WEB MATERIAL

Improving Propensity Score Estimators' Robustness to Model Misspecification Using Super Learner

Romain Pirracchio*, Maya L. Petersen, and Mark van der Laan

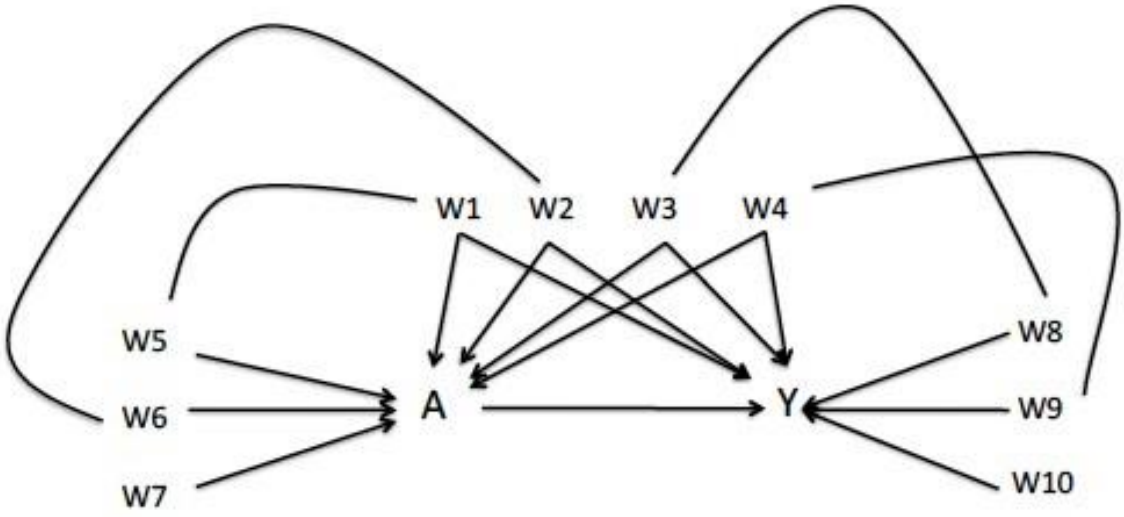
*Correspondence to Dr. Romain Pirracchio, Service d'Anesthésie-Réanimation, Hôpital Européen Georges Pompidou, 20 rue Leblanc, 75015 Paris, France (e-mail: romainpirracchio@yahoo.fr).

Web Table 1. Coefficients for Data Generation (Upper P anel) and Correlation Matrix (Lower Panel)

	True PS Model	Outcome Model
Intercept	0	−3.85
Coefficient 1	0.8	0.3
Coefficient 2	−0.25	−0.36
Coefficient 3	0.6	−0.73
Coefficient 4	−0.4	−0.2
Coefficient 5	−0.8	0.71
Coefficient 6	−0.5	−0.19
Coefficient 7	0.7	0.26

		Confounders				**Exposure Predictors**			**Outcome Predictor**		
		*X*1	*X*2	*X*3	*X*4	*X*5	*X*6	*X*7	*X*8	*X*9	*X*10
Confounders	*X*1	1									
	*X*2	0	1								
	*X*3	0	0	1							
	*X*4	0	0	0	1						
Exposure predictors	*X*5	0.2	0	0	0	1					
	*X*6	0	0.9	0	0	0	1				
	*X*7	0	0	0	0	0	0	1			
Outcome predictor	*X*8	0	0	0.2	0	0	0	0	1		
	*X*9	0	0	0	0.9	0	0	0	0	1	
	*X*10	0	0	0	0	0	0	0	0	0	1

Web Figure 1. Causal graph.



