A Simulation of Mediation, Confounding and Suppression Effects

by

David P. MacKinnon[1], Jennifer L. Krull[2], and Chondra M. Lockwood[1]

[1]Arizona State University

[2]University of Missouri

Simulation Addendum to MacKinnon, D. P., Krull, J. L, & Lockwood, C. M. (2002). Equivalence of the Mediation, Confounding and Suppression Effect. *Prevention Science, 1*(4), 173-181.

A simulation study was conducted to verify that tests for mediation are applicable in suppression and confounding models. Please see MacKinnon, Krull, and Lockwood (2002) for a full description of the equivalence of mediation, confounding and suppression effects and more complete model descriptions.

Methods

Simulation Description

The SAS® programming language was used to conduct the statistical simulations. The data were generated from the normal distribution (Box & Muller, 1958) transformation in the RANNOR function. The current time was used as the seed for each simulation. Seven different sample sizes of 25, 50, 100, 200, 500, 1000 and 5000 were chosen to reflect sample sizes in social science studies. Additionally, the independent variable in the simulation was treated in one of two ways: as a continuous variable or as a dichotomous variable (analogous to a treatment/no treatment distinction). Simulations for the continuous and dichotomous independent variable were identical except that the independent variable was dichotomized prior to the regression analysis in the latter case.

The parameters α and β which made up the third variable effect could each take on five different true values: $-\sqrt{.5}$, -.5, 0, .5, and $\sqrt{.5}$. The τ' parameter representing the direct effect could take on seven true values: $-\sqrt{.5}$, -.5, -.25, 0, .25, .5, and $\sqrt{.5}$. The parameter values were selected to correspond to values commonly observed in research studies and roughly correspond to small (.5 times .5), medium (.5 times $\sqrt{.5}$), and large ($\sqrt{.5}$ times $\sqrt{.5}$) effect sizes. The same parameter values were used for the case of a dichotomous independent variable, so the effect sizes are considerably smaller due to the reduced effect size when a variable is dichotomized (Cohen, 1983). The five values of α, five values of β and seven values of τ' parameter resulted in 175 different combinations of parameter values. Of the 175 combinations, 48 corresponded to consistent mediation or positive confounding models, 48 corresponded to inconsistent mediation or suppression models, and 79 corresponded to models where one of the three parameters equaled zero which we call the zero models. These zero models include situations of mediation or suppression depending on the sampling variability of the path equal to zero and the sign of the parameters in the model.

The population value of the variance associated with the continuous independent variable and the error variances associated with the mediator and the outcome variable were all set equal to 1. The variance of the independent variable, after dichotomization, was .25. Seven sample sizes, two independent variable types, and 175 parameter value combinations yielded 2450 simulations, with each consisting of 100 replications, providing a total of 245,000 observations.

Simulation Outcome Measures

Point estimates for the mediated effect, the direct effect, the total effects were calculated, as were the standard errors associated with each of these three measures (Hauck & Anderson, 1984). Two standard error formulas were evaluated, (1) first order Taylor series, and (2) second order Taylor series exact variance. The performance of each of these estimators was assessed with measures of bias (β_i) to compare estimates of the mediated effect ($\hat{w}$) to approximate true values (w) (Stone & Sobel, 1990) and mean square error (MSE_i).

$$\beta_i = \hat{w}_i - w_i$$

where $\hat{w}_i$ and w_i are the estimate and approximate true value at each replication. Additionally, the mean squared error (MSE_i) of each estimator was obtained by squaring the bias measure

$$MSE_i = (\hat{w}_i - w_i)^2$$

Confidence intervals associated with the mediated effect were examined by determining the proportion of times confidence intervals were to the left or right of the value of the mediated effect. The large sample 95% confidence limits were constructed using the mediated effect estimate plus and minus 1.96 times the estimate of the standard error of the mediated effect. With 100 replications, it is expected that 2.5 confidence intervals will be to the left of the true value of the mediated effect and 2.5 will be to the right of the true value, for a total of 5% of the confidence limits that will not include the true value.

