

RMediation: An R package for mediation analysis confidence intervals

Davood Tofighi · David P. MacKinnon

Published online: 13 April 2011
© Psychonomic Society, Inc. 2011

Abstract This article describes the RMediation package, which offers various methods for building confidence intervals (CIs) for mediated effects. The mediated effect is the product of two regression coefficients. The distribution-of-the-product method has the best statistical performance of existing methods for building CIs for the mediated effect. RMediation produces CIs using methods based on the distribution of product, Monte Carlo simulations, and an asymptotic normal distribution. Furthermore, RMediation generates percentiles, quantiles, and the plot of the distribution and CI for the mediated effect. An existing program, called PRODCLIN, published in *Behavior Research Methods*, has been widely cited and used by researchers to build accurate CIs. PRODCLIN has several limitations: The program is somewhat cumbersome to access and yields no result for several cases. RMediation described herein is based on the widely available R software, includes several capabilities not available in PRODCLIN, and provides accurate results that PRODCLIN could not.

Keywords Mediation · Indirect effect · R · Confidence intervals

D. Tofighi (✉)
School of Psychology,
Georgia Institute of Technology,
J.S. Coon Bldg. 654 Cherry Street,
Atlanta, GA 30332-0170, USA
e-mail: dtofighi@psych.gatech.edu

D. P. MacKinnon
Department of Psychology, Arizona State University,
950 S. McAllister,
Tempe, AZ 85287-1104, USA
e-mail: david.mackinnon@asu.edu

The use of mediation analysis in basic and applied research has been increasing (Baron & Kenny, 1986, has over 20,000 citations). In a mediation model, an independent variable (e.g., drug prevention intervention) is hypothesized to change a mediator (drug use norm among peers), which in turn changes an outcome (e.g., illicit drug use). Under certain assumptions, the mediated effect is the effect of the intervention on the outcome that is transmitted through the mediator. One important issue in mediation studies is to build confidence intervals (CIs) and test hypotheses regarding various effects (e.g., the mediated effect). There are several methods in the literature for computing CIs for the mediated effect. These methods can be roughly categorized into four groups: (1) the distribution of the product (e.g., MacKinnon, Fritz, Williams, & Lockwood, 2007; MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002), (2) the Monte Carlo method (MacKinnon, Lockwood, & Williams, 2004), (3) resampling methods (e.g., bootstrap resampling; MacKinnon et al., 2004), and (4) the asymptotic normal distribution method. Of these methods, the distribution of product has been shown to produce CIs with higher coverage rates, especially when the sample size is small (e.g., 50 or less; MacKinnon et al., 2002; MacKinnon et al., 2004). MacKinnon, Fritz, et al. adapted FORTRAN code¹ to form a computer program, called *PRODCLIN*, that computes the CIs for mediated effects, using the results in Meeker and

¹ Allen Miller wrote the FORTRAN code, named FNPROD, that uses the results by Meeker and Escobar (1994). PRODCLIN uses the FNPROD code.

Escobar (1994). To the best of our knowledge, PRODCLIN is the only computer program that produces CIs on the basis of the distribution-of-the-product method.

However, the PRODCLIN program (MacKinnon, Fritz, et al., 2007) has some limitations. First, popular statistical software packages, such as SPSS, SAS, and R (R Development Core Team, 2010), cannot directly run the PRODCLIN program. Instead, these packages need to run PRODCLIN as an external program and then upload the results to their “native” environments. This process can be cumbersome for users not familiar with running an external program from another statistical package. Furthermore, the PRODCLIN program is limited in that it does not produce CIs for some mediated effects for certain values of means and standard errors. Finally, the algorithm implemented in PRODCLIN has some limitations in producing CIs for the product of coefficients that are correlated. Note that we fixed the issues in the algorithm implemented in the PRODCLIN program. The new version of the PRODCLIN program is implemented in the RMediation package.²

The purpose of this study is to introduce an R (R Development Core Team, 2010) package called *RMediation*. The RMediation package provides a variety of methods for computing CIs, percentiles, and quantiles for the product of two normal random variables and the mediated effect. R is a freely available statistical software package that has become increasingly popular. R can be installed on various operating systems, such as different versions of MS Windows, Apple’s Mac OS X, and Linux-based systems such as Ubuntu. RMediation can readily be installed via the Internet onto any computer running the R software program. In addition, we conducted a small-scale simulation study that compared several methods of producing 95% CIs for mediated effects. These methods included the distribution-of-product method (MacKinnon et al., 2002), the Monte Carlo method (MacKinnon et al., 2004), the asymptotic normal distribution (AND) method, and three bootstrap methods: the percentile, bias-corrected bootstrap (BC), and accelerated bias-corrected bootstrap (BC_a).

