### Table S1

**Misspecified Nonlinear Propensity Score Models: Low-Degree Nonlinearity**

<table>
<thead>
<tr>
<th>Model</th>
<th>(N = 5,000)</th>
<th>(N = 800)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(RMPW)</td>
<td>(NRMPW\ 3 \times 3)</td>
</tr>
<tr>
<td><strong>Direct Effect Estimate</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Bias removal (a)</td>
<td>0.9106</td>
<td>0.8525</td>
</tr>
<tr>
<td>(b)</td>
<td>0.8842</td>
<td>0.8671</td>
</tr>
<tr>
<td>(c)</td>
<td>0.9160</td>
<td>0.8525</td>
</tr>
<tr>
<td>Relative efficiency (a)</td>
<td>0.9335</td>
<td>0.9529</td>
</tr>
<tr>
<td>(b)</td>
<td>1.0293</td>
<td>1.0695</td>
</tr>
<tr>
<td>(c)</td>
<td>0.8780</td>
<td>0.9628</td>
</tr>
<tr>
<td><strong>MSE</strong> (a)</td>
<td>0.0023</td>
<td>0.0030</td>
</tr>
<tr>
<td>(b)</td>
<td>0.0042</td>
<td>0.0042</td>
</tr>
<tr>
<td>(c)</td>
<td>0.0067</td>
<td>0.0099</td>
</tr>
<tr>
<td><strong>Indirect Effect Estimate</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Bias removal (a)</td>
<td>0.9092</td>
<td>0.8511</td>
</tr>
<tr>
<td>(b)</td>
<td>0.8731</td>
<td>0.8562</td>
</tr>
<tr>
<td>(c)</td>
<td>0.9104</td>
<td>0.8472</td>
</tr>
<tr>
<td>Relative efficiency (a)</td>
<td>1.5800</td>
<td>1.8114</td>
</tr>
<tr>
<td>(b)</td>
<td>0.6741</td>
<td>0.7422</td>
</tr>
<tr>
<td>(c)</td>
<td>0.7978</td>
<td>0.9910</td>
</tr>
<tr>
<td><strong>MSE</strong> (a)</td>
<td>0.0006</td>
<td>0.0013</td>
</tr>
<tr>
<td>(b)</td>
<td>0.0026</td>
<td>0.0026</td>
</tr>
<tr>
<td>(c)</td>
<td>0.0045</td>
<td>0.0079</td>
</tr>
<tr>
<td>Model</td>
<td>( N = 5,000 )</td>
<td>( N = 800 )</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>( \text{RMPW} )</td>
<td>( \text{NRMPW 3×3} )</td>
</tr>
<tr>
<td><strong>Direct Effect Estimate</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Bias removal</td>
<td>(a) 0.8412</td>
<td>0.8404</td>
</tr>
<tr>
<td></td>
<td>(b) 0.8029</td>
<td>0.8608</td>
</tr>
<tr>
<td></td>
<td>(c) 0.8703</td>
<td>0.8400</td>
</tr>
<tr>
<td>Relative efficiency</td>
<td>(a) 0.9688</td>
<td>0.9674</td>
</tr>
<tr>
<td></td>
<td>(b) 1.0423</td>
<td>1.0708</td>
</tr>
<tr>
<td></td>
<td>(c) 0.9459</td>
<td>0.9987</td>
</tr>
<tr>
<td><strong>MSE</strong></td>
<td>(a) 0.0028</td>
<td>0.0028</td>
</tr>
<tr>
<td></td>
<td>(b) 0.0049</td>
<td>0.0041</td>
</tr>
<tr>
<td></td>
<td>(c) 0.0082</td>
<td>0.0097</td>
</tr>
<tr>
<td><strong>Indirect Effect Estimate</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Bias removal</td>
<td>(a) 0.8397</td>
<td>0.8389</td>
</tr>
<tr>
<td></td>
<td>(b) 0.7900</td>
<td>0.8470</td>
</tr>
<tr>
<td></td>
<td>(c) 0.8656</td>
<td>0.8355</td>
</tr>
<tr>
<td>Relative efficiency</td>
<td>(a) 1.5569</td>
<td>1.7352</td>
</tr>
<tr>
<td></td>
<td>(b) 0.6859</td>
<td>0.7706</td>
</tr>
<tr>
<td></td>
<td>(c) 0.9208</td>
<td>1.0569</td>
</tr>
<tr>
<td><strong>MSE</strong></td>
<td>(a) 0.0012</td>
<td>0.0012</td>
</tr>
<tr>
<td></td>
<td>(b) 0.0035</td>
<td>0.0026</td>
</tr>
<tr>
<td></td>
<td>(c) 0.0060</td>
<td>0.0076</td>
</tr>
<tr>
<td>Model</td>
<td>N = 5,000</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-----------</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>RMPW</td>
<td>NRMPW 3×3</td>
</tr>
<tr>
<td>Direct Effect Estimate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Bias removal</td>
<td>(a)</td>
<td>0.9393</td>
</tr>
<tr>
<td></td>
<td>(b)</td>
<td>0.9249</td>
</tr>
<tr>
<td></td>
<td>(c)</td>
<td>0.9263</td>
</tr>
<tr>
<td>Relative efficiency</td>
<td>(a)</td>
<td>0.9394</td>
</tr>
<tr>
<td></td>
<td>(b)</td>
<td>1.0228</td>
</tr>
<tr>
<td></td>
<td>(c)</td>
<td>0.8877</td>
</tr>
<tr>
<td>MSE</td>
<td>(a)</td>
<td>0.0021</td>
</tr>
<tr>
<td></td>
<td>(b)</td>
<td>0.0038</td>
</tr>
<tr>
<td></td>
<td>(c)</td>
<td>0.0051</td>
</tr>
<tr>
<td>Indirect Effect Estimate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Bias removal</td>
<td>(a)</td>
<td>0.9378</td>
</tr>
<tr>
<td></td>
<td>(b)</td>
<td>0.9156</td>
</tr>
<tr>
<td></td>
<td>(c)</td>
<td>0.9231</td>
</tr>
<tr>
<td>Relative efficiency</td>
<td>(a)</td>
<td>1.4852</td>
</tr>
<tr>
<td></td>
<td>(b)</td>
<td>0.6531</td>
</tr>
<tr>
<td></td>
<td>(c)</td>
<td>0.7640</td>
</tr>
<tr>
<td>MSE</td>
<td>(a)</td>
<td>0.0005</td>
</tr>
<tr>
<td></td>
<td>(b)</td>
<td>0.0022</td>
</tr>
<tr>
<td></td>
<td>(c)</td>
<td>0.0029</td>
</tr>
</tbody>
</table>
### Table S4