The 175 combinations of parameter values produced three different types of models:

(1) consistent models, in which the true third-variable effect and direct effects had the same sign,

(2) inconsistent models, in which the true third-variable effect and direct effects had opposite signs,

and

(3) zero models, in which either the true third-variable effect or direct effect was equal to zero.

Results

Mediator, Confounder, and Suppressor Effects

As shown in Tables 1 and 2, point estimates of mediator, suppressor, and confounder effects and their standard errors were quite accurate for sample sizes of 50 or larger. In general more bias was observed in the case where the independent variable was binary but the point and standard error estimates remained accurate for sample sizes of 50 or larger.

Confidence Limits

The positions of true values relative to confidence limits are presented in Tables 3 and 4. For all models the true value is more often to the left of the confidence limits than the right when the third variable effect ($\alpha\beta = \tau - \tau'$) is negative. When sample sizes surpass 1000, the errors are generally balanced. When the true third-variable effect is positive, the true value is more often to the right of the confidence limits. Again, the errors to the left and to the right of the confidence limits are almost equal at samples of 1000 or greater.

Table 1. *Bias and Mean Squared Error by Type of Model and Sample Size for a Continuous Independent Variable*

		Sample Size						
		25	50	100	200	500	1000	5000
				Consistent models				
Direct Effect	Bias	-.0009	-.0020	.0001	-.0008	.0009	-.0000	-.0000
	MSE	.0663	.0302	.0142	.0067	.0028	.0013	.0003
Mediated Efect	Bias	.0007	.0021	.0010	.0002	.0002	.0001	-.0004
	MSE	.0351	.0167	.0079	.0039	.0015	.0007	.0002
Total Effect	Bias	-.0003	.0001	.0011	-.0006	.0011	.0001	-.0004
	MSE	.0622	.0301	.0145	.0068	.0028	.0014	.0003
SE_{Direct}	Bias	.0002	-.0002	.0001	-.0001	-.0000	-.0000	-.0000
	MSE	.0029	.0006	.0002	.0000	.0000	.0000	.0000
$SE_{Mediated}$	$Bias_1$	.0083	.0024	.0010	.0003	.0001	.0000	-.0000
	MSE_1	.0027	.0006	.0001	.0000	.0000	.0000	.0000
	$Bias_2$	.0083	.0024	.0010	.0003	.0001	.0000	-.0000
	MSE_2	.0026	.0006	.0001	.0000	.0000	.0000	.0000
				Inconsistent models				
Direct Effect	Bias	-.0011	-.0007	-.0010	.0022	-.0004	.0006	.0000
	MSE	.0645	.0310	.0140	.0070	.0028	.0013	.0003
Mediated Efect	Bias	-.0008	-.0006	-.0002	-.0018	.0000	-.0004	.0001
	MSE	.0379	.0166	.0077	.0038	.0015	.0007	.0001
Total Effect	Bias	-.0018	-.0013	-.0012	.0004	-.0004	.0002	.0001
	MSE	.0613	.0295	.0144	.0071	.0028	.0013	.0003
SE_{Direct}	Bias	-.0005	-.0004	-.0002	.0000	.0000	.0000	-.0000
	MSE	.0029	.0006	.0001	.0000	.0000	.0000	.0000
$SE_{Mediated}$	$Bias_1$	.0068	.0028	.0009	.0004	.0001	.0000	.0000
	MSE_1	.0028	.0006	.0001	.0000	.0000	.0000	.0000
	$Bias_2$	.0068	.0028	.0009	.0004	.0001	.0000	.0000
	MSE_2	.0027	.0006	.0001	.0000	.0000	.0000	.0000
				Zero models				
Direct Effect	Bias	-.0005	-.0053	-.0006	-.0009	-.0010	.0006	.0002
	MSE	.0571	.0261	.0126	.0062	.0025	.0012	.0002
Mediated Efect	Bias	-.0002	.0009	.0001	-.0007	.0001	-.0004	-.0000
	MSE	.0221	.0096	.0043	.0021	.0009	.0004	.0001
Total Effect	Bias	-.0007	-.0045	-.0005	-.0016	-.0009	.0002	.0001
	MSE	.0545	.0253	.0121	.0061	.0025	.0012	.0002
SE_{Direct}	Bias	.0002	-.0002	-.0000	-.0000	.0000	.0000	.0000
	MSE	.0026	.0006	.0001	.0000	.0000	.0000	.0000
$SE_{Mediated}$	$Bias_1$	.0133	.0049	.0018	.0008	.0003	.0001	.0000
	MSE_1	.0029	.0006	.0001	.0000	.0000	.0000	.0000
	$Bias_2$	.0111	.0038	.0013	.0006	.0002	.0001	.0000
	MSE_2	.0025	.0006	.0001	.0000	.0000	.0000	.0000

