F. Additional simulation results and additional simulation study using Highly Adaptive Lasso (HAL) (van der Laan, 2017)

F.1 Detailed simulation results from main text

[Table 1 about here.]

F.2 Additional simulation studies We conduct additional simulation studies to compare the performance of algorithms that use machine learning to estimate the nuisance functions for the parameters in an MSM. Specifically, we compare the proposed TMLE with IPW and ORE.

We simulated 1000 hypothetical cohorts of $n=(500,1000,5000)$ comprising the following variables: (U,L_1,L_2,A,B,C,Y), where $U\sim\text{Unif}[0,1]$ denotes an unmeasured baseline covariate, $L_1\sim\text{Ber}(0.5)$ and $L_2\sim\text{Unif}[0,1]$ denotes a measured baseline covariates, $(A,B,C)^T$ denotes a vector of exposures where $A\sim\text{Bin}(3,\text{expit}(-2+0.75L_1+0.5L_2+U))$, $B\sim\text{Bin}(3,\text{expit}(-1+0.5L_1-0.2\exp(0.5L_2)+U))$, $C\sim\text{Bin}(3,\text{expit}(-1+L_1+U))$, and $Y\sim\text{Ber}(\text{expit}(-1-A+B+C-2L_1+L_2-0.25L_2^2+0.25L_1L_2))$ denotes a binary outcome. We consider estimating the parameters in the MSM given by $E(Y^g|L;\theta)=\text{expit}(\theta_0+\theta_1L_2)$.

Since HAL can only handle `binomial`, `gaussian`, `poisson`, `cox` or `mgaussian` family objects, we will need to convert our multinomial exposure variables (e.g., for A, we have the $\text{supp}(A)=\{0,1,2,\dots,K\}$) to binary outcomes and estimating $P(A=k|L)$ for $k\geq 0$. This can be done by generating a series of logistic regression models as follows:

(1) Calculate $\hat{P}(A=0|L)$ and $\hat{P}(A\neq 0|L)=\hat{P}(A>0|L)$ using a logistic regression model for binary events $\{A=0\}$ and $\{A\neq 0\}$.
(2) For $k=1,2,...,K-1$ calculate $\hat{P}(A=k|A>k-1,L)$ and $\hat{P}(A\neq k|A>k-1,L)=\hat{P}(A>k|A>k-1,L)$ using logistic regression models.
(3) Using the fact that $P(A=k|L)=P(A=k|A>k-1,L)\prod_{j=0}^{k-1}P(A>j|A>j-1,L)$, calculate $\hat{P}(A=k|L)=\hat{P}(A=k|A>k-1,L)\prod_{j=0}^{k-1}\hat{P}(A>j|A>j-1,L)$ using the fitted logistic regression models from previous steps. For instance, for $k=2$, we can calculate $\hat{P}(A=2|L)=\hat{P}(A=2|A>1,L)\hat{P}(A>1|A>0,L)\hat{P}(A>0|L)$.
(4) For level K, $\hat{P}(A=K|L)=1-\sum_{k=0}^{K-1}\hat{P}(A=k|L)$

Equivalent methods can also be applied to B where $\text{supp}(B)=\{0,1,2,3\}$.

Table G.2 compares the performance of the 3 estimators, and the results are similar to those in the main manuscript. The ORE and IPW estimators show more bias compared with the TMLE as they are not expected to converge at $\sqrt{n}$ rates when machine learning is used for nuisance parameter estimation. TMLE, on the other hand, show little to no bias in all instances. This agrees with theory as TMLE allows the nuisance functions to converge at slower nonparametric rates. The estimated coverage probability of the confidence intervals for TMLE based on the asymptotic variance gets closer to the nominal 95% as sample size increases. For ($\hat{\beta}_0$,$\hat{\beta}_1$), the 95% coverage probability is [(91.6,90.4), (92.2,92.6), (94.3,93.7)] for n=(500, 1000, 5000), respectively.

[Table 2 about here.]

