Appendices for Kernel Mean Matching with Mahalanobis Distance for Causal Inference of Time-to-event Outcome

1st Qin Ma
Department of Computer Science and Engineering
Shanghai Jiao Tong University
Shanghai, China
m872384296@sjtu.edu.cn

2nd Lin Zeng
Department of Computer Science and Engineering
Shanghai Jiao Tong University
Shanghai, China
zenglin10@sjtu.edu.cn

3rd Shikui Tu
Department of Computer Science and Engineering
Shanghai Jiao Tong University
Shanghai, China
tushikui@sjtu.edu.cn

4th Lei Xu
Department of Computer Science and Engineering
Shanghai Jiao Tong University
Shanghai, China
leixu@sjtu.edu.cn

APPENDIX A
PROOFS

A. Lemma 1

Proof:

Since $\boldsymbol{\Sigma}_{\mathcal{F}}$ is a nonsingular matrix, it is positive definite. Then we have that $\boldsymbol{\Sigma}_{\mathcal{F}}^{-1}$ is also positive definite. Therefore, $\boldsymbol{\Sigma}_{\mathcal{F}}^{-1}$ can be always decomposed as $\boldsymbol{\Sigma}_{\mathcal{F}}^{-1} = \boldsymbol{U}^{\mathsf{T}}\boldsymbol{U}$, where $\boldsymbol{U}$ is an invertible matrix. The matrix $\boldsymbol{U}$ satisfies that $\Psi(\boldsymbol{x}) = \boldsymbol{U}\Phi(\boldsymbol{x})$, where $\Psi : \mathcal{X} \to \mathcal{F}$ is another map into a feature space $\mathcal{F}$. Meanwhile, the map Ψ satisfies that the inner product of the map is a Mahalanobis distance kernel, formally, $k_{\mathrm{M}}(\boldsymbol{x}, \boldsymbol{x}') = \langle \Psi(\boldsymbol{x}), \Psi(\boldsymbol{x}') \rangle$. Therefore, integrating the matrix $\boldsymbol{U}$ into the feature map Φ, and then the objective function in (4) can be written as

$$\left\| \mathbb{E}\left[\Psi(\boldsymbol{X}_{\mathrm{t}})\right] - \mathbb{E}\left[w(\boldsymbol{X}_{\mathrm{c}})\Psi(\boldsymbol{X}_{\mathrm{c}})\right] \right\|^2, \tag{A.1}$$

which is the exact form of KMM with Mahalanobis distance kernels. ■

B. Lemma 2

Proof:

Based on Lemma 1, we can rewrite (4) following the form of KMM given by (A.1). Huang et al. [1] have proven that if the feature space $\mathcal{F}$ is an RKHS with a universal kernel, then $\Pr(\boldsymbol{X}_{\mathrm{t}} = \boldsymbol{x}) = \Pr(\boldsymbol{X}_{\mathrm{c}} = \boldsymbol{x})$ is equivalent to the kernel mean embedding $\mathbb{E}\left[\Psi(\boldsymbol{X}_{\mathrm{t}})\right] = \mathbb{E}\left[\Psi(\boldsymbol{X}_{\mathrm{c}})\right]$. Here, we directly use their results. Because of the convexity of KMM problem, together with the constraints that guarantee the derived distributions to follow the law of probability, we finally have $\Pr(\boldsymbol{X}_{\mathrm{t}} = \boldsymbol{x}) = w(\boldsymbol{x}) \Pr(\boldsymbol{X}_{\mathrm{c}} = \boldsymbol{x})$. ■

C. Theorem 3

Proof:

Based on the law of conditional probability, the potential survival functions can be written as

$$\mathrm{S}_1(t|Z=1) = \prod_{j=1}^{m} \left[1 - \Pr\left(T(1) = t_j | T(1) \geq t_j, Z = 1\right)\right],$$

$$\mathrm{S}_0(t|Z=1) = \prod_{j=1}^{m} \left[1 - \Pr\left(T(0) = t_j | T(0) \geq t_j, Z = 1\right)\right].$$

Considering (1) and (2), we just need to prove that

- $\Pr\left(T(1) = t_j | T(1) \geq t_j, Z = 1\right)$ is consistently estimated by d_{j1}/Y_{j1} and
- $\Pr\left(T(0) = t_j | T(0) \geq t_j, Z = 1\right)$ is consistently estimated by $d_{j0}^{\mathrm{w}}/Y_{j0}^{\mathrm{w}}$.

The former one is straightforward:

$$\begin{aligned}
\operatorname{plim}\left(\frac{d_{j1}}{Y_{j1}}\right) &= \operatorname{plim}\left[\frac{\sum_{i=1}^{n}\mathbb{I}\left(\widetilde{T}_i=t_j,\delta_i=1,Z_i=1\right)\Big/n}{\sum_{i=1}^{n}\mathbb{I}\left(\widetilde{T}_i\geq t_j,Z_i=1\right)\Big/n}\right] \\
&= \frac{\Pr\left(\widetilde{T}=t_j,T\leq C,Z=1\right)}{\Pr\left(\widetilde{T}\geq t_j,Z=1\right)} && \text{(A.2)} \\
&= \frac{\Pr\left(T=t_j,C\geq t_j,Z=1\right)}{\Pr\left(T\geq t_j,C\geq t_j,Z=1\right)} && \text{(A.3)} \\
&= \frac{\Pr\left(T=t_j,Z=1\right)}{\Pr\left(T\geq t_j,Z=1\right)} && \text{(A.4)} \\
&= \Pr\left(T\left(1\right)=t_j|T\left(1\right)\geq t_j,Z=1\right). && \text{(A.5)}
\end{aligned}$$

Equality (A.2) can be derived through the law of large number. Since $\widetilde{T}=t_j\cap T\leq C$ is equivalent to $T=t_j\cap C\geq t_j$ and $\widetilde{T}\geq t_j$ is equivalent to $T\geq t_j\cap C\geq t_j$, (A.3) can thus be obtained. Equality (A.4) is based on the independent censoring assumption. Equality (A.5) is due to $T=t_j\subset T\geq t_j$ and $T=ZT\left(1\right)+\left(1-Z\right)T\left(0\right)$.