Note: The subscripts (1 and 2) correspond to the first order and second order Taylor series solutions for the variance of the mediated effect.

Table 2. *Bias and Mean Squared Error by Type of Model and Sample Size for a Dichotomous Independent Variable*

		Sample Size						
		25	50	100	200	500	1000	5000
				Consistent Models				
Direct Effect	Bias	-.0049	.0091	.0005	.0027	.0016	.0002	.0002
	MSE	.1922	.0924	.0439	.0219	.0088	.0045	.0009
Mediated Efect	Bias	-.0074	-.0026	.0022	.0002	.0003	-.0008	.0004
	MSE	.0885	.0390	.0196	.0092	.0038	.0018	.0004
Total Effect	Bias	-.0123	.0065	.0027	.0029	.0019	-.0006	.0007
	MSE	.2259	.1122	.0552	.0273	.0113	.0055	.0011
SE_{Direct}	Bias	-.0133	-.0041	-.0014	-.0004	-.0001	-.0000	-.0000
	MSE	.0057	.0012	.0003	.0001	.0000	.0000	.0000
$SE_{Mediated}$	$Bias_1$	.0094	.0020	.0011	.0003	.0001	.0000	-.0000
	MSE_1	.0078	.0019	.0005	.0001	.0000	.0000	.0000
	$Bias_2$	.0090	.0020	.0011	.0003	.0001	.0000	-.0000
	MSE_2	.0071	.0018	.0004	.0001	.0000	.0000	.0000
				Inconsistent Models				
Direct Effect	Bias	-.0128	.0039	-.0051	.0016	-.0015	.0013	.0002
	MSE	.1941	.0956	.0454	.0225	.0091	.0043	.0009
Mediated Efect	Bias	.0088	.0018	.0017	-.0016	.0015	.0001	-.0000
	MSE	.0928	.0412	.0187	.0095	.0038	.0019	.0004
Total Effect	Bias	-.0041	.0058	-.0033	.0000	-.0001	.0014	.0001
	MSE	.2339	.1143	.0553	.0284	.0112	.0055	.0011
SE_{Direct}	Bias	-.0119	-.0043	-.0014	-.0006	-.0001	-.0001	-.0000
	MSE	.0056	.0012	.0003	.0001	.0000	.0000	.0000
$SE_{Mediated}$	$Bias_1$	.0102	.0032	.0009	.0001	-.0000	-.0000	-.0000
	MSE_1	.0081	.0019	.0004	.0001	.0000	.0000	.0000
	$Bias_2$	.0097	.0032	.0009	.0001	-.0000	-.0000	-.0000
	MSE_2	.0074	.0018	.0004	.0001	.0000	.0000	.0000
				Zero models				
Direct Effect	Bias	.0018	-.0011	.0012	-.0003	.0014	-.0001	-.0003
	MSE	.1854	.0860	.0426	.0214	.0085	.0042	.0008
Mediated Efect	Bias	.0019	-.0001	-.0004	.0018	-.0010	-.0002	.0001
	MSE	.0521	.0238	.0106	.0053	.0021	.0010	.0002
Total Effect	Bias	.0037	-.0012	.0008	.0015	.0004	-.0003	-.0002
	MSE	.2027	.0983	.0472	.0242	.0095	.0048	.0010
SE_{Direct}	Bias	-.0124	-.0042	-.0016	-.0006	-.0001	-.0001	-.0000
	MSE	.0053	.0011	.0002	.0001	.0000	.0000	.0000
$SE_{Mediated}$	$Bias_1$	.0276	.0102	.0041	.0016	.0006	.0003	.0001
	MSE_1	.0085	.0020	.0005	.0001	.0000	.0000	.0000
	$Bias_2$	.0212	.0079	.0031	.0010	.0004	.0002	.0000
	MSE_2	.0066	.0017	.0004	.0001	.0000	.0000	.0000