Misspecified Non-additive Propensity Score Models: Moderate-Degree Non-additivity

<table>
<thead>
<tr>
<th>Model</th>
<th>( N = 5,000 )</th>
<th>( N = 800 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( RMPW )</td>
<td>( NRMPW \ 3\times3 )</td>
</tr>
<tr>
<td><strong>Direct Effect Estimate</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Bias removal</td>
<td>0.8757</td>
<td>0.8618</td>
</tr>
<tr>
<td>(b)</td>
<td>0.8439</td>
<td>0.8816</td>
</tr>
<tr>
<td>(c)</td>
<td>0.8500</td>
<td>0.8588</td>
</tr>
</tbody>
</table>

| Relative efficiency | 1.2890 | 0.9565 | 0.9600 | 0.9291 | 0.9051 | 0.8900 |
| (a) | 0.6931 | 1.0681 | 1.0810 | 1.0309 | 0.9939 | 0.9773 |
| (b) | 0.8917 | 0.9646 | 0.9464 | 0.9439 | 0.8747 | 0.8455 |

| **MSE** | 0.0026 | 0.0027 | 0.0023 | 0.0116 | 0.0120 | 0.0120 |
| (a) | 0.0045 | 0.0041 | 0.0037 | 0.0212 | 0.0214 | 0.0220 |
| (b) | 0.0077 | 0.0073 | 0.0055 | 0.0258 | 0.0273 | 0.0292 |

