Supporting Information for Net Benefit Index: Assessing the Influence of a Biomarker for Individualized Treatment Rules by Zhou Y, Song, PXK, Fu, H.

Outcome Weighted Learning (OWL)

Denote the space of the observed data as $(\boldsymbol{X}, A, B)$. The distribution of $(\boldsymbol{X}, A, B)$ is denoted as P and the expectation of B with respect to P is denoted as E. Given a specific decision rule D, the distribution of $(\boldsymbol{X}, A, B)$ where $A = D(\boldsymbol{X})$ is denoted as P^D, and the expectation of B with respect to P^D is denoted as E^D. The expected clinical benefit under the given decision rule D can be calculated as the value function $V(D)$:

$$\begin{aligned} V(D) = E^D(B) &= \int B dP^D = \int B \frac{dP^D}{dP} dP \\ &= E\left\{ \frac{I(A = D(\boldsymbol{X}))}{P(A_i|\boldsymbol{X}_i)} B \right\} \end{aligned} \tag{S1}$$

The optimal decision rule will be D^* such that:

$$D^* \in \arg\max_D E\left\{ \frac{I(A = D(\boldsymbol{X}))}{P(A_i|\boldsymbol{X}_i)} B \right\} \tag{S2}$$

which is equivalent to:

$$D^* \in \arg\min_D E\left\{ \frac{I(A \neq D(\boldsymbol{X}))}{P(A_i|\boldsymbol{X}_i)} B \right\} \tag{S3}$$

The term $\frac{I(A \neq D(\boldsymbol{X}))}{P(A_i|\boldsymbol{X}_i)} B$ is actually a weighted classification error. Therefore, OWL is actually a weighted classification problem. With a set of *iid* observations $(\boldsymbol{X}_i, A_i, B_i), i = 1, ..., n$, we can approximate the optimization problem in (S3) by the empirical value:

$$D^* \in \arg\min_D \frac{1}{n} \sum_{i=1}^{n} \frac{B_i}{P(A_i|\boldsymbol{X}_i)} I\{A_i \neq D(\boldsymbol{X}_i)\} \tag{S4}$$

Since $D(\boldsymbol{X})$ can always be represented as $\text{sign}(f(\boldsymbol{X}))$ for some decision function f, where:

$$D(\boldsymbol{X}) = \begin{cases} 1, & f(\boldsymbol{X}) > 0 \\ -1, & f(\boldsymbol{X}) < 0 \end{cases}$$

the optimization problem in (S4) is equivalent to:

$$f^* \in \arg\min_f \frac{1}{n} \sum_{i=1}^{n} \frac{B_i}{P(A_i|\boldsymbol{X}_i)} I\{A_i \neq \text{sign}(f(\boldsymbol{X}_i))\} \tag{S5}$$

We can first obtain the optimal decision function f^* based on the optimization of (S5), and then set $D^*(\boldsymbol{X}) = \text{sign}(f^*(\boldsymbol{X}))$ to get the optimal decision rule D^*.

SVM Solution to OWL

The optimization problem (S5) is a weighted summation of 0-1 loss, which is neither convex nor continuous. It makes the problem difficult to be solved directly. Therefore, OWL uses a convex surrogate hinge loss $x^+ = \max(0, x)$, which is commonly used in SVM, to replace the 0-1 loss. In order to further penalize the complexity of the decision function f to avoid overfitting, OWL adds a l_2 penalty into the optimization problem. The final function OWL aims to minimize is:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{B_i}{P(A_i|\boldsymbol{X}_i)}(1 - A_i f(\boldsymbol{X}_i))^+ + \lambda_n||f||_2 \tag{S6}$$

where λ_n is the regulization parameter. This optimization problem can be solved using the technique of SVM.