Next, we will prove the later one:

$$\begin{aligned}
\operatorname{plim}\left(\frac{d_{j0}^{\mathrm{w}}}{Y_{j0}^{\mathrm{w}}}\right) =& \operatorname{plim}\left[\frac{\frac{\sum_{i=1}^{n}w\left(\boldsymbol{X}_i\right)\mathbb{I}\left(\widetilde{T}_i=t_j,\delta_i=1,Z_i=0\right)}{n}\Big/\frac{\sum_{i=1}^{n}\mathbb{I}\left(Z_i=0\right)}{n}}{\frac{\sum_{i=1}^{n}w\left(\boldsymbol{X}_i\right)\mathbb{I}\left(\widetilde{T}_i\geq t_j,Z_i=0\right)}{n}\Big/\frac{\sum_{i=1}^{n}\mathbb{I}\left(Z_i=0\right)}{n}}\right] \\
=& \frac{\sum_{\boldsymbol{x}}w\left(\boldsymbol{x}\right)\Pr\left(\widetilde{T}=t_j,T\leq C,\boldsymbol{X}=\boldsymbol{x}|Z=0\right)}{\sum_{\boldsymbol{x}}w\left(\boldsymbol{x}\right)\Pr\left(\widetilde{T}\geq t_j,\boldsymbol{X}=\boldsymbol{x}|Z=0\right)} \\
=& \frac{\sum_{\boldsymbol{x}}w\left(\boldsymbol{x}\right)\Pr\left(T=t_j,\boldsymbol{X}=\boldsymbol{x}|Z=0\right)}{\sum_{\boldsymbol{x}}w\left(\boldsymbol{x}\right)\Pr\left(T\geq t_j,\boldsymbol{X}=\boldsymbol{x}|Z=0\right)} \\
=& \frac{\sum_{\boldsymbol{x}}\frac{\Pr\left(\boldsymbol{X}=\boldsymbol{x}|Z=1\right)}{\Pr\left(\boldsymbol{X}=\boldsymbol{x}|Z=0\right)}\cdot\Pr\left(T\left(0\right)=t_j|\boldsymbol{X}=\boldsymbol{x},Z=0\right)\cdot\Pr\left(\boldsymbol{X}=\boldsymbol{x}|Z=0\right)}{\sum_{\boldsymbol{x}}\frac{\Pr\left(\boldsymbol{X}=\boldsymbol{x}|Z=1\right)}{\Pr\left(\boldsymbol{X}=\boldsymbol{x}|Z=0\right)}\cdot\Pr\left(T\left(0\right)\geq t_j|\boldsymbol{X}=\boldsymbol{x},Z=0\right)\cdot\Pr\left(\boldsymbol{X}=\boldsymbol{x}|Z=0\right)} && \text{(A.6)} \\
=& \frac{\sum_{\boldsymbol{x}}\Pr\left(\boldsymbol{X}=\boldsymbol{x}|Z=1\right)\Pr\left(T\left(0\right)=t_j|\boldsymbol{X}=\boldsymbol{x},Z=1\right)}{\sum_{\boldsymbol{x}}\Pr\left(\boldsymbol{X}=\boldsymbol{x}|Z=1\right)\Pr\left(T\left(0\right)\geq t_j|\boldsymbol{X}=\boldsymbol{x},Z=1\right)} && \text{(A.7)} \\
=& \frac{\sum_{\boldsymbol{x}}\Pr\left(T\left(0\right)=t_j,\boldsymbol{X}=\boldsymbol{x}|Z=1\right)}{\sum_{\boldsymbol{x}}\Pr\left(T\left(0\right)\geq t_j,\boldsymbol{X}=\boldsymbol{x}|Z=1\right)} \\
=& \Pr\left(T\left(0\right)=t_j|T\left(0\right)\geq t_j,Z=1\right),
\end{aligned}$$

where (A.6) is due to Lemma 2, and (A.7) is based on the unconfoundedness assumption. Till now, the consistency of the proposed estimator is proved.