Note: The subscripts (1 and 2) correspond to the first order and second order Taylor series solutions for the variance of the mediated effect.

Table 3. *Position of True Value Relative to Confidence Limits by Type of Model and Sample Size for a Continuous Independent Variable*

		Sample Size						
		25	50	100	200	500	1000	5000
		Consistent Models						
TMed < 0	$Right_1$	.0067	.0108	.0138	.0183	.0204	.0183	.0229
	$Left_1$	.0696	.0546	.0496	.0392	.0396	.0300	.0292
	$Right_2$	.0063	.0100	.0133	.0183	.0196	.0183	.0229
	$Left_2$	.0600	.0504	.0483	.0383	.0396	.0300	.0292
TMed > 0	$Right_1$	.0646	.0567	.0450	.0333	.0304	.0308	.0325
	$Left_1$	.0088	.0138	.0138	.0158	.0192	.0204	.0221
	$Right_2$	.0533	.0529	.0442	.0329	.0296	.0308	.0325
	$Left_2$	.0079	.0129	.0133	.0150	.0192	.0200	.0221
TMed = 0	$Right_1$	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	$Left_1$	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	$Right_2$	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	$Left_2$	N/A	N/A	N/A	N/A	N/A	N/A	N/A
		Inconsistent Models						
TMed < 0	$Right_1$	.0063	.0067	.0121	.0179	.0183	.0246	.0267
	$Left_1$	.0767	.0629	.0408	.0383	.0346	.0250	.0296
	$Right_2$	.0058	.0063	.0117	.0179	.0183	.0246	.0267
	$Left_2$	.0688	.0592	.0404	.0379	.0342	.0250	.0296
TMed > 0	$Right_1$	.0817	.0538	.0354	.0404	.0238	.0275	.0238
	$Left_1$	.0067	.0104	.0125	.0129	.0204	.0213	.0213
	$Right_2$	.0696	.0504	.0350	.0396	.0238	.0271	.0238
	$Left_2$	.0063	.0104	.0125	.0129	.0204	.0208	.0213
TMed = 0	$Right_1$	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	$Left_1$	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	$Right_2$	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	$Left_2$	N/A	N/A	N/A	N/A	N/A	N/A	N/A
		Zero models						
TMed < 0	$Right_1$	.0150	.0125	.0088	.0138	.0225	.0238	.0200
	$Left_1$	.0650	.0713	.0400	.0325	.0350	.0300	.0288
	$Right_2$	.0125	.0125	.0088	.0125	.0225	.0238	.0200
	$Left_2$	.0550	.0688	.0400	.0313	.0350	.0300	.0288
TMed > 0	$Right_1$	.0688	.0550	.0488	.0438	.0325	.0288	.0363
	$Left_1$	.0113	.0075	.0088	.0100	.0250	.0238	.0238
	$Right_2$	.0500	.0525	.0475	.0438	.0325	.0288	.0363
	$Left_2$	.0113	.0075	.0088	.0100	.0250	.0238	.0238
TMed = 0	$Right_1$	.0070	.0113	.0159	.0176	.0210	.0240	.0224
	$Left_1$	.0070	.0092	.0162	.0205	.0198	.0213	.0230
	$Right_2$	.0059	.0105	.0148	.0168	.0208	.0240	.0224
	$Left_2$	.0060	.0079	.0144	.0202	.0195	.0210	.0230

Note: The 1 and 2 subscripts correspond to the first and second order Taylor series solution for the variance of the mediated effect. N/A means not available. Table entries are the proportion of times the true value was to the left or the right of the 95% confidence limits.