The RMediation package employs three methods for producing CIs for the product of two normal random variables (e.g., mediated effects): (1) the distribution-of-product approach introduced by MacKinnon et al. (2002), (2) the Monte Carlo method, and (3) the AND method. The

distribution-of-product method is implemented using two computer programs: PRODCLIN (MacKinnon, Fritz, et al., 2007), and R Distribution of Product (RDOP), which is an R program we wrote to implement the distribution-of-product method using the results in Meeker and Escobar (1994). A user can specify a significance level, the means and the standard errors for the random variables X and Y , and the correlation between the two variables. Furthermore, RMediation provides quantiles and percentiles for the distribution of the product of two normal random variables, using the distribution-of-product method and the Monte Carlo simulations. We present a method in RMediation for calculating the Monte Carlo error so that a user can modify the level of accuracy for the percentiles and quantiles.

We also fixed the error that caused the PRODCLIN program to yield no results for the mediated effects with certain means and standard errors and for cases in which the two coefficients were correlated. The improvement was implemented in the algorithm generating the upper and lower confidence limits. Finally, RMediation produces a kernel density plot of the empirical distribution of the mediated effect and an overlaid plot of the associated CI with error bars (see Fig. 1). Such plots can help researchers visualize the uncertainty associated with the estimated mediated effect.

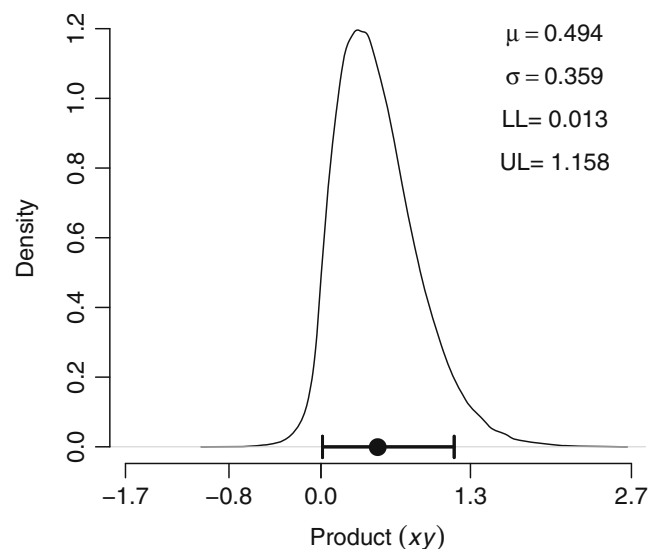


Fig. 1 Kernel density plot of the distribution of the product of two normal variables (i.e., mediated effect) and the 90% CI with error bars for the mediated effect, with $\hat{a} = 0.295$, $\hat{b} = 1.673$, $SE(\hat{a}) = 0.163$, and $SE(\hat{b}) = 0.695$. *LL* lower limit, *UL* upper limit

² The new version of MS DOS-based PRODCLIN program will be available for download from <http://www.public.asu.edu/davidpm/ripl/Prodclin/> and <http://amp.gatech.edu/RMediation>.