If we assume that the decision rule f is a linear function $f(\boldsymbol{X}) = \beta_0 + \boldsymbol{X}\boldsymbol{\beta}$, the optimization problem of OWL can be solved as follows by introducing the slack variable $\xi_i = (1 - A_i f(\boldsymbol{X}_i))^+$:

$$\begin{aligned} &\text{minimize} \quad \frac{1}{n}\sum_{i=1}^{n}\frac{B_i}{P(A_i|\boldsymbol{X}_i)}\xi_i + \lambda_n||\beta||_2 \\ &\text{subject to} \quad A_i(X_i^T\beta + \beta_0) \geqslant 1 - \xi_i \\ &\qquad\qquad\quad \xi_i \geqslant 0, \quad i = 1, ..., n \end{aligned} \tag{S7}$$

Let $\kappa = \frac{1}{2n\lambda_n}$, the optimization problem is transformed to:

$$\begin{aligned} &\text{minimize} \quad \frac{1}{2}||\beta||_2 + \kappa\sum_{i=1}^{n}\frac{B_i}{P(A_i|\boldsymbol{X}_i)}\xi_i \\ &\text{subject to} \quad A_i(X_i^T\beta + \beta_0) \geqslant 1 - \xi_i \\ &\qquad\qquad\quad \xi_i \geqslant 0, \quad i = 1, ..., n \end{aligned} \tag{S8}$$

Solve this problem by introducing the Lagrange Multiplier, we come to the dual problem:

$$\begin{aligned} \underset{\alpha}{\text{maximize}} \quad & \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i\alpha_j A_i A_j X_i \cdot X_j \\ \text{subject to} \quad & 0 \leqslant \alpha_i \leqslant \kappa \frac{B_i}{P(A_i|\boldsymbol{X}_i)} \\ & \sum_{i=1}^{n} \alpha_i A_i = 0 \end{aligned} \tag{S9}$$

where α_i is the Lagrange Multiplier. If we assume a nonlinear decision rule f, the optimization problem of OWL can be solved using the kernal function k:

$$\begin{aligned} \underset{\alpha}{\text{maximize}} \quad & \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i\alpha_j A_i A_j k(X_i, X_j) \\ \text{subject to} \quad & 0 \leqslant \alpha_i \leqslant \kappa \frac{B_i}{P(A_i|\boldsymbol{X}_i)} \\ & \sum_{i=1}^{n} \alpha_i A_i = 0 \end{aligned} \tag{S10}$$

Multiple-Variable-Based Decision Rule Evaluation When $\boldsymbol{X}_e = Null$

This simulation concerns a setting where $\boldsymbol{X}_e = Null$. We have multiple signal and noise candidate biomarkers, $X_j \sim U(0,1), j = 1, \ldots, 10$, in which only X_1 and X_2 are signal biomarkers involved in the optimal ITR. The correlation structure of the variables is that $\text{Corr}(X_1, X_3) = \text{Corr}(X_2, X_4) = 0.5$, and $\text{Corr}(X_s, X_t) = 0.2, s, t \in \{5, \ldots, 110\}, s \neq t$. B is generated from a normal distribution with mean $\mu = 0.5 + X_1 + 2.0Af(X)$, where $f(\boldsymbol{X})$ is given as follows:

7) (Linear) $f(\boldsymbol{X}) = 0.5(1 + 2X_1 - 4X_2)$;

8) (Binary) $f(\boldsymbol{X}) = 6\{I(X_1 > 0.29 \cap X_2 < 0.71) - 0.5\}$;

9) (Nonlinear) $f(\boldsymbol{X}) = (X_1 - 0.1)^+ - (X_2 - 0.22)^+$.

Summary statistics of variable selection based on the proposed NBI test method, SAS and riskRFE are included in Table S.1.

Additional Simulation Studies

Table S.2 lists the summary statistics of variable selection based on the proposed NBI method, SAS and riskRFE in the multiple-variable-based decision rule evaluation when $\boldsymbol{X}_e \neq Null$ and $n = 200$. Table S.3 lists the summary statistics of variable selection based on the proposed NBI test method, SAS and riskRFE in the multiple-variable-based decision rule evaluation when $\boldsymbol{X}_e = Null$ and NBI is calculated under the Gaussian kernel. In this simulation setting, the true decision rule f is set as $f(\boldsymbol{X}) = \exp(X_1^2) - \exp(X_2^2)$. Table S.4 is the comparison of prediction accuracy for the ITRs derived under NBI test and the standard OWL in the multiple-variable-based decision rule evaluation when $\boldsymbol{X}_e \neq Null$. Table S.5 lists the summary statistics of variable selection based on l_1-OWL in the multiple-variable-based decision rule evaluation when $\boldsymbol{X}_e = Null$.