■

APPENDIX B
TABLES AND FIGURES

TABLE B.1
BASIC DESCRIPTIVE STATISTICS FROM THE TCGA-CDR BRCA

		Total No. 838	Negative No. 202	Positive No. 636	P-value
Age		58 (48 - 66)	55 (47 - 62)	58 (49 - 67)	0.005
Gender					0.13
	Female	828 (99%)	202 (24%)	626 (76%)	
	Male	10 (1%)	0 (0%)	10 (100%)	
Race					< 0.001
	Not Evaluated	1 (0%)	0 (0%)	1 (100%)	
	American Indian or Alaska native	1 (0%)	1 (100%)	0 (0%)	
	Asian	51 (6%)	16 (31%)	35 (69%)	
	Black or African American	163 (19%)	66 (40%)	97 (60%)	
	White	622 (74%)	119 (19%)	503 (81%)	
Radiation Therapy		377 (45%)	94 (47%)	283 (44%)	0.63
Histology					< 0.001
	Infiltrating Carcinoma NOS	1 (0%)	1 (100%)	0 (0%)	
	Infiltrating Ductal Carcinoma	594 (71%)	178 (30%)	416 (70%)	
	Infiltrating Lobular Carcinoma	165 (20%)	4 (2%)	161 (98%)	
	Medullary Carcinoma	2 (0%)	2 (100%)	0 (0%)	
	Metaplastic Carcinoma	7 (1%)	6 (86%)	1 (14%)	
	Mixed Histology (please specify)	21 (3%)	4 (19%)	17 (81%)	
	Mucinous Carcinoma	12 (1%)	0 (0%)	12 (100%)	
	Other specify	36 (4%)	7 (19%)	29 (81%)	
Marigin Status					0.094
	Close	31 (4%)	10 (32%)	21 (68%)	
	Negative	749 (89%)	184 (25%)	565 (75%)	
	Positive	58 (7%)	8 (14%)	50 (86%)	
PR Status					< 0.001
	Indeterminate	3 (0%)	0 (0%)	3 (100%)	
	Negative	291 (35%)	189 (65%)	102 (35%)	
	Positive	544 (65%)	13 (2%)	531 (98%)	
HER2 Status					0.29
	Not Available	8 (1%)	2 (25%)	6 (75%)	
	Not Evaluated	101 (12%)	29 (29%)	72 (71%)	
	Equivocal	165 (20%)	33 (20%)	132 (80%)	
	Indeterminate	10 (1%)	1 (10%)	9 (90%)	
	Negative	445 (53%)	104 (23%)	341 (77%)	
	Positive	109 (13%)	33 (30%)	76 (70%)	
Stage					0.013
	Discrepancy	3 (0%)	1 (33%)	2 (67%)	
	Not Available	4 (0%)	3 (75%)	1 (25%)	
	Stage I	72 (9%)	14 (19%)	58 (81%)	
	Stage IA	75 (9%)	19 (25%)	56 (75%)	
	Stage IB	6 (1%)	0 (0%)	6 (100%)	
	Stage II	4 (0%)	1 (25%)	3 (75%)	
	Stage IIA	266 (32%)	85 (32%)	181 (68%)	
	Stage IIB	199 (24%)	36 (18%)	163 (82%)	
	Stage IIIA	124 (15%)	25 (20%)	99 (80%)	
	Stage IIIB	15 (2%)	5 (33%)	10 (67%)	
	Stage IIIC	52 (6%)	9 (17%)	43 (83%)	
	Stage IV	12 (1%)	3 (25%)	9 (75%)	
	Stage X	6 (1%)	1 (17%)	5 (83%)	
Tumor Status		754 (90%)	174 (86%)	580 (91%)	0.044
Status					0.1
	Alive	757 (90%)	176 (23%)	581 (77%)	
	Death	81 (10%)	26 (32%)	55 (68%)	

