


## A comparison of the type I error rates of three assessment methods for indirect effects

Pier-Olivier Caron

To cite this article: Pier-Olivier Caron (2019): A comparison of the type I error rates of three assessment methods for indirect effects, Journal of Statistical Computation and Simulation, DOI: [10.1080/00949655.2019.1577858](https://doi.org/10.1080/00949655.2019.1577858)

To link to this article: <https://doi.org/10.1080/00949655.2019.1577858>

 View supplementary material 

 Published online: 10 Feb 2019.

 Submit your article to this journal 

 View Crossmark data 



# A comparison of the type I error rates of three assessment methods for indirect effects

Pier-Olivier Caron 

Département des Sciences humaines, Lettres et Communications, Université TÉLUQ, Montréal, Canada

## ABSTRACT

Mediation analysis is a popular statistical analysis verifying the relation between an independent variable and a dependent variable through a mediator. There are three traditional tests to assess indirect effects: the Baron and Kenny test (BK), the Sobel test (ST) and the bootstrap method (BT). Previous studies have showed that the BT is more powerful and more conceptually appropriate. However, no study has systematically compared these tests regarding the type I error rate. A Monte-Carlo simulation is carried out with 19 scenarios varying paths (but no indirect effect), 9 scenarios varying the direct effect, and 6 sample sizes (1056 different scenarios). Results show that the BT had an overall good performance even for small sample size and whatever the effect sizes. The ST and the BK test were conservative, especially with small sample size and low effect sizes. In conclusion, these tests should be avoided, and the BT is recommended.

## ARTICLE HISTORY

Received 13 September 2018  
Accepted 30 January 2019


## KEYWORDS

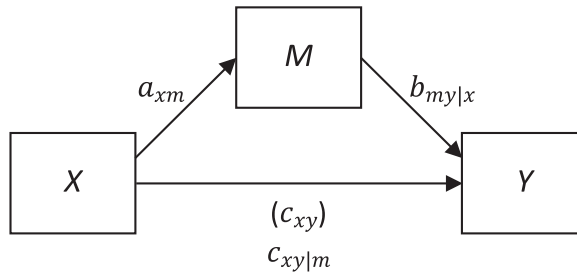
Mediation analysis; bootstrap; Sobel test; Baron and Kenny; multivariate delta method; Monte-Carlo simulations; path analysis

## 1. Introduction

Simple mediation analysis, a subset of path analysis [1], is an increasingly popular statistical analysis generating a lot of great interests and new developments in the methodological and statistical literature [2–6] as well-being widespread among applied researchers [7,8]. Within mediation analysis, the researcher is interested in the relation of the independent variable ( $x$ ) on the dependent variable ( $y$ ) through the mediator variable ( $m$ ). The variable  $m$  is the mechanism that underlies the relationship between  $x$  and  $y$ . The path diagram corresponding to a simple mediation model is presented in Figure 1. When there is no  $m$ , the existing relation between  $x$  and  $y$  is said to be the total effect, represented by  $c_{xy}$ . It corresponds to the bivariate correlation between  $x$  and  $y$ . The total effect can be divided into two other effects: the direct effect ( $c_{xy|m}$ ) and the indirect effect ( $ab$ ). The direct effect,  $c_{xy|m}$ , is the coefficient between  $x$  and  $y$  when the effect of  $m$  has been partialled out. The indirect effect,  $ab$ , is the product of  $a_{xm}$ , the regression coefficient of  $x$  predicting  $m$ , and  $b_{my|x}$ , the coefficient of between  $m$  and  $y$  when the effect of  $x$  has been partialled out. The indirect effect,  $ab$ , and the direct effect,  $c_{xy|m}$ , sum to the total effect  $c_{xy}$ , such that:  $a_{xm} \times$

**CONTACT** Pier-Olivier Caron  pocaron19@gmail.com  Département des Sciences humaines, Lettres et Communications, Université TÉLUQ, 5800, Rue Saint-Denis (Bureau 1105), Montréal, Québec, Canada H2S 3L5

 Supplemental data for this article can be accessed here. <https://doi.org/10.1080/00949655.2019.1577858>



**Figure 1.** Conceptual diagram of mediation analysis.

$b_{my|x} + c_{xy|m} = c_{xy}$ . The indirect effect is the effect of interest in mediation analysis. In mediation analysis, the product  $ab$  is tested to see if it is statistically different from 0 which would support a mediating effect.

There are three usual perspectives to determine if  $ab$  is statistically significant: (a) the causal step methods, the more popular being the Baron and Kenny test (BK; [9]), (b) the multivariate delta methods from which the Sobel test (ST; [10]) is a particular case; and (c) the bootstrap methods (BT; [11]). Each will be further detailed below.

BK is a three-step regression analysis. The first step is to check if the relation between  $x$  and  $y$ , i.e. assessing if  $c_{xy}$  is significant potentially meaning there is a relation to be explained by a mediator. The second step is to check if  $a_{xm}$  is significant, i.e. testing if there is a relation between the mediator and the predictor. Finally, the last step is to regress  $y$  on  $x$  and  $m$  to obtain  $b_{my|x}$  and  $c_{xy|m}$ . If  $b_{my|x}$  is significant then the method suggests that a mediation process occurred. If  $c_{xy|m}$  is no longer significant (compared to  $c_{xy}$ ), the mediation is said to be complete, otherwise, it is deemed partially mediated.

ST, which is a special case of the multivariate delta method, is a statistic with an asymptotically standard normal distribution ( $z$ -distribution) computed from the indirect effect as:

$$z = \frac{a_{xm}b_{my|x}}{SE} \tag{1}$$

where  $SE$  is the standard error of the indirect effect computed by:

$$SE = \sqrt{a_{xm}^2 s_{b_{my|x}}^2 + b_{my|x}^2 s_{a_{xm}}^2} \tag{2}$$

and where  $s^2$  represents the variance of the path  $a_{xm}$  and  $b_{mx|y}$ . The value  $z$  is then compared to an  $\alpha$  threshold like 1.64, for unilateral test or 1.96, for bicaudal test. This test has the assumption that the product of two regression coefficients is normally distributed, which is not always true in practice [12,13].

BT bootstraps the sample in order to build a 95% confidence interval (or any percentage actually) of the indirect effect and test if it entails the null hypothesis (i.e. the indirect effect is 0). Because it is as bootstrap methods, it is free from the statistical distribution assumption (more robustness) and is more powerful (less type II error) for mediation analysis [12]. The bootstrap method [14] is a computer-based method which treats the sample as a pseudo-population (that is, the sample distributions reflect the population distribution). It randomly selects with replacement subjects of the original sample in order to generate

another sample and compute the desired statistics. Then, it repeatedly does this last step a tremendous amount of time (for instance, over 5000) in order to create an empirical sampling distribution of the desired statistics. Confidence intervals can be computed from the sampling distribution and inference regarding hypothesis testing can be done. There is also the bias-corrected and accelerated (BCa) bootstrap interval, a method introduced to correct bias and skewness in the distribution of bootstrap estimates. BT is generally recommended because of its robustness, power and ease of use.

