Behaviour Research and Therapy 98 (2017) 39–57

Contents lists available at ScienceDirect

Behaviour Research and Therapy

journal homepage: www.elsevier.com/locate/brat

Regression-based statistical mediation and moderation analysis in clinical research: Observations, recommendations, and implementation

Andrew F. Hayes*, Nicholas J. Rockwood

Department of Psychology, The Ohio State University, 1835 Neil Avenue, Columbus, OH 43210, USA

ARTICLE INFO

Article history:
Received 23 August 2016
Received in revised form
2 November 2016
Accepted 4 November 2016
Available online 5 November 2016

Keywords:
Mediation analysis
Moderation
Interaction
Regression analysis
Mechanisms

ABSTRACT

There have been numerous treatments in the clinical research literature about various design, analysis, and interpretation considerations when testing hypotheses about mechanisms and contingencies of effects, popularly known as *mediation* and *moderation* analysis. In this paper we address the practice of mediation and moderation analysis using linear regression in the pages of *Behaviour Research and Therapy* and offer some observations and recommendations, debunk some popular myths, describe some new advances, and provide an example of mediation, moderation, and their integration as conditional process analysis using the PROCESS macro for SPSS and SAS. Our goal is to nudge clinical researchers away from historically significant but increasingly old school approaches toward modifications, revisions, and extensions that characterize more modern thinking about the analysis of the mechanisms and contingencies of effects.

© 2016 Elsevier Ltd. All rights reserved.

Clinical research is about more than establishing that an effect exists, such as whether a new form of therapy is more effective than existing methods for treating certain conditions, or whether people who have certain experiences in life such as psychological trauma are more likely to suffer later in life from certain symptoms such as posttraumatic stress. It is just as important to understand how such effects operate and the boundary conditions of those effects. The former refers to the mechanism by which an effect is transmitted, whereas the latter speaks to the circumstances, contexts, or types of people for whom an effect exists and for whom it does not. Establishing boundary conditions is particularly important in application, because such understanding provides insight into the types of people for whom a particular therapeutic method works or does not, or what dispositions or attitudes might influence how much a life experience has an effect positive or negative down the road. There have been numerous treatments in the clinical research literature (e.g., Breitborde, Srihari, & Pollard et al., 2010; Kraemer, 2016; Kraemer, Wilson, Fairburn, & Agras, 2002; Magill, 2011) of various design, analysis, and interpretation considerations when testing hypotheses about mechanisms and contingencies of effects, popularly known as *mediation* and *moderation* analysis, respectively.

Given the importance of understanding the mechanisms and contingencies of effects, and the diverse perspectives in the methodology literature about how to test questions about mediation and moderation, we were asked by guest editors of *Behaviour Research and Therapy (BRaT)* to write a pedagogically-oriented overview of the practice of mediation and moderation analysis, so as to provide authors and reviewers some guidance on how to implement the advice offered by methodologists who think about these questions for a living. We took this as a challenge, and started by scanning the pages of the last five years of this journal to see what researchers are actually doing, noting in particular the kinds of designs researchers use and how they go about analyzing their data so that we could make an informed assessment of the conventions and procedures used by researchers in this area.

It didn't take us long to appreciate that the task we were invited to perform was next to impossible. There is too much diversity and complexity in method and design in the pages of this journal for us to provide a coherent treatment of best practices and current recommendations. We could have exhausted our entire page budget discussing just one specific method (e.g., mediation analysis) and one specific type of design (e.g., longitudinal), but doing so would have limited the value of this paper to only those who use such

* Corresponding author.
E-mail addresses: hayes.338@osu.edu (A.F. Hayes), rockwood.19@osu.edu (N.J. Rockwood).

http://dx.doi.org/10.1016/j.brat.2016.11.001
0005-7967/© 2016 Elsevier Ltd. All rights reserved.

designs. Yet most of the methods used by clinical researchers in the pages of *BRaT* have one thing in common, and that is their reliance on linear modeling principles. Given that many of the published examples we found in our perusal of the journal are based on straightforward linear regression analysis (as discussed by, say, Baron & Kenny, 1986), yet sometimes seemed to reflect a lack of appreciation for or awareness of current advances and changes in thinking, we decided to keep things simple and focus our treatment on the fundamentals applied with ordinary least squares (OLS) regression with continuous outcomes. Restricting our discussion to this simpler problem also allowed us to broaden the audience while shaving off material that would have been required to bring the typical reader up to speed on a more complex method. Still, many of the recommendations we offer in the OLS regression context generalize to more complex methods used by clinical researchers. To satisfy the request for a pedagogical treatment, we have kept the mathematics to a minimum when possible and discuss implementation of some of these methods using the PROCESS macro available for SPSS and SAS (Hayes, 2013) that has become widely used by researchers interested in testing hypotheses about moderation and mediation.

Throughout we provide references to examples of some of the things we have seen, most published in the pages of *BRaT*, illustrating points we make or things we recommend doing or not doing. It is not our intention to finger-wag when we cite examples of things we suggest avoiding or that represent outdated thinking. We recognize that substantive researchers doing meaningful clinical research have more important things to do than staying up to date on recent innovations, nuances, and updates in methodology, and that there is always a time lag between movements in methodology and implementation by those doing the substantive work of the business. Our goal is to nudge clinical researchers a bit in a particular direction rather than question the quality or value of the work being done by contributors to this journal.

Before diving in, we want to make our position on the role of data analysis in research clear from the outset, to avoid unnecessary confusion, overconfidence in what statistics can do by those who adopt some of our recommendations, and to preempt accusations that we are oversimplifying a complex problem in science. There are some hardliners who say that to claim the existence of cause-effect relationships (and mediation is by definition a cause-effect process), one must engage in experimental manipulation with random assignment, collect data over time or, ideally, both. Furthermore, one must meet an overwhelming number of assumptions beyond those of linear modeling that go by such names as “sequential ignorability,” “stable unit treatment value” and others, many that are quite technical in nature or hard or impossible to test. Others argue that one cannot conduct a mediation analysis with merely correlational data, that moderators must be independent of presumed causes of effects, and the list of requirements goes on and on (see e.g., Emsley, Dunn, & White, 2010; Preacher, 2015, for a discussion of many of these assumptions). We feel that if these are taken as literal requirements rather than as just ideals or recommendations, most research would not be done because most researchers cannot meet these requirements (due to resource constraints, ethics, and a myriad list of other reasons). Indeed, the use of such a high standard for causal inference would render most of the natural sciences unable to say anything about cause-effect relationships, given that experimentation, manipulation, and the various assumptions that social scientists often impose on themselves are rarely used or met in the natural sciences (c.f., Darlington & Hayes, 2017, pp. 166–168). We would rather see more imperfect work conducted and published than see research slow to a trickle because investigators don't feel that their work will satisfy all critics and pass every test for valid causal inference.

Our position is a more relaxed one reflecting our laissez-faire attitude about the role of data analysis in science (see Hayes, 2013, pp. 15–18, for a more extended discussion). Here, we don't dwell on some of the philosophical debates that one can find in the methodology literature about what cause-effect means, the limitations of various research designs for entertaining cause-effect questions, and the boundaries of the value of regression analysis and statistical control. This article is about data analysis, but we see data analysis as a tool, only one of many in a researcher's arsenal, and ultimately secondary to theory, knowledge of the literature in one's substantive area, and solid logical argument. Statistical methods are agnostic, indeed, ignorant about the origins of the data with respect to measurement and design. Inferences about substantive meaning are made not with output from routines built into statistical software, but by researchers who are attempting to make sense of and interpret that output. Inferences are products of our minds, not our mathematics. Any statistical method can be used on data regardless of its source as a tool to help guide the researchers' thinking about their data and their findings. So we don't agree that one cannot conduct a mediation analysis with correlational data, or that moderators must be uncorrelated with independent variables in order to do a moderation analysis. You can do most anything you want with your data. Most any statistical tool can provide some insight into the story you ultimately end up telling with your data.

1. Statistical mediation analysis

Mediation analysis is used when a researcher seeks to test hypotheses about or better understand how an effect of X on Y operates. The causal antecedent X could be which of two forms of therapy a client receives, or it could be an individual difference measure such as exposure to various sources of trauma, or any other conceivable variable that has some kind of causal force on a consequent outcome variable. That consequent Y could be something like frequency or severity of symptoms of some ailment, or how much satisfaction a person gets from interpersonal interactions in the course of day-to-day life. A therapeutic method (X) might affect symptoms experienced after the termination of therapy (Y) because the method influences how people interpret negative events that occur in life (M), and those interpretations then influence the extent to which symptoms are manifested. Or traumatic experiences (X) might negatively influence happiness one gets from interpersonal interactions (Y) because traumatic experiences result in the manifestation of certain behaviors that others find uncomfortable to witness (M), and this in turn produces less pleasant interactions. In both of these examples, X affects Y because X affects the *mediator variable* M, and this causal effect then transmits X's effect to Y through the effect of M on Y. Thus, a mediation model is a set of two or more causal events chained together in sequence of the form $X \rightarrow M \rightarrow Y$. So by definition, mediator variable M must be causally located between X and Y. It must be affected by X, and it in turn must affect Y.

Although mediation analysis has been around in various forms for at least 70 years or so, Baron and Kenny (1986) popularized an approach using easy-to-understand regression analysis principles. The overarching purpose of the analysis by their approach, sometimes called the *causal steps approach*, is to determine whether M can be deemed a mediator of the effect of X on Y. They described a series of analytical steps or *criteria* required to establish mediation. Whether these criteria are met is determined by estimating regression coefficients for X and M in three regression models, two with Y as the dependent variable and one with M as the dependent variable:

$$\widehat{Y} = i_Y + cX \quad (1)$$

$$\widehat{M} = i_M + aX \quad (2)$$

$$\widehat{Y} = i_Y + c'X + bM \quad (3)$$

These equations are represented in the form of a path diagram in Fig. 1, panel A. The first criterion is that *X* (which can be either dichotomous or continuous) must be related to *Y* (assumed to be continuous or at least treated analytically as such) manifested by a statistically significant path *c*, estimated using equation (1). This is called the *total effect* of *X* on *Y*. Having established a total effect of *X* on *Y*, the second criterion requires that *X* is related to *M*, meaning that the regression coefficient denoted *a* in Fig. 1, panel A, estimated using equation (2), is statistically significant. Once this criterion is met, one must then establish a statistically significant association between *M* (also assumed to be continuous in our treatment here) and *Y* when *X* is statistically controlled. This is the *b* coefficient in equation (3), and path *b* in Fig. 1, panel A. If these three conditions are met, and if the effect of *X* on *Y* when *M* is held constant (coefficient *c'* in equation (3), called the *direct effect* of *X*) is closer to zero than is *X*'s effect without controlling for *M* (coefficient *c* in equation (1), the *total effect* of *X*), then *M* can be deemed a mediator of *X*'s effect on *Y*.

The effects represented by the regression coefficients in equations (1)–(3) can be estimated with any regression analysis or structural equation modeling (SEM) program, but throughout this article we discuss model estimation using PROCESS, a convenient, free, and easy-to-use computational add on for SPSS and SAS documented by Hayes (2013) that can be downloaded from www.processmacro.org. We illustrate a mediation analysis by estimating a model examining the effect of mindfulness behaviour relapse prevention (MBRP) therapy relative to a traditional cognitive behavioral therapy ("therapy as usual") on alcohol and other substance use at a four month follow up, with craving for substances measured at two month follow up as the mediator. The data (fabricated for this example but inspired by a study conducted by Witkiewitz & Bowen, 2010) come from 168 clients of a public service agency randomly assigned to receive MBRP therapy or therapy as usual (coded 1 and 0, respectively, in a variable named MBRP). Craving for substances was measured at the start of therapy (CRAVE0, with higher scores representing greater craving) as well as two months following the termination of therapy (CRAVE2), and substance use was measured at four month follow up (USE4, with higher scores reflecting more frequent substance use). Also available is the number of hours of therapy administered (TREATHRS). The terms in capital letters are variable names in the data file (available at www.afhayes.com) and are used in PROCESS code in the appendices.