TABLE B.2
SENSITIVITY ANALYSIS FOR RMSE OF THE ATT ESTIMATION ON SIMULATION DATA

	$\sigma = 0.7$					$\sigma = 0.3$				
Estimator	t_1	t_2	t_3	t_4	t_5	t_1	t_2	t_3	t_4	t_5
(i) $\gamma = 1\text{e}-4$										
Crude	0.240	0.242	0.215	0.158	0.089	0.138	0.155	0.151	0.118	0.069
AKME	0.228	0.230	0.205	0.151	0.085	0.127	0.143	0.139	0.109	0.065
OW	0.224	0.227	0.203	0.150	0.084	0.125	0.141	0.138	0.108	0.065
CBPS	0.233	0.235	0.210	0.154	0.086	0.131	0.148	0.144	0.112	0.067
OS TMLE	0.224	0.228	0.206	0.155	0.091	0.122	0.140	0.139	0.111	0.070
eKMM-surv	0.079	0.101	0.115	0.105	0.086	0.084	0.118	0.144	0.143	0.108
mKMM-surv	**0.053**	**0.062**	**0.065**	**0.059**	**0.055**	**0.050**	**0.065**	**0.073**	**0.070**	**0.059**
(ii) $\gamma = 1\text{e}-3$										
Crude	0.240	0.242	0.215	0.158	0.089	0.138	0.155	0.151	0.118	0.069
AKME	0.228	0.230	0.205	0.151	0.085	0.127	0.143	0.139	0.109	0.065
OW	0.224	0.227	0.203	0.150	0.084	0.125	0.141	0.138	0.108	0.065
CBPS	0.233	0.235	0.210	0.154	0.086	0.131	0.148	0.144	0.112	0.067
OS TMLE	0.224	0.228	0.206	0.155	0.091	0.122	0.140	0.139	0.111	0.070
eKMM-surv	0.044	0.060	0.068	0.071	0.076	0.043	0.052	0.062	0.066	0.072
mKMM-surv	**0.032**	**0.039**	**0.046**	**0.051**	**0.055**	**0.031**	**0.038**	**0.044**	**0.048**	**0.053**
(iii) $\gamma = 1\text{e}-2$										
Crude	0.240	0.242	0.215	0.158	0.089	0.138	0.155	0.151	0.118	0.069
AKME	0.228	0.230	0.205	0.151	0.085	0.127	0.143	0.139	0.109	0.065
OW	0.224	0.227	0.203	0.150	0.084	0.125	0.141	0.138	0.108	0.065
CBPS	0.233	0.235	0.210	0.154	0.086	0.131	0.148	0.144	0.112	0.067
OS TMLE	0.224	0.228	0.206	0.155	0.091	0.122	0.140	0.139	0.111	0.070
eKMM-surv	0.039	0.048	0.056	0.062	0.067	0.036	0.043	0.052	0.059	**0.064**
mKMM-surv	**0.035**	**0.042**	**0.051**	**0.057**	**0.061**	**0.035**	**0.042**	**0.051**	**0.058**	**0.064**
(iv) $\gamma = 1\text{e}-1$										
Crude	0.240	0.242	0.215	0.158	0.089	0.138	0.155	0.151	0.118	0.069
AKME	0.228	0.230	0.205	0.151	0.085	0.127	0.143	0.139	0.109	0.065
OW	0.224	0.227	0.203	0.150	0.084	0.125	0.141	0.138	0.108	0.065
CBPS	0.233	0.235	0.210	0.154	0.086	0.131	0.148	0.144	0.112	0.067
OS TMLE	0.224	0.228	0.206	0.155	0.091	0.122	0.140	0.139	0.111	0.070
eKMM-surv	**0.032**	**0.040**	**0.046**	**0.050**	0.059	**0.030**	**0.035**	**0.041**	**0.046**	0.050
mKMM-surv	0.046	0.050	0.051	**0.050**	**0.052**	0.035	0.039	0.043	0.047	**0.048**
(v) $\gamma = 1$										
Crude	0.240	0.242	0.215	0.158	0.089	0.138	0.155	0.151	0.118	0.069
AKME	0.228	0.230	0.205	0.151	0.085	0.127	0.143	0.139	0.109	0.065
OW	0.224	0.227	0.203	0.150	0.084	0.125	0.141	0.138	0.108	0.065
CBPS	0.233	0.235	0.210	0.154	0.086	0.131	0.148	0.144	0.112	0.067
OS TMLE	0.224	0.228	0.206	0.155	0.091	0.122	0.140	0.139	0.111	0.070
eKMM-surv	**0.029**	**0.036**	**0.042**	**0.047**	**0.052**	**0.038**	**0.050**	**0.060**	**0.064**	**0.060**
mKMM-surv	0.224	0.226	0.201	0.148	0.083	0.129	0.145	0.141	0.109	0.065

TABLE B.3
SENSITIVITY ANALYSIS FOR BIAS OF THE ATT ESTIMATION ON SIMULATION DATA

	$\sigma = 0.7$					$\sigma = 0.3$				
Estimator	t_1	t_2	t_3	t_4	t_5	t_1	t_2	t_3	t_4	t_5
(i) $\gamma = 1\text{e}-4$										
Crude	0.238	0.240	0.213	0.155	0.083	0.135	0.152	0.147	0.113	0.062
AKME	0.226	0.228	0.203	0.148	0.078	0.124	0.140	0.136	0.104	0.057
OW	0.222	0.224	0.200	0.146	0.078	0.122	0.138	0.134	0.103	0.056
CBPS	0.231	0.233	0.207	0.151	0.080	0.128	0.145	0.140	0.107	0.059
OS TMLE	0.222	0.226	0.204	0.152	0.084	0.119	0.137	0.135	0.106	0.061
eKMM-surv	−0.069	−0.088	−0.098	−0.085	−0.052	−0.079	−0.112	−0.136	−0.132	−0.086
mKMM-surv	**−0.044**	**−0.049**	**−0.048**	**−0.036**	**−0.018**	**−0.042**	**−0.055**	**−0.060**	**−0.053**	**−0.031**
(ii) $\gamma = 1\text{e}-3$										
Crude	0.238	0.240	0.213	0.155	0.083	0.135	0.152	0.147	0.113	0.062
AKME	0.226	0.228	0.203	0.148	0.078	0.124	0.140	0.136	0.104	0.057
OW	0.222	0.224	0.200	0.146	0.078	0.122	0.138	0.134	0.103	0.056
CBPS	0.231	0.233	0.207	0.151	0.080	0.128	0.145	0.140	0.107	0.059
OS TMLE	0.222	0.226	0.204	0.152	0.084	0.119	0.137	0.135	0.106	0.061
eKMM-surv	−0.012	−0.020	−0.025	−0.022	−0.015	**0.000**	**−0.001**	**0.000**	**0.000**	**0.000**
mKMM-surv	**0.000**	**0.005**	**0.009**	**0.013**	**0.012**	−0.003	**−0.001**	0.003	0.006	0.007
(iii) $\gamma = 1\text{e}-2$										
Crude	0.240	0.242	0.215	0.158	0.089	0.138	0.155	0.151	0.118	0.069
AKME	0.228	0.230	0.205	0.151	0.085	0.127	0.143	0.139	0.109	0.065
OW	0.224	0.227	0.203	0.150	0.084	0.125	0.141	0.138	0.108	0.065
CBPS	0.233	0.235	0.210	0.154	0.086	0.131	0.148	0.144	0.112	0.067
OS TMLE	0.224	0.228	0.206	0.155	0.091	0.122	0.140	0.139	0.111	0.070
eKMM-surv	0.039	0.048	0.056	0.062	0.067	0.036	0.043	0.052	0.059	**0.064**
mKMM-surv	**0.035**	**0.042**	**0.051**	**0.057**	**0.061**	**0.035**	**0.042**	**0.051**	**0.058**	**0.064**
(iv) $\gamma = 1\text{e}-1$										
Crude	0.238	0.240	0.213	0.155	0.083	0.135	0.152	0.147	0.113	0.062
AKME	0.226	0.228	0.203	0.148	0.078	0.124	0.140	0.136	0.104	0.057
OW	0.222	0.224	0.200	0.146	0.078	0.122	0.138	0.134	0.103	0.056
CBPS	0.231	0.233	0.207	0.151	0.080	0.128	0.145	0.140	0.107	0.059
OS TMLE	0.222	0.226	0.204	0.152	0.084	0.119	0.137	0.135	0.106	0.061
eKMM-surv	**0.002**	**0.001**	**0.000**	**0.001**	**0.005**	**0.007**	**0.005**	**0.003**	**−0.001**	**−0.001**
mKMM-surv	0.035	0.035	0.031	0.022	0.013	0.019	0.018	0.015	0.009	0.004
(v) $\gamma = 1$										
Crude	0.238	0.240	0.213	0.155	0.083	0.135	0.152	0.147	0.113	0.062
AKME	0.226	0.228	0.203	0.148	0.078	0.124	0.140	0.136	0.104	0.057
OW	0.222	0.224	0.200	0.146	0.078	0.122	0.138	0.134	0.103	0.056
CBPS	0.231	0.233	0.207	0.151	0.080	0.128	0.145	0.140	0.107	0.059
OS TMLE	0.222	0.226	0.204	0.152	0.084	0.119	0.137	0.135	0.106	0.061
eKMM-surv	**−0.008**	**−0.012**	**−0.015**	**−0.015**	**−0.009**	**−0.024**	**−0.035**	**−0.041**	**−0.041**	**−0.027**
mKMM-surv	0.223	0.224	0.199	0.144	0.077	0.127	0.142	0.137	0.104	0.057