We also report the simulation results (see Table S.6- S.8) for an additional single-variable-based nonlinear decision rule ($f(\boldsymbol{X}) = 1 + X_1 + X_2^3 - \exp(X_3)$) and an additional multiple-variable-based nonlinear decision rule ($f(\boldsymbol{X}) = 1.5\{1 + X_1 - \log(X_2+1) + 2X_3^3 - \exp(X_4)\}$). It is seen that NBI still gives good variable selection results in the single-variable-based decision rule evaluation, while in the multiple-variable-based decision rule evaluation, SAS outperforms NBI. It is interesting to note that in scenario $n = 1200$, SAS gives the perfect results with size=2.000, TDR=1.000 and MCC=1.000, with ZERO standard deviation. This does not seem to be an appropriate setting to reflect a real-world scenario. Such perfection may be resulted from too strong signal-to-noise ratio in the previously chosen nonlinear decision rule.

[Table 1 about here.]

[Table 2 about here.]

[Table 3 about here.]

[Table 4 about here.]

[Table 5 about here.]

[Table 6 about here.]

[Table 7 about here.]

[Table 8 about here.]

Computing Code

Our numerical calculations of NBI have been programmed with the Python coding language. One example of the Python code used in the simulation study in Section 4.2 to produce Table 3 is available at https://github.com/yiwangz/NBI for a free download.

Table S.1

Size, TDR, MCC, and CCR for variable selection based on NBI test, SAS, and riskRFE in the multiple-variable-based decision rule evaluation when $\boldsymbol{X}_e = Null$.

NBI					
scenario	n	size (sd)	TDR (sd)	MCC (sd)	CCR (sd)
linear	800	1.799 (0.761)	0.915 (0.192)	0.797 (0.221)	0.871 (0.068)
	1000	1.895 (0.777)	0.918 (0.181)	0.826 (0.207)	0.880 (0.067)
	1200	1.934 (0.707)	0.917 (0.184)	0.844 (0.211)	0.881 (0.069)
binary	800	2.115 (0.617)	0.935 (0.151)	0.921 (0.149)	0.905 (0.072)
	1000	2.145 (0.590)	0.937 (0.150)	0.933 (0.142)	0.909 (0.070)
	1200	2.136 (0.467)	0.948 (0.134)	0.953 (0.120)	0.914 (0.061)
nonlinear	800	1.957 (0.840)	0.898 (0.216)	0.814 (0.243)	0.840 (0.087)
	1000	2.045 (0.914)	0.893 (0.209)	0.828 (0.232)	0.847 (0.080)
	1200	1.977 (0.771)	0.912 (0.195)	0.842 (0.224)	0.855 (0.081)
SAS					
scenario	n	size(sd)	TDR(sd)	MCC(sd)	CCR(sd)
linear	800	2.096 (0.295)	0.968 (0.098)	0.977 (0.070)	0.988 (0.007)
	1000	2.034 (0.192)	0.989 (0.061)	0.992 (0.044)	0.989 (0.006)
	1200	2.022 (0.147)	0.993 (0.049)	0.995 (0.035)	0.991 (0.005)
binary	800	2.502 (0.712)	0.852 (0.190)	0.891 (0.143)	0.825 (0.011)
	1000	2.316 (0.584)	0.904 (0.166)	0.930 (0.123)	0.828 (0.011)
	1200	2.208 (0.457)	0.934 (0.139)	0.953 (0.101)	0.826 (0.012)
nonlinear	800	3.228 (1.065)	0.688 (0.219)	0.760 (0.179)	0.963 (0.015)
	1000	2.840 (0.903)	0.769 (0.213)	0.827 (0.166)	0.964 (0.015)
	1200	2.672 (0.826)	0.810 (0.207)	0.859 (0.158)	0.967 (0.012)
riskRFE					
scenario	n	size(sd)	TDR(sd)	MCC(sd)	CCR(sd)
linear	800	2.686 (0.955)	0.756 (0.229)	0.774 (0.217)	0.925 (0.042)
	1000	2.443 (0.829)	0.813 (0.228)	0.815 (0.227)	0.928 (0.045)
	1200	2.219 (0.676)	0.870 (0.196)	0.854 (0.198)	0.932 (0.035)
binary	800	3.312 (1.037)	0.646 (0.209)	0.714 (0.192)	0.859 (0.108)
	1000	2.985 (0.968)	0.710 (0.215)	0.763 (0.186)	0.883 (0.093)
	1200	2.638 (0.904)	0.783 (0.216)	0.810 (0.192)	0.885 (0.085)
nonlinear	800	2.969 (1.087)	0.672 (0.244)	0.691 (0.249)	0.842 (0.069)
	1000	2.645 (0.958)	0.741 (0.244)	0.746 (0.245)	0.845 (0.071)
	1200	2.401 (0.842)	0.789 (0.236)	0.773 (0.241)	0.853 (0.076)

Table S.2
Size, TDR, MCC, and CCR for variable selection based on NBI test, SAS and riskRFE in the multiple-variable-based decision rule evaluation when $n = 200$.