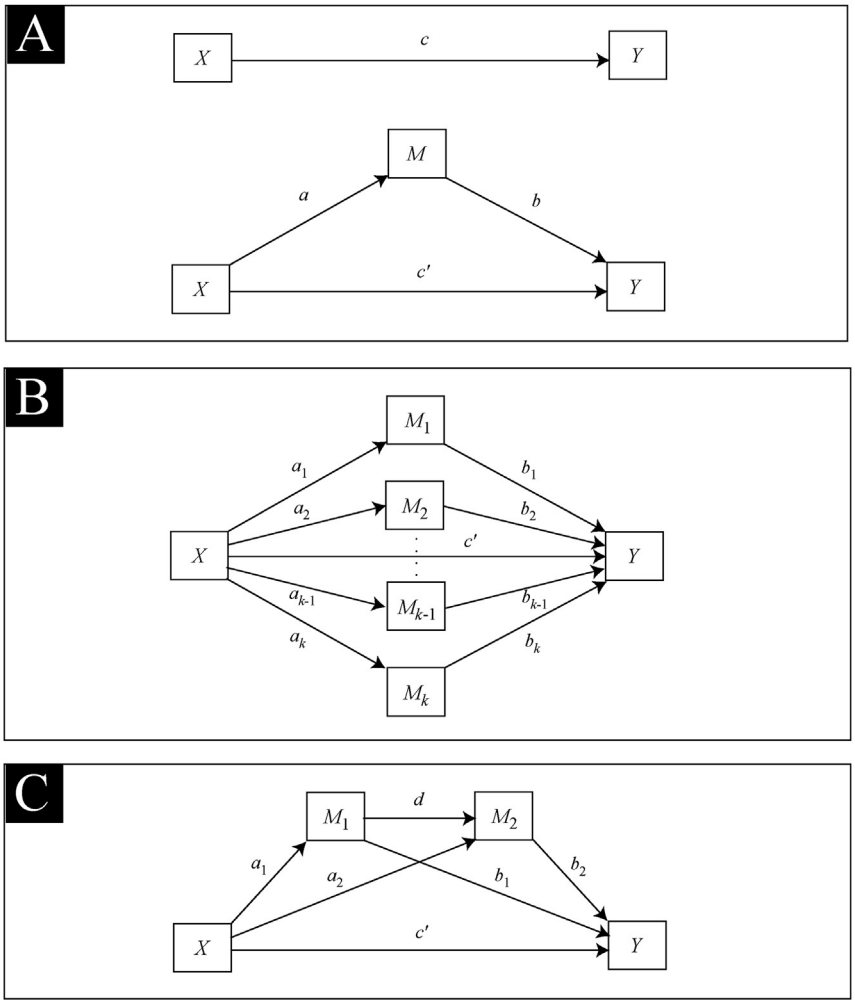

Fig. 1. Simple (panel A), parallel multiple (panel B), and serial multiple mediator models (panel C).

From our examination of the pages of *BRaT*, we noticed that clinical researchers often quantify change over time by measuring mediators and outcomes at some kind of baseline or "pretest" period and then "posttest," such as after the termination of therapy or otherwise after some period of time has elapsed. Most typically, difference scores are constructed by subtracting *M* or *Y* measured at the earlier time from *M* or *Y* measured at the later time (e.g., Boden, John, & Goldin et al., 2012; Pictet, Jermann, & Ceschi, 2016) and these difference scores as measures of change are used as *M* and *Y* in the analysis. Sometimes *X* is also a difference score constructed similarly. But we recommend avoiding the use of difference scores for any variable in this fashion unless also including the earlier measure as a covariate. Difference scores tend to be negatively correlated with initial state and positively correlated with final state (Campbell & Kenny, 1999, pp. 87–100) and this can make it appear that change has occurred differentially as a function of independent variables when regression to the mean, "self-selection," or other less interesting processes can explain trends in change just as well. Darlington and Hayes (2017, p. 135–144, 169–174) discuss and illustrate how it is generally better when modeling change over time to predict later measures from independent variables of interest while controlling for earlier measures. A mathematically equivalent approach is to model difference scores using earlier measures as covariates. Doing so generally produces more precise estimates of effects and more properly accounts for regression to the mean.

Accordingly, in our example, our mediator is craving at two month follow up rather than the difference in craving between follow up and baseline, with baseline craving used as a covariate in the model. We also use treatment hours as a covariate to account for the fact that those who received MBRP therapy received more therapy than those given therapy as usual. Unless strong theory would suggest otherwise or a covariate is likely caused by the outcome variable in the equation in which the covariate would otherwise be included (see Darlington & Hayes, 2017, pp. 538–541), we recommend including the covariates in all equations. Covariates are usually included in a model because a failure to include covariates leaves open alternative explanations for an association observed without them. But if the covariates are not included in each equation, then the associations representing the direct and indirect effects may still contain some component attributable to shared covariation with a covariate. So our example is defined by modified forms of equations (1)–(3), as such:

$$\widehat{Y} = i_Y + cX + c_1U_1 + c_2U_2 \tag{4}$$

$$\widehat{M} = i_M + aX + a_1U_1 + a_2U_2 \tag{5}$$

$$\widehat{Y} = i_Y + c'X + bM + b_1U_1 + b_2U_2 \tag{6}$$

where *X* is form of therapy received (0 = therapy as usual, 1 = MBRP therapy), *M* is craving at two month follow up, *Y* is use at four month follow up, and U_1 and U_2 are baseline craving and treatment hours, respectively. These equations are represented in path diagram form in Fig. 2. Using this approach, path *a* in equation (5) estimates how much more or less a person is craving substances at two month follow up, depending on form of therapy received, and relative to expected craving given the person's craving as measured at baseline while also accounting for differences between treatments in number of hours of treatment delivered. Path *b* in equation (6) quantifies how much more or less a person is using substances at four month follow up from craving at two months relative to expected use given initial craving and hours of treatment (as well as which treatment was given). So "change" in this analysis is measured as deviation from expectation rather than literally as the difference between measures.[1]

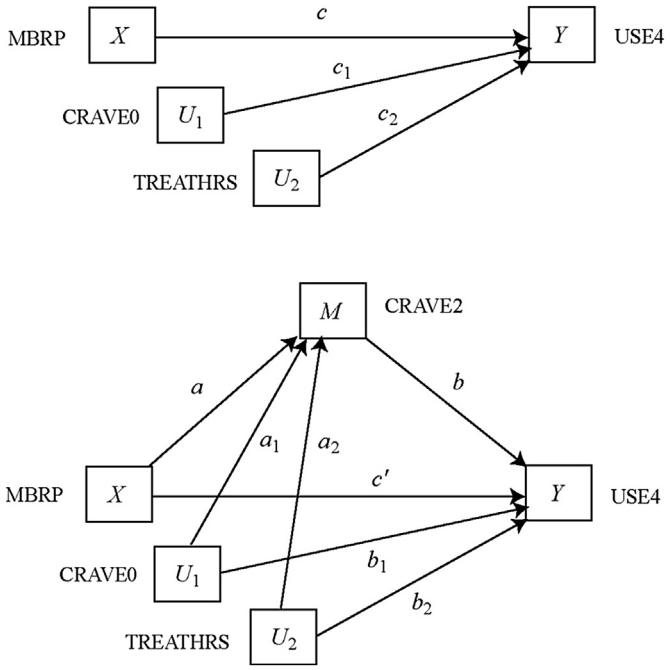

Fig. 2. A path diagram of a simple mediation model with 2 covariates representing equations (4)–(6) and corresponding to the mindfulness-based relapse prevention therapy example.

Appendix A contains PROCESS output for this model. It contains four main sections, one containing the model of craving at two month follow up (Equation (5), in the section labeled "Outcome: crave2"), one containing the model of substance use at four month follow up (Equation (6), the first section labeled "Outcome: use4"), one representing the total effect (Equation (4), under the heading "total effect model") and one containing a summary of the direct and indirect effects. The *a*, *b*, *c*, and *c*′ paths are labeled in the output. Though we discuss inference later, these results are consistent with the claim that MBRP therapy reduces later substance use through its effect on the reduction of craving (because *a* is negative, meaning those who received MBRP therapy craved less two months later, on average, relative to expected craving given initial craving and time in treatment), which in turn is related to a reduction in substance use (because *b* is positive; those who craved less at two month follow up used less after four months relative to expected from initial craving and treatment hours). We can say so even though the relationship between type of therapy and substance use, the total effect $c = -0.183$, is not statistically significant (more on this later). Controlling for craving at baseline and two month follow up as well as treatment hours, there is no direct effect of MBRP therapy on later substance use ($c' = 0.093$, but not statistically different from zero).

[1] Witkiewitz and Bowen (2010) did not have a baseline measure of substance use, so even though our data are fabricated for the purpose of this example, we respected the design as described in their article by excluding such a measure. Had such a measure been available, we would have used it as a covariate in equations (4)–(6) as well. Even so, the analysis we are doing here is not the same as the analysis reported in Witkiewitz and Bowen (2010).

1.1. A more modern framework

Reliance on the "criteria to establish mediation" logic described in Baron and Kenny (1986) can be found by researchers publishing in the pages of *BRaT* (see e.g., Boden, John, & Goldin et al., 2012; Gaynor, Ward, Garety, & Peters, 2013; Manuel & Wade, 2013; Smith, Scott, & Eshkevari et al., 2015; Staring, van den Berg, & Cath et al., 2016), but it is waning in popularity throughout the social sciences and largely no longer recommended by methodologists who think and write about mediation analysis. Although the mathematics of mediation analysis as described by Baron and Kenny (1986) and equations (1)–(3) (in our example, equations (4)–(6)) remains useful and important in more contemporary approaches, inference about mediation, at least since around the turn of the century, is now squarely focused on the *indirect effect* of *X* on *Y*. The indirect effect of *X* on *Y* through mediator *M* quantifies the estimated difference in *Y* resulting from a one-unit change in *X* through a sequence of causal steps in which *X* affects *M*, which in turn affects *Y*. Thus, it is the conjunction of the effect of *X* on *M* and the effect of *M* on *Y*. This indirect effect is estimated as the product of regression coefficients *a* and *b* in equations (2) and (3) (or 5 and 6). When using regression analysis to estimate the effects in a mediation model, *ab* is equal to $c - c'$ and therefore quantifies the difference between the effect of *X* on *Y* when *M* is controlled versus when it is not (this is also true when covariates are used so long as the same covariates are used in all the equations). A rejection of the null hypothesis that the indirect effect is zero (or an interval estimate that doesn't include zero) is sufficient to support a claim of mediation of the effect of *X* on *Y* through *M*. PROCESS does these computations automatically whenever a mediation model is specified in the command line. As can be seen toward the end of the output in Appendix A, the indirect effect of MBRP therapy on later substance use through craving is −0.276, which is the product of *a* (−0.574) and *b* (0.481) and also the difference between *c* and *c'* (−0.183−0.093 = −0.276). As discussed later, this indirect effect is statistically different from zero.

By contemporary thinking, tests of significance for the individual paths *a* and *b* are not required to determine whether *M* mediates the effect of *X* on *Y*, contrary to the causal steps logic which requires that both *a* and *b* are statistically significant. Indeed, one does not even need to establish that the total effect of *X* as quantified by *c* is different from zero, since the size of *c* does not determine or constrain the size of *ab*. Rather, all that matters is whether *ab* is different from zero by some kind of inferential standard such as a null hypothesis test or confidence interval. We recommend that researchers interested in testing a mediation hypothesis move away from the piecemeal causal steps logic popularized by Baron and Kenny (1986) by focusing attention on estimating the indirect effect *ab*, conducting an inference about that product, and interpreting the indirect effect by considering not only its sign, but also the sign of its constituent components (paths *a* and *b*). The significance or nonsignificance of *a* and *b* may be diagnostic of the likelihood that *ab* is significant, but these aren't the story or even important. What matters is *ab*, not *a* and *b*.

As this recommendation is likely to be perceived as controversial, a bit of elaboration is worthwhile. It is founded on three common-sense principles of inference (see Hayes, 2015). The first principle is that any empirical claim should be based on a quantification of the effect most directly relevant to that claim. If *X*'s effect on *Y* is mediated by *M*, this means that changing *X* by one unit will result in an estimated change in *Y* equal to *ab* through a sequence of causal steps whereby *X* affects *M* and *M* affects *Y*. So *ab* directly quantifies the movement of *Y* by *X* through *M*. Neither *a* nor *b* quantify this. Why would we base a test on mediation on tests of *a* and *b* when neither *a* nor *b* by themselves quantify the change in *Y* through *M* resulting from a change in *X*?

A rebuttal is that if both *a* and *b* are different from zero, then so too must the their product, and if either *a* or *b* is zero, then so too must be *ab*. Although this logic is sensible, the second principle counters this rebuttal: A claim should be based on as few inferential tests as required in order to support it. Statistical tests are merely tools used by researchers to make decisions under uncertainty. They are fallible by nature, and we all recognize that they can lead to mistakes (Type I and Type II errors). Given their imperfect nature, why require three imperfect statistical tests (one for *c*, one for *a*, and one for *b*) in order to test a mediation hypothesis when all this is needed is a single statistical test on *ab*?

The third principle is that when possible, we should convey information to the consumer of our work about the uncertainty attached to estimates of quantities used to support our claims (c.f., Kelley & Preacher, 2012). The causal steps criteria result in a dichotomous decision about mediation. Is *M* a mediator or not? But as discussed earlier, mediation manifests itself empirically in the form of an indirect effect, which is a quantity that can be estimated. Furthermore, using a variety of methods (discussed later), we can express the inference in the form of a confidence interval conveying the uncertainty attached to the estimate. This provides more information to the reader than does a simple claim as to whether *M* is functioning as a mediator or not.

