Appendices

1 Outcome Variable Distribution in FHS and BRFSS datasets

We presented the outcome variable distribution for both FHS and BRFSS datasets:

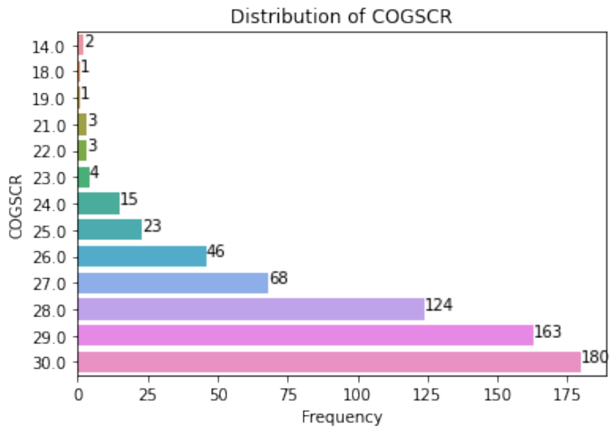

Figure 5: The outcome variable COGSCR distribution for FHS dataset.

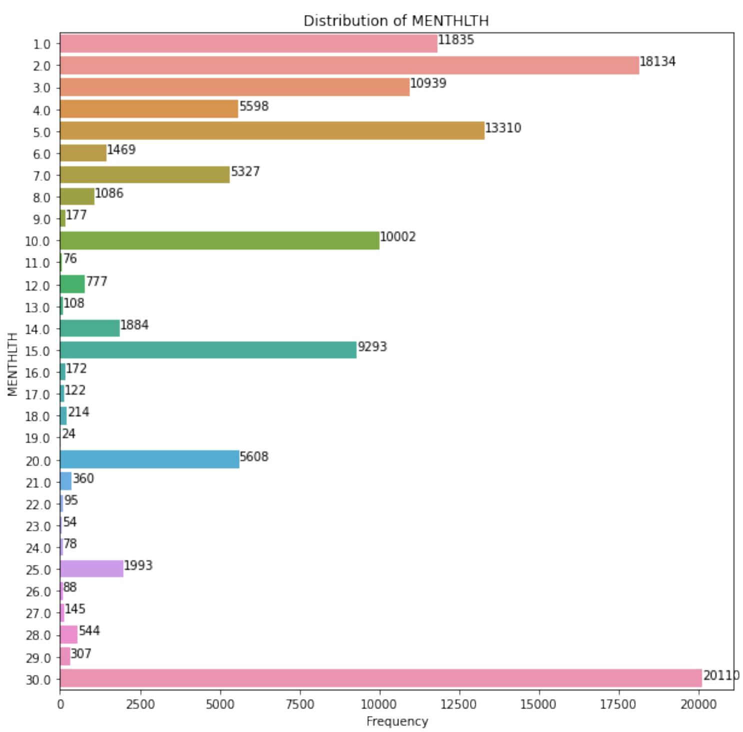

Figure 6: The outcome variable MENTHLH for BRFSS dataset.

2 Optimization in MTL Algorithm

Fast iterative shrinkage thresholding algorithm (FISTA) shown in Algorithm 1 is implemented to optimize the $l_{2,1}$-norm regularization problem in Eq.(4) with the general updating steps:

$$\Phi^{(l+1)} = \pi_P(S^{(l)} - \frac{1}{\gamma^{(l)}}\mathcal{L}'(S^{(l)})), \tag{9}$$

where l is the iteration index, $\frac{1}{\gamma^{(l)}}$ is the possible largest step-size that is chosen by line search and $\mathcal{L}'(S^{(l)})$ is the gradient of $\mathcal{L}(\cdot)$ at search point $S^{(l)}$. $S^{(l)} = \Phi^{(l)} + \alpha^{(l)}(\Phi^{(l)} - \Phi^{(l-1)})$ are the search points for each task, where $\alpha^{(l)}$ is the combination scalar. $\pi_P(\cdot)$ is $l_{2,1}-$regularized Euclidean projection shown as:

$$\pi_P(H(S^{(l)})) = \min_{\Phi} \frac{1}{2}||\Phi - H(S^{(l)})||_F^2 + \lambda||\Phi||_{2,1}, \tag{10}$$

where $H(S^{(l)}) = S^{(l)} - \frac{1}{\gamma^{(l)}}\mathcal{L}'(S^{(l)})$ is the gradient step of $S^{(l)}$. A sufficient scheme that solves Eq.(10) has been proposed as Theorem 1.

