



Published in final edited form as:

Eval Health Prof. 2022 March ; 45(1): 54–65. doi:10.1177/01632787211070811.

A Randomization Permutation Test for Single Subject Mediation

David P. MacKinnon, Ph.D.¹, Heather L. Smyth¹, Jennifer Somers, Ph.D.¹, Emily Ho, Ph.D.², Julia Norget³, Milica Mio evi , Ph.D.⁴

¹Department of Psychology, Arizona State University, Tempe, AZ, USA

²Department of Medical Social Sciences, Northwestern University, Chicago, IL

³Psychological Methods and Evaluation, Bielefeld University, Bielefeld, Germany

⁴Department of Psychology, McGill University, Montreal, QC, Canada

Abstract

In response to the importance of individual-level effects, the purpose of this paper is to describe the new randomization permutation (RP) test for a mediation mechanism for a single subject. We extend seminal work on permutation tests for individual-level data by proposing a test for mediation for one person. The method requires random assignment to the levels of the treatment variable at each measurement occasion, and repeated measures of the mediator and outcome from one subject. If several assumptions are met, the process by which a treatment changes an outcome can be statistically evaluated for a single subject, using the permutation mediation test method and the permutation confidence interval method for residuals. A simulation study evaluated the statistical properties of the new method suggesting that at least eight repeated measures are needed to control Type I error rates and larger sample sizes are needed for power approaching .8 even for large effects. The RP mediation test is a promising method for elucidating intraindividual processes of change that may inform personalized medicine and tailoring of process-based treatments for one subject.

Keywords

single subject; mediating mechanisms; active ingredients; idiographic methods; permutation tests

The purpose of this paper is to describe and apply a method for investigating a mediating process with data from a single participant. We first describe the background for mediation analysis and the motivation for single subject mediation analysis. We review randomization and permutation tests and describe an application of the randomization and permutation tests to the case of mediation analysis from one participant. The test is described as the RP (randomization and permutation) test because it includes a randomization test for the

Corresponding Author: David P. MacKinnon, Department of Psychology, Arizona State University, PO Box 871104, Mcalister ave, Tempe, AZ 85287-1104, USA., davidpm@asu.edu.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Supplemental Material

Supplemental material for this article is available online.

independent variable to mediator and a permutation test for the mediator to the outcome variable. The method is applied to several example data sets and strengths and limitations of the method are discussed.

Statistical mediation analysis helps answer questions related to how or why an intervention achieves an effect. Knowledge of mediational mechanisms can inform understanding of how treatments promote wellbeing and suggest opportunities for fine-tuning effective interventions to maximize their impact and minimize their costs. Mediation analysis is useful in addressing questions about whether the intervention changed the putative mechanism of interest and whether that mechanism influenced the targeted outcome (O'Rourke & MacKinnon, 2018). Mediation analysis decomposes the total effect of an intervention on a targeted outcome into a direct effect and an indirect effect through a mediator. The indirect effect represents the effect of an intervention on a mediating variable, and the effect of this mediator on an outcome. Examples of indirect effects are effects of an intervention on craving which then reduces drug use (Mayhugh et al., 2018), effects of an intervention on pain catastrophizing which then increases physical activity (Leeuw et al., 2007), among many other examples (MacKinnon, 2008; Maric, de haan, et al., 2015; Maric, Prins, & Ollendick, 2015). The importance of mediation analysis for theory and practice has made it a rapidly growing area of methods development and application (MacKinnon, 2008; VanderWeele, 2015). Whereas many statistical tests exist for evaluating mediational processes by which interventions achieve their effects on groups of individuals, there are few statistical tests for evaluating mediational processes for a single participant. Some methods are available that estimate effects for groups that can also assess mediation for individuals. For example, the 1-1-1 multilevel model includes both group-level and individual-level mediated effects (MacKinnon, 2008; Preacher et al., 2010). However, this method requires a data set with repeated measures of X, M, and Y for each participant in order to estimate the group-level and individual-level mediated effects. There are few methods for mediation analysis for one subject, although the topic of mediation analysis for single subjects is starting to receive more research attention (e.g., Gaynor & Harris, 2008; Geuke, Maric, Mio evi , Wolters, & de Haan, 2019; Maric, Prins, & Ollendick, 2015; Mio evi et al., 2020; Vuorre & Bolger, 2018) and the method described in this paper was originally presented at a meeting devoted to single subject causal mediation analysis (MacKinnon, 2019). In this paper, we propose a randomization permutation (RP) mediation test on data collected from single-case experimental designs in order to evaluate mediation for a single subject.

In the following sections, we provide background on mediation analysis. We describe the RP test for single-case experimental designs. We then provide motivating examples of single-case experiments with random assignment of treatment doses or conditions to treatment times, before describing the RP mediation test. We describe the results of a simulation study evaluating the performance of the RP method as a function of number of measurement occasions, effect size, and different amounts of dependency across time. SAS programs for generic RP tests are available in the Appendix in Taylor and MacKinnon (2012) and programs for the examples in this article are included in the Supplemental Material Appendix.

Current Approaches to Mediation

Intervention research, including analysis of mediators of intervention effects, typically draws inferences from an observed sample to generalize to the population. Nomothetic or variable-oriented approaches yield results at the population level, which can inform policy decisions and group-level intervention programs. Traditional mediation analysis from a variable-oriented approach assumes a homogeneous mediating process within the population; in other words, each person undergoing the mediation process is assumed to do so at the same rate and to the same degree. However, inferences about psychological processes made at the population level may not generalize to a specific individual (Cattell, 1952), a fallacy known as the ecological fallacy (Molenaar & Campbell, 2009).

Idiographic or person-centered methods may inform treatment decisions for a specific patient and may lead to more evidence-based clinical decision making (Barlow et al., 2009; Gast, 2010; Lei et al., 2012; Maric et al., 2015; Molenaar, 2004). Idiographic methods require collecting repeated measures from an individual subject, and there has been increased attention given to the importance of individual subject experimental studies for improving evidence-based therapeutic practices (Barlow et al., 2009). Echoing Gordon Allport's call to return to the individual (Allport, 1962), Barlow et al. (2009) detail experimental strategies designed to maximize the scientific yield from single subject studies. Evidence from single-case experimental designs (often referred to as "*N*-of-1 randomized controlled trials") is considered one of the most rigorous forms of evidentiary support for therapies, according to international evidence-based guidelines (Onghena et al., 2019). By randomly assigning treatment times to conditions or doses, experimental manipulation effects on both the outcome and the putative mediator can be achieved (Barlow et al., 2009; Maric et al., 2012). In this way, the single-case experimental design offers some advantages over traditional group-based designs. In traditional designs, even when participants are randomly assigned to different treatment conditions, and the assumption of temporal precedence is met, the mediator is not experimentally manipulated within persons, which limits inference regarding mediators as causal processes (MacKinnon, 2008). By contrast, randomized designs that involve direct manipulations of a mediating variable at different times offer an idiographic approach to understanding causal mediational processes for individuals.