NBI				
scenario	size (sd)	TDR (sd)	MCC (sd)	CCR (sd)
linear	1.321 (1.327)	0.549 (0.443)	0.414 (0.364)	0.707 (0.106)
binary	1.235 (1.186)	0.538 (0.447)	0.403 (0.371)	0.654 (0.081)
nonlinear	1.637 (1.685)	0.580 (0.431)	0.430 (0.372)	0.686 (0.108)
SAS				
scenario	size (sd)	TDR (sd)	MCC (sd)	CCR (sd)
linear	6.819 (2.154)	0.323 (0.140)	0.316 (0.239)	0.918 (0.039)
binary	7.626 (1.984)	0.262 (0.104)	0.189 (0.236)	0.723 (0.023)
nonlinear	7.246 (2.046)	0.295 (0.120)	0.271 (0.226)	0.889 (0.043)
riskRFE				
scenario	size (sd)	TDR (sd)	MCC (sd)	CCR (sd)
linear	5.325 (1.503)	0.330 (0.156)	0.296 (0.301)	0.747 (0.081)
binary	6.223 (1.418)	0.321 (0.102)	0.342 (0.204)	0.645 (0.100)
nonlinear	5.304 (1.570)	0.323 (0.156)	0.286 (0.291)	0.727 (0.081)

Table S.3

Size, TDR, MCC, and CCR for variable selection based on NBI test, SAS and riskRFE in the multiple-variable-based decision rule evaluation when NBI is calculated under Gaussian kernel.

NBI					
scenario	n	size(sd)	TDR(sd)	MCC(sd)	CCR(sd)
nonlinear	800	2.125 (0.413)	0.958 (0.116)	0.966 (0.090)	0.878 (0.062)
	1000	2.175 (0.506)	0.949 (0.132)	0.962 (0.099)	0.884 (0.061)
	1200	2.150 (0.410)	0.951 (0.124)	0.963 (0.092)	0.890 (0.052)
SAS					
scenario	n	size(sd)	TDR(sd)	MCC(sd)	CCR(sd)
nonlinear	800	2.327 (0.604)	0.902 (0.168)	0.928 (0.125)	0.990 (0.007)
	1000	2.183 (0.426)	0.941 (0.131)	0.958 (0.095)	0.991 (0.006)
	1200	2.105 (0.341)	0.967 (0.104)	0.976 (0.075)	0.992 (0.005)
riskRFE					
scenario	n	size(sd)	TDR(sd)	MCC(sd)	CCR(sd)
nonlinear	800	2.320 (0.467)	0.783 (0.279)	0.776 (0.338)	0.811 (0.109)
	1000	2.200 (0.400)	0.798 (0.293)	0.777 (0.360)	0.819 (0.106)
	1200	2.168 (0.374)	0.816 (0.277)	0.795 (0.340)	0.832 (0.101)

Table S.4
Prediction accuracy for ITRs derived by NBI and the standard OWL in the multiple-variable-based decision rule evaluation.

scenario	n	CCR (sd) NBI	CCR (sd) OWL
linear	800	0.835 (0.081)	0.837 (0.037)
	1000	0.852 (0.076)	0.850 (0.039)
	1200	0.870 (0.070)	0.862 (0.035)
binary	800	0.765 (0.098)	0.649 (0.121)
	1000	0.786 (0.095)	0.667 (0.125)
	1200	0.805 (0.090)	0.666 (0.125)
nonlinear	800	0.818 (0.081)	0.821 (0.042)
	1000	0.832 (0.075)	0.830 (0.041)
	1200	0.847 (0.073)	0.841 (0.037)

Table S.5
Size, TDR, MCC and CCR in the multiple-variable-based decision rule evaluation based on l_1-OWL when $\boldsymbol{X}_e = Null$.