Single-mediator model

In single-level randomized controlled trials with two groups (e.g., intervention vs. control), a single-mediator model is defined as follows: An independent variable (e.g., $X = 1$, if a person participates in resistance skill program; otherwise, 0) is hypothesized to change a mediator (e.g., $M =$ drug refusal skill) that, in turn, changes an outcome variable (e.g., $Y =$ frequency of drug use). Three equations³ used to assess quantities in a single-mediator model are shown below (Baron & Kenny, 1986; MacKinnon, 2008):

$$Y_i = d_1 + c X_i + \varepsilon_{1i} \quad (1)$$

$$M_i = d_2 + a X_i + \varepsilon_{2i} \quad (2)$$

$$Y_i = d_3 + c' X_i + b M_i + \varepsilon_{3i}, \quad (3)$$

where Y_i is the outcome variable measured on individual i , X_i is an indicator variable that represents whether the i^{th} person received the intervention (1 = program; 0 = control), and M_i is the mediator. The coefficient c in Eq. 1 represents the total effect of the prevention program on drug use. The coefficient c' in Eq. 3 represents the direct effect of the prevention program on drug use, controlling for the participants' refusal skills. The direct effect captures the difference between treatment and control group adjusted for participants' refusal skills and indicates the part of the program effect not accounted for by the mediator; the coefficient b describes the effect of refusal skills on drug use, controlling for the program effect; the coefficient a in Eq. 2 represents the degree to which the intervention increased refusal skills, relative to the control group. ε_{1i} , ε_{2i} , and ε_{3i} denote the residual terms; the coefficients d_1 , d_2 , and d_3 are the intercepts.

The magnitude of the effect of the prevention program on decreasing drug use mediated by the individuals' refusal skills is represented by $a b$ (MacKinnon & Dwyer, 1993). The total effect of the prevention program on decreasing drug use is $c = a b + c'$ (Alwin & Hauser, 1975). A key interest in prevention studies is to test the mediated effect $a b$. A significant mediated effect provides

evidence consistent with the theory: The preventive intervention changed the mediator, thereby altering the outcome.

The estimators of the parameters in Eqs. 1, 2 and 3 can be obtained using the least squares or the maximum likelihood method. The estimator of the mediated effect $a b$ is shown by $\hat{a} \hat{b}$, where “^” denotes the estimator of each respective parameter. Another estimator of the mediated effect is $\hat{c} - \hat{c}'$. Under certain conditions, the following expression holds: $\hat{a} \hat{b} = \hat{c} - \hat{c}'$ (MacKinnon, Warsi, & Dwyer, 1995). It is assumed that the equations represent the true underlying mediation relations satisfying statistical and inferential assumptions (see MacKinnon, 2008, Chaps. 3 and 13 for more on these assumptions).

Hypothesis testing

Testing hypotheses in a single-mediator model has received extensive attention (MacKinnon et al., 2002). In classical statistics, researchers are often interested in testing whether a parameter or a function of parameters is significantly different from zero. Researchers have recently emphasized using CIs, as well as reporting p values for hypothesis testing (Harlow, Mulaik, & Steiger, 1997; Wilkinson & the Task Force on Statistical Inference, 1999). While classical hypothesis testing provides reject/not-reject decision for null hypothesis using test statistics, CIs also provide an interval estimate that represents uncertainty in estimating the quantities of interest in a single-mediator model. CIs can also be used in hypothesis testing. This section discusses three methods for building CIs for the mediated effect that were implemented in the RMediation package.

Distribution of the product

MacKinnon et al. (2002) proposed the distribution-of-product method for building a CI for the mediated effect. In addition, MacKinnon, Fritz, et al. (2007) introduced the PRODCLIN program, which produced CIs for the mediated effect on the basis of the distribution-of-product method, using the analytical method proposed by Meeker and Escobar (1994). This section describes a few methods for evaluating the cumulative distribution function (CDF) for the distribution of the product of two normal random variables, including the one used in the RMediation package and the PRODCLIN program.