### 1.1. Objective

BT is recommended as the more powerful methods and the more conceptually appropriate method to assess the indirect effect [8,11–13]. Except power analysis, type I error rate is also a relevant property of statistical test. However, few studies [15,13] have evaluated the type I error rate of indirect effect tests in wide range of null indirect effects. For instance, Mackinnon, Lockwood, Hoffman, West, and Virgil [15] investigated 14 tests including the BK and ST, but not BT (the new recommended test) in few null-model scenarios (with no indirect effect). Thus, the purpose of the current paper is to systematically assess the type I error rate of BK, ST and BT on an extensive range of null-model scenarios ( $a_{xm} \times b_{my|x} = 0$ ).

## 2. Methods

A Monte Carlo simulation is carried out on Linux server with Intel i7-4770 processor running at 3.4 GHz. This CPU has 32kB of L1 cache, 256kB of L2 cache per core, and 8MB of L3 cache shared by all cores. The machine has 32 GB of RAM (DDR3-1600 with double-channel). Turbo Boost and power management was disabled forcing it to run at this base frequency. The simulation was carried in R [16] and took 11 days to complete.

The simulation included 19 null indirect effects where  $a_{xm}$  and  $b_{my|x}$  which were varied from 0 to .90 by an increment of .10, but where their product was always .00. There were also 9 conditions where  $c_{xy}$  was varied from 0 to .90 by increment of .10. All parameters were standardized coefficients; means of variable were 0 and variances were fixed to 1. Conditions where the sum of square of coefficients was over 1 (for instance,  $b_{my|x} = .80$  and  $c_{xy} = .70$ ) were excluded. Finally, there were 6 sample sizes,  $n = 50, 100, 150, 200, 250$  and 500 which are typical sample sizes in biological and social sciences [8,17–20]. Each condition was repeated 5000 times. In total, there were 1056 scenarios. The simulation was conducted as follows:

1. Choose a null indirect effect condition,  $a_{xm}$  and  $b_{my|x}$ ;
2. Choose a total direct effect,  $c_{xy}$ ;
3. If  $(b_{my|x}^2 + c_{xy}^2) > 1$ , return to step 1, otherwise continue;
4. Choose a sample size,  $n$ ;
5. Generate a data set with the above properties;
6. Evaluate the indirect effect with BK, ST and BT;
7. Record if the results was significant or not for each test;
8. Repeat steps 5, 6 and 7 a total of 5000 times;
9. Record the type I error rate (number of significant results divided by 5000) of each test;

10. Repeat steps 1 through 9 until all conditions have been done;

This section will now describe step 5 further. Data were generated as follows:

- 5.1 Generate a random sample of size  $n$  from a normal random variable  $X$  with mean of 0 and variance of 1,  $X \sim \mathcal{N}(0, 1)$ ;
- 5.2 Generate the variable  $M$  according to  $M = a_{xm}X + e_m$ , where the error of  $M$ ,  $e_m$ , is normally distributed with a mean of 0 and variance of  $1 - a_{xm}^2$ ;
- 5.3 Generate  $Y$  according to  $Y = c_{xy|m}X + b_{my|x}M + \text{sqrt}(e_y)$ , where the error of  $Y$ ,  $e_y$ , is normally distributed with a mean of 0 and a variance of  $1 - (c_{xy|m}^2 + b_{my|x}^2 + 2a_{xm}c_{xy|m}b_{my|x})$  (that is, the sum of two correlated variables is the sum of their variance plus twice their covariance; [21]).

The code to generate data and carry the indirect effect tests was inspired from Caron and Valois [22]. The script to carry the current simulation can be found as supplementary material with the current article.

### 3. Results

Table 1–18 shows all the results of the simulation. The tables are created as follow: the parameter  $a_{xm}$  is presented in decreasing order, from .90 to 0 and the parameter  $b_{my|x}$  is presented in ascending order from 0 to .90. All the products of  $a_{xm}$  and  $b_{my|x}$  equal .00. The middle of the table ( $a_{xm} = b_{my|x} = .00$ ) represents the harder condition, that is, the scenario in which an indirect effect is the less likely to occur by chance. Tables are placed to show results from BM, BK and, ST, and by increasing sample sizes. Within tables, light blue colored cells point out very low type I error rate defined as below the .025 threshold. Light red colored cells point out very high type I error rate, above de .075 threshold. Cells between

**Table 1.** Type I error rate of the Bootstrap test (BT) with sample size  $n = 50$ . (Color online).

Path	$a_{xm}$	$b_{my x}$	$c_{xy}$									
			.00	.10	.20	.30	.40	.50	.60	.70	.80	.90
.90	.00	.00	.067	.072	.068	.069	.069	.066	.070	.069	.069	.070
.80	.00	.00	.081	.074	.068	.081	.075	.071	.076	.065	.076	.073
.70	.00	.00	.073	.079	.074	.082	.080	.077	.080	.078	.083	.074
.60	.00	.00	.083	.087	.086	.084	.080	.083	.081	.088	.079	.089
.50	.00	.00	.085	.091	.087	.088	.087	.080	.081	.091	.091	.092
.40	.00	.00	.077	.076	.088	.085	.081	.077	.073	.078	.084	.085
.30	.00	.00	.061	.062	.059	.064	.058	.058	.057	.065	.062	.055
.20	.00	.00	.029	.036	.037	.034	.034	.031	.037	.033	.034	.039
.10	.00	.00	.013	.014	.016	.017	.013	.014	.013	.013	.015	.016
.00	.00	.00	.010	.007	.008	.008	.007	.010	.009	.007	.009	.011
.00	.10	.00	.016	.015	.016	.016	.018	.016	.018	.020	.024	.041
.00	.20	.00	.037	.032	.034	.034	.032	.038	.041	.057	.065	.085
.00	.30	.00	.059	.057	.057	.063	.061	.074	.074	.089	.083	.084
.00	.40	.00	.084	.083	.080	.074	.085	.093	.092	.085	.088	.073
.00	.50	.00	.085	.090	.084	.084	.086	.089	.082	.078	.069	–
.00	.60	.00	.083	.079	.084	.085	.090	.076	.073	.079	–	–
.00	.70	.00	.082	.073	.084	.079	.077	.077	.072	.070	–	–
.00	.80	.00	.071	.075	.075	.075	.070	.066	–	–	–	–
.00	.90	.00	.065	.067	.068	.069	.063	–	–	–	–	–