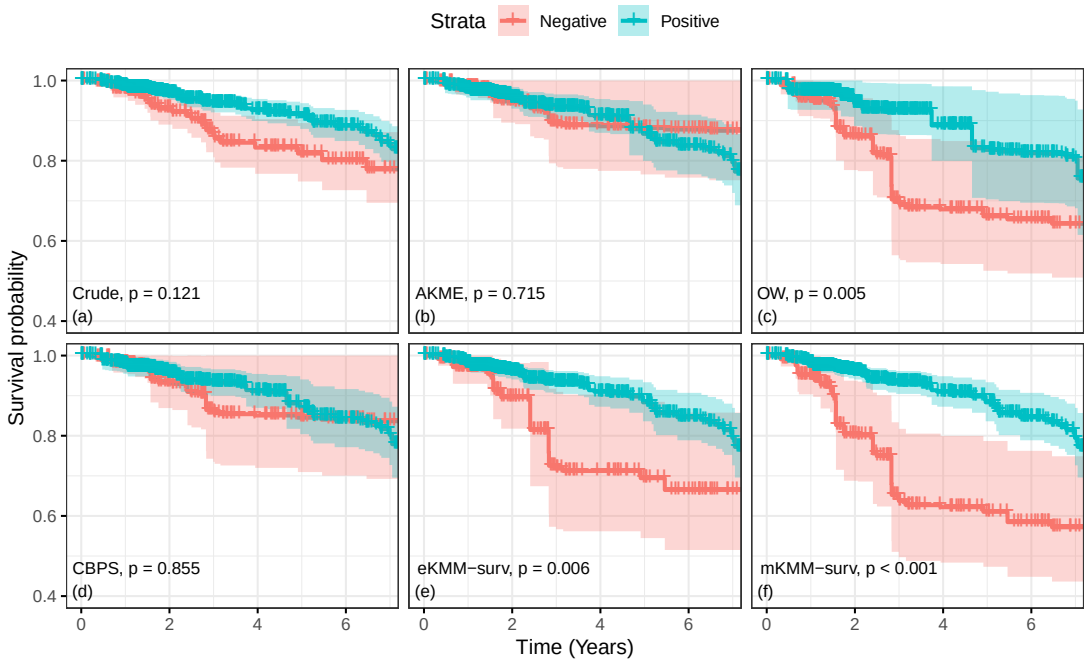

Fig. B.1. The adjusted KM curves for the endpoint of disease-specific survival with 95% confidence interval using different weighting methods. The hypothesis test method we use is the weighted log-rank test.