G. Additional sensitivity analysis of Nurses' Health Study

Under no diet intervention, the estimated two-year and four-year mortality risk are 1.97% (95% CI: 1.82% – 2.11%) and 5.37% (95% CI: 5.14%–5.60%), respectively. We estimated the two-year and four-year mortality risk under the dietary substitution intervention to be 1.60% (95% CI: 1.26% – 1.91%) and 4.43% (95% CI: 3.86% – 5.03%) using TMLE with parametric models, and 1.68% (95% CI: 1.42% – 1.94%) and 4.72% (95% CI: 4.12% – 5.25%) using TMLE with HAL, respectively. The estimates from both of the methods are similar (see Figure 6). Based on the results from TMLE with HAL, our estimates suggest that the intervention would slightly decrease the two-year mortality by an estimated 0.29% (95% CI: 0.06%–0.52%) and the four-year mortality by an estimated 0.65% (95% CI: 0.16%–1.15%).

Sensitivity analysis was performed using various values of x={8, 9, 10}, and the results closely resemble those obtained when x=7 (see Figures G.3 and G.3). That is, there is some evidence to suggest that the proposed substitution strategy would slightly decrease the two-year and four-year mortality rate.

[Figure 2 about here.]

[Figure 3 about here.]

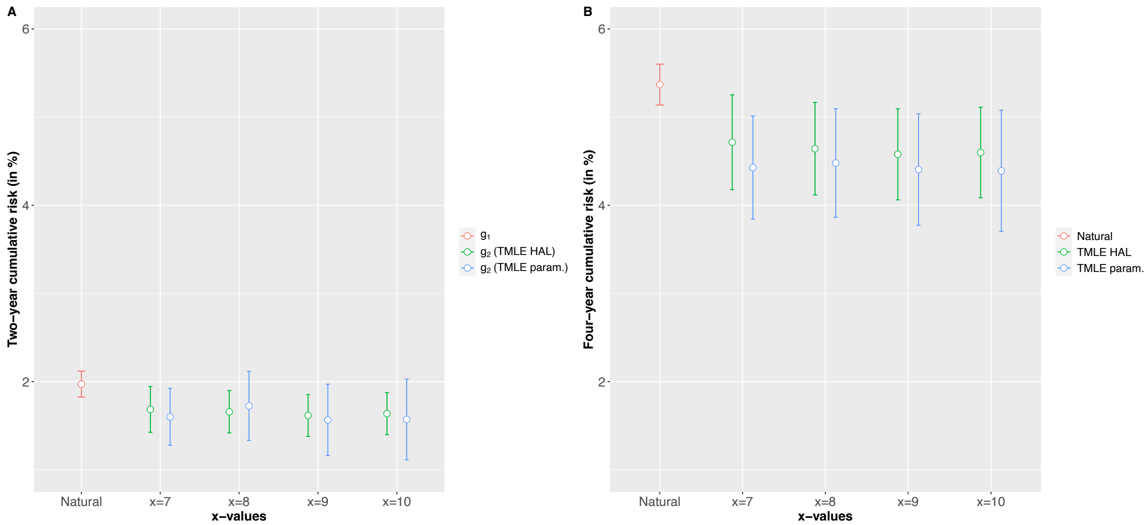

Figure G.2. Estimated two-year and four-year cumulative risk for x={7,8,9,10} using TMLE with parametric models for estimating nuisance functions and TMLE with HAL for estimating nuisance functions.