**Table 2.** Type I error rate of the Baron and Kenny test (BK) with sample size  $n = 50$ . (Color online).

Path		$c_{xy}$									
$a_{xm}$	$b_{my x}$	.00	.10	.20	.30	.40	.50	.60	.70	.80	.90
.90	.00	.003	.005	.013	.028	.042	.046	.054	.051	.048	.052
.80	.00	.002	.005	.014	.029	.045	.046	.050	.046	.049	.056
.70	.00	.002	.005	.013	.032	.041	.047	.052	.050	.057	.048
.60	.00	.004	.007	.015	.027	.041	.049	.048	.048	.046	.053
.50	.00	.002	.004	.016	.029	.039	.045	.047	.051	.047	.051
.40	.00	.002	.005	.014	.025	.034	.038	.039	.040	.046	.040
.30	.00	.001	.004	.008	.019	.021	.029	.027	.031	.033	.029
.20	.00	.001	.002	.006	.010	.013	.014	.015	.013	.013	.015
.10	.00	.000	.000	.001	.003	.005	.004	.005	.005	.005	.006
.00	.00	.000	.000	.000	.002	.002	.004	.002	.002	.002	.003
.00	.10	.000	.000	.002	.003	.003	.005	.007	.007	.010	.017
.00	.20	.001	.002	.005	.008	.010	.016	.019	.021	.031	.044
.00	.30	.002	.004	.009	.013	.022	.033	.031	.046	.042	.048
.00	.40	.003	.008	.015	.020	.030	.045	.045	.049	.054	.054
.00	.50	.009	.014	.016	.027	.035	.045	.046	.049	.049	-
.00	.60	.014	.016	.022	.026	.034	.044	.044	.059	-	-
.00	.70	.020	.018	.027	.027	.032	.039	.050	.051	-	-
.00	.80	.023	.023	.024	.025	.030	.037	-	-	-	-
.00	.90	.030	.026	.025	.026	.023	-	-	-	-	-

**Table 3.** Type I error rate of Sobel test (ST) with sample size  $n = 50$ . (Color online).

Path		$c_{xy}$									
$a_{xm}$	$b_{my x}$	.00	.10	.20	.30	.40	.50	.60	.70	.80	.90
.90	.00	.051	.053	.058	.055	.057	.057	.056	.050	.053	.053
.80	.00	.053	.051	.050	.049	.049	.051	.055	.052	.053	.049
.70	.00	.047	.045	.043	.049	.043	.039	.052	.049	.048	.044
.60	.00	.034	.038	.039	.042	.032	.032	.033	.035	.036	.038
.50	.00	.024	.027	.026	.023	.031	.025	.023	.026	.024	.026
.40	.00	.017	.016	.015	.015	.014	.015	.017	.017	.016	.016
.30	.00	.007	.006	.005	.008	.007	.008	.007	.006	.007	.006
.20	.00	.003	.002	.002	.002	.003	.002	.002	.002	.002	.001
.10	.00	.001	.001	.001	.000	.000	.001	.000	.000	.000	.001
.00	.00	.000	.000	.000	.001	.000	.000	.000	.001	.000	.000
.00	.10	.000	.000	.000	.001	.000	.000	.000	.002	.002	.002
.00	.20	.002	.001	.002	.004	.001	.003	.003	.005	.007	.019
.00	.30	.004	.004	.006	.008	.007	.009	.011	.018	.024	.041
.00	.40	.013	.015	.016	.019	.016	.019	.025	.033	.040	.055
.00	.50	.026	.028	.027	.027	.026	.029	.037	.044	.048	-
.00	.60	.038	.037	.034	.039	.046	.043	.047	.052	-	-
.00	.70	.047	.049	.044	.044	.044	.049	.055	.059	-	-
.00	.80	.050	.053	.049	.048	.054	.043	-	-	-	-
.00	.90	.047	.058	.051	.052	.049	-	-	-	-	-

.025 and .075 were deemed sufficiently near the .05  $\alpha$  threshold and are left without color. There is no value in the bottom right corner of tables because the sum of the variance of these corresponding standardized coefficients yield variance higher than 1.

There was no remarkable effect of  $c_{xy}$  on BT and ST, but a small one on BK depending whether the path was strong enough to be significant in function of the sample size (see columns  $c_{xy} = 0$  to .40 in Tables 2, 5, 8, 11, 14, and 17). This is expected as BK is highly dependent on the outcome that  $c_{xy}$  is significant to detect an indirect effect. There was also no evident effect of  $a_{xm}$  or  $b_{my|x}$  on all three tests. However, all tests had a low type I error

**Table 4.** Type I error rate of the Bootstrap test (BT) with sample size  $n = 100$ . (Color online).

Path		$c_{xy}$									
$a_{xm}$	$b_{my x}$	.00	.10	.20	.30	.40	.50	.60	.70	.80	.90
.90	.00	.064	.055	.062	.061	.063	.053	.060	.061	.059	.061
.80	.00	.055	.060	.062	.061	.066	.066	.063	.059	.058	.059
.70	.00	.060	.072	.066	.069	.067	.061	.060	.057	.069	.067
.60	.00	.072	.061	.062	.068	.068	.063	.068	.064	.068	.071
.50	.00	.078	.076	.070	.075	.076	.071	.075	.079	.074	.071
.40	.00	.084	.086	.081	.076	.078	.080	.078	.082	.077	.076
.30	.00	.074	.075	.078	.078	.075	.075	.076	.081	.081	.074
.20	.00	.051	.049	.055	.049	.056	.052	.055	.047	.050	.056
.10	.00	.019	.024	.018	.020	.021	.021	.014	.018	.016	.024
.00	.00	.006	.007	.006	.004	.007	.009	.009	.008	.006	.006
.00	.10	.018	.023	.019	.019	.025	.023	.023	.034	.042	.062
.00	.20	.056	.047	.053	.051	.063	.061	.068	.078	.084	.074
.00	.30	.075	.073	.083	.083	.085	.084	.075	.075	.075	.063
.00	.40	.080	.077	.081	.085	.077	.083	.069	.070	.063	.059
.00	.50	.073	.070	.076	.073	.073	.065	.063	.063	.063	-
.00	.60	.069	.070	.069	.069	.066	.063	.061	.060	-	-
.00	.70	.073	.065	.061	.063	.059	.062	.054	.059	-	-
.00	.80	.064	.064	.063	.064	.061	.063	-	-	-	-
.00	.90	.055	.059	.057	.061	.055	-	-	-	-	-