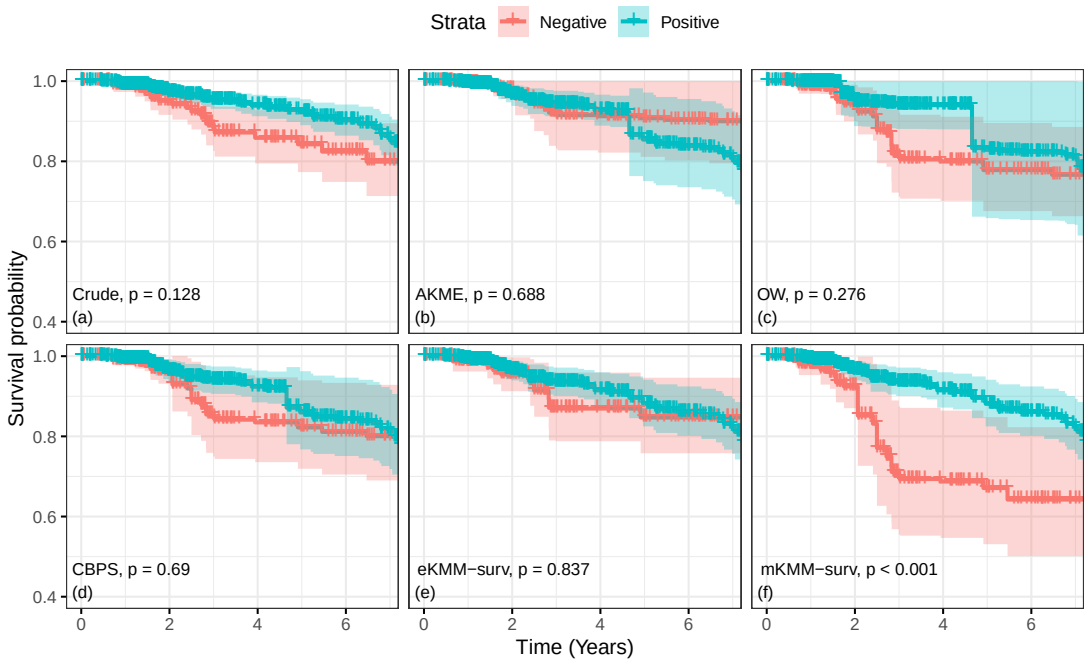

Fig. B.2. The adjusted KM curves for the endpoint of disease-free interval with 95% confidence interval using different weighting methods. The hypothesis test method we use is the weighted log-rank test.

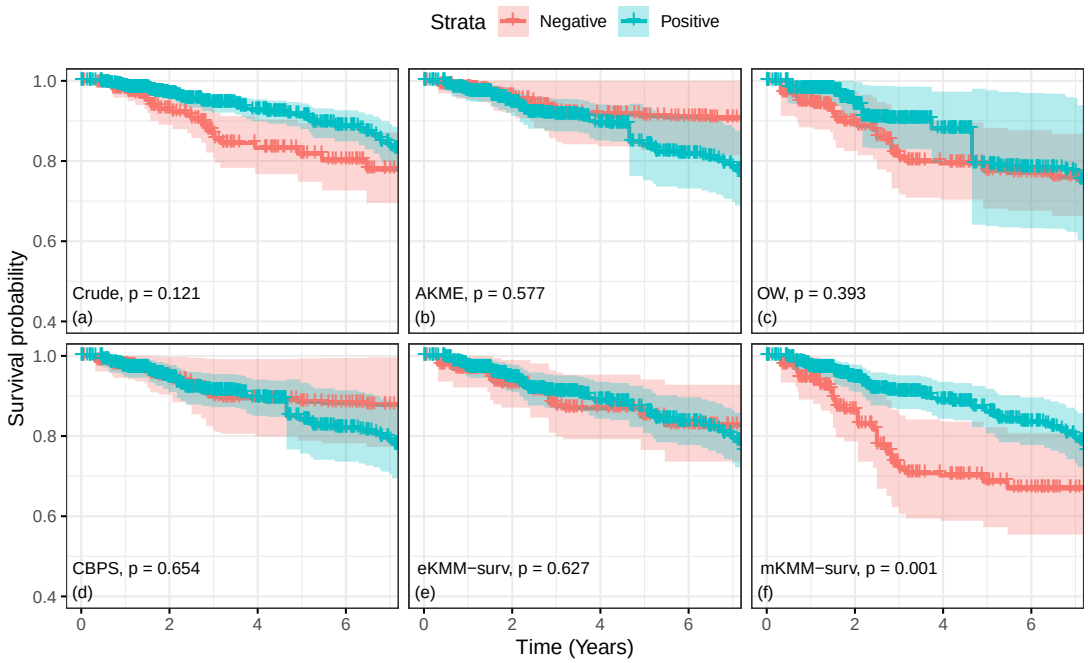

Fig. B.3. The adjusted KM curves for the endpoint of progression-free interval with 95% confidence interval using different weighting methods. The hypothesis test method we use is the weighted log-rank test.

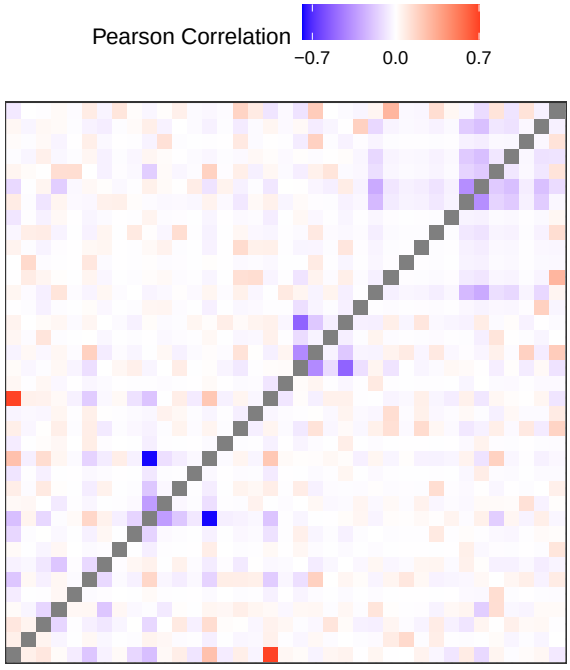

Fig. B.4. The heatmap for correlation matrix of the covariates (diagonal is suppressed). For multilevel categorical variable, we expand them to dichotomous dummy variables and then calculate the correlation coefficients.