Theorem 1 *$\hat{\Phi}$'s primal optimal point in Eq.(10) can be calculated with λ as:*

$$\hat{\Phi}_j = \begin{cases} \left(1 - \frac{\lambda}{\|H(S^{(l)})_j\|_2}\right) H(S^{(l)})_j & \text{if} \quad \lambda > 0, \| H(S^{(l)})_j \|_2 > \lambda \\ 0 & \text{if} \quad \lambda > 0, \| H(S^{(l)})_j \|_2 \leq \lambda \\ H(S^{(l)})_j & \text{if} \quad \lambda = 0, \end{cases} \tag{11}$$

where $H(S^{(l)})_j$ is the j^{th} row of $H(S^{(l)})$ and $\hat{\Phi}_j$ is the j^{th} row of $\hat{\Phi}$.

Algorithm 1: Fast iterative shrinkage thresholding algorithm (FISTA) for optimizing the $l_{2,1}$-norm regularization problem.

Input: Input variables $\{X_1, X_2, \cdots, X_T\}$, output variable Y across all T tasks, initialization of feature weights $\Phi^{(0)}$ and λ

Output: $\hat{\Phi}$

1 **Initialize**: $\Phi^{(1)} = \Phi^{(0)}$, $d_{-1} = 0$, $d_0 = 1$,$\gamma^{(0)} = 1$,$l = 1$;
2 **repeat**
3   Set $\alpha^{(l)} = \frac{d_{l-2}-1}{d_{l-1}}$, $S^{(l)} = \Phi^{(l)} + \alpha^{(l)}(\Phi^{(l)} - \Phi^{(l-1)})$;
4   **for** $j = 1, 2, \cdots J$ **do**
5   Set $\gamma = 2^j\gamma_{l-1}$;
6   Compute $\Phi^{(l+1)} = \pi_P(S^{(l)} - \frac{1}{\gamma^{(l)}}\mathcal{L}'(S^{(l)}))$;
7   Compute $Q_\gamma(S^{(l)}, \Phi^{(l+1)})$;
8   **if** $\mathcal{L}(\Phi^{(l+1)}) \leq Q_\gamma(S^{(l)}, \Phi^{(l+1)})$ **then**
9             $\gamma^{(l)} = \gamma$, **break** ;
10   **end**
11   **end**
12     $d_l = \frac{1+\sqrt{1+4d_{l-1}^2}}{2}$;
13     $l = l + 1$;
14 **until** *Convergence of* $\Phi^{(l)}$;
15 $\hat{\Phi} = \Phi^{(l)}$;

From the 4^{th} line to the 11^{th} line in Algorithm 1, the optimal $\gamma^{(l)}$ is chosen by the backtracking rule. And $\gamma^{(l)} \geq b$, where b is the Lipschitz constant of $\mathcal{L}(\cdot)$ at search point $S^{(l)}$, which means $\gamma^{(l)}$ is satisfied for $S^{(l)}$ and $\frac{1}{\gamma^{(l)}}$ is the possible largest step size.

At the 7^{th} line in Algorithm 1, tangential line of $\mathcal{L}(\cdot)$ at search point $S^{(l)}$, denoted as $Q_\gamma(S^{(l)}, \Phi^{(l+1)})$, is computed by:

$$\begin{aligned} Q_\gamma(S^{(l)}, \Phi^{(l+1)}) = \mathcal{L}(S^{(l)}) + \frac{\gamma}{2} \parallel \Phi^{(l+1)} - S^{(l)} \parallel^2 \\ + \langle \Phi^{(l+1)} - S^{(l)}, \mathcal{L}'(S^{(l)}) \rangle. \end{aligned}$$

3 Optimization in CMTL Algorithm

In Eq.(8), the equation is conjointly convex with respect to (w.r.t.) C and Φ, which is an convex unconstrained smooth optimization problem w.r.t. C. We iteratively update the gradient step of the aforementioned optimization problem in order to find the global optimum w.r.t. C:

$$G_\Phi = S - \frac{1}{\gamma}[\bigtriangledown \mathcal{L}(S_\Phi) + 2\rho_1\eta(1+\eta)(\eta I + C_S)^{-1}S^T], \tag{12}$$

where S_Φ is the search point of Φ that is defined as $S_\Phi^{(l)} = \Phi^{(l)} + \alpha^{(l)}(\Phi^{(l)} - \Phi^{(l-1)})$. The search point of C is denoted as C_S, which can be similarly updated as $C_S^{(l)} = C^{(l)} + \alpha^{(l)}(C^{(l)} - C^{(l-1)})$ at the l^{th} iteration. $\bigtriangledown \mathcal{L}(S)$ is the gradient of $\mathcal{L}(S)$ that is calculated as:

$$\bigtriangledown \mathcal{L}(S) = \left[\frac{l^{'}(S_1)}{N_1}, \frac{l^{'}(S_2)}{N_2}, \cdots, \frac{l^{'}(S_T)}{N_T} \right]. \tag{13}$$

Similarly, in the optimization of MTFL, FISTA is also implemented for optimizing the CMTL, except the line 6 is replaced with the corresponding proximal operator that is solved by the following steps. To optimize the convex set C, we need to solve a convex constrained minimization problem, which is formulated with its corresponding proximal operator and calculated using its gradient step, denoted as G_C, at the search point C_S:

$$\min_C \|C - G_C\|_F^2, \quad \text{s.t.} \quad \text{tr}(C) = K, C \preceq I, C \in \mathbb{S}_+^T. \tag{14}$$

We can compute the G_C by:

$$G_C = C_S + \frac{\rho_1\eta(1+\eta)}{\gamma} S^T S(\eta I + C_S)^{-2}. \tag{15}$$

A solution of Eq.(14) is proposed and summarized in the following theorem.