Because mediation at the level of the person is an intraindividual process that can occur differently across individuals (Collins et al., 1998), traditional statistical methods for mediation are not directly applicable for single-case experimental designs. Person-oriented mediation assumes that individuals react differently to the intervention as well as the mediators that are changed by the intervention. Unfortunately, there are few available statistical tests for individual-level mediation, and many approaches to evaluate results from single-case experimental designs rely solely on visual analysis (Gaynor & Harris, 2008) or are concerned with tests of univariate outcomes that cannot be applied to explicitly investigate mediators of intervention outcomes (Geuke et al., 2019). Other proposed methods based on piecewise regression analysis can be used to compute and directly test the significance of the indirect effect (e.g., baseline vs. treatment; Mio evi et al., 2020). These methods have not been extended to situations where the treatment times are randomly

assigned to conditions or doses, nor can they be used for continuous X. Here we propose a RP mediation test to evaluate the effects of an intervention on a specific person via the hypothesized mediator.

There are several challenges for estimating a mediation effect for an individual participant. Because the data are obtained repeatedly from the participant, there may be dependency between the observations. If the dependency is not modeled, then residuals from the individual participant are no longer independent and identically distributed. One common way to handle that dependency is to model the dependency with lagged coefficients and other approaches such as the autoregressive integrated moving average model (ARIMA) model. This challenge is evaluated for the proposed RP test by evaluating the method in a simulation study of data with and without dependency. Another related challenge for a single participant method is the extent to which each observation is comparable to observations taken at other times—an assumption of exchangeable observations. The RP method described herein assumes that observations are exchangeable.

Randomization Permutation Tests for One Subject

Randomization tests are a special class of permutation tests for randomized experiments, where experimental units are randomly assigned to treatment conditions or doses in order to evaluate treatment effects (Edgington & Onghena, 2007). In the case of single-case or *N*-of-1 designs, times are randomly assigned to treatments. Edgington, 1967 proposed a single-case randomization test by demonstrating that if repeated measures are taken from one subject and time of measurement is randomized to treatment, then a randomization test evaluates treatment effects. The subject is measured at each occasion and the treatment effect is computed as the difference between average outcome scores for two treatment conditions. The observed difference between the average in each group is compared to the distribution of group differences stemming from all possible random orderings of times to treatments that could have been observed. This comparison results in a randomization test *p*-value to test the null hypothesis that there is no relation between treatment and outcome scores.

For example, given a treatment with six measurement occasions (3 for placebo and 3 for intervention), there are $6!/3!3! = 20$ possible combinations of treatment/placebo orderings (often called permutations in the research literature) including the one combination that was actually observed (Edgington & Onghena, 2007). We calculate the difference in mean score on the outcome for treatment and for control for all 20 combinations including the one that actually occurred, forming a distribution of scores under the null hypothesis of no treatment effect. The probability value (*p*-value) is equal to the proportion of times that a value equal to or larger than the observed effect (i.e., observed difference between means for treatment and control) is observed in the distribution of all possible mean difference scores. If the *p*-value is equal to or smaller than a significance level such as .05, then the null hypothesis is rejected. The test assumes that if the treatment has no effect at all, then any one of the 20 permuted mean differences could have been obtained. Other measures besides the difference in means are possible for each data set including the correlation, regression coefficient, and mediated effect. Regardless of the quantity being estimated, this approach assumes that the

different possible randomizations to treatment are exchangeable, that is, all else being equal, when there is not an effect of the manipulation, each possible randomization scheme is equivalent to all other randomization schemes.

With the example of six measurement occasions and 20 possible orderings, the smallest one-tailed p -value would be $1/20 = .05$. To increase the number of possible orderings, a larger number of occasions can be used. For example, with 8 occasions with 4 placebo and 4 intervention conditions, there are $8!/4!4! = 70$ possible combinations of treatment/placebo orders. With equal numbers in each group for 10, 12, and 14 occasions there are 252, 924, and 3432 possible combinations of treatment/placebo orderings, respectively. It may also be possible to apply Monte Carlo data generation for the case with limited numbers of repeated measures (Ernst, 2004).

Permutation Tests for Mediation Analysis

MacKinnon (2008) described a group-based permutation test for mediation. Permutation tests for mediation analysis are more complicated because there are two coefficients of interest: the X-to-M relation (see equation (1)) and the M-to-Y relation adjusted for X (see equation (2)), where i represents intercepts and e represents residuals.

$$M = i_m + aX + e_m \quad (1)$$

$$Y = i_y + bM + c'X + e_y \quad (2)$$

The terms randomization test and permutation test are often used synonymously in the research literature even though for mediation only the X-to-M relation can be evaluated with a randomization test, when there is randomization of X. The permutation test for the M-to-Y relation is not a randomization test because there is not randomization. Onghena (2018) emphasizes the difference between the randomization test and the permutation test: whereas the randomization test is based on random assignment to experimental conditions, the permutation test is based on random sampling. Both tests allow the estimation of a sampling distribution for a test statistic (Onghena, 2018; Taylor & MacKinnon, 2012). Because the proposed test includes a randomization test for X to M and a permutation test for M to Y, we call it a RP test, following Onghena's recommendation.

Because there are two tests in mediation analysis and three variables, the number of possible randomized and permuted data sets can be very large. For the group-based randomization test with six cases and two variables, there are $6! = 720$ data sets. If a mediating variable is also measured for each participant, with three participants randomly assigned to each condition of X, there is now a total of 20 randomizations of X times 720 permutations for M and Y so the total number of RP datasets equals 14,400. A mediation analysis would be conducted for each of the 14,440 data sets, resulting in a distribution of mediated effects. The probability value for the mediated effect is the proportion of the data sets with a mediated effect as large as or larger than the observed mediated effect. However, there was evidence that this group-based RP test can have inflated type I error rates when one of

the two paths from X to M or M to Y was zero and the other path was nonzero in the population (MacKinnon, 2008; Taylor & MacKinnon, 2012) as has sometimes been found in other studies of the permutation test (Churchill & Doerge, 2008). An RP test for a group of participants assessing each path individually, X to M, and M to Y (called a joint significance test in single sample mediation analysis, MacKinnon et al., 2002), had accurate Type I error rates but did not provide an estimate or confidence interval for the indirect effect (Taylor & MacKinnon, 2012).

Taylor and MacKinnon (2012) demonstrated that a confidence interval could be estimated for the mediated effect by randomizing and permuting residuals rather than observations, based on work for regression by ter Braak (1992) and Manly (1997). Taylor and MacKinnon (2012) conducted an extensive simulation study showing that randomizing and permuting observations can have inflated Type I error rates for some population parameter combinations but that randomizing and permuting residuals was an accurate method for constructing confidence intervals and testing the statistical significance of the mediated effect. Taylor and MacKinnon (2012) showed that estimating the sampling distribution of the test statistic, rather than comparing the test statistic to a sampling distribution under the null hypothesis, offers a better test of mediation with accurate Type I error rates in a variable-centered approach. The number of possible combinations of data sets for the RP method is equal to $N!^2$ for continuous X, M, and Y, leading to a large number of possible data sets, for example, for 5, 6, 7, and 8 sample sizes (in this article this would correspond to 5, 6, 7, and 8 measurement occasions), there are 14,400, 518,400, 25,401,600, and 1,625,792,400 data sets, respectively. All of this earlier work on the RP test for mediation was for a group of participants, was variable-oriented, and did not consider a single subject design.

The purpose of this paper is to show how adding measures of a mediating variable at each of the repeated measurement occasions in a single-case experimental design can be incorporated in a RP test of mediation for one person. The method creates the RP data sets based on randomization and permutation of residuals. Like work on group designs for a single-case by Edgington, 1967, the new method also requires that experimental units (i.e., measurement occasions) are randomly assigned to treatments as well as several other assumptions related to confounding to be discussed later. Based on these assumptions, the mediating process by which a treatment changes an outcome can be investigated for a single subject.