First, let us define the CDF of the product of two normal random variables. Let variables X and Y have a bivariate normal distribution. Also, let μ_X and μ_Y be the means, σ_X and σ_Y be the standard deviations, and $-1 < \rho < 1$ be the correlation between X and Y . To simplify the derivation of

³ Only Eqs. 2 and 3 are required to estimate a single-mediator model. Equation 1 is used to estimate the total effect.

the distribution of product XY , we make the variables scale free by dividing each variable by its respective standard deviation. That is,

$$U = \frac{X}{\sigma_X}, \text{ and } V = \frac{Y}{\sigma_Y}.$$

Let $Z = UV$. The relationship between the CDF of the product XY and that of Z is as follows:

$$\Pr(XY \leq k) = \Pr\left\{\frac{X}{\sigma_X} \frac{Y}{\sigma_Y} \leq \frac{k}{\sigma_X \sigma_Y}\right\} = \Pr(Z \leq z),$$

where $z = k/(\sigma_X \sigma_Y)$. Note that $(U, V)^T$ has a bivariate normal distribution:

$$\begin{pmatrix} U \\ V \end{pmatrix} \sim BVN\left(\begin{pmatrix} \mu_U \\ \mu_V \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right) \tag{4}$$

where $\mu_U = \mu_X/\sigma_X$ and $\mu_V = \mu_Y/\sigma_Y$.

Now let $F_Z(q)$ be the CDF of Z . The CDF of Z is defined as follow:

$$F_Z(q) = \iint_A f_{U,V}(u, v|\boldsymbol{\mu}, \boldsymbol{\Sigma}) du dv, \tag{5}$$

where $A = \{(u, v) \in \mathbb{R}^2 : u \times v \leq q\}$, $f_{U,V}(u, v|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the bivariate normal probability density function (PDF) for (U, V) , and $q \in \mathbb{R}$ is a quantile.

There are several methods for evaluating the distribution of the product in Eq. 5. Craig (1936) provided an analytical method for evaluating the CDF of the product of two normal random variables in Eq. 5. According to Craig, the mean and variance of Z are as follows:

$$\begin{aligned} \mu_Z &= \mu_U \mu_V + \rho \\ \sigma_Z^2 &= \mu_U^2 + \mu_V^2 + 2\mu_U \mu_V \rho + 1 + \rho^2 \end{aligned} \tag{6}$$

When either X or Y has a mean of zero, the distribution of Z is approximately proportional to the Bessel function of the second kind of zero order with a purely imaginary argument. The shape of the distribution is symmetric around the mean of zero. On the other hand, when neither X or Y has a mean of zero and X and Y are independent ($\rho = 0$), the mean and variance of Z are as follows:

$$\begin{aligned} \mu_Z &= \mu_U \mu_V \\ \sigma_Z^2 &= \mu_U^2 + \mu_V^2 + 1 \end{aligned}$$

In addition, Meeker, Cornwell, and Aroian (1981) provided a numerical algorithm for evaluating the CDF of

Z . The numerical method directly evaluates the double integral in Eq. 5, using an adaptive Romberg integration method with an absolute error tolerance of $1.0E-10$. Meeker et al. also provided tables of quantiles for the distribution of a standardized variable:

$$W = \frac{Z - \mu_Z}{\sigma_Z}.$$

Finally, Meeker and Escobar (1994) provided a simpler method for evaluating the CDF in Eq. 5 (note that both RMediation and PRODCLIN employ this method). They simplified Eq. 5 as follows:

$$F_Z(q) = \int_{-\infty}^{\infty} \phi(u - \mu_U) \Phi\left[\text{sign}(u) \frac{q/u - \mu_{V|u}}{\sqrt{1 - \rho^2}} du\right] \tag{7}$$

where ϕ and Φ are the PDF and CDF of the standard normal distribution, respectively. $\mu_{V|u}$ is the conditional mean of V , which equals $\mu_V + \rho(u - \mu_U)$. $\text{sign}(\cdot)$ is the sign function, and $-1 < \rho < 1$.

Monte Carlo method

Another method for evaluating the CDF of a product of two normal variables in Eq. 5 is to use the Monte Carlo method. In this section, we also present a method for calculating the associated Monte Carlo error. Using the Monte Carlo method to evaluate the CDF of the product of two normal variables requires reformulating Eq. 5 as follows:

$$\begin{aligned} F_Z(q) &= \iint_A f_{U,V}(u, v|\boldsymbol{\mu}, \boldsymbol{\Sigma}) du dv \\ &= \iint I_A(u, v) f_{U,V}(u, v|\boldsymbol{\mu}, \boldsymbol{\Sigma}) du dv, \end{aligned}$$

where $I_A(u, v)$ is the indicator function defined as follows:

$$I_A(u, v) = \begin{cases} 1 & \text{if } (u, v) \in A, \\ 0 & \text{if } (u, v) \notin A. \end{cases}$$