**Table 5.** Type I error rate of the Baron & Kenny test (BK) with sample size  $n = 100$ . (Color online).

Path		$c_{xy}$									
$a_{xm}$	$b_{my x}$	.00	.10	.20	.30	.40	.50	.60	.70	.80	.90
.90	.00	.003	.008	.026	.047	.051	.045	.048	.050	.051	.051
.80	.00	.002	.008	.025	.041	.049	.053	.050	.053	.049	.051
.70	.00	.001	.009	.027	.045	.051	.048	.046	.045	.054	.054
.60	.00	.002	.009	.022	.044	.050	.047	.049	.047	.050	.056
.50	.00	.004	.010	.024	.044	.051	.050	.050	.055	.049	.048
.40	.00	.004	.010	.028	.039	.042	.049	.047	.057	.049	.046
.30	.00	.003	.007	.024	.033	.045	.046	.043	.043	.047	.042
.20	.00	.001	.003	.016	.021	.027	.028	.029	.022	.026	.028
.10	.00	.000	.002	.003	.007	.008	.007	.005	.006	.009	.010
.00	.00	.000	.001	.001	.001	.003	.003	.003	.003	.002	.002
.00	.10	.000	.001	.002	.006	.014	.011	.009	.016	.021	.032
.00	.20	.004	.005	.014	.021	.032	.032	.038	.044	.052	.050
.00	.30	.004	.008	.022	.038	.041	.046	.045	.050	.050	.048
.00	.40	.004	.011	.027	.042	.049	.053	.047	.051	.047	.050
.00	.50	.009	.015	.025	.036	.048	.049	.049	.050	.052	-
.00	.60	.012	.018	.026	.037	.048	.048	.049	.049	-	-
.00	.70	.017	.020	.022	.031	.041	.052	.047	.050	-	-
.00	.80	.022	.023	.023	.026	.043	.052	-	-	-	-
.00	.90	.028	.023	.025	.026	.039	-	-	-	-	-

rate when  $a_{xm} = b_{my|x} = 0$  or when one of the path was small (below .30). This effect was dependent on sample size, lower values increase the trend, whereas higher value makes it disappear, excepting in the  $a_{xm} = b_{my|x} = 0$  scenarios where all tests were strongly conservative. ST was more similar to BT. BK was overly conservative in scenarios where  $a_{xm}$  or  $b_{my|x}$  were low. In  $c_{xy} < .20$  scenarios, BK was strongly conservative regardless of the sample size (especially, when  $c_{xy} = .00$ ). In contrary, BT has an acceptable type I error rate. ST was somewhat between both in terms of performance.

**Table 6.** Type I error rate of Sobel test (ST) with sample size  $n = 100$ . (Color online).

Path		$c_{xy}$									
$a_{xm}$	$b_{my x}$	.00	.10	.20	.30	.40	.50	.60	.70	.80	.90
.90	.00	.050	.051	.051	.047	.045	.052	.056	.049	.051	.049
.80	.00	.052	.055	.052	.048	.045	.048	.050	.051	.053	.053
.70	.00	.048	.052	.048	.049	.046	.050	.052	.046	.047	.053
.60	.00	.048	.041	.043	.038	.045	.043	.047	.051	.043	.049
.50	.00	.037	.039	.040	.031	.036	.033	.037	.040	.038	.044
.40	.00	.030	.026	.024	.032	.028	.025	.029	.028	.031	.032
.30	.00	.017	.015	.015	.014	.016	.014	.015	.015	.014	.011
.20	.00	.006	.007	.005	.004	.004	.005	.006	.004	.004	.005
.10	.00	.000	.001	.000	.000	.001	.001	.001	.000	.001	.001
.00	.00	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
.00	.10	.001	.000	.001	.001	.001	.001	.001	.002	.004	.008
.00	.20	.006	.004	.003	.005	.005	.006	.008	.014	.021	.032
.00	.30	.014	.013	.015	.013	.015	.016	.021	.029	.036	.052
.00	.40	.023	.031	.025	.027	.031	.032	.039	.044	.052	.055
.00	.50	.039	.036	.036	.040	.044	.045	.045	.050	.053	-
.00	.60	.047	.048	.045	.043	.045	.053	.045	.052	-	-
.00	.70	.050	.047	.054	.048	.052	.052	.049	.051	-	-
.00	.80	.053	.050	.048	.054	.051	.054	-	-	-	-
.00	.90	.046	.052	.052	.052	.049	-	-	-	-	-

**Table 7.** Type I error rate of the Bootstrap test (BT) with sample size  $n = 150$ . (Color online).

Path		$c_{xy}$									
$a_{xm}$	$b_{my x}$	.00	.10	.20	.30	.40	.50	.60	.70	.80	.90
.90	.00	.057	.059	.064	.058	.055	.059	.056	.052	.053	.059
.80	.00	.061	.059	.055	.056	.060	.062	.054	.057	.055	.058
.70	.00	.060	.058	.059	.063	.061	.057	.058	.059	.061	.055
.60	.00	.062	.067	.063	.066	.062	.065	.060	.059	.067	.062
.50	.00	.068	.071	.067	.066	.068	.066	.063	.069	.067	.067
.40	.00	.077	.067	.070	.067	.074	.074	.073	.067	.071	.074
.30	.00	.083	.078	.084	.076	.069	.084	.073	.077	.083	.081
.20	.00	.068	.063	.057	.064	.062	.060	.059	.056	.064	.059
.10	.00	.023	.026	.023	.020	.023	.021	.028	.023	.027	.022
.00	.00	.006	.006	.005	.007	.008	.007	.007	.008	.007	.008
.00	.10	.024	.025	.025	.025	.027	.029	.029	.038	.045	.076
.00	.20	.065	.060	.063	.072	.065	.065	.080	.082	.073	.077
.00	.30	.079	.082	.076	.078	.078	.072	.074	.071	.061	.063
.00	.40	.072	.076	.068	.067	.067	.068	.070	.064	.061	.056
.00	.50	.063	.064	.064	.067	.062	.066	.055	.056	.056	-
.00	.60	.057	.060	.056	.070	.060	.054	.059	.062	-	-
.00	.70	.061	.060	.053	.064	.062	.056	.060	.057	-	-
.00	.80	.060	.057	.060	.056	.051	.057	-	-	-	-
.00	.90	.059	.058	.058	.059	.056	-	-	-	-	-

Regarding the effect of sample sizes, as the value increases, BT was less likely to commit false positives. Still, the high type error rates (red colored cells pattern) never exceeded .10. The pattern of high type I error rate seemed otherwise random. Tables 2–3 show that the BK and ST have really low type I error rate when  $n = 50$ . In other terms, they were conservative and were unlikely to find an indirect effect. They probably lack of power with sample sizes as low. In contrary, BT had high but not excessive, type I error in the  $n = 50$  scenarios.