APPENDIX C
RELATED WORK

Adjusted Kaplan–Meier Estimator (AKME) [2] based on Inverse Probability Weighting (IPW) was used extensively to adjust confounders for estimating KM curves and test differences of survival functions between groups. A similar work by Cole and Hernán [3] suggested using stabilized IPW for treatment groups to obtain nonconfounded survival curves and fit null Cox proportional hazard regression models. The IPW weights are applied on each individual. Specifically, the weights inverse to the propensity score are applied on the treatment individuals and that inverse to $(1 -$ propensity score$)$ are applied on the control individuals. The treatment individual whose probability of receiving treatment is large receives relatively small weight while the control individual whose probability of receiving treatment is large receives relatively large weight. This weighting strategy can reduce the difference of the covariates distribution between the treatment and control groups. The weighted pseudo population mimics a randomized experiment, including all the individuals from the available sample [4], [5]. However, the propensity scores when being close to 0 or 1 would result in extreme weights, leading to high variance of the estimates of causal effects [6]–[8].

To alleviate the issue of extreme weights, an OW based method [9] was recently proposed to estimate the counterfactual survival functions, which combine Inverse Probability of Censoring Weighting (IPCW) [10] and propensity score weighting. OW applies the propensity scores to the opposite group so that the weights are bounded within an interval of $[0, 1]$, which eradicates extreme weights encountered in AKME. Li et al. [11] proved theoretically in OW that the weighted means of the covariates in two groups are exactly identical, when the propensity scores are estimated under a logistic regression model. The above weighting methods all require the propensity score model to be correctly specified. Otherwise, substantial bias may be introduced to the causal effect estimation.

Covariate Balancing Propensity Score (CBPS) [12] was proposed to avoid misspecification of the propensity score model, which optimizes the covariates balancing while modeling treatment assignment. The CBPS method uses the dual characteristics of propensity score as the conditional probability of treatment assignment and covariates balancing score. CBPS have been used in numerous studies including survival analysis fields [13].

A recent presented one-step Targeted Maximum Likelihood Estimator (OS TMLE) was used to tackle the issue that the adjusted survival curves are not monotone [14]. The method they proposed is built upon the OS TMLE theory [15]. Due to the joint targeting, this powerful framework estimates the entire survival curve and respects the monotonically decreasing shape of the estimand.

APPENDIX D
SIMULATION DESIGN

The covariates vector follows a multivariate normal distribution $\boldsymbol{X}_i \sim \mathcal{N}(0, \boldsymbol{\Sigma})$, where the (r, c)th entry of the covariance matrix $\boldsymbol{\Sigma}$ is given by $\Sigma_{r,c} = \max(r, c) \cdot \sigma^{|r-c|}$. We consider two covariance matrix settings, $\sigma = 0.7, 0.3$, representing strong dimensional correlation and weak dimensional correlation, respectively. These two settings are verified to be positive definite. To mimic the complicated real-world data-generating process, we introduced the underlying unobserved covariates $\boldsymbol{X}_i^* = \left(X_{i,1}^*, \ldots, X_{i,p^*}^*\right)^\mathsf{T}$, where $X_{i,k}^*$ is the nonlinear transforms of specific dimensions of $\boldsymbol{X}_i$, given by $X_{i,k}^* = (X_{i,2k-1} + X_{i,2k} + 1)^2$ for $k = 1, \ldots, p^*$. The dimension of the unobserved covariates is determined by $p^* = p/2$, where p is set to $p = 10$ in this simulation experiment. The outcome and treatment assignment variables are generated by the unobserved covariates $\boldsymbol{X}_i^*$.

We assume that the time-to-event outcome generating process follows the Cox proportional hazards model, and the baseline hazard function is assumed to be constant. Given above, we generated a time-to-event outcome for each individual using a data-generating process described by Bender [16]: $T_i = -\log U \Big/ \lambda \exp\left(Z_i \beta_{\text{treat}} + \boldsymbol{X}_i^{*\mathsf{T}} \boldsymbol{\beta}\right)$, where $U \sim \text{Uniform}(0, 1)$ is a uniformly distributed random variable, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{p^*})^\mathsf{T}$ is the vector of regression coefficients of the unobserved covariates while in each dimension the coefficient is given by $\beta_k = 0.005k$ for $k = 1, \ldots, p^*$; β_{treat} is the regression coefficient of treatment indicator and $\lambda = 0.05$ is the constant baseline hazard function. By applying this data-generating process, uniformly distributed random variable U can be transformed into event times following a Cox-exponential model with the constant baseline hazard function λ. In this simulation, the treatment do not have any effect on the outcome, in other words, the causal effect of the treatment is 0. Consequently, the regression coefficient of the treatment indicator Z_i is given by $\beta_{\text{treat}} = 0$. Incorporating the censoring time C_i, which is considered to be uniformly distributed and generated from $C_i \sim \text{Uniform}(0, 20)$, the possibly censoring event time $\widetilde{T}_i$ and the censoring indicator δ_i are finally obtained. Throughout this simulation, the censoring rate is around 30%.