Autocorrelation in N-of-1 Designs

N-of-1 designs may exhibit autocorrelation, or the tendency for repeated measures to be correlated with each other (Busk & Marascuilo, 1988). Thus, the assumption of independently and identically distributed residuals is often not satisfied. In a meta-analysis of 800 *N*-of-1 designs, Shadish and Sullivan (2011) found, after adjusting for bias, that the mean autocorrelation estimate was $r_j = 0.20$. If the autocorrelation is positive, estimated standard errors will be smaller, leading to an increase in Type I error rate. If the autocorrelation is negative, estimated standard errors will be larger leading to an increase in Type II error rates (Kazdin, 2011). Thus, a method is needed that does not require the

assumption of independence. Alternatively, there are statistical methods that account for autocorrelation. Because of the importance of autocorrelation for single subject analyses, the simulations presented later evaluate effects for different autocorrelation values using lagged effects. For example, autoregressive parameters could be included in equations (1) and (2) to model dependency in the data. The applications of the method to the data examples also include lagged values as predictors to adjust for possible autocorrelation, that is, lagged M at the previous time point is another predictor in equation (1) and lagged Y at the previous time and lagged M at the previous time is another predictor of Y in equation (2).

Randomization Permutation Mediation Test

The RP mediation test evaluates the extent to which treatment affects a mediator that affects an outcome for one subject. It is assumed that there is random assignment of treatment times to treatment conditions or doses.

Steps

The RP confidence interval test for the mediated effect is based on the permutation of residuals described by ter Braak (1992) and Manly (1997) and the RP mediation test described by Taylor and MacKinnon (2012). In this test, a sampling distribution of the mediated effect is created through the RP of regression residuals, and the confidence intervals is obtained from the distribution of mediated effects in the RP distribution. Although the concept is straightforward, clarification of several steps in the process and notation are important to distinguish observed, residual, calculated, and permuted values. If autocorrelation were expected, the lagged value of the dependent variable (or another method to model dependency) would be included in the equation for M in Step 1 and for Y in Step 6. The Supplemental Material Appendix for this paper summarize the notation used in this manuscript for the RP confidence interval method.

1. Estimate a model regressing M on X as in equation (1). This results in an observed estimate of the *a*-path and intercept in the single mediator model.
2. Calculate predicted values of M using the estimates obtained in Step 1, labeled \hat{M} , and residuals, labeled e_m . Note when X is binary, there are two residuals. When X is continuous, it is possible that there is a different residual for each repeated measure.
3. Create a large number of datasets by randomizing the residuals in every possible combination, or a large sample (e.g., 2000) of the possible combinations, and reassigning them to the original, unpermuted data. The reassigned randomized residual is labeled e_m^* .
4. Calculate new values of M, labeled M^* , by summing the predicted value of \hat{M} and the permuted residual for each repeated measure, e_m^* : $M^* = \hat{M} + e_m^*$
5. Estimate a model regressing M^* on X as in equation (1) for all the datasets. This results in randomized estimates of the *a*-path, which are labeled a^* .
6. Repeat steps 1–5 by estimating a model regressing Y on M and X, as in equation (2). Step 5 will result in permuted estimates of the *b*-path, labeled b^* . Because

M is a continuous measure, there can conceivably be a different residual for each repeated measure.

7. Combine the product ab in the observed data and the product of a^*b^* in each of the RP datasets. These products make up a sampling distribution for the mediated effect.
8. Find the values in the distribution that correspond to the lower and upper limit of the desired confidence interval, for example, 2.5% and 97.5% percentile for 95% confidence intervals.
9. Examine the confidence interval to assess whether zero is in the confidence interval for significance testing and to consider the range of the possible effect.

If autocorrelation is expected, the lagged value of the dependent variable would be included in the equations for M and Y. For example, in the examples, we estimated the lagged M variable as a predictor of M and the lagged M and lagged Y variable as predictors of Y. In practice, a researcher may include autoregressive parameters based on theory and prior research, for example, it may be possible that the mediation equation does not require an autoregressive parameter but the outcome equation does or vice versa.

Assumptions of the RP Mediation Test

The results of the single subject RP test apply to the individual in the study. The usual mediation assumptions apply to this model including a self-contained model with no omitted influences, correct functional form for the relations of variables in the mediating process, reliable and valid measures, uncorrelated errors across equations, correct temporal precedence, and correct timing of measurement to capture the mediation process (MacKinnon, 2008). For the application in this paper, we also assume that the relation of M to Y does not change across measurement occasions so there is not an XM interaction in equation (2). In addition, four no unmeasured confounding assumptions identify direct and indirect effects by consideration of possible covariates that reduce the plausibility of unmeasured confounders of mediation relations (Pearl, 2001; VanderWeele & Vansteelandt, 2009; Valeri & VanderWeele, 2013):

1. No unmeasured confounders of the effect of the independent variable X on the dependent variable Y conditional on covariates.
2. No unmeasured confounders of the effect of the mediator M on the dependent variable Y conditional on the independent variable X and covariates.
3. No unmeasured confounders of the effect of the independent variable X on the mediator M conditional on covariates.
4. No measured or unmeasured confounders of the effect of the mediator M on the dependent variable Y that are affected by the independent variable X .

The single subject design raises important issues about the mediation assumptions. The assumption of a self-contained model indicates that there are no omitted variables in the statistical analysis. Because one participant is in the study, omitted variables that are consistent over time in the individual are addressed. However, it is possible that confounders

may change over time owing to boredom, maturation, etc. By randomizing times to conditions, assumptions 1 and 3 (i.e., the effect of X on the mediator and the effect of X on the outcome) are satisfied, at least with a large number of measurement occasions. However, randomization of times does not remove possible confounding of the M to Y relation (assumption 2) and also the extent to which change in X changes a mediator that affects Y (assumption 4). Assumptions 2 and 4 are not satisfied even if X represents assignment to levels of a randomized treatment because an individual self-selects their values on the mediator in response to treatment, given their observed level of the treatment and covariates (Holland, 1988; Imai et al., 2010; MacKinnon, 2008; MacKinnon & Pirlott, 2015). In other words, just like mediation designs with many subjects, mediation analysis with a single subject does not have a causal interpretation without further assumptions even when times are randomized in the single subject design (Holland, 1988; MacKinnon, 2008; Robins & Greenland, 1992; VanderWeele & Vansteelandt, 2009).

Clinical Examples

The RP test can evaluate mediation for a single person. Although there are different single-case experimental designs (e.g., withdrawal/reversal, alternating treatments, multiple baselines, and changing criterion designs; Onghena et al., 2019), we focus on the randomized treatment design, which is appropriate for treatments that have a rapid onset and offset of effects and can be administered in a double-blind, placebo-controlled, single-case experiment. We provide one example in which a subject's measurement occasions are randomly assigned to either treatment or placebo (i.e., binary X). Next, because researchers may be interested in dose-response treatment effects (McVay et al., 2019), we provide two examples in which measurement occasions are randomly assigned to alternating treatment doses (i.e., continuous X). More information on these examples including plots are included with the Supplemental Material Appendix for this paper.