**Table 8.** Type I error rate of the Baron & Kenny test (BK) with sample size  $n = 150$ . (Color online).

Path		$c_{xy}$									
$a_{xm}$	$b_{my x}$	.00	.10	.20	.30	.40	.50	.60	.70	.80	.90
.90	.00	.002	.013	.035	.051	.048	.053	.053	.047	.047	.050
.80	.00	.001	.011	.034	.045	.052	.055	.047	.050	.048	.051
.70	.00	.003	.011	.034	.050	.050	.050	.049	.051	.050	.045
.60	.00	.002	.010	.036	.053	.051	.055	.051	.045	.054	.050
.50	.00	.003	.014	.035	.047	.049	.051	.047	.055	.052	.050
.40	.00	.003	.011	.031	.046	.050	.049	.051	.047	.048	.050
.30	.00	.001	.012	.036	.045	.045	.047	.048	.047	.052	.050
.20	.00	.002	.010	.025	.034	.034	.035	.030	.028	.037	.031
.10	.00	.000	.005	.008	.009	.011	.010	.011	.009	.012	.009
.00	.00	.000	.000	.001	.004	.003	.002	.002	.003	.003	.003
.00	.10	.000	.003	.007	.012	.011	.012	.014	.021	.021	.040
.00	.20	.003	.009	.021	.033	.036	.036	.047	.045	.047	.053
.00	.30	.005	.014	.032	.042	.048	.044	.049	.046	.047	.051
.00	.40	.006	.015	.030	.043	.047	.052	.053	.052	.051	.051
.00	.50	.010	.018	.030	.045	.048	.051	.045	.051	.050	-
.00	.60	.010	.021	.027	.051	.046	.045	.049	.054	-	-
.00	.70	.017	.023	.025	.043	.052	.050	.052	.051	-	-
.00	.80	.022	.022	.027	.037	.048	.050	-	-	-	-
.00	.90	.030	.028	.025	.035	.046	-	-	-	-	-

**Table 9.** Type I error rate of Sobel test (ST) with sample size  $n = 150$ . (Color online).

Path		$c_{xy}$									
$a_{xm}$	$b_{my x}$	.00	.10	.20	.30	.40	.50	.60	.70	.80	.90
.90	.00	.050	.053	.053	.054	.059	.049	.044	.057	.051	.052
.80	.00	.050	.049	.055	.047	.050	.050	.048	.054	.052	.052
.70	.00	.050	.054	.049	.048	.049	.051	.048	.050	.051	.048
.60	.00	.047	.047	.045	.045	.046	.047	.052	.049	.053	.047
.50	.00	.041	.039	.042	.047	.045	.042	.042	.040	.047	.045
.40	.00	.036	.034	.037	.030	.035	.035	.037	.036	.042	.035
.30	.00	.024	.020	.019	.020	.023	.024	.020	.026	.022	.022
.20	.00	.008	.007	.006	.007	.007	.009	.009	.007	.007	.008
.10	.00	.001	.001	.001	.001	.002	.001	.001	.002	.001	.002
.00	.00	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
.00	.10	.001	.002	.002	.000	.001	.002	.002	.002	.004	.010
.00	.20	.008	.006	.007	.006	.009	.006	.013	.021	.025	.039
.00	.30	.027	.019	.021	.023	.023	.028	.037	.040	.036	.049
.00	.40	.032	.033	.034	.035	.034	.039	.044	.036	.046	.048
.00	.50	.039	.042	.040	.042	.045	.040	.050	.050	.051	-
.00	.60	.044	.050	.051	.046	.052	.044	.046	.054	-	-
.00	.70	.047	.048	.049	.050	.055	.048	.051	.051	-	-
.00	.80	.049	.049	.046	.053	.046	.050	-	-	-	-
.00	.90	.053	.054	.054	.050	.055	-	-	-	-	-

Table 19 shows the summary of the simulation. In general, BT had a good type I error rate (few colored cells). For instance, BT had 24.6% of cells being colored (either blue or red) whereas ST and BK had 32.5% and 34.7% of colored cells (only blue). The ST and BK exhibit the same kind of performance. Both never exceeded a type I error rate of .060 (see max value). They were both unlikely to make false positive. As such, the performance of BT seems better than its counterpart. BT showed no systematic pattern to commit high or low false positive (except the  $a_{xm} = b_{my|x} = 0$  cases). As previously mentioned, patterns

**Table 10.** Type I error rate of the Bootstrap test (BT) with sample size  $n = 200$ . (Color online).

Path		$c_{xy}$									
$a_{xm}$	$b_{my x}$	.00	.10	.20	.30	.40	.50	.60	.70	.80	.90
.90	.00	.059	.054	.057	.051	.056	.053	.062	.058	.055	.061
.80	.00	.061	.054	.058	.059	.052	.054	.057	.052	.062	.058
.70	.00	.053	.057	.054	.058	.052	.055	.056	.059	.054	.060
.60	.00	.061	.058	.057	.061	.055	.060	.056	.055	.058	.061
.50	.00	.056	.065	.062	.060	.062	.065	.065	.055	.058	.066
.40	.00	.065	.071	.070	.066	.067	.063	.065	.060	.069	.069
.30	.00	.069	.079	.071	.078	.074	.073	.076	.070	.075	.078
.20	.00	.072	.074	.071	.077	.075	.065	.068	.079	.066	.073
.10	.00	.028	.026	.030	.031	.024	.029	.029	.030	.029	.030
.00	.00	.005	.005	.006	.006	.007	.005	.005	.006	.007	.007
.00	.10	.024	.029	.030	.029	.031	.033	.041	.048	.054	.078
.00	.20	.073	.075	.071	.070	.078	.079	.079	.083	.078	.062
.00	.30	.080	.077	.078	.080	.077	.071	.072	.062	.058	.059
.00	.40	.065	.071	.061	.068	.064	.064	.064	.059	.054	.061
.00	.50	.061	.066	.059	.063	.066	.054	.050	.057	.055	-
.00	.60	.061	.059	.061	.061	.054	.059	.061	.057	-	-
.00	.70	.059	.059	.059	.056	.052	.052	.054	.050	-	-
.00	.80	.056	.055	.057	.051	.054	.058	-	-	-	-
.00	.90	.052	.050	.053	.051	.052	-	-	-	-	-

**Table 11.** Type I error rate of the Baron & Kenny test (BK) with sample size  $n = 200$ . (Color online).

Path		$c_{xy}$									
$a_{xm}$	$b_{my x}$	.00	.10	.20	.30	.40	.50	.60	.70	.80	.90
.90	.00	.002	.016	.043	.047	.051	.049	.054	.053	.050	.059
.80	.00	.003	.013	.041	.050	.046	.049	.049	.048	.056	.053
.70	.00	.003	.013	.040	.049	.047	.050	.048	.051	.047	.048
.60	.00	.002	.016	.038	.051	.047	.052	.049	.046	.049	.051
.50	.00	.002	.013	.042	.048	.049	.050	.054	.043	.047	.053
.40	.00	.002	.018	.046	.048	.046	.048	.046	.045	.050	.051
.30	.00	.003	.014	.038	.051	.047	.050	.048	.044	.048	.049
.20	.00	.001	.009	.036	.046	.043	.039	.041	.046	.039	.040
.10	.00	.000	.003	.015	.016	.014	.015	.013	.014	.015	.013
.00	.00	.000	.001	.002	.003	.003	.002	.002	.002	.003	.003
.00	.10	.001	.007	.014	.014	.015	.017	.023	.023	.028	.045
.00	.20	.002	.013	.027	.040	.048	.046	.049	.054	.055	.047
.00	.30	.005	.016	.040	.051	.052	.051	.050	.048	.050	.054
.00	.40	.007	.020	.035	.054	.047	.048	.053	.049	.049	.056
.00	.50	.007	.021	.033	.048	.054	.045	.042	.047	.050	-
.00	.60	.013	.022	.037	.046	.045	.051	.053	.052	-	-
.00	.70	.016	.024	.031	.048	.044	.047	.047	.048	-	-
.00	.80	.022	.024	.026	.039	.049	.050	-	-	-	-
.00	.90	.027	.023	.026	.039	.049	-	-	-	-	-

of high or low type I error rate of BT were likely to be random whereas BK and ST showed systematic trends.