To illustrate, suppose that we simulate a random sample of $(u^1, v^1), \dots, (u^m, v^m)$ from the bivariate normal distribution in Eq. 4. The Monte Carlo estimate of the percentile p for the quantile q is given by

$$p = \frac{\sum_{j=1}^m I_A(u^j, v^j)}{m}.$$

The associated Monte Carlo (simulation) error of the percentile estimate p is given by

$$SE_p = \sqrt{\frac{\sum_{j=1}^m (I_A(u^j, v^j) - p)^2}{m(m - 1)}}.$$

Note that the Monte Carlo error depends on m , which is controlled by the user. As m becomes larger, the Monte Carlo error becomes smaller.

Asymptotic normal distribution method

Another approach to testing the mediated effect and producing CIs is to use the asymptotic properties of the ML estimator of a b and form a z test statistic. In this approach, $z = (\hat{a}\hat{b})/SE(\hat{a}\hat{b}) \sim N(0, 1)$, where \sim means approximately and $SE(\hat{a}\hat{b})$ is the standard error of the estimator of ab . As the sample size increases, z converges in distribution to the standard normal distribution. There are various methods for calculating $SE(\hat{a}\hat{b})$ (MacKinnon, 2008). RMediation uses the variance of the product of two normal random variables presented in Eq. 6. For the mediated effect, because the covariance between \hat{a} and \hat{b} is zero (Tofighi, MacKinnon, & Yoon, 2009), the standard error of the mediated effect is simplified as follows:

$$SE(\hat{a}\hat{b}) = \sqrt{(a SE(\hat{b}))^2 + (b SE(\hat{a}))^2 + SE(\hat{a})^2 SE(\hat{b})^2}.$$

The asymptotic 95% CI for the mediated effect is $\hat{a}\hat{b} \pm 1.96 \times SE(\hat{a}\hat{b})$.

RMediation package

The RMediation package provides functions for computing $(1 - \alpha)\%$ CIs, percentiles, and quantiles for the distribution of the product of two normal random variables. To install

```
> medci(mu.x = 0.295, mu.y = 1.673, se.x = 0.163, se.y = 0.695, rho = 0, alpha = 0.1, type = "prodclin")
```

```
5 %    95%
0.01341 1.15797
```

The arguments `mu.x` and `mu.y` refer to the means for the first and second variables, respectively, which correspond to the estimates of a and b paths, respectively. The arguments `se.x` and `se.y` specify the standard deviations for the first and second variables, respectively, which correspond to the standard errors for the estimates for a and b paths, respectively. The argument `rho` specifies the correlation between two variables with the default value of 0. The argument `alpha` is the significance level for the CI with the default value of .05. The argument `type` takes on the values “`prodclin`”

the RMediation in R (R Development Core Team, 2010), one needs to be connected to the Internet. To install RMediation within R, use the following function: `install.packages("RMediation")`. The name of the package should be specified in quotation marks. Also note that the commands used in the R environment are called *functions* and are case sensitive. Each function (command) accepts *arguments* to be specified in parentheses after the name of the function. Arguments modify the behavior of a function. If there is more than one argument, the arguments need to be separated by commas. To assign a value to an argument, “=” is placed after the name of the argument and before the value.

To use RMediation, load the package into the R environment. To do that, use the R function `library`. One of the arguments of the function `library` is `package`, whose value must be set to the name of the package to be loaded:

```
> library(package = "RMediation").
```

To illustrate, consider an example where the previous version of the PRODCLIN program did not yield the desired results. Suppose that we want to find a 90% CI for a mediated effect where $\hat{a} = 0.295$, $\hat{b} = 1.673$, $SE(\hat{a}) = .163$, and $SE(\hat{b}) = 0.695$. In RMediation, the function `medci` produces CIs for the product of two normal variables and mediated effects. The following command produces a 90% CI, using the new version of the PRODCLIN program:

(default, the PRODCLIN program, MacKinnon, Fritz, et al., 2007), “`DOP`” (the RDOP program), “`MC`” (the Monte Carlo approach), “`asymp`” (the AND method), or “`all`” (using all four methods). It is important to note that the values for the argument `type` must be enclosed in single or double quotation marks.