Although a single inference about the indirect effect as quantified with *ab* is all that is needed to test a mediation hypothesis, this doesn't mean we can just ignore *a* and *b* when interpreting the results. Suppose you predict a positive indirect effect of *X* on *Y* through *M*. Recognize that $ab > 0$ can result from *a* and $b < 0$, or *a* and $b > 0$. Yet these two patterns of signs would have completely different substantive interpretations. Your theory might predict a positive indirect effect because *X* should be positively related to *M* and *M* should be positively related to *Y*. It certainly would be inappropriate to claim support for your theory based on a positive indirect effect if both *a* and *b* turned out to be *negative*. Similarly, a negative indirect effect can result from $a > 0$ and $b < 0$, or from $a < 0$ and $b > 0$. So it is important to consider the signs of *a* and *b* when interpreting an indirect effect and whether one's mediation hypothesis is supported. But there is no need to require that both *a* and *b* are different from zero by a hypothesis test or confidence interval standard.

What about mediation without evidence of a total effect of *X*? Many researchers find it difficult to fathom how an indirect effect can exist if a total effect does not. The problem resides in the mistaken belief that *c* is always a substantively useful measure of the effect of *X* on *Y*. More specifically, the problem is the misunderstanding that if *c* is zero, this means *X* has no effect on *Y*. But $c = 0$ means only that on the aggregate, when all paths of influence between *X* and *Y* are added up, *X* and *Y* are linearly unrelated. This does not mean *X* doesn't affect *Y*. As Bollen (1989) long ago articulated, "correlation is neither necessary nor sufficient for causation" (p. 52). In the algebra of path analysis expressed in equations (1)–(3) (as well as 4–6), $c = c' + ab$.[2] So, for example, *X* may positively affect *Y* indirectly, but if *X* negatively affects *Y* directly and this direct effect is of the same magnitude as the indirect effect but opposite in sign, then $c = 0$. For an example of this in the clinical research literature, see Seehuus, Clifton, and Rellini (2015). Some people call this *inconsistent mediation* or evidence of *suppression*. But labels don't help explain such a phenomenon; only substantive understanding of the theory and literature can, and usually it is not

[2] This is true in equation (4) through (6) if the same covariates are included in all three of the equations. If you distribute covariates differently across the three equations, then typically $c \neq c' + ab$.

difficult to come up with substantively sensible interpretations of direct and indirect effects that are opposite in sign. However, such competition between direct and indirect effects is not required for this phenomenon to occur. For example, Fillo, Alfano, and Paulus et al. (2016) found a statistically significant negative indirect effect of sleep disturbance on relapse situation self-efficacy through emotion dysregulation but nonsignificant negative direct and total effects. And in more complex models, the total effect is the sum of the direct effect and all possible indirect effects, of which there may be many, and various combinations of sizes of direct and indirect effects can produce a total effect equal to zero. We discuss this more later and provide an example.

Given that *c* does not have to be and often won't be larger than *ab* and can even be much smaller or even of opposite sign, it is hard for us to recommend the ratio of *ab* to *c* as a measure of effect size in mediation analysis. Researchers who report this measure (in *BRaT*, e.g., Boden et al., 2012; Manuel & Wade, 2013; McLean, Yeh, Rosen, & Foa, 2015) typically interpret this ratio as the proportion (or percentage when expressed on a 0 to 100 metric) of the effect of *X* on *Y* that is mediated. But a proportion must be between 0 and 1, a property that *ab/c* does not have. This *ratio* (a better term than "proportion" or "percentage") can be any real number. For example, in the MBRP therapy example, $ab/c = -1.508$, which clearly is not a proportion. Because 0 and 1 are not the upper and lower bounds of *ab/c,* it cannot be interpreted as a proportion; indeed, a number such as 0.25 does not have any sensible interpretation without a lower or upper bound, i.e., there is no context for interpreting whether 0.25 is large or small. We recommend avoiding this measure of effect size or discussing mediation results in terms of the "proportion" of the effect that is mediated. Other measures of effect size exist that are better, though none are without limitations. If *X* and *Y* are on meaningful metrics, *ab* can be inherently meaningful as an effect size measure. For experiments with a dichotomous *X* with *X* coded such that the two groups differ by one unit (e.g., 0 and 1, or −0.5 and 0.5), the partially standardized indirect effect (defined as *ab* divided by the standard deviation of *Y*) is a decent measure. Hayes (2013, p. 184–193) and Preacher and Kelley (2011) discuss some other options.[3]

So it is outdated to insist that one has evidence of an association between *X* and *Y* prior to testing a mediation hypothesis. This perspective is now widely disseminated in various fields (see e.g., Cerin & MacKinnon, 2009; Hayes, 2009; Rucker, Preacher, Tormala, & Petty, 2011; Shrout & Bolger, 2002; Zhao, Lynch, & Chen, 2010), though the practice of conditioning a mediation analysis on a statistically significant total effect of *X* remains common among followers of the Baron and Kenny logic. Recently, Kenny and Judd (2014) provided a new argument that should convince any doubters. They show that the standard test on the total effect (i.e., a hypothesis test or confidence interval for *c* in equation (1) or 4) is generally conducted with less power than are tests of an indirect effect of equal size. So it makes no sense to condition one's decision to seek evidence of an indirect effect on a test of the total effect that is less trustworthy (i.e., less powerful) than tests on the indirect effect.

1.2. Approaches to inference about the indirect effect

As already discussed, we recommend inference about mediation be based on an inference about the indirect effect of *X* on *Y* estimated as *ab* rather than individual tests of the paths of the model (*c*, *a*, and *b*). There are many approaches to inference about the product of regression coefficients, and much research comparing their validity and power. One of the earlier tests, the Sobel test (Sobel, 1982), is sometimes used in the pages of *BRaT* (see e.g., Armstrong, Zald, & Olatunji, 2011; Gaynor et al., 2013; Pictet et al., 2016). It relies on an estimate of the standard error of *ab*. Assuming the sampling distribution of the ratio of *ab* to its standard error is normal, a *p*-value or confidence interval can be constructed using the standard normal distribution. Though very popular, research shows that the Sobel test is lower in power than alternatives, primarily because the sampling distribution of *ab* is typically not normal (e.g., Hayes & Sharkow, 2013). Using an inappropriate reference distribution for the sampling distribution of a statistic can lead to decision errors and poor confidence intervals. Although PROCESS will produce the Sobel test (see the documentation), we recommend avoiding this test.

There are alternatives that research shows perform better than the Sobel test without making any assumption about the shape of the sampling distribution of *ab*. These include the bootstrap confidence interval (Shrout & Bolger, 2002; Preacher & Hayes, 2004, 2008), the Monte Carlo confidence interval (Preacher & Selig, 2012), and the Bayesian credible interval (Yuan & MacKinnon, 2009).[4] Research shows these all perform well while also outperforming the Sobel test (Fritz, Taylor, & MacKinnon, 2012; Hayes & Scharkow, 2013; Preacher & Selig, 2012; Williams & MacKinnon, 2008; Yuan & MacKinnon, 2009). Of these, the bootstrap confidence interval has become especially popular, perhaps because it is widely implemented in macros written for SPSS, SAS, and R (e.g., Hayes, 2013; Tofighi & MacKinnon, 2009), including PROCESS, and is also built into some SEM programs such as Mplus and AMOS.

A bootstrap confidence interval for the indirect effect is constructed by randomly resampling *n* cases from the data with replacement, where *n* is the original sample size in the study, and estimating the model and resulting indirect effect *ab* in this *bootstrap sample*. Repeated thousands of times, an empirical representation of the sampling distribution of *ab* is built and a confidence interval for the indirect effects constructed using various percentiles of the bootstrap distribution. For example, the 2.5th and 97.5th percentiles of the bootstrap distribution of *ab* define the upper and lower bounds of a 95% bootstrap confidence interval for the indirect effect. If the interval is entirely above or below zero, this supports a claim of mediation, whereas a confidence interval straddling zero does not provide definitive evidence that *X*'s effect on *Y* operates through *M*. PROCESS constructs a bootstrap confidence interval automatically for any model that includes an indirect effect. As can be seen at the end of the output in Appendix A, the bootstrap confidence for the indirect effect of MBRP therapy on substance use through craving is −0.502 to −0.125. As this does not straddle zero, this is evidence of a statistically significant indirect effect.[5]

In our perusal of *BRaT*, we found several instances where researchers employed more than one test of mediation. For example, Boden et al. (2012), Smith, Scott, & Eshkevari et al. (2015), and Staring et al. (2016) conducted an analysis relying on the causal

[3] Preacher and Kelley introduce kappa-squared as a new and seemingly sensible measure of effect size in mediation analysis. But kappa-squared has since been discredited by Wan and Fan (2015), who pointed out some mathematical errors made in its derivation.

[4] Some consider the *test of joint significance* a test of the indirect effect that makes no assumption about the shape of the distribution of *ab*. The test of joint significance rejects the null that the indirect effect is equal to zero if two null hypotheses, one that $a = 0$ and one that $b = 0$, are both rejected. Although the joint significance test performs well, it violates all three of the principles of inference described earlier. Thus, we don't recommend using this test.

[5] Technically, we can't say $p < 0.05$ because the bootstrap distribution of *ab* is not constructed conditioned on a true null hypothesis. But substantively, the conclusion is the same. We can be pretty sure (95% confident) that the indirect effect is negative.

steps criteria to establish mediation and then followed up by the use of a bootstrap confidence interval for the indirect effect. Boden et al. (2012) threw in an additional Sobel test for the indirect effect for good measure. And Sandin, Sanchez-Arribas, Chorot, and Valiente (2015) report bootstrap confidence intervals and the Sobel test without relying on the causal steps criteria. We see little point to this redundancy and consider it analytical overkill. Only one test of mediation is required, and we recommend the use of a bootstrap confidence interval or another method that makes no assumption about the shape of the sampling distribution of *ab*.

1.3. On "complete" or "full mediation"

Researchers frequently attempt to label a mediation process as *partial* or *complete*, language frequently used in mediation analyses reported in *BRaT* (e.g., Deacon, Kemp, & Dixon et al., 2013; Manuel & Wade, 2013; Morgan, MacKinnon, & Jorm, 2013; Staring et al., 2016; Vincent & Walsh, 2013). Partial mediation refers to a pattern of findings where mediation is established in the presence of significant total effect of *X* and the direct effect of *X* (c') is statistically different from zero. With partial mediation, the effect of *X* on *Y* is not completely explained by the $X \rightarrow M \rightarrow Y$ sequence of events. But if the direct effect is not statistically different from zero, then *M* is deemed a complete mediator (also called *full mediation*). That is, by this reasoning, all of the effect of *X* on *Y* is carried through the mediation process, meaning that $ab = c$ and thus $c' = 0$.

Hayes (2013, pp. 170–172) and Rucker et al. (2011) argue the complete and partial mediation concepts have little value and should be abandoned. Here we focus only on complete mediation. If we take as a given that being able to completely explain an effect is a more desirable and more impressive empirical outcome than only partially explaining an effect, then we should be more impressed by empirical outcomes based on smaller samples, and we should seek smaller samples so as to maximize the likelihood of being able to claim complete rather than partial mediation. As justification for this position, consider two investigators who have identical results (i.e., equivalent total, direct, and indirect effects) but their studies are based on different sample sizes. In that situation, the investigator with the smaller sample size will be more likely to be able to claim complete mediation than the investigator with the larger sample because the direct effect will be tested in the former investigator's study with less power. So we will be more impressed by the findings of the first investigator even though he or she has less data. By this reasoning, it is to an investigator's advantage to minimize the sample size so that there is sufficient power to detect the indirect effect, but insufficient power to detect the direct effect. Of course, this is crazy and contrary to the way we think about collecting data and evaluating results. All other things being equal, we should have more faith in and be more impressed by studies based on more data, and we should aspire to collect as much data as possible. The problem is that the distinction between complete and partial mediation is empirically determined primarily by the power of the test of the direct effect, which of course is determined in part by sample size.

Another problem with complete mediation is that such a claim suggests only a single variable can completely mediate the effect of *X* on *Y*. That is, if one investigator finds evidence of complete mediation of the effect of *X* on *Y* by his or her *M*, this implies that no additional research is needed to explain the effect of *X*, that no additional theory is needed to account for the process by which *X* affects and *Y*, and no other theory will do so as adequately. Yet Rucker et al. (2011) show that this isn't true. Two different investigators can both completely account for the effect of *X* on *Y* with their favored mediator. But if two investigators can both completely account for the effect of *X* on *Y* through their favored mediator, what value is there in the claim that any specific mediator completely accounts for the effect of *X*?

Given these problems with complete mediation as a concept, we recommend that researchers avoid using this term when articulating mediation hypotheses or attempting to label or interpret the results of a mediation analysis.

1.4. More than one mediator

Most causal effects probably operate through more than one mechanism, and so it is worth testing multiple mechanisms when theory or hypothesis suggest a more complex process than can be modeled with a single mediator. For instance, Vincent and Walsh (2013) examined the effect of a computerized cognitive behavioral therapy on quality of sleep through the effects of the therapy on pre-sleep arousal, time in bed prior to sleep, and consistency in awake and arise times. And Arch (2014) found that pregnant women were less willing to receive pharmacotherapy because pregnant women perceived it as less credible than nonpregnant controls and also had more concerns about its use, both of which were related to willingness to combine therapy with drugs.