Web Appendix

R Code for Super Learner-Based Propensity Score Matching

```
require(SuperLearner) require(Matching)

outc_SL<-baz$z.a # z.a: outcome; baz: dataset expl_SL<-
(baz[,c(1,2,3,4,5,6,7,8,9,10)]) # explanatory variables

# PS modelling using the Super Learner
        # 10-fold cross-validation
        # loss function: non negative least square

loss function = non-negative least square
cv.fit <- CV.SuperLearner(Y = outc_SL, X = expl_SL, V=10, family="binomial",
SL.library=SL.library, verbose=FALSE,
cvControl=list(stratifyCV=TRUE,shuffle=TRUE,V=10),method = "method.NNLS")

predictions <- cbind(cv.fit$SL.predict,cv.fit$library.predict) ps<-predictions[,1]

# PS matching
match<-Match(Y=baz$y.a, Tr=baz$z.a, X=baz$score, caliper=0.2,
M=1,estimand="ATE",ties=TRUE,version="standard",replace=TRUE)

# Balance diagnosis
## Full Balance Diagnosis for matching

Balance<-MatchBalance(z.a ~ w1+w2+w3+w4+w5+w6+w7+w8+w9+w10,
                match.out = match, digits=4, paired=T, data=baz)

sdiff.before<-sdiff.1.after<-NULL for (i in 1:10)
{
   sdiff.before[i]<-Balance$BeforeMatching[[i]]$sdiff sdiff.after[i]<-
   Balance$AfterMatching[[i]]$sdiff
}
```

Web Table 2. Covariate Balance Diagnosis (*X*1, *X*2, *X*3, *X*4: Confounders; *X*5, *X*6, *X*7: Exposure Predictors; *X*8, *X*9, *X*10: Outcome Predictors)

Scenario A

	Naive	Logit Matching	SL Matching	Logit IPTW	SL IPTW
*X*1	28.4	0.5	5.6	−0.1	4.1
*X*2	−35.2	−0.1	−5.2	0.1	−3.4
*X*3	23.7	−0.2	5.6	−0.1	4.4
*X*4	−32.4	−0.3	−4.9	0.1	−3.4
*X*5	−28.1	−0.2	−5.7	0.1	−4.1
*X*6	−35.3	0.2	−5.2	0.1	−3.6
*X*7	58.5	0.8	8.2	0.4	4.7
*X*8	2.8	−0.1	0.4	−0.1	0.3
*X*9	−23.2	0.1	−2.7	0.2	−1.4
*X*10	0.3	0.2	−0.2	0.1	−0.1

Scenario B

	Naive	Logit Matching	SL Matching	Logit IPTW	SL IPTW
*X*1	26.1	0.1	2.5	−0.2	4.3
*X*2	−33.1	0.6	0.2	1.0	−3.0
*X*3	22.5	0.1	2.5	−0.2	4.4
*X*4	−30.4	0.7	−0.7	1.0	−3.0
*X*5	−25.9	−0.2	−2.5	0.4	−4.3
*X*6	−31.6	−0.1	−1.2	0.5	−3.4
*X*7	34.8	0.5	0.3	−3.1	6.3
*X*8	3.2	0.1	0.1	−0.1	0.2
*X*9	−20.3	0.2	0.6	0.5	−1.4
*X*10	0.5	0.1	−0.1	−0.1	−0.1

Scenario C

	Naive	Logit Matching	SL Matching	Logit IPTW	SL IPTW
*X*1	43.0	1.2	6.7	0.3	4.7
*X*2	−40.0	0.5	−5.2	2.0	−2.7
*X*3	37.2	0.8	6.2	0.1	4.3
*X*4	−37.2	0.5	−4.6	2.0	−2.7
*X*5	−25.8	0.1	−5.4	0.7	−4.0
*X*6	−38.8	0.8	−4.7	1.5	−2.7
*X*7	38.4	1.4	6.5	1.0	4.4
*X*8	4.4	−0.1	0.4	−0.1	0.5
*X*9	−27.2	0.5	−2.1	1.4	−0.6
*X*10	−0.3	0.1	−0.1	0.1	−0.2

Scenario D

	Naive	Logit Matching	SL Matching	Logit IPTW	SL IPTW
*X*1	59.6	9.1	3.7	1.7	12.8
*X*2	−35.5	8.0	−1.6	21.9	−4.0
*X*3	49.7	6.9	5.1	0.5	12.5
*X*4	−35.5	12.5	2.9	21.9	−4.0
*X*5	−18.1	5.0	−0.9	9.7	−3.2
*X*6	−32.9	7.0	1.0	13.0	−3.9
*X*7	−3.4	0.7	−0.2	2.5	−0.8
*X*8	6.1	0.4	−0.4	−0.3	1.1
*X*9	−25.7	7.0	4.0	12.9	−1.4
*X*10	−0.3	−0.4	−0.1	−0.1	−0.2

Abbreviations: IPTW, inverse probability of treatment weighting; SL, Super Learner.