Randomization to Treatment Condition

Our first example describes a single-case experimental design for understanding processes that may confer risk for developing an alcohol use disorder in a high-risk individual. We pattern the hypothetical data based on a study by Mayhugh et al. (2018) who examined patterns of momentary stress and craving among non-dependent, moderate-heavy alcohol users who drank as usual for 3 days and abstained from drinking for 3 days. Mayhugh et al. (2018) found evidence that, on average, drinking relieved stress relative to abstinence, and higher stress was associated with greater cravings. Data based on one participant in this study (Participant 18) were used to create a synthetic dataset, which is shown in the Supplemental Material Appendix. The participant was randomly assigned days in which they could drink or had to abstain, in order to investigate the effect of abstinence on craving levels via increased stress. Results of the RP test suggest that, for this participant, abstinence reduced stress, which in turn predicted higher craving levels (median mediated effect of $-.63$ with 95% confidence limits: $[-1.36, -0.01]$). The estimates include adjustment for dependency in stress (lag 1 estimate = 0.07 , $p = 0.67$) in the mediator equation and the dependency in cravings (lag 1 estimate = 0.29 , $p = 0.07$) and stress (lag 1 estimate = -0.22 , $p = 0.20$) in the outcome equation. Note that the mediated effect for this hypothetical

participant is opposite from the overall results of Mayhugh et al. (2018) demonstrating how single subject analysis may obtain different results for different participants.

Randomization to Treatment Dose

Randomization permutation tests can also be employed to examine dose-response relationships. Examining dose-response relationships may uncover for whom certain foods or food additives have an adverse effect, as well as the process through which these individuals are affected. For example, although the majority of children do not have a behavioral response to food additives, some children do exhibit a significant increase in their negative behaviors following consumption of artificial food colors (Weiss et al., 1980). Randomization permutation tests can evaluate potential mechanisms, such as the secretion of histamine, by which food additives influence behavior for each affected child (Stevens et al., 2013). We present data in Supplemental Material Appendix A based on a hypothetical study where the participant was randomly assigned to different doses of food additives on different measurement occasions, and repeated measures of plasma histamine and behavior problems were assessed. Results of the RP test suggest that, for this participant, food additives were related to plasma histamine, which in turn predicted more behavior problems (median mediated effect = .043 with 95% confidence limits: [0.02, 0.08]). The estimates include adjustment for dependency in plasma histamine (lag 1 estimate = 0.02, $p = 0.90$) in the mediator equation, and the dependency in behavior problems (lag 1 estimate = 0.02, $p = 0.41$) and histamine (lag 1 estimate = $-.24$, $p = 0.22$) in the outcome equation.

Similarly, single-case experimental designs may be useful in evaluating the mechanisms through which short-acting pharmacological interventions achieve their effects. Randomization permutation tests could evaluate whether the effect of oxytocin, administered as a nasal spray, on increased cooperation operates via increased recognition of happy faces (Gossen et al., 2012; Rilling et al., 2012; Schulze et al., 2011; Shin et al., 2015, 2018). Supplemental Material Appendix A shows data from a hypothetical study where treatment times were randomly assigned to intranasally-administered oxytocin doses, and measures of a participant's facial emotion recognition and prosocial behavior were taken at each treatment occasion. Results of the RP test suggest that, for this participant, oxytocin was not related to prosocial behavior via facial emotion recognition because the confidence interval contains zero (median mediated effect = .02 with 95% confidence limits: $[-0.00, 0.05]$). The estimates include adjustment for dependency in facial emotion recognition, (lag 1 estimate = -0.05 , $p = 0.75$) in the mediator equation and the dependency in prosocial behavior (lag 1 estimate = 0.00, $p = 0.98$) and facial emotion recognition (lag 1 estimate = $-.15$, $p = .55$) in the outcome equation.

Simulation Study

We conducted a simulation study to systematically evaluate the performance of the proposed mediation analysis for a single subject as a function of number of measurement occasions and the influence of different amounts of dependency in repeated measures from the single participant. We focused on the power and Type I error rates of the RP confidence intervals for the indirect effect.

Methods

Data were simulated with two types of dependency: (1) no dependency, and (2) autocorrelation using lagged effects for both M and Y. Data were generated in SAS software version 9.4. The predictor X was generated as a random binary variable. The mediator M and outcome Y were generated according to equations (1) and (2) with the residuals generated from a normal distribution. In order to evaluate the effects of autocorrelation of M and Y, lagged M and Y variables from the generated data were added as predictors lagged by one time occasion. Lagged effects of 0.3, 0.5, and 0.8 were specified, and the lagged effect was the same for M and Y in all conditions. The autocorrelated data were analyzed by both modeling and not modeling the lagged M and Y variables. Modeling the lagged effect was accomplished by including a lag 1 predictor of M and Y in the equation for M and Y, respectively. The values of the a , b , and c' parameters were set to either 0, .59 (a large effect corresponding approximately to a correlation of .5), or 1 (a very large effect corresponding approximately to a correlation of .7) in seven combinations of a , b , and c' . There were three combinations with a nonzero population mediated effect to assess power: large, large, and zero; large, large, and large; very large, very large, and zero; for a , b , and c' , respectively. There were four combinations with a zero population mediated effect to assess Type I error rates: zero, large, and zero; large, zero, and zero; zero, large, and large; large, zero, and large, for a , b , and c' , respectively. The value of 1 represents a very large effect size, exceeding the large effect size of .59 often used in mediation simulation studies. The very large effect size was used in the simulation because of the possible reduced power to detect effects in single subject designs. There were 10 levels of repeated measurements, ranging from $N=4$ to $N=100$ occasions ($N=4, 6, 8, 10, 12, 16, 20, 40, 50, 100$). For empirical power, there were three parameter combinations times 10 sample sizes times seven analysis models with and without modeling lagged effects for 210 conditions reported in Table 1. For empirical Type I error rates, there were four parameter combinations, 10 sample sizes times seven analysis models with and without modeling lagged effects for 280 conditions reported in Table 2. Each condition used 2000 RPs except for the conditions in the smallest sample size, $N=4$. When $N=4$, only 576 RPs were used, which is the total number of possible RPs at this sample size. Each condition was replicated 500 times. The empirical Type I error rate and empirical statistical power were computed for each condition. The Type I error rate was equal to the proportion of confidence intervals for the indirect effect that excluded 0 when the true indirect effect was 0, and values above 0.075 were considered excessive according to Bradley's (1978) robustness criterion. Power was equal to the proportion of confidence intervals for the indirect effect that excluded 0 when the true indirect effect was nonzero and empirical power values of 0.8 or higher were deemed adequate, following conventions for statistical power.

Results

Power

When there is no autocorrelation for M and Y in the data generating process, then 40 repeated observations are required for power greater than 0.8 when a , and b are both very large; $N=100$ repeated observations are required when a and b are large (Table 1).

When there is an unmodeled autocorrelation of 0.3, $N = 100$ repeated observations are sufficient for power greater than 0.8 when the effects of a and b are both large. When the autocorrelation of 0.3 is modeled, there is power greater than 0.8 for $N = 50$ with large a and b . No conditions had power greater than 0.8 for an unmodeled autocorrelation of 0.5 or 0.8, but modeling the autocorrelation did have greater than 0.8 power for N of 50 for very large a and very large b and N -of-100 starting at large a and large b . The power for a modeled autocorrelation of 0.8 along with very large a and very large b effects was reduced to .257 likely owing to the multicollinearity of very large a , very large b effects, and a large autocorrelation of 0.8 for M and Y ; such very large effect sizes are perhaps unlikely to occur with real single subject data.