In these two examples, the mediation processes were estimated as operating "in parallel." A *parallel multiple mediator* with *k* mediators is displayed in Fig. 1, panel B. In such a model, mediators may be and often are **c**orrelated, but nothing in the model allows one mediator to causally influence another. A more complex variant is a *serial multiple mediator model*, depicted in Fig. 1 panel C. In a serial multiple mediator model, mediators are linked together in a causal chain, with one mediator allowed to influence one or more mediators causally downstream. For example, Newby, Williams, and Andrews (2014) estimated such a model of the effect of internet-delivered cognitive behavioral therapy for treatment of anxiety and depression on changes in positive meta-cognitive beliefs, which in turn was proposed to reduce repetitive negative thinking, which in turn would then result in an improvement in symptoms.

We found some examples in *BRaT* where investigators postulated more than one mechanism at work carrying the effect of *X* on *Y* but then estimated a set of two or more simple mediation models, each with a single mediator (e.g., Beadman et al., 2015; Kuckertz, Amir, & Boffa et al., 2014). We recommend that, instead, investigators interested in mediation through more than one mediator do so by estimating all the indirect effects in one multiple mediator model. There are several rationales for our position (see Hayes, 2013; Preacher & Hayes, 2008). First, if more than one mechanism is in operation, as actually theorized or hypothesized, estimating a set of simple mediation models with one mediator reduces the correspondence between theory and model or method. If the more complex process hypothesized or theorized is actually operating, each of the simpler models is inherently misspecified. Although all models are misspecified in some way, why not minimize the misspecification as much as possible by maximizing the correspondence between theory and mathematical model?

Second, when proposed mediators are correlated, as they often are, at least one of the indirect effects revealed by a set of single-mediator analyses may be *epiphenomenal*. Like spuriousness, epiphenomenality is an explanation for an association between two variables. If *X* causes *Y*, and *X* is correlated with *Z* (for whatever reason), then *Z* may be correlated with *Y* as well. One might then be inclined to think of *Z* as a potential cause of *Y* even though it is merely correlated with the true cause of *Y* and not causally influencing *Y* itself. The apparent causation by *Z* is ephiphenomenal. Applied to mediation analysis, if your proposed mediator M_1 is not actually mediating the effect of *X* on *Y* yet is correlated with M_2, which *is* a mediator of the effect of *X* on *Y*, a mediation analysis with

M_1 but not M_2 in the model may nevertheless reveal a significant indirect effect of X on Y through M_1. Including both M_1 and M_2 as mediators in the model reduces this problem, because each indirect effect estimates the component of the indirect effect of X on Y through a specific mediator that is *unique* to that mediator (i.e., after controlling for the effect of the other mediators in the model on Y).

Third, and perhaps the most valuable benefit of estimating a model with multiple mediators, it is possible to compare the size of indirect effects through different mediators (see Hayes, 2013; MacKinnon, 2000; Preacher & Hayes, 2008). This can be useful for competitive theory testing. Theory A may propose that the effect of X on Y operates through M_1, whereas theory B proposes that M_2 is the mediator transmitting X's effect on Y. With both theories' mediators in the model simultaneously, you can conduct a statistical test of the difference in the indirect effects through each mediator. It may be, for example, that both indirect effects are different from zero, but the indirect effect through theory A's mediator is larger than the indirect effect through theory B's. Such a comparison is possible even if the two mediators are on different measurement scales because the scaling of a mediator is not a part of the scaling of an indirect effect through that mediator. The mathematics behind this can be found in Preacher and Hayes (2008). Intuitively, this can be understood by recognizing that an indirect effect is interpreted purely in the units of X and Y. So if two indirect effects share X and Y, they are on comparable measurement scales.

The estimation of indirect effects in a multiple mediator model is only a little more complicated than for the simple mediation model. In a parallel multiple mediator model with k mediators, there are k *specific indirect effects*, one for each mediator. The specific indirect effect of X on Y through mediator j is estimated as the product of a_j in equation (7) and b_j in equation (8):

$$\widehat{M}_j = i_{M_j} + a_j X \tag{7}$$

$$\widehat{Y} = i_Y + c'X + \sum_{j=1}^{k} b_j M_j \tag{8}$$

In other words, to get the indirect effect of X on Y through M_j, regress M_j on X (equation (7)) and Y on X and all k of the mediators (equation (8)). With k mediators, there are k versions of equation (7) differing by which mediator is on the left side of the equation, and therefore k a paths, one for each mediator, which is multiplied by that mediator's corresponding effect on Y in equation (8) (i.e., b_j) to get the specific indirect effect of X on Y through M_j. The sum of the k specific indirect effects, when added to the direct effect of X (c' in equation (8)) yields the total effect of X (c from equation (1)).

For the serial multiple mediator model, a modification is needed to estimate the specific indirect effects. If mediators are arranged in a causal sequence, such that M_1 is causally prior to M_2, M_2 causally prior to M_3, and so forth, then the model of mediator j should include X and all mediators causally prior to M_j. For instance, in a serial multiple mediator model with 2 mediators, one estimates M_1 from only X, as in equation (7), but M_2 is estimated from X and M_1:

$$\widehat{M}_2 = i_{M_2} + a_2 X + dM_1 \tag{9}$$

Finally, equation (8) is used to estimate all the effects of the Ms on Y, generating the b paths and c', the direct effect of X. This model has three specific indirect effects, one through M_1 only (estimated as a_1b_1), one through M_2 only (estimated a_2b_2), and one through M_1 and M_2 in serial (estimated as a_1db_2). As with the parallel multiple mediator model, these specific indirect effects, when summed and added to the direct effect c' from equation (8), produce the total effect c estimated with equation (1).

Inference for a specific indirect effect in the parallel and serial multiple mediator models can proceed as discussed for the simple mediator model. A bootstrap confidence interval for the specific indirect effect through a mediator or sets of mediators that does not include zero is sufficient to support a claim of mediation of the effect of X on Y through that specific pathway. The PROCESS macro for SPSS or SAS provides not only model coefficients but also bootstrap confidence intervals for inference about indirect effects in parallel and serial multiple mediator models. An SEM program that can conduct an inference for functions of model coefficients can also be used.

A bootstrap confidence interval can be used to test the equality of any two specific indirect effects, thereby allowing a claim as to whether one indirect effect is different than another. For instance, in a parallel multiple mediator model, the difference between the indirect effect through mediator M_i and mediator M_j is $a_ib_i - a_jb_j$, and the distribution of this difference can be used for inference about the difference between the indirect effects. Spinhoven, Pennnix, & Krempeniou et al. (2015) used this approach when comparing the indirect effects of trauma rumination on PTSD onset through trauma-related affect (anxiety) and trauma-related cognitions (beliefs about impact on life). They found the indirect effect was larger through affect than through cognitions, suggesting anxiety is the more pronounced or stronger of the two mechanisms.

The mathematics of the multiple mediator model provide another context for understanding why it is a mistake to ask questions about mediation only if the total effect of X on Y is statistically significant. The total effect c is the sum of the direct effect of X and all specific indirect effects. Each specific indirect effect represents one mechanism or pathway of influence. These pathways may be similar or different in size or sign. But if their sum adds up to something close to the direct effect c' but opposite in sign (or if the sum is zero and so is the direct effect) then the total effect is zero. One could even find that every specific indirect effect is statistically different from zero even though the total effect is not. If this seems far-fetched, consider a study of U.S. Army medics conducted by Pitts and Safer (2016). They found no statistically significant relationship between combat experience (X) and depression (Y). Yet a parallel mediation model with two mediators (feelings of threat and positivity of their views of the combat experience) revealed two statistically significant indirect effects, one positive and one negative. Medics who had more combat experience perceived greater threat to self during combat (M_1), and this greater threat was related to higher depression (and thus a positive indirect effect). At the same time, those who experienced more combat had more positive views about their combat experiences (M_2), which was *negatively* related to depression (and hence a negative indirect effect). The absence of a direct effect, when added to two indirect effects comparable in size but different in sign, resulted in a total effect that was small and not significantly different from zero.

1.5. Multicategorical independent variables

The path analysis algebra and all the equations described thus far are based on the requirement that X is either dichotomous or continuous. But sometimes X is multicategorical, such as in an experiment or clinical trial that involves at least two treatment conditions and one or more control groups. For example, Beadman et al. (2015) conducted a study examining the direct and indirect effects (through credibility and expected success) of three strategies for interpreting and responding to cravings (diffusion, reappraisal, and suppression) on changes in smoking behaviour after

exposure to the strategies. In this circumstance, researchers not familiar with alternatives may end up resorting to such practices as collapsing some groups into one so as to produce a dichotomous *X*, or doing separate analyses on only two groups after discarding the other groups from the analysis (as did Beadman et al., 2015; also see Trompetter, Bohlmeijer, Fox, & Schreurs, 2015) thereby reducing power for detecting any effects. Alternatively, researchers may rely on the causal steps logic using analysis of variance rather than regression analysis.

These approaches are problematic, and none of them are necessary. As many users of regression analysis are aware, a multicategorical variable with k categories can be included in a regression model if it is properly represented using a coding system with $k - 1$ variables representing groups. *Dummy* or *indicator coding* is one common strategy for representing groups, but there are others (see Darlington & Hayes, 2017). Hayes and Preacher (2014) describe a procedure for regression-based mediation analysis when *X* is multicategorical. We have not yet seen this method used in any study published in *BRaT* and recommend researchers interested in doing a mediation analysis with a multicategorical *X* familiarize themselves with this procedure, which is easy to do using PROCESS.

2. Moderation analysis

Whereas mediation analysis focuses on *how* a causal effect operates, moderation analysis is used to address, *when*, or under *what circumstances*, or for *what types of people* that effect exists or does not and in what magnitude. Fig. 3, panel A, graphically depicts the concept of moderation. In this figure, the arrow linking *W* to the effect of *X* on *Y* denotes that *X*'s effect on *Y* depends in some way on *W*. More specifically, *X*'s effect on *Y* is said to be moderated by *W* if the size or sign of *X*'s effect on *Y* varies with *W*. For example, compared to a traditional therapy, a new therapy (*X*) might be effective at reducing symptoms of depression (*Y*), but that effect might be smaller, or perhaps the therapy is even *harmful*, among people suffering from anxiety (*W*). Or traumatic experiences (*X*) may reduce how satisfied people are with their relationships (*Y*), but perhaps having an empathetic social support network of people who have also experienced trauma (*W*) provides a buffer against the negative effects of trauma.

Moderation is also popularly known as *interaction.* If *X*'s effect on *Y* is moderated by *W*, then *X* and *W interact*. Most researchers are familiar with the concept of interaction through exposure to analysis of variance, and moderation hypotheses are regularly tested by clinical researchers who examine whether the effect of an intervention or individual difference (*X*) varies over time (e.g., with *Y* measured at baseline versus a follow up versus, perhaps, another follow up). In such an analysis, time is the proposed moderator *W*. But anything can be a moderator, whether measured discretely or continuously, and analysis of variance is just a special case of linear regression analysis with categorical independent variables. Interaction or moderation hypotheses can be tested with regression analysis without requiring that *X* and *W* are categorical.

In a regression model of the form $\hat{Y} = b_0 + b_1X + b_2W$, *X*'s effect on *Y* is fixed to be the same—b_1—regardless of the value of *W*. This is a serious analytical constraint, and the opposite of what is desired when testing a moderation hypothesis. In order to use regression analysis to test a moderation hypothesis, this constraint on *X*'s effect must be released. This can be accomplished by specifying *X*'s effect to be a *function* of *W*. A linear function is commonly used, though others are possible. If we substitute $b_1 + b_3W$ for b_1 in the above equation, we get

$$\widehat{Y} = i_Y + (b_1 + b_3W)X + b_2W \tag{9}$$

which is mathematically equivalent to

$$\widehat{Y} = i_Y + b_1X + b_2W + b_3XW \tag{10}$$

A test of linear moderation is conducted with a hypothesis test or confidence interval for the regression coefficient for *XW*, which is equivalent to an inference about whether the weight for *W* in the linear function defining *X*'s effect ($b_1 + b_3W$) is equal to zero. If b_3 is different than zero, this means that *X*'s effect on *Y* varies with *W*. This procedure works whether *W* is dichotomous or continuous. If *X* or *W* are multicategorical, modifications to the regression math are needed. See Hayes and Montoya (2016) and Darlington and

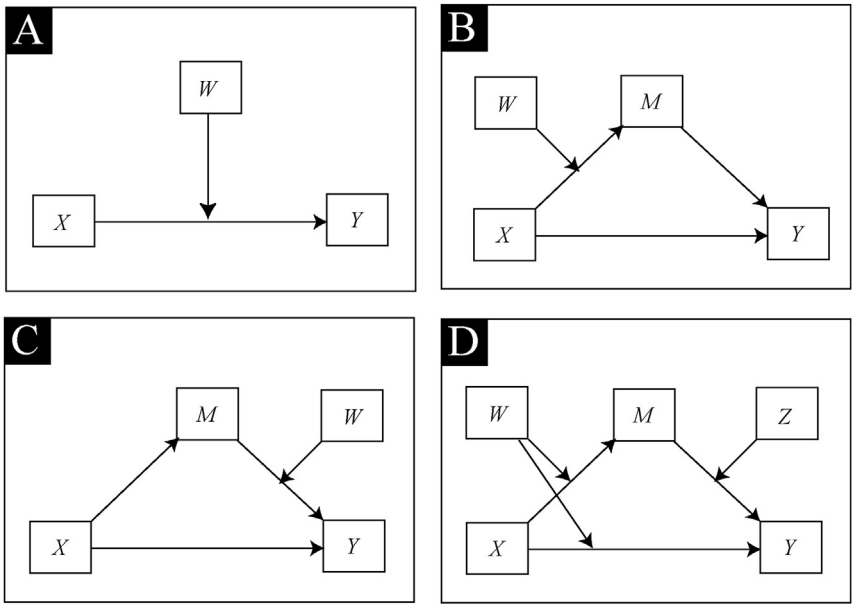

Fig. 3. Simple moderation (panel A), and three conditional process models illustrating first stage (panel B), second stage (panel C), and first and second stage moderated mediation combined with moderation of the direct effect (panel D).