Table 4. *Position of True Value Relative to Confidence Limits by Type of Model and Sample Size for a Dichotomous Independent Variable*

		Sample Size						
		25	50	100	200	500	1000	5000
		Consistent Models						
TMed < 0	$Right_1$	.0067	.0108	.0138	.0183	.0204	.0183	.0229
	$Left_1$	.0696	.0546	.0496	.0392	.0396	.0300	.0292
	$Right_2$	.0063	.0100	.0133	.0183	.0196	.0183	.0229
	$Left_2$	.0600	.0504	.0483	.0383	.0396	.0300	.0292
TMed > 0	$Right_1$	.0646	.0567	.0450	.0333	.0304	.0308	.0325
	$Left_1$	.0088	.0138	.0138	.0158	.0192	.0204	.0221
	$Right_2$	.0533	.0529	.0442	.0329	.0296	.0308	.0325
	$Left_2$	.0079	.0129	.0133	.0150	.0192	.0200	.0221
TMed = 0	$Right_1$	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	$Left_1$	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	$Right_2$	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	$Left_2$	N/A	N/A	N/A	N/A	N/A	N/A	N/A
		Inconsistent Models						
TMed < 0	$Right_1$	.0063	.0067	.0121	.0179	.0183	.0246	.0267
	$Left_1$	.0767	.0629	.0408	.0383	.0346	.0250	.0296
	$Right_2$	.0058	.0063	.0117	.0179	.0183	.0246	.0267
	$Left_2$	.0688	.0592	.0404	.0379	.0342	.0250	.0296
TMed > 0	$Right_1$	.0817	.0538	.0354	.0404	.0238	.0275	.0238
	$Left_1$	.0067	.0104	.0125	.0129	.0204	.0213	.0213
	$Right_2$	.0696	.0504	.0350	.0396	.0238	.0271	.0238
	$Left_2$	.0063	.0104	.0125	.0129	.0204	.0208	.0213
TMed = 0	$Right_1$	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	$Left_1$	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	$Right_2$	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	$Left_2$	N/A	N/A	N/A	N/A	N/A	N/A	N/A
		Zero models						
TMed < 0	$Right_1$	.0150	.0125	.0088	.0138	.0225	.0238	.0200
	$Left_1$	.0650	.0713	.0400	.0325	.0350	.0300	.0288
	$Right_2$	.0125	.0125	.0088	.0125	.0225	.0238	.0200
	$Left_2$	.0550	.0688	.0400	.0313	.0350	.0300	.0288
TMed > 0	$Right_1$	.0688	.0550	.0488	.0438	.0325	.0288	.0363
	$Left_1$	.0113	.0075	.0088	.0100	.0250	.0238	.0238
	$Right_2$	.0500	.0525	.0475	.0438	.0325	.0288	.0363
	$Left_2$	.0113	.0075	.0088	.0100	.0250	.0238	.0238
TMed = 0	$Right_1$	.0070	.0113	.0159	.0176	.0210	.0240	.0224
	$Left_1$	.0070	.0092	.0162	.0205	.0198	.0213	.0230
	$Right_2$	.0059	.0105	.0148	.0168	.0208	.0240	.0224
	$Left_2$	.0060	.0079	.0144	.0202	.0195	.0210	.0230

Note: The 1 and 2 subscripts correspond to the first and second order Taylor series solution for the variance of the mediated effect. N/A means not available.