Type I Error Rate

Results indicate that with no autocorrelation for M and Y in the data generating process, $N = 8$ is sufficient for attaining Type I error rates below the upper bound of Bradley's (1978) robustness criterion (Table 2). Furthermore, Type I error rates remain acceptable at larger sample sizes. With unmodeled autocorrelations of 0.3, 0.5, and 0.8 in both M and Y , the Type I error rates are acceptable starting at $N = 8$, but there were excessive Type I error rates when the autocorrelation was not modeled for larger sample sizes, as high as 0.27 for $N = 100$ and large a and zero b . When the autocorrelation was modeled, there were no occasions when the Type I error rates exceeded Bradley's criterion.

Discussion

The RP mediation test provides an option for research on mediating processes with repeated measures data from a single subject. The strengths of the method include its acknowledgment of individual differences by focusing on one individual at a time, a long history of research evaluating randomization and permutation methodology, and how it makes the maximum use of data from one subject. The simulation study demonstrated that under ideal conditions with 40 repeated measures, the proposed method has 0.8 power to detect a very large effect of X to M and M to Y . For large effects of X to M and M to Y , 100 repeated measures are required for 0.8 power. Fewer repeated measures would require larger effect sizes and more repeated measures would require smaller effect sizes for adequate power though it may be possible to combine Monte Carlo data generation with the RP test with limited numbers of repeated measures (Ernst, 2004). The proposed method provides an important complement to visual inspection of data from one individual across time and conditions. The simulation study also demonstrated that it is best to model dependency in the mediator and outcome over time. The need to model dependency in time series is recognized in other approaches for single subject time series such as the Autoregressive Integrated Moving Average (ARIMA) model (Ljung & Box, 1978). The role of dependency in single subject data is an area of active discussion (Barnard-Brak, Watkins, & Richman, 2021) and warrants further research. The RP method described herein can accompany methods to model dependency. Like many simulation studies, this simulation study demonstrated the RP method in ideal conditions where the data generating model and the data analysis model coincided. With real data, the dependency is unknown, so it is important that researchers obtain a model that adequately deals with the dependency in the

data. Nevertheless, the results in this paper provide proof of concept information for the RP test.

The single subject RP mediation method has several limitations. Some of the limitations apply to mediation analysis in general and others are specific to conducting analysis with a single subject. The proposed RP method is based on residuals from regression equations, which may be different than RP of observations, though the RP of residuals had better statistical properties than RP of observations in Taylor and MacKinnon (2012). Assumptions of group-based methods regarding valid and reliable measurement of constructs apply to single subject designs. We assume that constructs are measured without error. It may be possible to include measurement models for variables in the single subject design, but this will add more complexity to model estimation and is a topic for future research. A primary limitation of the RP method is that the results of the analysis apply to one participant and may not apply to other participants. Methods for meta-analysis of small N designs (Shadish et al., 2008) and Bayesian methods combining information across studies (Wurpts et al., 2021) may be used to combine results across individuals.

Several confounding assumptions are important for the single subject design (see Josephy et al., 2015 for more on counterfactual models for within-subject designs). In particular, the assumption of no confounders of M to Y illustrates several limitations of the proposed method. First, the relation of the intervention to the mediator is randomized, but the relation of the mediator to the outcome is not. There can be confounders of the M to Y relation for the single subject RP test as for the group-based mediation methods, though there is some confounding control of individual differences because the single participant provides the data. Similarly, there may be effects of X that change a variable that confounds the M to Y relation. For example, in the example relating the abstinence or drinking condition to stress, and then stress to craving for alcohol, there can be confounders of the stress to craving relation. An external variable, such as weekend day, may predict both stress and craving. Approaches to sensitivity analysis for unmeasured confounders and adjustment for measured confounders address violations of this assumption (MacKinnon & Pirlott, 2015). Measured confounders could be included as additional predictors in the regression models (Kroehl et al., 2020), for example, coding weekend day or not as an additional predictor. Treatment may also change more than one mediator, which violates the assumption that X does not change a mediator that changes the M to Y relation. In principle, it is possible to add an additional mediator to address this aspect of X induced confounding of the M to Y relation, but this model is beyond the scope of this paper and would require additional assumptions regarding how the mediators are related (VanderWeele & Vansteelandt, 2013).

Another limitation of the single subject method proposed here extends from the long history of criticism of within-subjects designs including carryover effects, history effects, and maturation effects that provide alternative explanations of results. For the mediation model, this limitation applies to both the mediator and the outcome variable. Even with these effects, successful randomization of treatment to measurement occasions reduces alternative explanations for estimates of the X to M and X to Y but not for the M to Y relation. Carryover effects, history, and maturation could be alternative explanations of the M to Y relation in the single subject design. The method assumes that with randomization of

treatment to times, each condition is exchangeable—that the same results would be obtained with different actual orders of the assigned treatment. Discussion of these assumptions are an important part of reporting results from the RP mediation analysis.

There are several important future extensions of this method. The performance of different approaches to modeling dependency, especially across research areas, is an important research topic. The need for modeling dependency over time may differ, for example, between designs with a short time lag compared to a long time lag between measurements. The method described in this article is for randomization of occasions to treatment. The RP mediation test could also be developed for other designs, including designs in which randomization of treatment to days follows other methods such as an ABAB design or more complicated design (Onghena et al., 2019). The method can be applied to any design, even designs where treatment was not randomized to conditions, but the RP method would represent statistical associations and not causal effects. The methods can be extended to more complicated designs but performance for such designs will depend on the successful randomization of treatments to measurement occasions or groups of measurement occasions and whether carryover effects could explain results. An extension of the RP method could include additional subjects in the RP analysis with dummy codes and blocking participants. Adding more participants may improve the conclusions of the study, though if all participants yield the same result, then analysis methods at the group-level are likely the preferred method of choice.

In conclusion, the combination of the randomization (for the relation of X to M) and the permutation (for the relation of M to Y adjusted for X) in the RP mediation test is a promising solution for investigating mediation with single subject data. The current method requires that measurement occasions are randomized to treatments and there is adequate modeling of dependency in the data. The results of the analysis apply to the participant under study and not necessarily other participants. With the RP method, it is possible to observe different, perhaps even opposite, results across individual participants, consistent with the goals of idiographic theory.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The National Institute on Drug Abuse (R37DA09757 and UH2DA041713) supported this research in part. Portions of this paper were presented at the 2019 Single Subject Causal Mediation Analysis Workshop at the Lorentz Center in Leiden, Netherlands.

References

- Allport GW (1962). The general and the unique in psychological science. *Journal of Personality*, 30(3), 405–422. 10.1111/j.1467-6494.1962.tb02313.x [PubMed: 13860641]
- Barlow DH, Nock MK, & Hersen M (2009). *Single case experimental designs: Strategies for studying behavior change* (3rd ed.). Pearson.