| **Indirect Effect Estimate** | | | | | | |
| % Bias removal | 0.8742 | 0.8604 | 0.9064 | 0.8718 | 0.8611 | 0.8725 |
| (a) | 0.8292 | 0.8662 | 0.9118 | 0.8262 | 0.8661 | 0.8489 |
| (b) | 0.8479 | 0.8566 | 0.8997 | 0.8419 | 0.8450 | 0.8269 |

| Relative efficiency | 1.5349 | 1.7647 | 1.6173 | 1.2890 | 1.2282 | 1.0362 |
| (a) | 0.7229 | 0.7570 | 0.7467 | 0.6931 | 0.6949 | 0.6848 |
| (b) | 0.9438 | 0.9704 | 0.9262 | 0.8917 | 0.7942 | 0.7436 |

| **MSE** | 0.0009 | 0.0011 | 0.0006 | 0.0024 | 0.0026 | 0.0028 |
| (a) | 0.0030 | 0.0025 | 0.0021 | 0.0117 | 0.0112 | 0.0116 |
| (b) | 0.0056 | 0.0051 | 0.0032 | 0.0132 | 0.0140 | 0.0158 |
Online Supplement of Stata code for RMPW Analyses

I. Parametric RMPW analysis for a binary mediator using Generalized Method of Moments (GMM) to account for estimation error in the weights

Note: the procedure below does not exclude observations for which there is no common support.

*** Generate a constant to be used in the GMM command

generate cons = 1

*** Specify moment equations, storing them in “locals”
*** deltaC is the control group mean outcome; deltaE the experimental group mean outcome
*** deltaStar1 is the counterfactual mean outcome of the experimental group
*** deltaStar0 is the counterfactual mean outcome of the control group

local equation1 ( Z - (( 1 / ( 1 + \exp(-{xb1: X1 X2 X3 X4 X5 X6 X7 X8 X9 cons}) )))) * A
local equation2 ( Z - (( 1 / ( 1 + \exp(-{xb2: X1 X2 X3 X4 X5 X6 X7 X8 X9 cons}) )))) * (1 - A)
local equation3 ( Y - {deltaC} ) * ( 1 - A)
local equation4 ( Y - {deltaE} ) * A
local equation5 ( Y - {deltaStar1} ) * ///
( (( Z * (( 1 / ( 1 + \exp(-{xb2:})) ))/( 1 / ( 1 + \exp(-{xb1:})) )))) + ((1-Z) * ///
(\exp(-{xb2:}) / ( 1 + (\exp(-{xb2:})) ))) / ( \exp(-{xb1:}) / ( 1 + (\exp(-{xb1:})))))) ) * A
local equation6 ( Y - {deltaStar0} ) * ///
( (( Z * (( 1 / ( 1 + \exp(-{xb1:})) ))/( 1 / ( 1 + \exp(-{xb2:})) )))) + ((1-Z) * ///
(\exp(-{xb1:}) / ( 1 + (\exp(-{xb1:})) ))) / ( \exp(-{xb2:}) / ( 1 + (\exp(-{xb2:})))))) ) * (1 – A)

*** Specify “instruments,” storing them in locals

local equation1inst X1 X2 X3 X4 X5 X6 X7 X8 X9
local equation2inst X1 X2 X3 X4 X5 X6 X7 X8 X9

*** Execute GMM command

gmm (eq1: `equation1') (eq2: `equation2') (eq3: `equation3') (eq4: `equation4') (eq5: `equation5') ///
(eq6: `equation6'), instruments(eq1: `equation1inst') instruments(eq2: `equation2inst')
instruments(eq3: ) instruments(eq4: ) instruments(eq5: ) instruments(eq6: ) winitial(identity) onestep

*** Estimate Natural Indirect Effect

lincom _b[/deltaE] - _b[/deltaStar1]

*** Estimate Natural Direct Effect

lincom _b[/deltaStar1] - _b[/deltaC]
***Estimate Pure Indirect Effect

\[ \text{lincom } \_b[/\delta Star0] - \_b[/\delta C] \]

*** Estimate Treatment-by-Mediator Interaction Effect

\[ \text{lincom } (_b[/\delta E] - _b[/\delta Star1]) - (_b[/\delta Star0] - _b[/\delta C]) \]
II. Parametric RMPW analysis for a binary mediator

Note: this code does not account for estimation error in the RMPW weights, but does exclude observations due to lack of common support.