Hayes (2017) for a discussion.

To illustrate, we examine whether MBRP therapy's effect on craving relative to therapy as usual is dependent on a person's level of depression at the start of therapy. Assuming the dependency between the effect of MBRP therapy and depression is linear, we estimate equation (10), where X is treatment condition, Y is craving at two month follow up, and W is score on the Beck Depression Inventory at the start of therapy (BDI0 in the data). As in our mediation analysis, we statistically control for craving at the start of therapy (U_1) and treatment hours (U_2) by including them in the model. PROCESS does all the work using OLS regression, including the construction of the product of X and W. The PROCESS command and output can be found in Appendix B.[6] The resulting model, found in the first section of the PROCESS output, is

$$\hat{Y} = 1.031 + 0.599X + 0.054W - 0.045XW + 0.192U_1 - 0.018U_2 \quad (11)$$

with a weight for XW, $b_3 = -0.045$, that is statistically significant. We conclude that the effect of MBRP therapy relative to therapy as usual on later craving, and relative to expected craving given a person's craving prior to the start of therapy, depends on depression level at the start of therapy.

2.1. Visualizing interaction

A test of moderation is usually a vague test. If the regression coefficient for the product of XW is statistically significant, this means only that the effect of X on Y depends on W. Interaction can take many forms, and it is important to visualize the model in order to make sense of how X's effect varies with W. We recommend visualizing any interaction, both for yourself and for the reader, that represents a nontrivial component of your substantive findings.

When X and W are both categorical and there are no covariates, visualizing an interaction is a fairly simple task. One simply calculates the mean of Y in the various "cells" of the design (e.g., each cell mean in a 2×2 design) and then represents these in some form, such as a bar or line plot. But when X or W is a continuum or when the model includes covariates, visualization can be tricky. The temptation is to categorize X or W when continuous (e.g., using a median or mean split or through trichotomization or an alternative method), construct means for Y for the various artificial groups this procedure creates, and then proceed to plot as one would if reporting an interaction following an analysis of variance. We don't recommend this approach, as it does not respect the mathematics of the model that generates the test of interaction you are trying to visualize and so won't properly visualize the interaction you are reporting. That is, the test of the interaction is conducted using the variables in their continuous form and plotting the data in a categorized form may display a different pattern than the model produces. Further, different categorization methods can result in different plots. There is a better procedure that represents the interaction as tested in the model with the variables in the original (i.e., uncategorized) form. A detailed description of this procedure is available in Hayes (2013, p. 231–234) that we only summarize here.

The regression equation is a model that generates estimates of Y for various combinations of X and W. To visualize a regression model with an interaction, choose various combinations of X and W that are within the range of the data and plug these combinations into the regression equation to produce estimates of Y for these combinations of X and W. If the model includes covariates, set them to their sample means (this is legitimate even for covariates that are dichotomous and the numerical codes for the two groups are arbitrary). Be sure to use values of X and W that make sense. This is easy if X or W is dichotomous. Just use the two values coding the two groups. For continuous X or W, choose values that aren't too extreme in the distribution, such as the 25th, 50th, and 75th percentile, or perhaps the mean, a standard deviation below the mean, and a standard deviation above the mean. Avoid using the maximum and minimums unless you have many values in the data at these extremes, and also make sure if you use a standard deviation below or above the mean that these are within the range of the observed data (i.e., not above the maximum or below the minimum of observed values). Once you have selected these values and generated estimates of Y using the regression equation and these combinations of X and W, then plot using whatever graphing software you prefer.

The PROCESS macro makes this entire procedure especially easy, as it will choose values of X and W, generate estimates of Y for combinations of these values, and the SPSS version of PROCESS will even write SPSS syntax to produce a plot of the model. See the section of PROCESS output at the end of Appendix B. We used this code and then edited the resulting figure to produce the diagram of equation (11) found in Fig. 4, panel A. Although you probably see two lines with different slopes, what you should focus on instead is

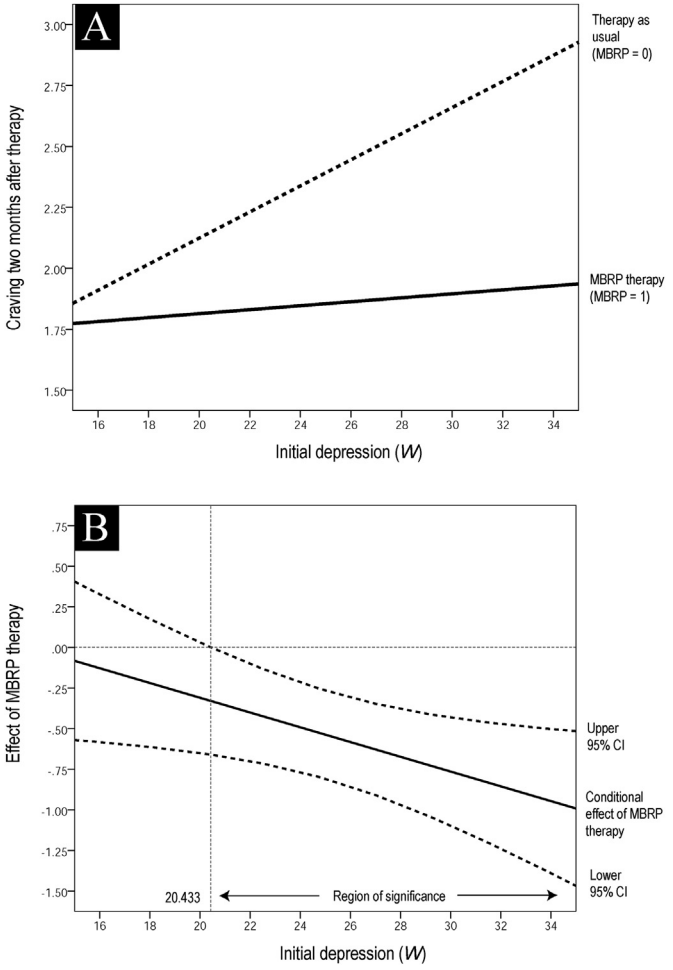

Fig. 4. A depiction of the interaction between therapy type and initial depression in the model of craving at two month follow up (panel A) and a Johnson-Neyman plot representing the same interaction (panel B).

[6] In PROCESS model 1 (the model number used to estimate a simple moderation model), the syntax requires that we label the moderator M rather than W, and so what we are calling W here is labeled M in the PROCESS code and output.

the distance between the two lines at values of initial depression, as we are interested in how the effect of therapy on craving varies with depression. As can be seen, it appears that the effect of MBRP therapy on later craving is larger among those higher in initial depression, represented by the growing gap between the two lines with increasing pre-therapy depression.

2.2. Some myths and misunderstandings

A variety of myths are circulating in the literature and practice about how to properly test a moderation hypothesis using linear regression analysis in this way, as well as widespread misunderstanding about interpretation of the regression coefficients in such a model. One of the most pervasive myths is that *X* and *W* should be standardized or mean centered prior to constructing the product so as to "reduce the negative effects of multicollinearity" (we are paraphrasing here; the expression and propagation of this myth takes many verbal forms in journal articles. For examples in *BRaT*, see e.g., Conklin, Cassiello-Robbins, & Brake et al., 2015; Dennis-Tiwary, Egan, Babkirk, & Denefrio, 2016; Lebowitz, Shic, & Campbell at al. 2015). This myth has been repeatedly debunked though nevertheless doggedly persists (debunkers include Friedrich, 1982; Hayes, 2013; Irwin & McClelland, 2001; Kromrey & Foster-Johnson, 1998). In a model of the form in equation (10), *XW* will often be highly correlated with *X*, *W*, or both. Although mean centering *X* and *W* will reduce the correlation between *XW* and its components *X* and *W* (this is not a myth) it turns out that this simply has no effect on the test of interaction, for reasons described by Hayes (2013, pp. 282–288) and others. Centering or standardization of *X* and *W* is not required to test a moderation hypothesis, though doing so can have certain benefits unrelated to what centering does to the correlation between variables in the model. The coefficient for *XW*, its standard error, *p*-value, and confidence interval will be the same regardless of whether one mean centers *X* and *W* (without standardization) prior to constructing the product. Although standardization will change the regression coefficient and its standard error and confidence interval, the ratio of the regression coefficient to its standard error will not be changed, and so the hypothesis test for the interaction is unaffected. Standardization merely changes the metric of measurement such that "1-unit" on the measurement scales corresponds to 1 standard deviation.

A second myth is that a proper test of moderation requires hierarchical variable entry, with *X* and *W* entered into a regression model first, followed by *XW*. A statistically significant change in R^2 when *XW* is added to the model is then used as evidence of moderation of *X*'s effect on *Y* by *W*. Many investigators reporting a moderation analysis describe using a hierarchical entry procedure such as this (e.g., Bailey & Wells, 2016; Lebowitz et al., 2015). But the *p*-value for the change in R^2 when *XW* is added to the model is the same as the *p*-value for the regression coefficient for *XW* when testing the null hypothesis that the regression coefficient for the product equals zero. These are mathematically identical tests and so will produce identical results for the test of moderation. Indeed, notice in the PROCESS output in Appendix B that an *F*-test for the change in R^2 when the product is added to the model is provided, but its *p*-value is the same as the *p*-value of the regression coefficient for the product. Hierarchical entry of the product is not required to test moderation using regression analysis unless the interaction requires more than one regression coefficient to estimate it, such as when *X* or *W* is a multicategorical variable (see Hayes & Montoya, 2016). Hierarchical entry does produce the change in R^2, and this can be a useful statistic to report (it is mathematically equivalent to an effect size measure known in the analysis of variance literature as *eta-squared*). However, the change in R^2 when *XW* is added to the model is equivalent to the squared semipartial correlation for *XW*. Most regression programs have options for outputting the semipartial correlation for each variable in a regression model, so hierarchical entry isn't required even to get a measure of the change in model fit.

A third myth of moderation analysis is that a proposed moderator *W* of *X*'s effect on *Y* must be uncorrelated with *X*, and thus a moderation analysis is not appropriate if *X* and *W* are not statistically independent. Although we found no instances in our perusal of *BRaT* articles of authors stating they had forgone a moderation analysis because *X* and *W* were correlated, we have heard this position taken in questions we have received in our consulting work, and it is the position taken by the so-called "McArthur School" as discussed in the work of Kraemer and colleagues (e.g., Kraemer, Kiernan, Essex, & Kupfer, 2008; Kraemer et al., 2002) that has been influential among clinical researchers. In the world of experimentation (e.g., clinical trials), where participants are randomly assigned to levels of *X*, independence between *X* and *W* is all but assured (i.e., should be no more strongly correlated than would be expected by chance) unless *W* is affected by *X* (which suggests *W* could possibly be a mediator rather than a moderator of *X*'s effect). Independence between *X* and *W* is interpretationally convenient because when *X* is determined through random assignment and one finds a difference between groups defined by *X* on dependent variable *Y* that can't be attributed to various design flaws or procedural mishaps, it is clear that this difference is attributable to *X* and not to *W*. In other words, we can say that *X* is a cause of *Y* and that any variation in *X*'s effect on *Y* related to *W* is due to *W*'s role as a moderator rather than *W*'s role as the sole cause of *Y*.

But interpretational convenience does not translate into a mathematical requirement. Earlier we discussed that by representing *X*'s effect on *Y* in $\hat{Y} = b_0 + b_1X + b_2W$ as a linear function $b_1 + b_3W$, the result after substitution is equation (10), the classic linear moderation model. There is nothing in this algebra that requires *W* and *X* to be statistically independent for $b_1 + b_3W$ to represent the relationship between *W* and *X*'s effect on *Y*. Indeed, outside of the experimenter's laboratory or the design purity of a clinical trial, we think it is probably quite common for a moderator of *X*'s effect to be correlated with *X*. For instance, more exposure to mass media exemplars of the "thin-as-ideal" standard for women could enhance the likelihood of disordered eating or self-esteem problems in women, more so among women whose peer group buys into this social portrayal of how women should look. But it seems likely that the correlation between a woman's exposure through the media and her peer group's endorsement of the standard would be positive. That doesn't mean that the beliefs of the peer group couldn't be moderating the effect of media exposure on something like disordered eating, though it could be that peer group beliefs is functioning as a mediator instead of (or in addition to; see Hayes, 2013, pp. 399–402) moderator if peer group beliefs are caused by exposure to mass media exemplars. Furthermore, correlation between exposure and peer groups beliefs means it becomes harder to disentangle the potential causal effects of exposure through media channels versus influence from the peer group. But again, interpretational convenience does not imply mathematical or substantive necessity.

