

Technical Appendix for: “Comparing logistic and log-binomial models for causal mediation analyses of binary mediators and rare binary outcomes: evidence to support cross-checking of mediation results in practice”

This technical appendix is structured as follows.

In the first section, we provide specific details on the Valeri-VanderWeele macro *%mediation* (Valeri and VanderWeele, 2013) and how it can be applied. In the second section, we describe our macro *%OR_exact_formulas* which implements the exact odds ratio inference for mediation presented in the main paper. The third section introduces the macro *%simul_data* which serves to generate simple mediation data (exposure A , mediator M , and outcome Y) on which the *%mediation* and *%OR_exact_formulas* macros can be applied. The fourth and last section presents the results of mediation analyses performed on two large datasets of size $n = 100,000$ generated with the *%simul_data* macro on the basis of Scenarios 1 and 2 (scenarios described in Table 1 of the manuscript). We refer the reader to the SAS code contained in the file “simulation_execution.sas” to replicate the results found in this last section; for each scenario, the file “simulation_execution.sas” performs first the data generation and then executes, in order, the mediation analyses using the *%mediation* and *%OR_exact_formulas* macros.

1. Valeri-VanderWeele macro *%mediation*

The file “all_macros.sas” contains the original Valeri-VanderWeele macro *%mediation* with the following minimal modifications:

1. In the step `BOOTSTRAP PROCEDURE`, the argument `seed=0` in the `ranuni` function was changed to 1234 to fix the state with which a pseudo-random number generator was initialized in order to obtain reproducible results (if `seed=0`, the time of day is used to initialize the seed stream; see SAS documentation <https://support.sas.com/documentation/cdl/en/lrdict/64316/HTML/default/viewer.htm#a000202926.htm>);
2. The intercept was set to -4 in the log-binomial (log-linear) outcome model to increase the chance that the outcome probability falls into the interval (0, 1) (Fang, 2011).

Before each mediation analysis using the *%mediation* macro, the user has to close the SAS session in order to delete all temporary SAS files due to the previous execution of the *%mediation* macro (i.e., each *%mediation* macro execution requires to open a new SAS session) (Valeri and VanderWeele, 2013). The user can invoke this macro from the “all_macros.sas” file as follows:

```
filename macros "...\\all_macros.sas";
```

```
%include macros;
```

It is recommended to add the following statement before to run the *%mediation* macro to suppress printing to the log:

```
options nonotes nosource nosource2 errors=0;
```

To change the options back to the default and start printing to the log again, the user has to submit the following `OPTIONS` statement (see SAS code “`simulation_execution.sas`” for an example):

```
options notes source source2 errors=20;
```

We refer the readers to [Valeri and VanderWeele \(2013\)](#) for the details on how to assign values to the macro variables in the macro *%mediation* by these authors.

2. Macro *%OR_exact_formulas*

The SAS macro *%OR_exact_formulas* is developed for binary outcome, mediator and exposure. It provides estimates of the natural direct and indirect effects (NDE and NIE), and of the total effect (TE) on the OR scale using exact logistic-based expressions (refer to expressions (8-9) in the main paper). The user can invoke this macro from the “`all_macros.sas`” file as follows:

```
filename macros "...\\all_macros.sas";
%include macros;
```

This file also contains three auxiliary sub-macros called automatically by the principal macro *%OR_exact_formulas* (directly or via other sub-macros) during the different execution steps. Before running the macro *%OR_exact_formulas*, the user needs to delete the data sets “`crude_boot_estimates`” and “`adjusted_boot_estimates`” from the library `WORK`. The following statements can be used (see also SAS code “`simulation_execution.sas`” for an example):

```
proc datasets library = WORK noprint;
delete crude_boot_estimates adjusted_boot_estimates;
run; quit;
```

The user needs to assign values to the macro variables for the principal macro *%OR_exact_formulas* only (no user-specifications are required for the sub-macros). See Technical Appendix Table 1 for macro variables’ description and their possible values.

Technical Appendix Table 1: Macro `%OR_exact_formulas` variables

Macro variable	Description	Remarks
<code>mydata</code>	Name of the input data	
<code>exposure</code>	Name of the binary exposure (0/1 coding); must be in the input data	
<code>mediator</code>	Name of the binary mediator (0/1 coding); must be in the input data	
<code>outcome</code>	Name of the binary outcome (0/1 coding); must be in the input data	
<code>adjusted</code>	<code>adjusted=0</code> for unadjusted mediation analysis; <code>adjusted=1</code> for covariate adjusted mediation analysis	
<code>cvar</code>	If adjusted analysis is required (<code>adjusted=1</code>), the user needs to specify the list of adjustment variables (covariates) as follows: <code>cvar = covariate₁ covariate₂</code> <code>... covariate_n</code>	Use space to separate covariates' names. All categorical variables need to be entered as dummy variables.
<code>interaction</code>	<code>interaction=1</code> if an exposure-mediator interaction term is included in the model; <code>interaction=0</code> otherwise	
<code>iter</code>	Number of bootstrap samples (e.g., 1000)	Bootstrap (percentile method) is the default option to calculate the 95% confidence intervals for NIE, NDE, and TE.