*** Run binary logit for the control group

logit Z X1 X2 X3 X4 X5 X6 X7 X8 X9 if A==0

*** Generate predicted probability, that is pr(Z=1|X, A=0), for both experimental and control groups.

* Note that this will be an in-sample prediction for those in the control group
* and an out-of-sample prediction for those in the experimental group.

predict p0, pr
predict xb0, xb

*** Run binary logit for the experimental group

logit Z X1 X2 X3 X4 X5 X6 X7 X8 X9 if A==1

*** Generate predicted probability, that is pr(Z=1|X, A=1). Note that this will be an
* in-sample prediction for those in the experimental group
* and an out-of-sample prediction for those in the control group.

predict p1, pr
predict xb1, xb

*** Variables xb0 and xb1 are the logit scores of the respective propensity models for
* the experimental and control groups.

* Loop over logit scores (a= 0, 1), treatment groups (i= 0, 1), and mediator values (j=0, 1)

forvalues a=0(1)1  {

* Calculate the standard deviation of each logit score, to be used below.

qui sum xb`a'
sca sd`a’=r(sd)

forvalues i=0(1)1  {

forvalues j=0(1)2  {
qui sum xb`a' if A==`i' & Z==`j'

* Calculate the "minimum" and “maximum” of each logit score for each treatment-by-mediator group.
* Where the "maximum"("minimum") is actually 20% of a standard deviation of the logit score
* above (below) the actual maximum (minimum).

sca max`a'`i'`j'=(max) + .2*sd`a'
sca min`a'`i'`j'=(min) - .2*sd`a'
}
}
}

*** Generate an "exclude" indicator

* Loop over each logit score.

gen exclude=0

forvalues a=0(1)1  {
replace exclude=1 if xb`a'<max(min`a'00, min`a'01, min`a'10, min`a'11)
replace exclude=1 if xb`a'>min(max`a'00, max`a'01, max`a'10, max`a'11)
}

*** Generate parametric RMPW

gen rmpw=1 if exclude==0

replace rmpw=p0/p1 if A==1 & Z==1 & exclude==0
replace rmpw=(1-p0)/(1-p1) if A==1 & Z==0 & exclude==0

*** Generate a unique identifier, called "obs", for each person. This will allow
* duplicates to have the same identifier, which will be necessary for obtaining the correct
* standard errors.

gen obs=_n

*** Generate duplicate observations for the experimental group, where D1 is the indicator
* for duplicate. D1=0 for all control group observations and original experimental group
* observations.

expand 2 if A==1, gen(D1)

*** Make sure duplicates get a weight=1. Note that duplicate observations receive a different * weight than their original.

replace rmpw=1 if D1==1 & exclude==0

*** Outcome model.

* This command weights each observation and clusters standard errors at the person level, * adjusting for correlation in errors within each set of duplicates.

*** Optional adjustment for covariate X1, centered at its sample mean, for improving precision.

reg Y A D1 X1 [pweight=rmpw], vce(cluster obs)

*** To decompose natural indirect effect into the pure indirect effect and the natural *treatment-by-mediator interaction effect

*Create a duplicate set of the control group, which will be weighted

expand 2 if A==0, gen(D0)

* Generate a new set of weights for the duplicate control group

replace rmpw = p1/p0   if  A==0 & Z==1 & D0==1 & exclude==0
replace rmpw = (1-p1)/(1-p0)   if  A==0 & Z==0 & D0==1 & exclude==0

*** Outcome model to estimate the pure indirect effect and the natural treatment-by-mediator * interaction effect

*** Optional adjustment for covariate X1, centered at its sample mean, for improving precision.

reg Y A D1 D0 X1 [pweight=rmpw], vce(cluster obs)

lincom D1 – D0

*** The coefficient for D0 represents the pure indirect effect. The coefficient for D1 represents * the total indirect effect.