- Barnard-Brak L, Watkins L, & Richman DM (2021). Autocorrelation and estimates of treatment effect size for single-case experimental design data. *Behavioral Interventions*, 36(3), 595–605. 10.1002/bin.1783
- Bradley JV (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, 31(2), 144–152. 10.1111/j.2044-8317.1978.tb00581.x
- Busk PL, & Marascuilo LA (1988). Autocorrelation in single-subject research: A counterargument to the myth of no autocorrelation. *Behavioral Assessment*, 10(3), 229–242.
- Cattell RB (1952). The three basic factor-analytic research designs—their interrelations and derivatives. *Psychological Bulletin*, 49(5), 499–520. 10.1037/h0054245 [PubMed: 12993927]
- Churchill GA, & Doerge RW (2008). Naive application of permutation testing leads to inflated type I error rates. *Genetics*, 178(1), 609–610. 10.1534/genetics.107.074609 [PubMed: 18202402]
- Collins LM, Graham JW, & Flaherty BP (1998). An alternative framework for defining mediation. *Multivariate Behavioral Research*, 33, 295–312. [PubMed: 26771887]
- Edgington ES (1967). Statistical inference from $n = 1$ experiments. *The Journal of Psychology*, 65(2), 195–199. 10.1080/00223980.1967.10544864.
- Edgington ES, & Onghena P (2007). *Randomization tests* (4th ed.). Chapman & Hall/CRC: Taylor & Francis Group.
- Ernst MD (2004). Permutation methods: a basis for exact inference. *Statistical Science*, 19(4), 676–685. 10.1214/088342304000000396
- Gast DL (2010). *Single subject research methodology in behavioral sciences*. Routledge.
- Gaynor ST, & Harris A (2008). Single-participant assessment of treatment mediators: Strategy description and examples from a behavioral activation intervention for depressed adolescents. *Behavior Modification*, 32(3), 372–402. 10.1177/0145445507309028 [PubMed: 18391053]
- Geuke GGM, Maric M, Mio evi M, Wolters LH, & Haan E (2019). Testing mediators of youth intervention outcomes using single-case experimental designs Randomized controlled trials (RCTs) in clinical and community settings: Challenges, Alternatives and supplementary designs. *New Directions for Child and Adolescent Development*, 2019(167), 39–64. 10.1002/cad.20310
- Gossen A, Hahn A, Westphal L, Prinz S, Schultz RT, Gründer G, & Spreckelmeyer KN (2012). Oxytocin plasma concentrations after single intranasal oxytocin administration – A study in healthy men. *Neuropeptides*, 46(5), 211–215. 10.1016/j.npep.2012.07.001 [PubMed: 22884888]
- Holland PW (1988). Causal inference, path analysis, and recursive structural equation models. *Sociological Methodology*, 18, 449–484. 10.2307/271055.
- Imai K, Keele L, & Tingley D (2010). A general approach to causal mediation analysis. *Psychological Methods*, 15(4), 309–334. 10.1037/a0020761 [PubMed: 20954780]
- Joseph H, Vansteelandt S, Vanderhasselt M, & Loeys T (2015). Within-subject mediation analysis in AB/BA crossover designs. *International Journal of Biostatistics*, 11(1), 1–22. 10.1515/ijb-2014-0057
- Kazdin A (2011). *Single-case research designs: Method for clinical and applied settings* (2nd ed.). Oxford University Press.
- Kroehl ME, Lutz S, & Wagner BD (2020). Permutation-based methods for mediation analysis in studies with small sample sizes. *PeerJ*, 8(■■■■), e8246. 10.7717/peerj.8246 [PubMed: 32002321]
- Leeuw M, Houben RM, Severeijns R, Picavet HSJ, Schouten EG, & Vlaeyen JW (2007). Pain-related fear in low back pain: A prospective study in the general population. *European Journal of Pain*, 11(3), 256–266. 10.1007/s10865-006-9085-0 [PubMed: 16546425]
- Lei H, Nahum-Shani I, Lynch K, Oslin D, & Murphy SA (2012). A “SMART” design for building individualized treatment sequences. *Annual Review of Clinical Psychology*, 8(1), 21–48. 10.1146/annurev-clinpsy-032511-143152
- Ljung GM, & Box GEP (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2), 297–303. 10.1093/biomet/65.2.297
- MacKinnon DP (2008). *Introduction to statistical mediation analysis*. Lawrence Erlbaum Associates.
- MacKinnon DP (2019). Causal mediation analysis and applications to single case studies. Workshop presentation. Lorentz Center Workshop.

- MacKinnon DP, Lockwood CM, Hoffman JM, West SG, & Sheets V (2002). A comparison of methods to test mediated and other intervening variable effects. *Psychological Methods*, 7(1), 83–104. 10.1037/1082-989x.7.1.83. [PubMed: 11928892]
- MacKinnon DP, & Pirlott A (2015). Statistical approaches to enhancing the causal interpretation of the M to Y relation in mediation analysis. *Personality and Social Psychology Review*, 19(1), 30–43. 10.1177/1088868314542878 [PubMed: 25063043]
- Manly BF (1997). *Randomization, bootstrap and Monte Carlo methods in biology*. CRC Press.
- Maric M, de haan E, Hogendoorn SM, Wolters LH, & Huizenga HM (2015). Evaluating statistical and clinical significance of intervention effects in single-case experimental designs: An SPSS method to analyze univariate data. *Behavior Therapy*, 46(2), 230–241. 10.1016/j.beth.2014.09.005 [PubMed: 25645171]
- Maric M, Prins PJM, & Ollendick TH (2015). *Moderators and mediators of youth treatment outcomes*. Oxford University Press.
- Maric M, Wiers RW, & Prins PJM (2012). Ten ways to improve the use of statistical mediation analysis in the practice of child and adolescent treatment research. *Clinical Child and Family Psychology Review*, 15(3), 177–191. 10.1007/s10567-012-0114-y [PubMed: 22418896]
- Mayhugh RE, Rejeski WJ, Petrie MR, Laurienti PJ, & Gauvin L (2018). Differing patterns of stress and craving across the day in moderate-heavy alcohol consumers during their typical drinking routine and an imposed period of alcohol abstinence. *Plos One*, 13(4), e0195063. 10.1371/journal.pone.0195063 [PubMed: 29668736]
- McVay MA, Bennett GG, Steinberg D, & Voils CI (2019). Does-response research in digital health interventions: Concepts, considerations, and challenges. *Health Psychology*, 38(12), 1168–1174. 10.1037/hea0000805 [PubMed: 31580127]
- Mio evi M, Klaassen F, Geuke G, Moeyaert M, & Maric M (2020). Using bayesian methods to test mediators of intervention outcomes in single-case experimental designs. *Evidence-Based Communication Assessment and Intervention*, 14(1–2), 52–68. 10.1080/17489539.2020.1732029
- Molenaar PCM (2004). A manifesto on psychology as idiographic science: bringing the person back Into scientific psychology, this time forever. *Measurement: Interdisciplinary Research and Perspectives*, 2(4), 201–218. 10.1207/s15366359mea0204_1
- Molenaar PCM, & Campbell CG (2009). The new person-specific paradigm in psychology. *Current Directions in Psychological Science*, 18(2), 112–117. 10.1111/j.1467-8721.2009.01619.x
- Onghena P (2018). Randomization tests or permutation tests? A historical and terminological clarification. In Berger V(Ed.) *Randomization, masking, and allocation concealment*, (pp. 209–227). Chapman & Hall/CRC Press.
- Onghena P, Taniou R, De TK, & Michiels B (2019). Randomization tests for changing criterion designs. *Behaviour Research and Therapy*, 117(■■■■), 18–27. 10.1016/j.brat.2019.01.005 [PubMed: 30670306]
- O’Rourke HP, & MacKinnon DP (2018). Reasons for testing mediation in the absence of an intervention effect: A research imperative in prevention and intervention research. *Journal of Studies on Alcohol and Drugs*, 79(2), 171–181. 10.15288/jsad.2018.79.171. [PubMed: 29553343]
- Pearl J (2001). Direct and indirect effects. In Breese J, & Koller D (Eds.) *Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence*, San Francisco, CA, August 2–5, 2001, (pp. 411–420). Morgan Kaufmann.
- Preacher KJ, Zyphur MJ, & Zhang Z (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, 15(3), 209–233. 10.1037/a0020141 [PubMed: 20822249]
- Rilling JK, DeMarco AC, Hackett PD, Thompson R, Ditzen B, Patel R, & Pagnoni G (2012). Effects of intranasal oxytocin and vasopressin on cooperative behavior and associated brain activity in men. *Psychoneuroendocrinology*, 37(4), 447–461. 10.1016/j.psyneuen.2011.07.013 [PubMed: 21840129]
- Robins JM, & Greenland S (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, 3(2), 143–155. 10.1097/00001648-199203000-00013 [PubMed: 1576220]