In a moderation model such as equation (10), b_1 and b_2 are easily misinterpreted. Contrary to the beliefs of some, the regression coefficients for *X* and *W* in equations (9) and (10) are not "main effects." A main effect is a concept from analysis of variance that does not generalize to most regression models that include a product of predictors. In analysis of variance, a main effect is an *unweighted average simple effect*. It quantifies the effect of one categorical variable averaged across all levels of a second categorical variable. But in a regression model with a product, b_1 and b_2 are *conditional effects*. In this model, b_1 estimates the effect of *X* on *Y* when $W = 0$, and b_2

estimates the effect of W on Y when $X = 0$ (see Hayes, 2013, pp. 228–230). As such, b_1 and b_2 are closer to *simple effects* in the lingo of analysis of variance. If zero is not a meaningful point on the measurement scale or coding system for W, then b_1 and its test of significance and confidence interval is not meaningful. Likewise, if $X = 0$ is not meaningful, then b_2 and its test and confidence interval are not meaningful. Thus, be cautious when interpreting b_1 and b_2. They may not have a substantively meaningful interpretation. It is possible to set up a regression model so that b_1 and b_2 do quantify "main effects" when both X and W are dichotomous. This requires special coding procedures. See Hayes (2013, pp. 271–279) for a discussion.

The conditioning of b_1 and b_2 just described is one reason one might choose to mean center or standardize X and/or W prior to estimating a model with a product of XW. When W is mean centered or standardized, then b_1 estimates the effect of X among those average on W. And when X is mean centered, then b_2 estimates the effect of W among those average on X. These usually are meaningful, but they still aren't quite main effects as the term is used in analysis of variance. Importantly, b_1 and b_2 and their standard errors, p-values, and confidence intervals will typically be different when X and W are centered or standardized compared to when they are not. But this has nothing whatsoever to do with the reduction in the correlation between the product XW and X and W that occurs when centering or standardizing. The change is not due not to the solving of some mythical problem produced by collinearity. Rather, the change is the result of the conditioning of the estimates of X and W's effects (zero versus the mean of the other variable).

2.3. Probing an interaction

A visual depiction of an interaction, as in Fig. 4, panel A, conveys how X's effect on Y varies with W. This visual representation of the model is based on the sample regression coefficients. Given that each of the regression coefficients is estimated with sampling error, different samples would generate different visual depictions of the moderation. So it is insufficient to eyeball a visual representation of the model to determine where X's effect on Y is large or small, or to make claims where X has an effect on Y and where it does not. It is common to go a step further and formally "probe" the interaction, dissecting the model with the goal of determining not only how large X's effect on Y is at certain values of the moderator, but also to formally test a hypothesis about the size of the effect of X on Y at those values.

The most popular approach to probing an interaction goes by various names, including a *simple slopes analysis* (Cohen, Cohen, West, & Aiken, 2003), the *pick-a-point approach* (Bauer & Curran, 2005), or a *spotlight analysis* (Spiller, Fitzsimons, Lynch, & McClelland, 2013). This procedure involves picking two or more values of the moderator, estimating the conditional effect of X on Y at those values, and then testing whether those conditional effects are different from zero.

Earlier we saw that by including XW in the model of Y along with X and W, X's effect on Y is a linear function of W defined by the function $b_1 + b_3W$. Using this function, the effect of X on Y can be quantified for any value of W you choose by plugging that value into the function and doing the math. To complete the inference (i.e., is the conditional effect of X on Y at that value of W statistically different from zero?) a standard error is needed. The standard error falls out of standard covariance algebra and is estimated as

$$\sqrt{se^2_{b_1} + 2W(COV_{b_1b_3}) + W^2se^2_{b_3}},$$

where $se^2_{b_1}$ and $se^2_{b_3}$ are estimates of the squared standard errors of b_1 and b_3 and $COV_{b_1b_3}$ is the covariance between b_1 and b_3 (this covariance is not printed by most statistics programs by default, but most will print this if requested to do so). The ratio of $b_1 + b_3W$ to its standard error is distributed as $t(df_{residual})$ under the null hypothesis that the effect of X on Y at that value of W is equal to zero, where $df_{residual}$ is the residual degrees of freedom for the model. Alternatively, a confidence interval for the conditional effect of X can be constructed as the point estimate plus or minus about two standard errors.

Regression books that discuss interactions provide examples of these computations (e.g., Cohen et al., 2003). But we don't recommend doing them by hand, as the potential for mistake is high. It is too easy to put the wrong number in the wrong location in the computations, and unless the computations are done to a large number of decimals of precision, rounding error can propagate through them. Using programmed spreadsheets on your personal computer or web forms available online doesn't solve these problems. Fortunately, the computations can be done automatically with your regression program using what Hayes (2013; Darlington & Hayes, 2017) calls the *regression centering* implementation. This involves centering W around the value chosen prior to computing the product and then regressing Y on X, centered W, and X multiplied by the centered W. In the resulting model, the coefficient for X is the conditional effect of X on Y when W equals the chosen value that W was centered around, and its standard error, t- and p-values, and confidence interval produced by the regression program will correspond to the computations just described, but without the potential for error that hand computation or entering numbers into spreadsheets has.

An even simpler approach is to use existing macros for SPSS and SAS that automatically conduct a simple slopes analysis, such as MODPROBE (Hayes & Matthes, 2009), RLM (Darlington & Hayes, 2017), or PROCESS (Hayes, 2013). The PROCESS output in Appendix B, in the section of output titled "Conditional effect of X at value(s) of the moderator(s)," contains the conditional effects of MBRP therapy on later craving when initial depression is set to a standard deviation below the mean ($W = 18.508$), the mean ($W = 25.214$) and a standard deviation above the mean ($W = 31.920$), along with standard errors, t- and p-values, and confidence intervals. The negative estimates reflect that those given MBRP therapy craved substances less on average two months later than those given therapy as usual. This difference is statistically significant among those "moderate" or "relatively high" in initial depression but not among those "relative low" in initial depression, as defined using the distribution of BDI scores in the sample.

Although simple slopes analysis is common in the pages of *BRaT* (indeed, it seems to be the dominant approach in the clinical psychology literature and elsewhere), it suffers from some drawbacks as often implemented. First, when W is continuous, typically there is no clear rationale for choosing specific values of the moderator, and so researchers fall back on conventions such as the mean, a standard deviation below the mean, and a standard deviation above the mean of the sample, such as implemented in PROCESS. Such values are commonly use to operationalize low, moderate, and high on the moderator (e.g., Conklin, Cassiello-Robbins, & Brake et al., 2015; Lebowitz et al., 2015). But these are entirely arbitrary. Other values could be used. Importantly, the choice made will influence the results of the probing exercise. You might find that X is significantly related to Y among people relatively low on a moderator as defined by, say, one standard deviation below the mean, but a different operational definition of relatively low, such as the 25th percentile, which is just as sensible, could produce a nonsignificant effect of X.

Second, the use of arbitrary conventions for probing an interaction that rely on characteristics of the sample invites inconsistencies in findings across investigators conducting otherwise identical studies. Consider two investigators examining the effectiveness of a new therapy on some kind of psychological condition,

with that effectiveness thought to be moderated by an individual difference or background characteristic such as depression. Suppose that the investigators have conducted the study in exactly the same way, using the same procedures and measurement systems, but investigator A's study includes a vast and representative range of people who vary widely in depression, whereas investigator B's sample includes people who tend to be fairly high in depression by clinical standards. In this scenario, what investigator A calls "relatively high" in depression (one standard deviation above his or her sample mean; e.g., a score of 20 on the measurement scale) might correspond to what investigator B calls "relatively low" (one standard deviation below his or her sample mean, also 20 on the measurement scale) because investigator B's sample contains participants who tend to be more depressed than investigator A's. If depression moderates the effect of the therapy, investigator A might find the treatment is particularly effective (or particularly ineffective) among people "relatively high" in depression, whereas investigator B finds a similar effect but among people "relatively low" in depression. Their findings might be identical otherwise, but hypothesis tests on the conditional effects result in a seemingly different pattern of findings because they are operationalizing relatively low and relatively high differently. This argument applies even if the two investigators are using different measures of the same construct. Definitions of low, moderate, and high based on information about the sample may produce results that don't generalize across samples.

One solution is to use conventions that are not defined based on the distribution of the moderator in the sample but, instead, using population norms. Someone who is at the 25th percentile of population norms on a measure such as the Beck Depression Inventory could be construed as "relatively low" regardless of the distribution of BDI scores in any particular sample. If a therapy works well (or not well at all) among people relatively low in depression as defined by a normative absolute rather than a relative standard, we would expect greater consistency in findings across investigators examining that therapeutic method and using the same moderator variable, regardless of the distribution of the moderator in their respective studies. But the problem with this suggestion is that population norms may not be available for the moderator you are using, or your study may not include any participants within that range of the moderator. Indeed, such norms are probably not available for many measures that researchers use as moderators, especially if those moderators are constructed ad hoc for the specific study. But when sensible norms are available, PROCESS allows the user to choose any value or values of moderator for implementing this analysis, making it easy to implement alternative choices for the moderator. See the documentation.

An alternative approach is to use the Johnson-Neyman (JN) technique (Bauer & Curran, 2005; Hayes & Matthes, 2009; Johnson & Fey, 1950; Johnson & Neyman, 1936). Also called a *floodlight analysis* (Spiller et al., 2013), the JN technique does not require an investigator to pick values of *W* prior to estimating the conditional effect of *X* on *Y*. Instead, it analytically derives the values of *W*, if they exist, that identify points of transition along the continuum of the moderator between a statistically significant and nonsignificant effect of *X*. These points of transition define *regions of significance* for the effect of *X*—the range or ranges of the moderator *W* where *X* is significantly related to *Y* and where it is not. Because the investigator does not choose the values of *W* prior to probing the interaction, the results of the probing exercise will not be dependent on the values of *W* the investigator chooses.

The JN technique is not nearly as widely used as simple slopes analysis, but there is reason to believe the dominance of simple slopes analysis is slowly waning. Historically, the JN method has been difficult to implement for the typical researcher, but computational obstacles have been overcome in the last 10 years and the JN method is now implemented in regression-based moderation analysis tools one can find on the web (e.g., Preacher, Curran, & Bauer, 2006) as well as in macros for SPSS and SAS such as PROCESS, RLM (Darlington & Hayes, 2017), and MODPROBE (Hayes & Matthes, 2009). Perhaps as a consequence, its use appears to be accelerating in frequency (for some examples in *BRaT*, see Allen, Austin, Waldron, & Ollendick, 2015; Keng, Seah, Tong, & Smoski, 2016). Until recently, the JN method was available only when *X* is dichotomous or continuous. But Montoya (2016; Hayes & Montoya, 2016) developed and describe a macro for SPSS and SAS that implements the JN technique when *X* is multicategorical (three or more categories), further extending the utility of this method.

Toward the end of the PROCESS output in Appendix B can be found output from the JN technique applied to the MBRP therapy example. The output identifies the score of 20.443 on the BDI as a point of transition between a statistically significant and a nonsignificant effect of MBRP therapy on later craving. An examination of the table in the output shows that above this value (above which 78.6% of the sample resides), the effect of MBRP therapy is statistically significant and negative, but below this value it is not statistically significant. So the region of significance of the effect of MBRP therapy on later craving is initial BDI scores greater than about 20 or so. Above this value, MBRP therapy seems to result is less craving than does therapy as usual.

We have seen researchers using both simple slopes and the JN technique in the same analysis by choosing values of *W* but also providing regions of significance (e.g., Brandt, Bakhshaie, & Zvolensky et al., 2015; Poon, Turpyn, & Hansen et al., 2016). This is redundant in our opinion. The JN technique gives information about the statistical significance (or interval estimates) of the effect of *X* for *all* values of *W* one could choose. No new information is provided with *p*-values for the effect of *X* on specific and arbitrary values of *W*. Better would be to provide a graph that plots the relationship between *W* and the effect of *X* on *Y* along with a 95% confidence-band, as in Allen et al. (2015) and Fig. 4, panel B. In this figure, the slope of the line is b_3, the points on the line represent simple slopes or conditional effects of *X* (in this example, the gap between the lines in Fig. 4, panel A), and the dashed lines represent the confidence interval for the simple slope at a particular value of the moderator represented on the horizontal axis. Such a plot allows the reader who prefers a simple slopes analysis to look at the plot and determine whether the confidence intervals for *X*'s effect at the values of *W* he or she chooses contain zero. For details on construction of such a plot, see Hayes (2013, pp. 241–244) or Preacher et al. (2006).