In the example above, a user can also choose to not specify a value for the optional arguments `type` and `rho`, because the default values for these arguments are “`prodclin`” and 0, respectively. The previous command can be shortened as follows:

```
> medci(mu.x = 0.295, mu.y = 1.673, se.x = 0.163, se.y = 0.695, alpha = 0.1).
```

On the other hand, if a user needs values other than the defaults for the optional arguments, the person needs to explicitly specify these arguments. Suppose that you want the 90% CI for the product of two normal random variables

with the means equal to 0.2 and 0.4, standard deviations equal to 1 for both, and the correlation equal to .1, using all the available methods in the RMediation package. The specifications of `medci` arguments are as follows:

```
> medci(mu.x = 0.2, mu.y = 0.4, se.x = 1, se.y = 1, rho = 0.1, alpha = 0.1, type = "all")
```

```
$PRODCLIN
```

```
5 % 95 %
```

```
-1.393 2.121
```

```
$`Distribution of Product`
```

```
5 % 95 %
```

```
-1.393 2.121
```

```
$`Monte Carlo`
```

```
5 % 95 %
```

```
-1.389 2.135
```

```
$`Asymptotic Normal`
```

```
5 % 95 %
```

```
-1.741 1.901
```

Another capability of the function `medci` is to produce a graph for the distribution of the product. The plot uses the kernel density method with a standard normal distribution as the kernel function to estimate the PDF of the product of two normal random variables.⁴ To obtain a density plot, one

needs to set the argument `plot=TRUE`. At the same time, the argument `plotCI=TRUE` in the `medci` function overlays the plot of $(1-\alpha)\%$ CI with error bars on the density plot. The following command produces the density plot with an overlaid plot of the 90% CI, as shown in Fig. 1:

```
> medci(mu.x = 0.295, mu.y = 1.673, se.x = 0.163, se.y = 0.695, alpha = 0.1, plot = TRUE, plotCI = TRUE)
```

```
5 % 95 %
```

```
0.01341 1.15797
```

In addition, the RMediation package provides quantiles and percentiles of the distribution of the product of two normal random variables. The function `qprodnormal` computes the quantile for the distribu-

tion of product. The argument `type` in `qprodnormal` takes on the following values: “`prodclin`” (default, the PRODCLIN program; MacKinnon, Fritz, et al., 2007), “`DOP`” (the RDOP program), “`MC`” (the Monte Carlo approach), and “`all`” (using all three methods). To illustrate, suppose that we want to find quantiles for the probability $p = .975$ for the mediated effect, where $\hat{a} = 0.2$, $\hat{b} = 0.4$, $SE(\hat{a}) = 1$, and $SE(\hat{b}) = 1$. The following command produces the quantiles corresponding to $p =$

⁴ For more information about the kernel density estimation method, type `help(density)` within R.

.975, using all three methods with the associated numerical errors:

```

> qprodnormal(p = 0.975, mu.x = 0.2, mu.y = 0.4, se.x = 1, se.y = 1,
type = "all")

$PRODCLIN
$PRODCLIN$q
[1] 2.587

$PRODCLIN$error
[1] 2.239e-15

$`Distribution of Product`
$`Distribution of Product`$q
[1] 2.587

$`Distribution of Product`$error
[1] 1.340e-06

$`Monte Carlo`
$`Monte Carlo`$q
[1] 2.564

$`Monte Carlo`$error
[1] 1.561e-06

```

For example, the quantile for the probability $p = .975$ using both “`prodclin`” and “`DOP`” is equal to 2.587.