#### 4. Discussion

The purpose of the current study is to assess the type I error rate of BK, ST and BT on an extensive range of scenarios where the indirect effect is null. This question was evaluated via a Monte Carlo simulation including 1056 different scenarios of null indirect effect thus

**Table 12.** Type I error rate of Sobel test (ST) with sample size  $n = 200$ . (Color online).

Path		$c_{xy}$									
$a_{xm}$	$b_{my x}$	.00	.10	.20	.30	.40	.50	.60	.70	.80	.90
.90	.00	.051	.047	.044	.056	.047	.052	.055	.044	.056	.051
.80	.00	.052	.048	.049	.059	.044	.055	.044	.051	.055	.051
.70	.00	.044	.055	.055	.053	.045	.047	.048	.053	.046	.048
.60	.00	.042	.047	.046	.047	.054	.044	.049	.049	.050	.051
.50	.00	.046	.046	.048	.046	.041	.041	.046	.045	.044	.041
.40	.00	.041	.041	.037	.039	.038	.036	.041	.040	.039	.037
.30	.00	.029	.026	.030	.031	.028	.026	.030	.023	.026	.029
.20	.00	.010	.011	.013	.011	.014	.011	.012	.013	.012	.012
.10	.00	.001	.002	.001	.002	.003	.002	.001	.001	.001	.002
.00	.00	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
.00	.10	.001	.001	.002	.001	.002	.003	.002	.003	.007	.014
.00	.20	.009	.011	.010	.017	.015	.016	.018	.028	.033	.045
.00	.30	.028	.021	.031	.027	.033	.035	.036	.042	.047	.050
.00	.40	.041	.042	.035	.044	.044	.042	.047	.047	.051	.051
.00	.50	.042	.046	.045	.047	.044	.040	.051	.046	.052	-
.00	.60	.050	.047	.045	.049	.046	.048	.048	.053	-	-
.00	.70	.043	.043	.053	.048	.051	.045	.055	.049	-	-
.00	.80	.049	.046	.051	.053	.046	.050	-	-	-	-
.00	.90	.050	.059	.052	.048	.055	-	-	-	-	-

**Table 13.** Type I error rate of the Bootstrap test (BT) with sample size  $n = 250$ . (Color online).

Path		$c_{xy}$									
$a_{xm}$	$b_{my x}$	.00	.10	.20	.30	.40	.50	.60	.70	.80	.90
.90	.00	.059	.054	.052	.055	.052	.054	.051	.051	.053	.057
.80	.00	.056	.053	.054	.054	.048	.046	.062	.061	.050	.058
.70	.00	.053	.057	.053	.057	.055	.058	.059	.061	.057	.057
.60	.00	.056	.052	.057	.056	.058	.058	.057	.058	.056	.054
.50	.00	.054	.059	.057	.059	.058	.057	.061	.055	.065	.065
.40	.00	.066	.063	.055	.059	.059	.061	.065	.060	.061	.061
.30	.00	.070	.069	.069	.070	.079	.072	.076	.072	.067	.071
.20	.00	.075	.071	.077	.079	.077	.070	.075	.080	.079	.078
.10	.00	.033	.030	.031	.035	.033	.036	.033	.035	.032	.037
.00	.00	.007	.007	.004	.006	.007	.005	.008	.003	.008	.005
.00	.10	.037	.031	.033	.037	.040	.037	.047	.053	.066	.074
.00	.20	.070	.079	.072	.073	.070	.078	.078	.063	.069	.068
.00	.30	.072	.072	.076	.068	.070	.072	.065	.065	.055	.056
.00	.40	.062	.070	.069	.061	.062	.065	.057	.058	.055	.053
.00	.50	.061	.061	.059	.056	.067	.057	.055	.058	.055	-
.00	.60	.055	.064	.051	.060	.061	.056	.061	.054	-	-
.00	.70	.056	.059	.056	.063	.053	.058	.057	.044	-	-
.00	.80	.051	.053	.058	.053	.055	.052	-	-	-	-
.00	.90	.056	.054	.056	.062	.056	-	-	-	-	-

giving a systematic portrait of type I error rate of most used statistical tests of an indirect effect.

The results suggest that BT is the most reasonable method to detect an indirect effect with, on average, a type I error rate approximately equals to the  $\alpha$  threshold held at .05. BK and ST were both conservative and very few times did they commit to false positives. For instance, BK has a correct type I error rate when  $c_{xy}$  was moderate ( $> .2$ ) or when sample sizes were high enough. Otherwise, it was conservative. Moreover, at the sample size of 50, BK had very low type I error which indicates lack of power. In comparison, ST

**Table 14.** Type I error rate of the Baron & Kenny test (BK) with sample size  $n = 250$ . (Color online).

Path		$c_{xy}$									
$a_{xm}$	$b_{my x}$	.00	.10	.20	.30	.40	.50	.60	.70	.80	.90
.90	.00	.003	.016	.042	.050	.048	.045	.049	.047	.049	.053
.80	.00	.003	.018	.044	.052	.047	.041	.058	.054	.050	.053
.70	.00	.002	.018	.043	.052	.048	.054	.051	.053	.051	.046
.60	.00	.002	.014	.046	.051	.049	.050	.047	.047	.052	.048
.50	.00	.003	.017	.042	.049	.050	.046	.054	.046	.052	.053
.40	.00	.002	.018	.041	.046	.049	.047	.053	.052	.049	.049
.30	.00	.003	.017	.042	.048	.047	.052	.054	.050	.051	.053
.20	.00	.003	.013	.043	.044	.048	.044	.045	.046	.048	.050
.10	.00	.001	.004	.016	.017	.017	.019	.017	.019	.017	.020
.00	.00	.000	.001	.002	.003	.003	.002	.004	.001	.002	.002
.00	.10	.000	.005	.015	.018	.019	.018	.024	.026	.038	.047
.00	.20	.003	.015	.035	.042	.043	.047	.053	.042	.050	.054
.00	.30	.004	.018	.046	.045	.047	.051	.053	.056	.049	.053
.00	.40	.007	.023	.045	.046	.053	.053	.049	.048	.048	.049
.00	.50	.008	.021	.038	.044	.060	.047	.049	.048	.049	-
.00	.60	.010	.032	.034	.049	.058	.054	.057	.049	-	-
.00	.70	.015	.024	.031	.053	.047	.054	.052	.042	-	-
.00	.80	.025	.026	.030	.048	.048	.046	-	-	-	-
.00	.90	.028	.025	.025	.055	.049	-	-	-	-	-