2.4. Multicategorical focal predictor or moderator

When *X* or *W* is multicategorical and the other is continuous, testing a moderation hypothesis using regression analysis is a bit more complex, but the PROCESS macro has features that simplifies the analysis. We discourage researchers with multicategorical variables from throwing out one or more groups so that the focal predictor or moderator is dichotomous, or collapsing groups into one to make the analytical problem simpler. It is not that much more difficult compared to when *X* and *W* are continuous or dichotomous, and an easy-to-read tutorial on the topic is available (Hayes & Montoya, 2016) that discusses the estimation, visualizing, and probing of interactions in linear regression analysis when the focal predictor or moderator is multicategorical, along with instructions using the PROCESS macro. The topic is also covered in books on regression analysis that cover models with interactions (e.g., Cohen et al., 2003; Darlington & Hayes, 2017).

2.5. On interaction between independent variable (X) and mediator (M) in mediation analysis

Having discussed interaction, we now briefly return to the topic of mediation analysis before integrating moderation and mediation in the final section of this paper. It is possible that independent variable X and mediator M interact in the model of Y, meaning that the effect of M on Y varies with X. If M's effect on Y depends on X, then b in equation (3) (and b in equation (6) from our example) is an inappropriate estimator of M's effect. Furthermore, c' in equation (3) (and equation (6)) also misestimates the direct effect of X, as X's direct effect on Y would depend on M (by the symmetry property of interactions; see Darlington & Hayes, 2017). Kraemer et al. (2002, 2008) recommend that researchers interested in mediation replace equation (3) with one that includes the product of X and M in the model of Y

$$\hat{Y} = i_Y + c'X + b_1M + b_2XM \tag{12}$$

They also modify the Baron and Kenny criteria by requiring the centering of X prior to model estimation and replacing the requirement that the effect of M must be significant in the original equation (3) with the requirement that either M's effect on Y (b_1) *or* the interaction between X and M (b_2) must be significant in equation (12). More recently, the "causal inference," "counterfactual," or "potential outcomes" tradition in mediation analysis that has been gaining some traction in the literature (though less in practice) allows X and M to interact (Imai, Keele, & Tingley, 2010; Muthén & Asparouhov, 2015; Valeri & VanderWeele, 2013). In this framework, new terminology is used to distinguish between *natural* or *pure* and *controlled* direct and indirect effects, which estimate different things. Preacher, Rucker, and Hayes (2007; also see Hayes, 2013, 2015) also discuss models that allow the X by M interaction, which makes the indirect effect of X a function of X and the direct effect of X a function of M.

Although we think that it is worth testing for interaction between X and M when conducting a mediation analysis, we don't feel that inclusion of the XM product in the model of Y is necessary unless there is clear evidence that this interaction exists and that a failure to ignore it would result in a misleading set of claims. We see the Kraemer et al. modification of the Baron and Kenny criteria to be a sideways step rather than a step forward, as even with this modification, this approach still violates the common-sense principles of inference we discussed earlier. If X and M do interact, we recommend consulting the counterfactual mediation analysis literature if X is dichotomous, or using the approach described in Preacher et al. (2007) and Hayes (2015) when X is continuous. As the latter literature shows, X by M interaction means that X becomes a moderator of its own indirect effect through M, and that makes such a mediation model a *conditional process model* (discussed in the following section).

In our opinion, there is no reason to give the possible moderation of the effect of M on Y by X in a mediation model special empirical status relative to the possibility that *any* of the paths in a mediation model could be moderated. Indeed, it is safe to assume than when estimating any effect, whether using analysis of variance, regression, or any other method, that the effect is probably, in fact, moderated by *something* whether or not your model explicitly allows it. Indeed, we routinely assume the absence of interactions when we do data analysis by not including all possible two, three, and higher-order interactions in a model (c.f., Darlington & Hayes, 2017, pp. 442–443). But if your theory or hypothesis explicitly predicts interaction involving one or more of the paths in a mediation process, methods exist that allow for the estimation of indirect effects that then become conditional on a moderator or moderators. We address this topic next.

3. Integrating moderation and mediation: Conditional process analysis

We have seen that effects can operate *indirectly* through mediators, and that the size of effects can be dependent on other variables. As an indirect effect (mediation) is an effect, and effects can be contingent (moderation), it follows that an indirect effect can be contingent. In other words, the size of an indirect effect can be *dependent* on another variable—*moderated mediation*. For instance, a therapeutic method might indirectly influence later symptoms by changing people's cognitive appraisals, which in turn influences symptoms experienced, but this mechanism might be stronger in people who have more social support.

Methodologists have pondered the idea that mediation can be moderated since the early days of mediation analyses, though it wasn't until after the turn of the century that methodologists started writing in a systematic way about procedures for estimating and interpreting models that allow indirect effects to be dependent on other variables (e.g., Edwards & Lambert, 2007; Fairchild & MacKinnon, 2009; Muller, Judd, & Yzerbyt, 2005; Preacher et al., 2007). Hayes (2013; Hayes & Preacher, 2013) introduced the term *conditional process analysis* to refer to the collection of analytical strategies that focuses on examining the contingencies of mechanisms and test hypotheses about how processes can vary between people or across contexts.

Mediation and moderation can be analytically integrated as a conditional process model in a multitude of ways, depending on which stage of a mediation process is moderated, the number of mediators, the number of moderators, and whether or not the direct effect is also moderated. Fig. 3 panels B, C, and D depict a few examples. Panel B is a first stage conditional process model that allows the indirect effect of X on Y through M to be dependent on W, with this moderation operating in the first stage of the mediation process (i.e., the effect of X on M). Panel C depicts a second stage conditional process model, with the mediation of X's effect on Y through M being moderated through the moderation of the effect of M on Y. Panel D illustrates a more complex model with moderation in the first and second stage of the mediation process, with each stage moderated by a different variable, while also allowing the direct effect to be moderated by the variable moderating the first stage of the indirect effect. These are only a few of the many possibilities.

It is not difficult to find examples in the empirical literature of clinical researchers testing models that allow an indirect effect to be moderated (e.g., Gaume et al., 2016; Torres & Taknint, 2015), though such examples are largely nonexistent in the pages of *BRaT*, at least in the last five years or so. We don't believe this reflects that no one who publishes in this journal has pondered the contingencies of mechanisms. Rather, perhaps it reflects that existing methods have not yet widely disseminated among this population of researchers. If that is the explanation, hopefully our illustration will help to grow awareness of what is analytically possible.

As in any mediation model, a conditional process model has direct and indirect effects of X. But a hallmark of a conditional process model is that the indirect effect of X on Y through one or more mediators M is no longer fixed to be a single number but, instead, becomes a function of one or more moderators. That function can be simple or complex, depending on the complexity of the model. As most typically practiced, and in most examples in the literature, the indirect effect is a *linear* function of a moderator. But the relationship could be a nonlinear function if a single moderator is allowed to determine both the first and second stage components of the mediation process (see Hayes, 2015, for a discussion), or if

one of the paths is a multiplicative function of two moderators (see Hayes, 2016). The direct effect of *X* can be fixed or a function of a moderator, depending on theory, hypothesis, or intuition. In the end, the analyst's goal is to determine whether the indirect effect is moderated and, if so, probe and substantively interpret the meaning of the moderation of the indirect effect in terms of the theory or hypothesis being tested.

Conceptualizing a conditional process model can be a complex procedure requiring many decisions informed by theory, knowledge of the substantive literature, and perhaps a bit of intuition. But once done, estimation of the model is not difficult with the right software, nor is interpretation. In the rest of this section, we illustrate a conditional process analysis, emphasizing the PROCESS macro as the computational engine. Our example will combine the two analyses conducted earlier that focused on craving as a mediator of the effect of MBRP therapy on later substance use, and depression at the start of therapy as a moderator of the effect of MBRP therapy on craving. When these are integrated, the result is a first stage conditional process model as in Fig. 3, panel B, where *X* is therapy received, *M* is craving at two months, *Y* is substance use at four months, *W* is depression at the start of therapy, and craving at the start of therapy (U_1) and treatment hours (U_2) are covariates.

This model requires two equations to estimate the direct and indirect effects, one for *M* and one for *Y*. These equations are

$$\widehat{M} = i_M + a_1X + a_2W + a_3XW + a_4U_1 + a_5U_2 \tag{13}$$

$$\widehat{Y} = i_Y + c'X + b_1M + b_2U_1 + b_3U_2 \tag{14}$$

and follow directly from the mechanics of mediation and moderation analysis discussed earlier. This model is represented in path diagram form in Fig. 5. When the regression coefficients of this model are estimated using the PROCESS macro, the resulting equations are

$$\widehat{M} = 1.031 + 0.599X + 0.054W - 0.045XW + 0.192U_1 - 0.018U_2 \tag{15}$$

$$\widehat{Y} = 1.130 + 0.093X + 0.481M - 0.088U_1 - 0.020U_2 \tag{16}$$

These numbers come from the PROCESS output in Appendix C. There are three major sections of output. The first is the model of craving (under the heading “Outcome: crave2”) and the second is the model of substance use (under the heading “Outcome: use4”). The information in these two sections can be obtained from any OLS regression program using its built in regression routine, but PROCESS does it all in a single command and packages it in one output. The third section contains information about the direct and indirect effects and a test of moderation of the indirect effect. This information cannot be obtained without the assistance of PROCESS unless the model is set up and properly programmed in a SEM program that can estimate functions of coefficients (such as Mplus; see Hayes & Preacher, 2013). For observed variable models (i.e., no latent variables), we recommend using PROCESS, as it is much simpler to use and will produce the same results.

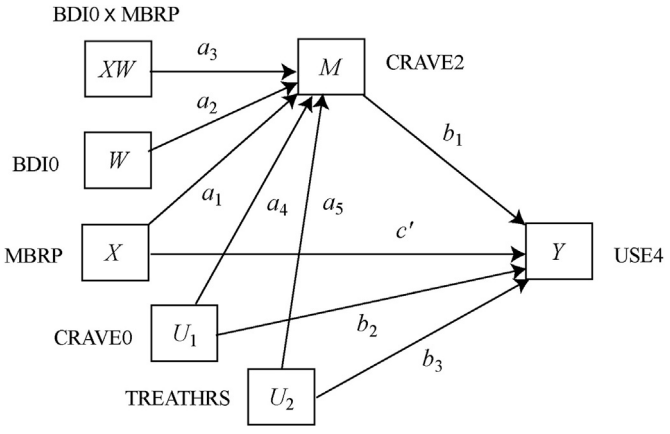

Fig. 5. A path diagram of a conditional process model with 2 covariates representing equations (13) and (14) and corresponding to the mindfulness-based relapse prevention therapy example.

The first question is whether there is evidence of moderation of the indirect effect. The indirect effect of MBRP therapy on use through craving is the conjunction of the effect of *X* on *M* and the effect of *M* on *Y*, just as in any mediation analysis. But in this model, the effect of MBRP therapy on craving is not a single number but, instead, a linear function of depression at the start of therapy because the model of craving is a linear moderation model. From equations (13) and (15) and the principles discussed earlier, the conditional effect of MBRP therapy on craving is $a_1 + a_3W = 0.599 - 0.045W$. The effect of craving on substance use is fixed to be a constant. It is estimated with the regression coefficient for craving in the model of substance use, b_1 in equations (14) and (16). Here, $b_1 = 0.481$. When these are multiplied together, the result is the indirect effect of *X* on *Y* through *M*:

$$\begin{aligned}(a_1 + a_3W)b_1 &= a_1b_1 + a_3b_1W \\ &= 0.599(0.481) + (-0.045)(0.481)W \\ &= 0.288 - 0.022W\end{aligned} \tag{17}$$

which is a linear function of *W*. So the indirect effect of MBRP therapy on substance use through craving depends linearly on depression at the start of therapy.

In equation (17), $a_3b_1 = -0.022$ quantifies the relationship between initial depression and the indirect effect of MBRP therapy. So with each one-unit increase in initial BDI score, the indirect effect of MBRP therapy declines by 0.022 units. Hayes (2015) calls a_3b_1 the *index of moderated mediation* for this model. To test moderation of the indirect effect, Hayes (2015) recommends a bootstrap confidence interval for the index of moderated mediation. This test is implemented by PROCESS and can be found in the output under the header “Index of Moderated Mediation.” As can be seen, the confidence interval for this index is −0.045 to −0.005. As this is entirely below zero, we conclude the indirect effect of MBRP therapy is negatively related to initial depression. The mediation is moderated. Notice that we have not concerned ourselves whether the first stage is moderated (i.e., in the sense that we haven't looked at or interpreted the *p*-value for a_3 in the model of *M*), nor whether b_1, the effect of *M* on *Y*, is statistically significant. What matters is the index of moderated mediation and whether it is different from zero. See Hayes (2015) for a discussion.

