parallel computing to speed up computation