**Table 15.** Type I error rate of Sobel test (ST) with sample size  $n = 250$ . (Color online).

Path		$c_{xy}$									
$a_{xm}$	$b_{my x}$	.00	.10	.20	.30	.40	.50	.60	.70	.80	.90
.90	.00	.056	.051	.050	.052	.051	.053	.049	.048	.053	.052
.80	.00	.053	.057	.053	.050	.052	.049	.050	.053	.046	.048
.70	.00	.047	.045	.046	.049	.050	.050	.052	.048	.050	.049
.60	.00	.047	.046	.049	.053	.052	.048	.050	.047	.045	.047
.50	.00	.046	.044	.040	.042	.046	.045	.046	.039	.049	.049
.40	.00	.038	.042	.042	.045	.043	.044	.040	.039	.041	.037
.30	.00	.027	.033	.031	.035	.033	.032	.030	.035	.035	.030
.20	.00	.015	.014	.014	.015	.013	.013	.015	.015	.013	.013
.10	.00	.002	.001	.003	.002	.001	.002	.002	.002	.002	.002
.00	.00	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
.00	.10	.003	.002	.001	.002	.003	.004	.004	.006	.009	.020
.00	.20	.013	.015	.013	.018	.016	.020	.023	.028	.036	.044
.00	.30	.035	.031	.035	.031	.034	.043	.039	.043	.051	.043
.00	.40	.042	.042	.042	.045	.041	.043	.046	.046	.046	.054
.00	.50	.048	.043	.046	.048	.046	.050	.049	.045	.050	-
.00	.60	.045	.051	.051	.045	.048	.051	.052	.055	-	-
.00	.70	.055	.047	.051	.048	.049	.048	.048	.051	-	-
.00	.80	.048	.047	.050	.047	.048	.050	-	-	-	-
.00	.90	.054	.048	.043	.051	.050	-	-	-	-	-

was modestly conservative compared to BK while more conservative than BT. BK also had a very low type I error rate at  $n = 50$ , indicating the same problem of power with small sample sizes. Finally, BT had the most accurate performance of type I error rate.

Table 19 shows that BT has an average (across all scenarios) 5.7% of committing type I error whereas BK and ST each had 3.3%. Moreover, BT had 24.6% of cells being colored (either blue or red) whereas ST and BK had 32.5% and 34.7% of colored cells (only blue). The ST and BK never exceeded a type I error rate of .075, which means they were unlikely to make false positives. As such, performance of BT seemed better, because scenario where it

**Table 16.** Type I error rate of the Bootstrap test (BT) with sample size  $n = 500$ . (Color online).

Path		$c_{xy}$									
$a_{xm}$	$b_{my x}$	.00	.10	.20	.30	.40	.50	.60	.70	.80	.90
.90	.00	.056	.055	.049	.054	.054	.053	.059	.051	.049	.054
.80	.00	.052	.048	.051	.044	.049	.053	.058	.054	.051	.054
.70	.00	.053	.053	.053	.046	.051	.053	.050	.050	.048	.051
.60	.00	.056	.053	.052	.058	.054	.057	.053	.049	.053	.052
.50	.00	.055	.056	.052	.054	.055	.053	.055	.055	.051	.054
.40	.00	.052	.057	.052	.051	.062	.060	.059	.059	.061	.062
.30	.00	.067	.060	.062	.065	.067	.059	.068	.062	.065	.057
.20	.00	.077	.071	.074	.067	.070	.070	.073	.072	.078	.065
.10	.00	.050	.053	.056	.047	.054	.054	.049	.054	.052	.058
.00	.00	.007	.004	.005	.004	.006	.006	.005	.006	.006	.006
.00	.10	.053	.051	.052	.054	.057	.069	.060	.078	.074	.071
.00	.20	.070	.076	.071	.071	.077	.069	.065	.067	.060	.054
.00	.30	.061	.063	.059	.057	.054	.056	.058	.058	.055	.053
.00	.40	.061	.053	.055	.057	.055	.054	.062	.056	.057	.056
.00	.50	.058	.053	.058	.053	.052	.053	.048	.051	.052	-
.00	.60	.054	.055	.052	.055	.055	.056	.056	.054	-	-
.00	.70	.053	.055	.056	.052	.055	.050	.051	.048	-	-
.00	.80	.053	.053	.055	.049	.054	.051	-	-	-	-
.00	.90	.054	.052	.049	.055	.058	-	-	-	-	-

**Table 17.** Type I error rate of the Baron & Kenny test (BK) with sample size  $n = 500$ . (Color online).

Path		$c_{xy}$									
$a_{xm}$	$b_{my x}$	.00	.10	.20	.30	.40	.50	.60	.70	.80	.90
.90	.00	.003	.029	.048	.053	.054	.049	.055	.047	.051	.051
.80	.00	.003	.030	.050	.044	.048	.050	.055	.051	.048	.052
.70	.00	.002	.032	.050	.042	.048	.049	.048	.046	.047	.045
.60	.00	.002	.031	.049	.049	.052	.050	.050	.045	.049	.048
.50	.00	.002	.030	.046	.048	.050	.046	.049	.050	.046	.051
.40	.00	.002	.031	.045	.045	.054	.054	.053	.052	.050	.055
.30	.00	.002	.030	.049	.053	.055	.045	.055	.046	.049	.046
.20	.00	.003	.030	.048	.047	.047	.044	.048	.051	.057	.045
.10	.00	.003	.019	.033	.028	.031	.033	.029	.031	.028	.034
.00	.00	.000	.002	.002	.002	.003	.003	.001	.003	.003	.002
.00	.10	.003	.020	.029	.028	.032	.038	.034	.049	.044	.049
.00	.20	.005	.029	.048	.051	.054	.052	.048	.054	.051	.049
.00	.30	.006	.031	.046	.047	.043	.047	.048	.048	.051	.052
.00	.40	.006	.027	.045	.050	.048	.049	.055	.052	.054	.053
.00	.50	.009	.026	.053	.048	.047	.050	.045	.048	.055	-
.00	.60	.013	.027	.046	.051	.051	.053	.052	.052	-	-
.00	.70	.017	.027	.048	.049	.053	.049	.049	.045	-	-
.00	.80	.023	.023	.050	.046	.053	.049	-	-	-	-
.00	.90	.029	.026	.044	.054	.056	-	-	-	-	-

had high false-positive looked random, whereas low type I error rates of BK and ST looked systematic.

From what is expected from the methodological literature on mediation analysis, results show that effects of  $a_{xm}$  and  $b_{my|x}$  look symmetric. Contrary to what MacKinnon et al. [15] reported, the parameter  $c_{xy}$  had an non-negligible influence on BK but was minor for the two other tests. This does not cast doubt on previous studies nevertheless: this result is more apparent because of the systematic investigation carried out herein.