*** The post-estimation command estimates, and does a significance test on, the natural * treatment-by-mediator interaction effect, which is the total indirect effect less the pure indirect * effect.
III. Nonparametric RMPW analysis for a binary mediator

Note: this code does not account for estimation error in the RMPW weights, but does exclude observations due to lack of common support.

*** Run binary logit for the control group.

logit Z X1 X2 X3 X4 X5 X6 X7 X8 X9 if A==0

*** Generate logit score (not probability), which will be used later to create a categorical * variable used in creating nonparametric weights.

predict xb0, xb

*** Run binary logit for the experimental group.

logit Z X1 X2 X3 X4 X5 X6 X7 X8 X9 if A==1

*** Generate logit score, which will be used later to create a categorical variable used in
* creating nonparametric weights.

predict xb1, xb

*** Identify common support: the code is the same as that under parametric analysis

*** Generate 3×3 nonparametric RMPW

*** Place all observations into three equal-sized categories based on their logit score
* from the experimental group model.

* Generate categorical variable h1 = (0, 1, 2) based on terciles in xb1.

egen h1=cut(xb1), group(3)

*** Within each category of h1, generate categorical variables (h00, h01, h02) = (0, 1, 2)
* based on terciles in xb0.

* Loop through each category of h1.

forvalues j=0(1)2  {

egen h0\'j\'=cut(xb0) if h1==\'j\', group(3)
*** Generate a strata variable to place each observation into one of 9 strata, * based on joint distribution of xb0 and xb1.

gen strata=.
replace strata=0 if h1==0 & h00==0
replace strata=1 if h1==0 & h00==1
replace strata=2 if h1==0 & h00==2
replace strata=3 if h1==1 & h01==0
replace strata=4 if h1==1 & h01==1
replace strata=5 if h1==1 & h01==2
replace strata=6 if h1==2 & h02==0
replace strata=7 if h1==2 & h02==1
replace strata=8 if h1==2 & h02==2

*** Calculate probabilities P(Z=1| A, strata). “Prij” is the probability that Z=1 in treatment * group i and strata j. Loop over treatment groups (i = 0, 1) and strata (j = 0, 1, . . . , 8).

forvalues i=0(1)1  {
    forvalues j=0(1)8  {
        qui sum Z if A==`i' & strata==`j' & exclude==0
        sca pr`i'`j'=r(mean)
    }
}

*** Generate nonparametric RMPW weights based on these calculated probabilities, * treatment group membership, strata membership, and Z.

gen nrmpw=1 if exclude==0

* Loop over strata categories (j= 0, 1, . . . , 8).
forvalues j=0(1)8  {
replace nrmpw = pr0`j'/pr1`j' if A==1 & strata==`j' & Z==1 & exclude==0
replace nrmpw = (1-pr0`j')/(1-pr1`j') if A==1 & strata==`j' & Z==0 & exclude==0
}

*** Use the same process here as in the parametric case to create a person-specific identifier * and generate duplicate observations.

gen obs=_n
expand 2 if A==1, gen(D1)
* Ensure duplicates receive a weight equal to 1.
replace nrmpw=1 if D1==1 & exclude==0

*** Outcome model.

*Command weights each observation and clusters standard errors at the person level, * adjusting for correlation in errors within each set of duplicates.