- Schulze L, Lischke A, Greif J, Herpertz SC, Heinrichs M, & Domes G (2011). Oxytocin increases recognition of masked emotional faces. *Psychoneuroendocrinology*, 36(9), 1378–1382. 10.1016/j.psyneuen.2011.03.011 [PubMed: 21477929]
- Shadish WR, Rindskopf DM, & Hedges LV (2008). The state of the science in the meta-analysis of single-case experimental designs. *Evidence-Based Communication Assessment and Intervention*, 2(3), 188–196. 10.1080/17489530802581603
- Shadish W, & Sullivan K (2011). Characteristics of single-case designs used to assess intervention effects in 2008. *Behavior Research Methods*, 43(4), 971–980. 10.3758/s13428-011-0111-y [PubMed: 21656107]
- Shin NY, Park HY, Jung WH, & Kwon JS (2018). Effects of intranasal oxytocin on emotion recognition in Korean male: A dose-response study. *Psychiatry Investigation*, 15(7), 710–716. 10.30773/pi.2018.02.19 [PubMed: 29898580]
- Shin NY, Park HY, Jung WH, Park JW, Yun JY, Jang JH, Kim SN, Han HJ, Kim SY, Kang DH, & Kwon JS (2015). Effects of oxytocin on neural response to facial expressions in Patients with schizophrenia. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, 40(8), 1919–1927. 10.1038/npp.2015.41 [PubMed: 25666311]
- Stevens LJ, Kuczek T, Burgess JR, Stochelski MA, Arnold LE, & Galland L (2013). Mechanisms of behavioral, atopic, and other reactions to artificial food colors in children. *Nutrition Reviews*, 71(5), 268–281. 10.1111/nure.12023 [PubMed: 23590704]
- Taylor AB, & MacKinnon DP (2012). Four applications of permutation methods to testing a single-mediator model. *Behavior Research Methods*, 44(3), 806–844. 10.3758/s13428-011-0181-x [PubMed: 22311738]
- ter Braak CJF (1992). Permutation versus bootstrap significance tests in multiple regression and ANOVA. In Jockel K-H, Rothe G, & Sandler W (Eds) *Bootstrapping and related techniques*, (pp. 79–85). Springer.
- Valeri L, & VanderWeele TJ (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. 10.1037/a0031034. [PubMed: 23379553]
- VanderWeele T (2015). *Explanation in causal inference: Methods for mediation and interaction*. Oxford University Press.
- VanderWeele TJ, & Vansteelandt S (2009). Conceptual issues concerning mediation, interventions and composition. *Statistics and its Interface (Special Issue on Mental Health and Social Behavioral Science)*, 2, 457–468.
- VanderWeele TJ, & Vansteelandt S (2013). Mediation analysis with multiple mediators. *Epidemiologic Methods*, 2(1), 95–115. 10.1515/em-2012-0010
- Vuorre M, & Bolger N (2018). Within-subject mediation analysis for experimental data in cognitive psychology and neuroscience. *Behavior Research Methods*, 50(5), 2125–2143. 10.3758/s13428-017-0980-9 [PubMed: 29247385]
- Weiss B, Williams JH, Margen S, Abrams B, Caan B, Citron LJ, Cox C, McKibben J, Ogar D, & Schultz S (1980). Behavioral responses to artificial food colors. *Science*, 207(4438), 1487–1489. 10.1126/science.7361103 [PubMed: 7361103]
- Wurpts I, Mio evi M, & MacKinnon DP (2021). Sequential bayesian data synthesis for mediation and regression analysis. *Prevention Science*.

Table 1.

Empirical Power by dependency, sample size, and population values of a, b, and c'.

n	a, b, and c'	Baseline - no lag for M and Y			Unmodeled lag for M and Y			Modeled lag for M and Y		
		Lagged coefficient=0	Lagged coefficient=-3	Lagged coefficient=-5	Lagged coefficient=-3	Lagged coefficient=-5	Lagged coefficient=-8	Lagged coefficient=-3	Lagged coefficient=-5	Lagged coefficient=-8
4	Large, large, zero	0.21	0.11	0.12	0.12	0.11	0.14	0.37	0.41	0.40
	Large, large, large	0.22	0.16	0.11	0.11	0.11	0.12	0.41	0.45	0.37
	Very large, very large, zero	0.29	0.18	0.20	0.20	0.18	0.17	0.49	0.45	0.47
6	Large, large, zero	0.10	0.05	0.06	0.06	0.05	0.06	0.14	0.12	0.11
	Large, large, large	0.11	0.08	0.07	0.07	0.08	0.07	0.15	0.14	0.13
	Very large, very large, zero	0.23	0.12	0.11	0.11	0.09	0.09	0.26	0.29	0.26
8	Large, large, zero	0.09	0.08	0.08	0.08	0.08	0.08	0.12	0.10	0.12
	Large, large, large	0.09	0.07	0.05	0.05	0.07	0.07	0.12	0.09	0.11
	Very large, very large, zero	0.26	0.17	0.12	0.12	0.10	0.10	0.28	0.25	0.27
10	Large, large, zero	0.10	0.08	0.06	0.06	0.06	0.06	0.11	0.10	0.12
	Large, large, large	0.11	0.09	0.04	0.04	0.09	0.09	0.12	0.13	0.12
	Very large, very large, zero	0.34	0.19	0.14	0.14	0.09	0.09	0.30	0.28	0.31
12	Large, large, zero	0.11	0.08	0.08	0.08	0.08	0.08	0.10	0.14	0.09
	Large, large, large	0.09	0.12	0.08	0.08	0.06	0.06	0.12	0.13	0.12
	Very large, very large, zero	0.38	0.22	0.17	0.17	0.09	0.09	0.32	0.34	0.33
16	Large, large, zero	0.18	0.11	0.09	0.09	0.05	0.05	0.14	0.15	0.13
	Large, large, large	0.16	0.09	0.07	0.07	0.06	0.06	0.15	0.18	0.15
	Very large, very large, zero	0.54	0.29	0.15	0.15	0.09	0.09	0.43	0.44	0.48
20	Large, large, zero	0.24	0.12	0.09	0.09	0.06	0.06	0.19	0.21	0.22
	Large, large, large	0.21	0.15	0.10	0.10	0.05	0.05	0.23	0.18	0.19
	Very large, very large, zero	0.59	0.40	0.18	0.18	0.10	0.10	0.55	0.54	0.53
40	Large, large, zero	0.47	0.28	0.14	0.14	0.08	0.08	0.44	0.42	0.42
	Large, large, large	0.44	0.21	0.16	0.16	0.07	0.07	0.42	0.47	0.44
	Very large, very large, zero	0.89	0.59	0.31	0.31	0.08	0.08	0.82	0.82	0.82
50	Large, large, zero	0.54	0.28	0.19	0.19	0.08	0.08	0.53	0.56	0.51
	Large, large, large	0.56	0.31	0.17	0.17	0.09	0.09	0.53	0.50	0.52