The function `pprodnormal` produces percentiles for the distribution of the product in the RMediation package by specifying the argument `type` to the following values:

“`prodclin`” (default, the PRODCLIN program, MacKinnon, Fritz, et al., 2007), “`DOF`” (RDOP program), “`MC`” (the Monte Carlo approach), or “`all`” (using all three methods). Let us find the percentile for the quantile $q = 2.587$ in the previous example:

```

> pprodnormal(q = 2.587, mu.x = 0.2, mu.y = 0.4, se.x = 1, se.y = 1)

$p
[1] 0.975

$error
[1] 2.128e-15

```

As was expected, the percentile for $q = 2.587$ is equal to 0.975.

Simulation study

We conducted a simulation study to compare 95% CIs produced by the methods in the RMediation program with

the CIs based on the three bootstraps methods in terms of the Type I error rates and the length of CIs. The simulation study followed a design similar to the one used by MacKinnon, Fritz, et al. (2007). Data for the single-mediator model were generated when $a = 0$ and $b = 0, .14$ (small), $.39$ (medium), and $.59$ (large). The value of c' was fixed at zero. The sample size took on the values 50, 100, and 200.

Data generation was performed in R (R Development Core Team, 2010). Independent random data for X , ε_{i2} , and ε_{i3} were generated from the standard normal distribution using the `rnorm` function in R. Values for M and Y were generated on the basis of the mediation model in Eqs. 2 and 3. Within each condition, 1,000 data sets were created. Each data set was analyzed using the model in Eqs. 2 and 3, and seven sets of 95% CIs were computed for each data set. The bootstrap methods used 1,000 bootstrap samples of each data set.

We used the standardized length of a CI, which is defined as d/SE , where d is the length of the CI and SE is the standard error for a particular method. Note that d for the AND CIs is constant and equals 3.920. For the PRODCLIN, RDOP, Monte Carlo, and AND methods, SE is computed using the formula in Eq. 6. For the bootstrap methods, SE is the standard error of a sample of 1,000 bootstrap mediated effect estimates. The results are shown in Table 1.

Note that the results for the PRODCLIN, RDOP, and Monte Carlo methods were similar, as were those for the BC and BC_a bootstrap methods. To save space, we presented the results for only the PRODCLIN and BC bootstrap methods. As can be seen in Table 1, the AND method yielded the most conservative CIs across all conditions. For the conditions $b = .14$ (small), the Type I error rates of CIs for the PRODCLIN, the percentile bootstrap, and the BC bootstrap were below the nominal value of .05. For the conditions $b = .39$ (medium) and .59 (large), the PRODCLIN Type I error rates were closest to the nominal value of .05, followed by the percentile and then the BC bootstrap method. The BC (BC_a) had inflated

Type I error rates above the nominal value of .05. As for the length of the CIs, the AND method had the shortest length, followed by PRODCLIN, the percentile bootstrap, and then the BC bootstrap method.

Conclusion

The present study provided a tutorial on how to use the RMediation package, using hypothetical numerical examples. The RMediation package provides functions to compute CIs, percentiles, and quantiles for the distribution of the product of two normal random variables based on the results of MacKinnon et al. (2002). In addition, the RMediation package produces a plot of the empirical distribution of the mediated effect and the overlaid plot of associated CI with error bars (see Fig. 1). These plots can help researchers visualize the uncertainty associated with the mediated effect. The RMediation package can be used in any situation where aspects of the distribution of the product of two random variables is of interest, such as the distribution of interaction variables formed by the product of two main effect variables and the distribution of scales formed by the product of two individual scales.

Overall, we recommend the distribution-of-product method over the AND and the bootstrap methods, especially for smaller sample sizes (e.g., 50). The bootstrap methods are not recommended because the analytical solution for testing the mediated effect already exists and is implemented in the RMediation package. In addition, for sample sizes less than 100, the bootstrap methods may