3. Macro `%simul_data`

This macro simulates a sample of n observations for the purpose of unadjusted mediation analysis for a binary outcome (Y), mediator (M) and exposure (A). The data are generated nonparametrically according the assigned values for the marginal probability $P(A = 1)$, and the conditional probabilities $P(M = 1|A = i)$, $P(Y = 1|A = i, M = j)$, $i, j \in \{0, 1\}$. To simulate data using `%simul_data` macro, the user has to specify the values for the number n

of observations to generate, the random seed, the exposure marginal probability $P(A = 1)$, and the mediator and outcome conditional probabilities $P(M = 1|A = 0)$, $P(M = 1|A = 1)$, $P(Y = 1|A = 0, M = 0)$, $P(Y = 1|A = 0, M = 1)$, $P(Y = 1|A = 1, M = 0)$, $P(Y = 1|A = 1, M = 1)$. Firstly, the macro `%simul_data` generates the binary exposure A (i.e., it generates n observations from a *Bernoulli*(p_A), where $p_A = P(A = 1)$). Secondly, for each generated value of A , the macro generates the binary mediator from a *Bernoulli*(p_M), where $p_M = P(M = 1|A = i)$, $i \in \{0, 1\}$. Thirdly, for each previously generated pair (A, M) , the binary outcome Y is generated from a *Bernoulli*(p_Y), where $p_Y = P(Y = 1|A = i, M = j)$, $i, j \in \{0, 1\}$. The user can invoke this macro from the “all_macros.sas” file as follows:

```
filename macros "...\\all_macros.sas";
%include macros;
```

See Technical Appendix Table 2 for macro variables’ description and their possible values.

Technical Appendix Table 2: Macro `%simul_data` variables

Macro variable	Description	Remarks
<code>n</code>	Number of observations to simulate	
<code>seed</code>	Random seed to initialize a pseudorandom number generator	The seed value must be a positive integer
<code>pr_A</code>	$P(A = 1)$	Real number in the interval (0, 1)
<code>pr_M_0</code>	$P(M = 1 A = 0)$	Real number in the interval (0, 1)
<code>pr_M_1</code>	$P(M = 1 A = 1)$	Real number in the interval (0, 1)
<code>pr_Y_00</code>	$P(Y = 1 A = 0, M = 0)$	Real number in the interval (0, 1)
<code>pr_Y_01</code>	$P(Y = 1 A = 0, M = 1)$	Real number in the interval (0, 1)
<code>pr_Y_10</code>	$P(Y = 1 A = 1, M = 0)$	Real number in the interval (0, 1)
<code>pr_Y_11</code>	$P(Y = 1 A = 1, M = 1)$	Real number in the interval (0, 1)

4. Mediation effects estimated from two large synthetic datasets generated on the basis of Scenario 1 and 2

See Technical Appendix Table 3 for results.

Technical Appendix Table 3: Results of mediation analyses performed on two datasets of size $n = 100,000$ generated on the basis of Scenarios 1 and 2 with the `%simul_data` macro^a

Effect	True RR ^b	App. OR	Delta method 95% CI	Bootstrap 95% CI	RR	Delta method 95% CI	Bootstrap 95% CI	Exact OR	Bootstrap 95% CI	Relative difference (%)		
										RR vs. true RR ^c	Exact OR vs. true RR ^d	App. OR vs. true RR ^e
Scenario 1												
NDE	1.52	1.64	1.57, 1.71	1.57, 1.71	1.55	1.49, 1.61	1.49, 1.61	1.62	1.54, 1.68	1.84	6.31	7.82
NIE	1.05	1.06	1.05, 1.07	1.05, 1.07	1.05	1.04, 1.05	1.04, 1.05	1.05	1.05, 1.06	-0.29	0.29	0.76
TE ^f	1.60	1.74	1.66, 1.81	1.66, 1.81	1.62	1.56, 1.69	1.56, 1.68	1.70	1.62, 1.78	1.63	6.70	8.77
Scenario 2												
NDE	1.35	2.37	2.18, 2.57	2.19, 2.58	1.34	1.29, 1.39	1.29, 1.39	1.38	1.32, 1.44	-1.11	2.07	75.09
NIE	1.28	1.37	1.32, 1.42	1.32, 1.42	1.26	1.23, 1.30	1.22, 1.30	1.31	1.26, 1.35	-1.71	1.71	6.78
TE ^f	1.74	3.25	2.96, 3.56	2.99, 3.57	1.69	1.61, 1.76	1.62, 1.76	1.80	1.72, 1.89	-2.77	3.92	87.09

Abbreviations: app., approximate; CI, confidence interval; NDE, natural direct effect; NIE, natural indirect effect; OR, odds ratio; RR, relative risk; TE, total effect.

a: The data-generation mechanism is summarized in Table 1 of the manuscript.

b: True RR calculated using the probabilities from Table 1 of the manuscript.

c: $(RR - \text{true RR})/\text{true RR}$.

d: $(\text{exact OR} - \text{true RR})/\text{true RR}$.

e: $(\text{approximate OR} - \text{true RR})/\text{true RR}$.

f: TE=NDE-NIE.

References

- Fang, J. (2011). Using SAS procedures FREQ, GENMOD, LOGISTIC, and PHREG to estimate adjusted relative risks –A case study. In *Proceedings of the SAS Global Forum*.
- Valeri, L. and VanderWeele, T. J. (2013). Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. *Psychological Methods*, 18(2):137–150.