<i>n</i>	<i>a</i> , <i>b</i> , and <i>c</i> '	Baseline - no lag for M and Y			Unmodeled lag for M and Y			Modeled lag for M and Y		
		Lagged coefficient=0	Lagged coefficient=-3	Lagged coefficient=-5	Lagged coefficient=-8	Lagged coefficient=-3	Lagged coefficient=-5	Lagged coefficient=-8	Lagged coefficient=-3	Lagged coefficient=-5
100	Very large, very large, zero	0.95	0.66	0.36	0.10	0.88	0.89	0.88	0.89	0.88
	Large, large, zero	0.85	0.49	0.29	0.07	0.82	0.81	0.84	0.81	0.84
	Large, large, large	0.83	0.53	0.26	0.09	0.80	0.81	0.79	0.81	0.79
	Very large, very large, zero	1.00	0.91	0.56	0.12	0.99	1.00	0.27	1.00	0.27

Note: For *a*, *b*, and *c*' large = .59, very large = 1.00. Column 1 ($\rho_1 = 0$) refers to a baseline condition with no dependency. Columns 2-7 refer to conditions where dependency was induced via lagged coefficients of equal value for M and Y. The lagged M and Y variables were not modeled in columns 2-4 and were modeled in columns 5-7.

Table 2.

Empirical Type I error rates by dependency, sample size, and population values of a, b, and c'.

n	a, b, and c'	Baseline - no lag for M and Y			Unmodeled lag for M and Y			Modeled lag for M and Y		
		Lagged coefficient = 0	Lagged coefficient = 3	Lagged coefficient = 8	Lagged coefficient = 3	Lagged coefficient = 5	Lagged coefficient = 8	Lagged coefficient = 3	Lagged coefficient = 5	Lagged coefficient = 8
4	Zero, large, zero	0.17	0.14	0.11	0.11	0.11	0.11	0.37	0.32	0.30
	Large, zero, zero	0.18	0.08	0.08	0.09	0.09	0.09	0.40	0.39	0.41
	Zero, large, large	0.19	0.11	0.12	0.15	0.15	0.15	0.34	0.36	0.37
6	Large, zero, large	0.21	0.11	0.10	0.10	0.10	0.10	0.42	0.39	0.41
	Zero, large, zero	0.09	0.03	0.07	0.10	0.10	0.10	0.12	0.12	0.09
	Large, zero, zero	0.06	0.05	0.02	0.04	0.04	0.04	0.11	0.09	0.10
8	Zero, large, large	0.07	0.06	0.06	0.06	0.06	0.06	0.09	0.14	0.10
	Large, zero, large	0.08	0.05	0.05	0.03	0.03	0.03	0.12	0.10	0.10
	Zero, large, zero	0.03	0.05	0.06	0.05	0.05	0.05	0.06	0.07	0.05
10	Large, zero, zero	0.05	0.03	0.02	0.03	0.03	0.03	0.07	0.07	0.06
	Zero, large, large	0.05	0.04	0.05	0.06	0.06	0.06	0.08	0.07	0.07
	Large, zero, large	0.04	0.04	0.04	0.03	0.03	0.03	0.05	0.04	0.07
12	Zero, large, zero	0.03	0.04	0.05	0.06	0.06	0.06	0.04	0.05	0.04
	Large, zero, zero	0.04	0.04	0.02	0.02	0.02	0.02	0.05	0.04	0.03
	Zero, large, large	0.03	0.04	0.05	0.07	0.07	0.07	0.07	0.04	0.05
16	Large, zero, large	0.03	0.02	0.02	0.03	0.03	0.03	0.04	0.03	0.03
	Zero, large, zero	0.03	0.06	0.06	0.06	0.06	0.06	0.04	0.05	0.05
	Large, zero, zero	0.02	0.01	0.02	0.03	0.03	0.03	0.04	0.04	0.05
20	Zero, large, large	0.04	0.04	0.05	0.06	0.06	0.06	0.04	0.06	0.06
	Large, zero, zero	0.03	0.03	0.03	0.01	0.01	0.01	0.04	0.05	0.02
	Zero, large, zero	0.02	0.05	0.05	0.06	0.06	0.06	0.04	0.04	0.04
24	Large, zero, zero	0.02	0.03	0.02	0.03	0.03	0.03	0.02	0.02	0.03
	Zero, large, large	0.04	0.05	0.04	0.07	0.07	0.07	0.04	0.05	0.04
	Large, zero, large	0.01	0.02	0.02	0.01	0.01	0.01	0.03	0.03	0.01
28	Zero, large, zero	0.04	0.05	0.04	0.07	0.07	0.07	0.03	0.03	0.03
	Large, zero, zero	0.03	0.02	0.03	0.04	0.04	0.04	0.02	0.04	0.04
	Zero, large, zero	0.03	0.02	0.03	0.04	0.04	0.04	0.02	0.04	0.04

n	a, b, and c'	Baseline - no lag for M and Y			Unmodeled lag for M and Y			Modeled lag for M and Y			
		Lagged coefficient = 0	Lagged coefficient = -3	Lagged coefficient = -5	Lagged coefficient = -8	Lagged coefficient = -3	Lagged coefficient = -5	Lagged coefficient = -8	Lagged coefficient = -3	Lagged coefficient = -5	Lagged coefficient = -8
40	Zero, large, large	0.04	0.05	0.05	0.08	0.04	0.06	0.06	0.06	0.06	0.06
	Large, zero, large	0.01	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.03	0.03
	Zero, large, zero	0.04	0.04	0.09	0.07	0.06	0.07	0.07	0.07	0.09	0.09
	Large, zero, zero	0.03	0.08	0.07	0.03	0.02	0.01	0.01	0.01	0.02	0.02
50	Zero, large, large	0.07	0.06	0.07	0.07	0.05	0.05	0.05	0.05	0.07	0.07
	Large, zero, large	0.02	0.07	0.07	0.05	0.03	0.04	0.04	0.04	0.02	0.02
	Zero, large, zero	0.07	0.08	0.04	0.06	0.07	0.06	0.06	0.06	0.07	0.07
	Large, zero, zero	0.03	0.08	0.10	0.07	0.03	0.03	0.03	0.03	0.03	0.03
100	Zero, large, large	0.07	0.06	0.05	0.06	0.05	0.04	0.04	0.04	0.06	0.06
	Large, zero, large	0.03	0.08	0.08	0.06	0.02	0.05	0.05	0.05	0.04	0.04
	Zero, large, zero	0.04	0.05	0.06	0.06	0.05	0.06	0.06	0.06	0.07	0.07
	Large, zero, zero	0.06	0.27	0.22	0.07	0.03	0.03	0.03	0.03	0.04	0.04
	Zero, large, large	0.07	0.06	0.08	0.06	0.04	0.04	0.04	0.04	0.06	0.06
	Large, zero, large	0.03	0.25	0.25	0.06	0.04	0.03	0.03	0.03	0.04	0.04

Note: The nominal Type I error rate is .05. For a, b, and c', large = .59. Column 1 ($\rho_1 = 0$) refers to a baseline condition with no dependency. Columns 2-7 refer to conditions where dependency was induced via lagged coefficients of equal value for M and Y. The lagged M and Y variables were not modeled in columns 2-4 and were modeled in columns 5-7.