Table 1 Proportion coverage and standardized length for 95% CIs ($a = c' = 0$)

b	N	Percentile	BC	PRODCLIN	AND
0	50	0.998(4.172)	0.984(4.405)	0.998(4.212)	1(3.920)
0	100	0.999(4.211)	0.996(4.441)	0.999(4.222)	1(3.920)
0	200	1(4.219)	0.997(4.465)	1(4.226)	1(3.020)
0	500	1(4.233)	0.996(4.479)	1(4.230)	1(3.920)
0.14	50	0.991(4.145)	0.982(4.367)	0.994(4.175)	0.998(3.920)
0.14	100	0.982(4.128)	0.968(4.336)	0.988(4.137)	1(3.920)
0.14	200	0.983(4.085)	0.951(4.272)	0.986(4.090)	0.997(3.920)
0.14	500	0.968(4.030)	0.940(4.120)	0.968(4.029)	0.998(3.920)
0.39	50	0.966(4.037)	0.940(4.157)	0.973(4.046)	0.996(3.920)
0.39	100	0.944(4.006)	0.922(4.056)	0.943(4.001)	0.979(3.920)
0.39	200	0.946(3.964)	0.925(3.982)	0.949(3.969)	0.971(3.920)
0.39	500	0.949(3.945)	0.945(3.946)	0.949(3.942)	0.955(3.920)
0.59	50	0.935(3.995)	0.915(4.037)	0.951(3.993)	0.971(3.920)
0.59	100	0.947(3.969)	0.932(3.980)	0.945(3.964)	0.965(3.920)
0.59	200	0.947(3.942)	0.939(3.946)	0.956(3.943)	0.958(3.920)
0.59	500	0.942(3.932)	0.934(3.933)	0.942(3.929)	0.949(3.920)

The numbers in parenthesis are the standardized lengths of the CIs. RMediation produced the CIs for all the methods except for the bootstrap. CI confidence interval, BC bias-corrected bootstrap, AND asymptotic normal distribution

result in undercoverage (i.e., coverage less than 95%) for the CIs, since the confidence limits vary considerably across the bootstrap samples (Good, 2006, Chap. 2). The undercoverage of the bootstrap methods has been corroborated by our small-scale simulation study and is consistent with past research (MacKinnon et al., 2004).

Author Note The research was supported by Award F31DA027336 from the National Institute on Drug Abuse to Davood Tofighi and PHS DA09757 to David P. MacKinnon.

References

- Alwin, D. F., & Hauser, R. M. (1975). Decomposition of effects in path analysis. *American Sociological Review*, *40*, 37–47.
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182.
- Craig, C. C. (1936). On the frequency function of xy . *Annals of Mathematical Statistics*, *7*, 1–15.
- R Development Core Team (2010). R: A language and environment for statistical computing [Computer software]. Vienna, Austria. Available from <http://www.R-project.org>
- Good, P. I. (2006). *Resampling methods: A practical guide to data analysis* (3rd ed.). Boston: Birkhäuser.
- Harlow, L., Mulaik, S. A., & Steiger, J. H. (Eds.). (1997). *What if there were no significance tests?* Mahwah: Erlbaum.
- MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York: Erlbaum.
- MacKinnon, D. P., & Dwyer, J. H. (1993). Estimating mediated effects in prevention studies. *Evaluation Review*, *17*, 144–158.
- MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology*, *58*, 593–614.
- MacKinnon, D. P., Fritz, M. S., Williams, J., & Lockwood, C. M. (2007). Distribution of the product confidence limits for the indirect effect: Program PRODCLIN. *Behavior Research Methods*, *39*, 384–389.
- MacKinnon, D. P., Lockwood, C. M., Hoffinan, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*, *7*, 83–104.
- MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, *39*, 99–128.
- MacKinnon, D. P., Warsi, G., & Dwyer, J. H. (1995). A simulation study of mediated effect measures. *Multivariate Behavioral Research*, *30*, 41–62.
- Meeker, W. Q., Cornwell, L. W., & Aroian, L. A. (1981). *Selected tables in mathematical statistics: The product of two normally distributed random variables*. Providence: American Mathematical Society.
- Meeker, W. Q., & Escobar, L. A. (1994). An algorithm to compute the cdf of the product of two normal random variables. *Communications in Statistics: Simulation and Computation*, *23*, 271–280.
- Tofighi, D., MacKinnon, D. P., & Yoon, M. (2009). Covariances between regression coefficient estimates in a single mediator model. *The British Journal of Mathematical and Statistical Psychology*, *62*, 457–484.
- Wilkinson, L., & the Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *The American Psychologist*, *54*, 594–604.