To simulate the imbalanced covariates in the observational studies, the logit of the probability to be assigned to the treatment group π_i is linear with the unobserved covariates $\boldsymbol{X}_i^*$. Formally, the treatment assignment generating procedure is given by $Z_i \sim \text{Bernoulli}(\pi_i)$, where $\pi_i = \text{logit}^{-1}\left(\boldsymbol{X}_i^{*\mathsf{T}} \boldsymbol{\alpha} + \alpha_0\right)$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{p^*})^\mathsf{T}$ with each dimension $\alpha_k = 0.003k$. α_0

represents the intercept which adjust the probability to roughly equate the numbers of the two groups. Under the two correlation settings, the intercepts are set to $\alpha_0 = 1, 0.8$, respectively.

For each settings above, we generate 1000 independent replications with the sample size of 1000. Throughout this simulation experiment, for mKMM-surv, the kernel function we select is $\exp\left[-\gamma\left(\boldsymbol{x}-\boldsymbol{x}'\right)^{\mathsf{T}}\boldsymbol{S}^{-1}\left(\boldsymbol{x}-\boldsymbol{x}'\right)\right]$ which is a generalization of the usually used Gaussian kernel, where γ is the kernel size. For eKMM-surv, we use the Gaussian kernel $\exp\left(-\gamma\left\|\boldsymbol{x}-\boldsymbol{x}'\right\|^2\right)$. Since the choice of kernel size straightforwardly affects the balancing performance, we conduct sensitivity analysis to show the scope of capability of the proposed method. We test a set of kernel sizes for both methods above, which are $\gamma = 1\mathrm{e}-4, 1\mathrm{e}-3, 1\mathrm{e}-2, 1\mathrm{e}-1, 1$.

REFERENCES

[1] J. Huang, A. Smola, A. Gretton, K. Borgwardt, and B. Schölkopf, "Correcting sample selection bias by unlabeled data," Tech. Rep. CS-2006-44, University of Waterloo, Waterloo, ON, Canada, 2006.

[2] J. Xie and C. Liu, "Adjusted kaplan–meier estimator and log-rank test with inverse probability of treatment weighting for survival data," *Statistics in Medicine*, vol. 24, no. 20, pp. 3089–3110, 2005.

[3] S. R. Cole and M. A. Hernán, "Adjusted survival curves with inverse probability weights," *Computer Methods and Programs in Biomedicine*, vol. 75, no. 1, pp. 45–49, 2004.

[4] F. Li, L. E. Thomas, and F. Li, "Addressing extreme propensity scores via the overlap weights," *American Journal of Epidemiology*, vol. 188, no. 1, pp. 250–257, 2018.

[5] L. Thomas, F. Li, and M. Pencina, "Using propensity score methods to create target populations in observational clinical research," *JAMA*, vol. 323, no. 5, pp. 466–467, 2020.

[6] J. M. Robins, A. Rotnitzky, and L. P. Zhao, "Analysis of semiparametric regression models for repeated outcomes in the presence of missing data," *Journal of the American Statistical Association*, vol. 90, no. 429, pp. 106–121, 1995.

[7] D. O. Scharfstein, A. Rotnitzky, and J. M. Robins, "Adjusting for nonignorable drop-out using semiparametric nonresponse models," *Journal of the American Statistical Association*, vol. 94, no. 448, pp. 1096–1120, 1999.

[8] J. D. Y. Kang and J. L. Schafer, "Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data," *Statistical Science*, vol. 22, no. 4, pp. 523–539, 2007.

[9] C. Cheng, F. Li, L. E. Thomas, and F. F. Li, "Addressing extreme propensity scores in estimating counterfactual survival functions via the overlap weights," *American Journal of Epidemiology*, 2022.

[10] A. E. Hubbard, M. J. van der Laan, and J. M. Robins, "Nonparametric locally efficient estimation of the treatment specific survival distribution with right censored data and covariates in observational studies," in *Statistical Models in Epidemiology, the Environment, and Clinical Trials* (M. E. Halloran and D. Berry, eds.), (New York, NY), pp. 135–177, Springer New York, 2000.

[11] F. Li, K. L. Morgan, and A. M. Zaslavsky, "Balancing covariates via propensity score weighting," *Journal of the American Statistical Association*, vol. 113, no. 521, pp. 390–400, 2018.

[12] K. Imai and M. Ratkovic, "Covariate balancing propensity score," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 76, no. 1, pp. 243–263, 2014.

[13] C. Chao, Y. Qian, X. Li, C. Sang, B. Wang, and X. ying Zhang, "Surgical survival benefits with different metastatic patterns for stage iv extrathoracic metastatic non-small cell lung cancer: A seer-based study," *Technology in Cancer Research & Treatment*, vol. 20, p. 15330338211033064, 2021.

[14] W. Cai and M. J. van der Laan, "One-step targeted maximum likelihood estimation for time-to-event outcomes," *Biometrics*, vol. 76, no. 3, pp. 722–733, 2020.

[15] M. van der Laan and S. Gruber, "One-step targeted minimum loss-based estimation based on universal least favorable one-dimensional submodels," *The International Journal of Biostatistics*, vol. 12, no. 1, pp. 351–378, 2016.

[16] R. Bender, T. Augustin, and M. Blettner, "Generating survival times to simulate cox proportional hazards models," *Statistics in Medicine*, vol. 24, no. 11, pp. 1713–1723, 2005.