Theorem 2 *Let $G_T = V\hat{\Sigma}V^T$ be the eigen-decomposition of gradient step $G_C \in \mathbb{S}^{T\times T}$, where $\hat{\Sigma} = diag(\hat{\sigma}_1, \cdots, \hat{\sigma}_T) \in \mathbb{R}^{T\times T}$ and $V \in \mathbb{R}^{T\times T}$ is orthonormal. The optimization problem is formulated as:*

$$\begin{aligned} &\mathit{min}_{\{\sigma_m\}} && \sum_{t=1}^T (\sigma_t - \hat{\sigma}_t)^2. \\ &\mathit{s.\ t.} && \sum_{t=1}^T \sigma_t = K,\ 0 \le \sigma_t \le 1,\ \forall t = 1, \cdots, T \end{aligned} \tag{16}$$

Let $\Sigma^* = diag(\sigma_1^*, \cdots, \sigma_T^*) \in \mathbb{R}^{T\times T}$, so that the optimal solution of the above optimization problem is $\{\sigma_1^*, \cdots, \sigma_T^*\}$. As a result, the proximal operator's optimal solution in Eq.(14) is calculated as $\hat{T} = V\Sigma^*V^T$.

4 Risk Factor Analysis - Additional Results

We presented the additional results of risk factor analysis for FHS and BRFSS datasets:

	MTL	LinearRegression	Ridge	Lasso	LassoLars	BayesianRidge	ElasticNet
1	CHF117	CHF011D	CHF115	MAXCOG	MAXCOG	CHF094	MAXCOG
2	CHF052	CHF079	CHF031	EXAM	EXAM	CHF027	EXAM
3	CHF116	CHF094	CHF048	EFLT50	EFLT50	CHF008D	EFLT50
4	CHF039	CHF101	CHF040	CHF117	CHF117	CHF109	CHF117
5	CHF029	CHF095	CHF087	CHF116	CHF116	CHF037	CHF116
6	CHF062	CHF091	CHF057	CHF115	CHF115	CHF038	CHF115
7	CHF094	CHF082	CHF039	CHF114	CHF114	CHF015D	CHF114
8	CHF075	CHF115	CHF029	CHF113	CHF113	CHF011D	CHF113
9	CHF091	CHF048	CHF052	CHF112	CHF112	CHF093	CHF112
10	CHF087	CHF080	CHF091	CHF111	CHF111	CHF099	CHF111

Figure 7: Top 10 selected RFs and their corresponding category numbers from our proposed MTL method and seven STL methods for FHS dataset. Please zoom in for clear visualization.

	MTL	LinearRegression	Ridge	Lasso	LassoLars	BayesianRidge	ElasticNet
1	PHYSHLTH	CHCKDNY1	CHCKDNY1	HLTHPLN1	_URBSTAT	CHCKDNY1	HLTHPLN1
2	_AGE80	_RFSEAT2	HLTHPLN1	_AGE_G	_TOTINDA	_RFSEAT2	_TOTINDA
3	_AGE_G	_AGE_G	_AGE_G	_TOTINDA	_SMOKER3	CHECKUP1	_AGE_G
4	ADDEPEV2	CHECKUP1	_AGE65YR	_URBSTAT	_RFSMOK3	_AGE_G	_AGE65YR
5	_AGEG5YR	_PHYS14D	DIFFWALK	_SMOKER3	_RFSEAT3	_PHYS14D	_AGE80
6	_HISPANC	_MENT14D	_RACE	_RFSMOK3	_RFSEAT2	_MENT14D	_EXTETH3
7	DECIDE	DIFFWALK	_MENT14D	_RFSEAT3	_RFHLTH	DIFFWALK	ADDEPEV2
8	_INCOMG	HLTHPLN1	CHECKUP1	_RFSEAT2	_RFDRHV6	HLTHPLN1	_INCOMG
9	_SMOKER3	SLEPTIM1	_INCOMG	_RFHLTH	_RFBMI5	CHCSCNCR	INCOME2
10	MEDCOST	CHCSCNCR	_RACE_G1	_RFDRHV6	_RFBING5	CVDINFR4	CHCKDNY1

Figure 8: Top 10 selected RFs and their corresponding category numbers from our proposed CMTL method and six STL methods for BRFSS dataset. Please zoom in for clear visualization.