scenario	n	size (sd)	TDR (sd)	MCC (sd)	CCR (sd)
linear	800	6.031 (3.358)	0.212 (0.197)	0.022 (0.288)	0.771 (0.161)
	1000	6.337 (3.203)	0.190 (0.159)	-0.013 (0.289)	0.785 (0.162)
	1200	6.744 (3.040)	0.190 (0.133)	-0.005 (0.271)	0.802 (0.156)
binary	800	5.950 (3.416)	0.187 (0.167)	-0.002 (0.276)	0.653 (0.144)
	1000	6.226 (3.163)	0.189 (0.159)	-0.012 (0.289)	0.665 (0.147)
	1200	6.757 (3.143)	0.199 (0.139)	0.005 (0.274)	0.679 (0.147)
nonlinear	800	5.887 (3.353)	0.196 (0.179)	0.003 (0.280)	0.731 (0.115)
	1000	6.427 (3.211)	0.199 (0.165)	-0.010 (0.290)	0.758 (0.110)
	1200	6.543 (2.968)	0.197 (0.144)	-0.004 (0.297)	0.769 (0.107)

Table S.6

Discovery rates for X_3 and X_4 in the single-variable-based decision rule evaluation for the additional nonlinear setting. (Discovery rate for X_4 equals 1-specificity.)

		$\rho = 0.0$		$\rho = 0.2$		$\rho = 0.5$		$\rho = 0.8$	
scenario	n	X_3	X_4	X_3	X_4	X_3	X_4	X_3	X_4
nonlinear	800	0.977	0.055	0.977	0.054	0.965	0.058	0.855	0.051
	1000	0.981	0.057	0.985	0.057	0.969	0.054	0.872	0.045
	1200	0.991	0.052	0.989	0.037	0.98	0.045	0.918	0.048

Table S.7

NBI values for X_3 and X_4 in the single-variable-based decision rule evaluation for the additional nonlinear decision rule.

		$\rho = 0.0$		$\rho = 0.2$	
scenario	n	X_3 mean (sd)	X_4 mean (sd)	X_3 mean (sd)	X_4 mean (sd)
nonlinear	800	1.443 (0.609)	-0.151 (0.633)	1.424 (0.604)	-0.153 (0.662)
	1000	1.405 (0.584)	-0.124 (0.613)	1.404 (0.572)	-0.112 (0.613)
	1200	1.421 (0.517)	-0.122 (0.595)	1.379 (0.520)	-0.168 (0.562)
		$\rho = 0.5$		$\rho = 0.8$	
scenario	n	X_3 mean (sd)	X_4 mean (sd)	X_3 mean (sd)	X_4 mean (sd)
nonlinear	800	1.266 (0.642)	-0.158 (0.646)	0.863 (0.646)	-0.143 (0.642)
	1000	1.233 (0.646)	-0.110 (0.600)	0.819 (0.649)	-0.139 (0.623)
	1200	1.221 (0.560)	-0.159 (0.595)	0.867 (0.571)	-0.123 (0.611)

Table S.8
Size, TDR, MCC, and CCR for variable selection based on NBI test, SAS and riskRFE in the multiple-variable-based decision rule evaluation for the additional nonlinear decision rule.

NBI					
scenario	n	size (sd)	TDR (sd)	MCC (sd)	CCR (sd)
nonlinear	800	2.064 (0.588)	0.936 (0.161)	0.912 (0.179)	0.910 (0.007)
	1000	2.085 (0.506)	0.943 (0.149)	0.932 (0.156)	0.912 (0.007)
	1200	2.085 (0.465)	0.946 (0.141)	0.938 (0.148)	0.912 (0.007)
SAS					
scenario	n	size (sd)	TDR (sd)	MCC (sd)	CCR (sd)
nonlinear	800	2.011 (0.104)	0.996 (0.035)	0.997 (0.025)	0.919 (0.008)
	1000	2.002 (0.045)	0.999 (0.015)	1.000 (0.011)	0.919 (0.010)
	1200	2.000 (0.000)	1.000 (0.000)	1.000 (0.000)	0.919 (0.009)
riskRFE					
scenario	n	size (sd)	TDR (sd)	MCC (sd)	CCR (sd)
nonlinear	800	3.062 (0.949)	0.713 (0.212)	0.782 (0.171)	0.878 (0.027)
	1000	2.802 (0.848)	0.772 (0.208)	0.829 (0.163)	0.883 (0.028)
	1200	2.561 (0.709)	0.833 (0.192)	0.878 (0.144)	0.894 (0.021)