*** Optional adjustment for covariate X1, centered at its sample mean, for improving precision.

reg Y A D1 X1 [weight=nrmpw], vce(cluster obs)

*** To decompose the indirect effect into the pure indirect effect and the natural *treatment-by-mediator interaction effect

*Create a duplicate set of the control group, which will be weighted
expand 2 if A==0, gen(D0)
* Generate new set of weights for the duplicate control group
* Loop over strata categories (j= 0, 1, . . . , 8).
forvalues j=0(1)8  {
replace nrmpw = pr1`j'/pr0`j' if A==0 & strata==`j' & Z==1 & D0==1 & exclude==0
replace nrmpw = (1-pr1`j')/(1-pr0`j') if A==0 & strata==`j' & Z==0 & D0==1 & exclude==0
}
*** Outcome model to estimate the pure indirect effect and the natural treatment-by-mediator interaction effect

*** Option adjustment for covariate X1, centered at its sample mean, for improving precision.

```
reg Y A D1 D0 X1 [pweight=nrmpw], vce(cluster obs)
lincom D1 – D0
```
IV. Parametric RMPW analysis for a three-category mediator

Note: this code does not account for estimation error in the RMPW weights, but does exclude observations due to lack of common support.

***** Run ordered logit  ***

*** Same as binary parametric analysis except that we have an ordered logit, where Z= 0, 1, 2 * with three predicted probabilities under each treatment.

ologit Z X1 X2 X3 X4 X5 X6 X7 X8 X9 if A==0
predict p00 p01 p02, pr
ologit Z X1 X2 X3 X4 X5 X6 X7 X8 X9 if A==1
predict p10 p11 p12, pr

*** Run ordered logit for each treatment group and generate logit scores.

ologit Z X1 X2 X3 X4 X5 X6 X7 X8 X9 if A==0
predict xbo0, xb
ologit Z X1 X2 X3 X4 X5 X6 X7 X8 X9 if A==1
predict xbo1, xb

* Loop over logit scores (a=0, 1), treatment groups “i” and mediator values “j”.

forvalues a=0(1)1  {

* Calculate the standard deviation of each logit score.

qui sum xbo`a'
sca sdo`a'=`r(sd)
forvalues i=0(1)1  {
forvalues j=0(1)2  {

* Calculate "minimums" and "maximums" for each treatment-by-mediator group as above.

qui sum xbo`a' if A==`i' & Z==`j'

sca max a'i'j'=r(max) + .2*sdo'a'

sca min a'i'j'=r(min) - .2*sdo'a'

*** Identify common support; generate an "exclude" indicator

* Loop over each logit score xbo “a”.

gen exclude3=0

forvalues a=0(1)1  {
replace exclude3=1 if xbo`a'>min(max`a'00, max`a'01, max`a'02, max`a'10, max`a'11, max`a'12)
replace exclude3=1 if xbo`a'<max(min`a'00, min`a'01, min`a'02, min`a'10, min`a'11, min`a'12)
}

*** Generate weight

gen rmpw3=1 if exclude3==0

forvalues j=0(1)2  {
replace rmpw3=p0`j'/p1`j' if A==1 & Z==`j' & exclude3==0
}

*** Create a person-specific identifier, generate duplicate LFA observations, and give duplicates a weight equal to 1.

gen obs=_n

expand 2 if A==1, gen(D1)

replace rmpw3=1 if D1==1 & exclude3==0

*** Outcome model

*** Optional adjustment for covariate X1, centered at its sample mean, for improving precision.
reg Y A D X1 [weight=rmpw3], vce(cluster obs)

*** To decompose natural indirect effect into the pure indirect effect and the natural treatment-by-mediator interaction effect

* Create a duplicate set of the control group, which will be weighted

expand 2 if A==0, gen(D0)

* Generate a new set of weights for the duplicate control group

forvalues j=0(1)2  {
replace rmpw3=p1’j’/p0’j’ if A==0 & Z==’j’ & exclude3==0
}

*** Outcome model to estimate the pure indirect effect and the natural treatment-by-mediator interaction effect

*** Optional adjustment for covariate X1, centered at its sample mean, for improving precision.

reg Y A D1 D0 X1 [pweight=rmpw], vce(cluster obs)

lincom D1 – D0

*** The coefficient for D0 represents the pure indirect effect. The coefficient for D1 represents the total indirect effect.

*** The post-estimation command estimates, and does a significance test on, the natural treatment-by-mediator interaction effect, which is the total indirect effect less the pure indirect effect.