**Table 18.** Type I error rate of the Sobel test (ST) with sample size  $n = 500$ . (Color online).

Path		$c_{xy}$									
$a_{xm}$	$b_{my x}$	.00	.10	.20	.30	.40	.50	.60	.70	.80	.90
.90	.00	.049	.044	.053	.050	.053	.046	.047	.053	.047	.048
.80	.00	.052	.049	.053	.050	.053	.052	.051	.047	.052	.050
.70	.00	.053	.046	.048	.045	.047	.053	.050	.054	.054	.052
.60	.00	.050	.048	.049	.047	.056	.049	.048	.047	.047	.052
.50	.00	.048	.042	.042	.047	.052	.042	.048	.057	.048	.045
.40	.00	.043	.049	.045	.046	.041	.042	.044	.041	.049	.043
.30	.00	.039	.046	.042	.040	.041	.042	.034	.040	.039	.040
.20	.00	.027	.028	.030	.026	.029	.028	.031	.026	.029	.025
.10	.00	.006	.006	.004	.006	.006	.004	.006	.006	.006	.005
.00	.00	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
.00	.10	.004	.005	.005	.007	.007	.008	.008	.012	.022	.031
.00	.20	.029	.028	.029	.028	.028	.032	.043	.041	.043	.045
.00	.30	.040	.042	.036	.042	.044	.040	.040	.044	.047	.050
.00	.40	.047	.041	.041	.040	.050	.047	.049	.049	.050	.052
.00	.50	.047	.048	.044	.048	.047	.049	.052	.046	.050	-
.00	.60	.052	.054	.043	.049	.052	.050	.048	.052	-	-
.00	.70	.050	.050	.048	.049	.053	.051	.049	.054	-	-
.00	.80	.051	.054	.050	.047	.051	.051	-	-	-	-
.00	.90	.051	.045	.054	.052	.050	-	-	-	-	-

**Table 19.** Summary of results.

Methods	Statistics					
	> .025	< .075	Mean	S.E.	Min	Max
Bootstrap test	.099	.147	.057	.019	.003	.093
Baron and Kenny test	.347	.000	.033	.019	.000	.060
Sobel test	.325	.000	.033	.019	.000	.059

The strength of the current study is the systematic evaluation (1056 scenarios) of three well-known and used statistical tests of an indirect effect. As well as supporting the previous methodological literature, it strengthens the conclusion that BT is the best methods and BK and ST should be avoided in most cases. A weakness of the current study, however, is that only simple mediation (a single independent variable, mediator and dependent variable relation) was tested. Future studies should try comparing the BT and the ST on more complex mediation models regarding their power and their type I error rate.

In conclusion, the BT methods is recommended as actually the best (among test used in the current study) to identify an indirect effect. It was the test with the most appropriate type I error rate when the sample size was at  $n = 50$ . As the sample size increases, the probability to commit false positive over .075 looks random. Otherwise, the BT was not overly conservative compared to BK and ST. In situations where a more conservative test would be necessary, it would be preferable to use BT with a stringent  $\alpha$  threshold than use other methods. As a reminder, strong statistical analyses do not supersede a strong experimental methodology. If the BT is recommended as a best practice in mediation analyses, methodology is also of importance.

**Disclosure statement**

No potential conflict of interest was reported by the author.

## ORCID

Pier-Olivier Caron  <http://orcid.org/0000-0001-6346-5583>

## References

- [1] Wright S. The method of path coefficients. *Ann Math Stat.* 1934;5:161–215.
- [2] Fritz MS, Kenny DA, Mackinnon DP. The combined effects of measurement error and omitting confounders in the single-mediator model. *Multivariate Behav Res.* 2016;51:281–697.
- [3] Huang J, Uuan Y. Bayesian dynamic mediation analysis. *Psychol Methods.* 2017;22:667–686.
- [4] Kelcey B, Dong N, Spybrook J, et al. Experimental power for indirect effects in group-randomized studies with group-level mediators. *Multivariate Behav Res.* 2017;52:699–719.
- [5] Lachowicz MJ, Preacher KJ, Kelley K. A novel measure of effect size for mediation analysis. *Psychol Methods.* 2018;23:244–261.
- [6] Montoya AK, Hayes AF. Two-condition within-participant statistical mediation analysis: a path-analytic framework. *Psychol Methods.* 2017;22:6–27.
- [7] Mackinnon DP, Fairchild AJ, Fritz MS. Mediation analysis. *Annu Rev Psychol.* 2007;58:593–614.
- [8] Rucker DD, Preacher KJ, Tormala ZL, et al. Mediation analysis in social psychology: current practices and new recommendations. *Soc Personal Psychol Compass.* 2011;5:359–371.
- [9] Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol.* 1986;51:1173–1182.
- [10] Sobel ME. Asymptotic confidence intervals for indirect effects in structural equation models. In: S Leinhardt, editor. *Social methodology.* Washington (DC): American Sociological Association; 1982.
- [11] Hayes AF. *Introduction to mediation, moderation and conditional process analysis.* New York (NY): Guilford; 2013.
- [12] Preacher KJ, Rucker DD, Hayes AF. Assessing moderated mediation hypotheses: theory, methods and prescriptions. *Multivariate Behavioral Research.* 2007;42:185–227.
- [13] Williams J, Mackinnon DP. Resampling and distribution of the product methods for testing indirect effects in complex models. *Struct Equ Modeling.* 2008;15:23–51.
- [14] Efron B, Tibshirani R. *An introduction to the bootstrap.* New York (NY): Chapman & Hall; 1979.
- [15] Mackinnon DP, Lockwood CM, Hoffman JM, et al. A comparison of methods to test mediation and other intervening variable effects. *Psychol Methods.* 2002;7:83–104.
- [16] R Core Team. *R: a language and environment for statistical computing.* Vienna: R Foundation for Statistical Computing, Vienna, Austria; 2017. Available at <http://www.Rproject.org/>
- [17] Achim A. Testing the number of required dimensions in exploratory factor analysis. *Quant Methods Psychol.* 2017;13:64–74.
- [18] Caron P-O. A Monte Carlo examination of the broken-stick distribution to identify components to retain in principal component analysis. *J Stat Comput Simul.* 2016;86:2405–2410.
- [19] Mackinnon DP, Warsi G, Dwyer JH. A simulation study of mediated effect measures. *Multivariate Behav Res.* 1995;30:41–62.
- [20] Peres-Neto PR, Jackson DA, Somers KM. How many principal components? stopping rules for determining the number of non-trivial axes revisited. *Comput Stat Data Anal.* 2005;49:974–997.
- [21] Howell DC. *Statistical methods for psychology.* Belmont: Wadsworth; 2012.
- [22] Caron P-O, Valois P. A computational description of simple mediation analysis. *Quant Methods Psychol.* 2018;14:147–158.