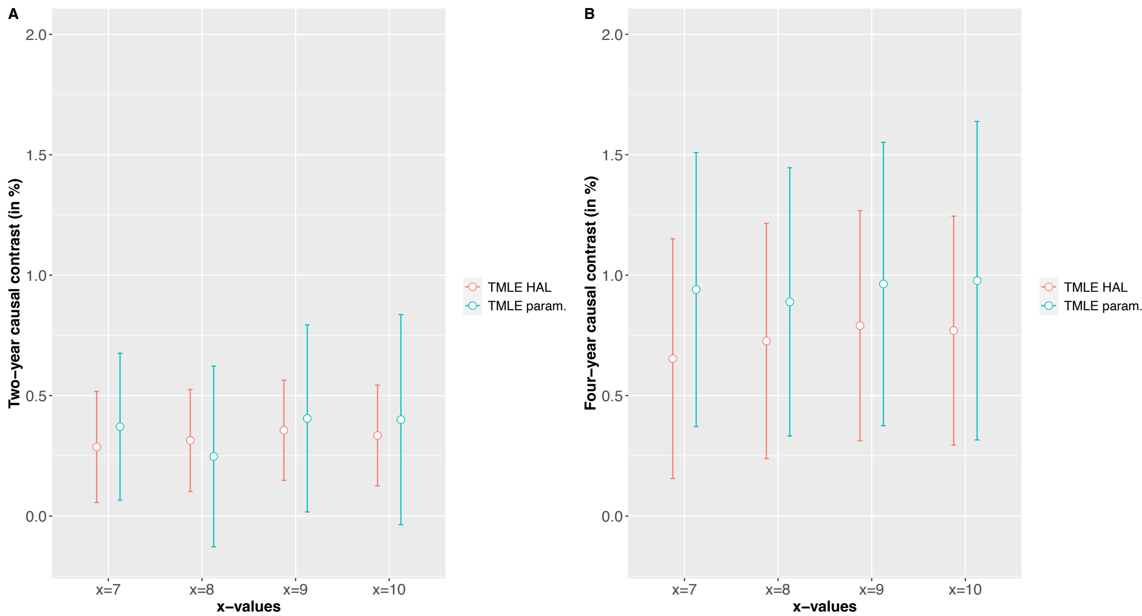

FIGURE G.3. Estimated two-year and four-year causal contrast for $x = \{7,8,9,10\}$ using TMLE with parametric models for estimating nuisance functions and TMLE with HAL for estimating nuisance functions.

	(i) Correctly Specified				(ii) Misspecified outcome model				(iii) Misspecified exposure model			
ORE	Bias	SE	MSE	CP	Bias	SE	MSE	CP	Bias	SE	MSE	CP
θ_0	0.003	0.065	0.004	95.6	−0.861	0.051	0.745	0.00	0.003	0.065	0.004	95.6
θ_1	−0.002	0.076	0.006	94.5	0.734	0.088	0.546	0.00	−0.002	0.076	0.006	94.5
IPW	Bias	SE	MSE	CP	Bias	SE	MSE	CP	Bias	SE	MSE	CP
θ_0	0.006	0.090	0.008	95.1	0.006	0.090	0.008	95.1	0.154	0.090	0.032	64.2
θ_1	−0.003	0.175	0.031	92.2	−0.003	0.175	0.031	92.2	−0.568	0.131	0.340	0.50
TMLE	Bias	SE	MSE	CP	Bias	SE	MSE	CP	Bias	SE	MSE	CP
θ_0	0.006	0.090	0.008	95.1	0.006	0.090	0.008	95.1	0.006	0.090	0.008	95.5
θ_1	−0.006	0.175	0.031	92.2	0.000	0.175	0.031	92.3	−0.009	0.122	0.015	95.1

TABLE G.1. Results for simulation study for n=5000 using parametric models for nuisance functions: Bias, standard error (SE), mean squared error (MSE) and 95% coverage probability (CP; $\times$100). True value of $(\theta_0,\theta_1)=(2.347,-1.635)$.

TABLE G.2. Results for simulation study for n=(500,1000,5000) using nonparametric models for nuisance functions: Bias, standard error (SE), and mean squared error (MSE). True value of (θ_0,θ_1)= (0.996,0.822).

	n=500			n=1000			n=5000		
ORE	Bias	SE	MSE	Bias	SE	MSE	Bias	SE	MSE
θ_0	-0.085	0.256	0.075	-0.044	0.178	0.042	-0.024	0.086	0.020
θ_1	-0.062	0.398	0.188	-0.044	0.293	0.109	-0.011	0.141	0.036
IPW	Bias	SE	MSE	Bias	SE	MSE	Bias	SE	MSE
θ_0	0.049	0.562	0.344	0.050	0.384	0.179	0.037	0.171	0.051
θ_1	-0.069	1.093	1.226	-0.076	0.756	0.596	-0.080	0.342	0.154
TMLE	Bias	SE	MSE	Bias	SE	MSE	Bias	SE	MSE
θ_0	0.003	0.474	0.242	0.001	0.337	0.133	-0.006	0.156	0.038
θ_1	0.045	0.889	0.800	0.029	0.665	0.446	0.016	0.312	0.112