With evidence of moderation of the indirect effect, the next step is to probe it. Preacher et al. (2007) discuss pick-a-point and JN approaches. But the derivation of JN regions of significance for the indirect effect that they describe requires the assumption of normality of the sampling distribution of the indirect effect, which is known to be false. This leaves the pick-a-point approach as the only alternative. To implement this approach, one chooses values of the moderator, estimates the indirect effect of *X* on *Y* through *M* at those values, and then conducts an inference. Equation (17) provides the formula for the indirect effect in this model; it is a function of initial depression. One could choose values of the Beck Depression Inventory, plug it into the formula, and either estimate the standard error or construct a confidence interval for the indirect effect at that value, which in this context is called a *conditional*

indirect effect since it is conditioned on a value of a moderator. As not much can be done with the standard error for forming an inference without making assumptions about the shape of the distribution of the conditional indirect effect, we recommend a bootstrap confidence interval. If a confidence interval for the conditional indirect effect does not straddle zero, this can be used as statistical evidence that *M* mediates the effect of *X* on *Y* at that value of the moderator. But if it straddles zero, then the evidence does not warrant a definitive claim of mediation at that value of the moderator.

PROCESS implements this procedure automatically. By default, it conditions the indirect effect on values of the moderator corresponding to the mean, a standard deviation below the mean, and a standard deviation above the mean, but alternative values can be used (see the documentation). This information is in the output labeled "Conditional indirect effect(s) of X on Y at values of the moderator(s)." As can be seen, at relatively low values of initial depression ($W = 18.508$), the indirect effect of MBRP therapy on substance use through depression is not definitively different from zero (−0.117, with a 95% bootstrap confidence interval from −0.296 to 0.023). But among those moderate in initial depression ($W = 25.214$), the indirect effect is negative (−0.263), with a 95% bootstrap confidence interval of −0.467 to −0.124. Similarly, the indirect effect is negative (−0.410, 95% bootstrap confidence interval of −0.719 to −0.183) among those relatively high in initial depression ($W = 31.920$). We know that the effect of craving on substance use is positive, so the negative indirect effects observed reflect the fact that among those moderate to high in initial depression, the effect of MBRP therapy relative to therapy as usual is more effective in reducing craving, and this reduced craving seems to translates to less substance use.

A mediation model has a direct effect too. In this example of conditional process analysis, the direct effect has been fixed to be constant across values of initial depression and, indeed, any other potential moderator. As a result, there is only a single estimate of the direct effect. This is *c′* in equations (14) and (16), which in this example is 0.093 and not statistically different from zero (see the last section of PROCESS output in Appendix C). So independent of the mechanism through craving at two month follow up, and controlling for treatment hours and initial craving, there is no statistically significant effect of MBRP therapy on later substance use.

4. Summary

In this article, using nothing more complicated than the principles of ordinary least squares regression analysis, we have described the analysis of mediation and moderation effects and their integration as conditional process analysis. We also illustrated implementation using PROCESS, a freely available computational tool for SPSS and SAS that takes much of the computational burden off the researcher's shoulders. We discussed some new developments in thinking over the last several years, debunked some myths, while noting the value of such procedures as the Johnson-Neyman technique for probing interactions, comparing specific indirect effects in models with more than one mediator, and a formal test of the moderation of a mechanism using the index of moderated mediation.

We conclude by again emphasizing the limited role that data analysis can play in answering questions of cause and effect. Establishing mediation, or that an effect is contingent (i.e., moderated), is far more than clicking "OK" in a dialog box in your statistical software and looking at the resulting output. Statistics can be a *part* of the argument, but it isn't *the* argument. The burden of putting the complete argument together falls on you, the researcher, to tell a compelling story illuminated by insights your analysis provides, but you should be aware of the limitations of your data, the alternative interpretations that exist for your findings, and how a potential critic might perceive the language you use to describe them. But most important of all is that your causal story is justified by existing theory, current literature, and logical reasoning, especially when the data collection methods leave much ambiguity about causality, direction of causal flow, and so forth (as they usually do). This crucial part of the scientific story telling process isn't always given the attention it deserves. As we said at the outset, inferences are products of our mind and not our mathematics, and how you use your mind to make sense of your findings is more important than the mathematical tools you used in attempt to support the claims you ultimately make.

Appendix A

Output from the PROCESS macro for SPSS from a simple mediation model.

```
process vars=crave2 use4 mbrp treathrs crave0/y=use4/x=mbrp/m=crave2/total=1
 /boot=10000/seed=10000/model=4.

************** PROCESS Procedure for SPSS Release 2.16.1 *****************

     Written by Andrew F. Hayes, Ph.D.    www.afhayes.com
  Documentation available in Hayes (2013). www.guilford.com/p/hayes3

**************************************************************************
Model = 4
  Y = use4
  X = mbrp
  M = crave2

Statistical Controls:
CONTROL= treathrs crave0

Sample size
    168

**************************************************************************
Outcome: crave2

Model Summary
       R      R-sq       MSE         F       df1       df2         p
   .4340     .1884     .7928   12.6887    3.0000  164.0000     .0000

Model
             coeff         se         t         p       LLCI       ULCI
constant    2.2469     .3786    5.9342    .0000    1.4992    2.9945
mbrp        -.5735     .1432   -4.0038    .0001    -.8563    -.2906   <- path a
treathrs    -.0189     .0107   -1.7614    .0800    -.0400     .0023
crave0       .2596     .0744    3.4901    .0006     .1128     .4065

**************************************************************************
Outcome: use4

Model Summary
       R      R-sq       MSE         F       df1       df2         p
   .7304     .5335     .2105   46.6070    4.0000  163.0000     .0000

Model
             coeff         se         t         p       LLCI       ULCI
constant    1.1298     .2150    5.2545    .0000     .7052    1.5544
crave2       .4810     .0402   11.9547    .0000     .4015     .5604   <- path b
mbrp         .0926     .0773    1.1979    .2327    -.0601     .2453   <- path c'
treathrs    -.0199     .0056   -3.5720    .0005    -.0309    -.0089
crave0      -.0884     .0397   -2.2246    .0275    -.1668    -.0099

************************ TOTAL EFFECT MODEL ****************************
Outcome: use4

Model Summary
       R      R-sq       MSE         F       df1       df2         p
   .3529     .1245     .3926    7.7757    3.0000  164.0000     .0001

Model
             coeff         se         t         p       LLCI       ULCI
constant    2.2105     .2664    8.2962    .0000    1.6844    2.7366
mbrp        -.1832     .1008   -1.8177    .0709    -.3822     .0158   <- path c
treathrs    -.0290     .0075   -3.8439    .0002    -.0439    -.0141
crave0       .0365     .0524     .6972    .4867    -.0669     .1399

***************** TOTAL, DIRECT, AND INDIRECT EFFECTS ********************

Total effect of X on Y
   Effect         SE          t          p       LLCI       ULCI
   -.1832     .1008    -1.8177     .0709    -.3822     .0158   <- path c

Direct effect of X on Y
   Effect         SE          t          p       LLCI       ULCI
    .0926     .0773     1.1979     .2327    -.0601     .2453   <- path c'

Indirect effect of X on Y
          Effect   Boot SE   BootLLCI   BootULCI
crave2   -.2758     .0945    -.5017     -.1246   <- ab with 95% bootstrap confidence interval

****************** ANALYSIS NOTES AND WARNINGS ************************

Number of bootstrap samples for bias corrected bootstrap confidence intervals:
  10000

Level of confidence for all confidence intervals in output:
  95.00
```

Appendix B

Output from the PROCESS macro for SPSS from a simple linear moderation model.

```
process vars=crave2 mbrp treathrs crave0 bdi0/y=crave2/x=mbrp/m=bdi0/
  jn=1/plot=1/model=1.

Model = 1
  Y = crave2
  X = mbrp
  M = bdi0

Statistical Controls:
CONTROL= treathrs crave0

Sample size
    168

**************************************************************************
Outcome: crave2

Model Summary
       R      R-sq       MSE         F       df1       df2         p
   .5144     .2646     .7273   11.6574    5.0000  162.0000    .0000

Model
             coeff         se          t          p       LLCI       ULCI
constant    1.0311     .4705     2.1915     .0298     .1020     1.9603
bdi0         .0536     .0131     4.0760     .0001     .0276     .0795     <- b2
mbrp         .5986     .5253     1.1394     .2562    -.4388    1.6360     <- b1
int_1       -.0454     .0201    -2.2560     .0254    -.0852    -.0057     <- b3
treathrs    -.0177     .0103    -1.7220     .0870    -.0380     .0026
crave0       .1917     .0735     2.6100     .0099     .0467     .3368

Product terms key:

int_1      mbrp     X   bdi0        <- PROCESS constructs the product on its own and includes it in the model

R-square increase due to interaction(s):
        R2-chng         F       df1       df2         p        <- Test of change in R2 when the product is added to the model.
int_1     .0231    5.0894    1.0000  162.0000    .0254

**************************************************************************

Conditional effect of X on Y at values of the moderator(s):
     bdi0      Effect         se          t          p       LLCI       ULCI
  18.5082     -.2424     .1921    -1.2613     .2090    -.6218     .1371
  25.2143     -.5470     .1374    -3.9806     .0001    -.8184    -.2757
  31.9204     -.8517     .1932    -4.4081     .0000   -1.2333    -.4702

Values for quantitative moderators are the mean and plus/minus one SD from mean.
Values for dichotomous moderators are the two values of the moderator.

b1 + b3BDI0                  <- Conditional effects of X ("simple slopes") on Y for low, moderate, and high values of the moderator along with t- and p-values and confidence intervals

********************* JOHNSON-NEYMAN TECHNIQUE **************************

Moderator value(s) defining Johnson-Neyman significance region(s)
     Value      % below      % above
   20.4432      21.4286      78.5714   <- Region of significance from the JN method.

Conditional effect of X on Y at values of the moderator (M)
      bdi0      Effect         se          t          p       LLCI       ULCI
    .0000      .5986     .5253     1.1394     .2562    -.4388     1.6360
   2.2500      .4963     .4817     1.0303     .3044    -.4550     1.4476
   4.5000      .3941     .4385      .8987     .3701    -.4718     1.2600
   6.7500      .2919     .3957      .7376     .4618    -.4896     1.0733
   9.0000      .1896     .3536      .5364     .5924    -.5086      .8879
  11.2500      .0874     .3123      .2799     .7799    -.5293      .7042
  13.5000     -.0148     .2724     -.0544     .9567    -.5527      .5230
  15.7500     -.1170     .2344     -.4994     .6182    -.5798      .3458
  18.0000     -.2193     .1994    -1.0994     .2732    -.6131      .1746
  20.2500     -.3215     .1695    -1.8967     .0596    -.6562      .0132
  20.4432     -.3303     .1672    -1.9747     .0500    -.6605      .0000
  22.5000     -.4237     .1476    -2.8704     .0046    -.7152     -.1322
  24.7500     -.5259     .1377    -3.8198     .0002    -.7978     -.2540
  27.0000     -.6282     .1422    -4.4161     .0000    -.9091     -.3473
  29.2500     -.7304     .1600    -4.5636     .0000   -1.0465     -.4143
  31.5000     -.8326     .1874    -4.4439     .0000   -1.2026     -.4626
  33.7500     -.9349     .2207    -4.2362     .0000   -1.3706     -.4991
  36.0000    -1.0371     .2577    -4.0246     .0001   -1.5459     -.5282
  38.2500    -1.1393     .2970    -3.8360     .0002   -1.7258     -.5528
  40.5000    -1.2415     .3378    -3.6751     .0003   -1.9087     -.5744
  42.7500    -1.3438     .3797    -3.5393     .0005   -2.0935     -.5940
  45.0000    -1.4460     .4222    -3.4246     .0008   -2.2798     -.6122

**************************************************************************

Data for visualizing conditional effect of X on Y
Paste text below into a SPSS syntax window and execute to produce plot.

DATA LIST FREE/mbrp bdi0 crave2.
BEGIN DATA.
     .0000    18.5082     2.0440
    1.0000    18.5082     1.8017
     .0000    25.2143     2.4033
    1.0000    25.2143     1.8563
     .0000    31.9204     2.7626
    1.0000    31.9204     1.9109
END DATA.
GRAPH/SCATTERPLOT=bdi0 WITH crave2 BY mbrp.

<- PROCESS for SPSS writes syntax that will produce a scatterplot depicting the model. Cut and paste into a syntax window and execute.

* Estimates are based on setting covariates to their sample means.

****************** ANALYSIS NOTES AND WARNINGS ************************

Level of confidence for all confidence intervals in output:
  95.00
```

Appendix C

Output from the PROCESS macro for SPSS from a conditional process model.

```
process vars=crave2 use4 mbrp treathrs crave0 bdi0/y=use4/x=mbrp/m=crave2
  /w=bdi0/model=7/boot=10000.

Model = 7
  Y = use4
  X = mbrp
  M = crave2
  W = bdi0

Statistical Controls:
CONTROL= treathrs crave0

Sample size
    168

**************************************************************************
Outcome: crave2

Model Summary
       R      R-sq       MSE         F       df1       df2         p
   .5144     .2646     .7273   11.6574    5.0000  162.0000    .0000

Model
             coeff         se          t          p       LLCI       ULCI
constant    1.0311     .4705     2.1915     .0298     .1020     1.9603
mbrp         .5986     .5253     1.1394     .2562    -.4388    1.6360     <- a1
bdi0         .0536     .0131     4.0760     .0001     .0276     .0795
int_1       -.0454     .0201    -2.2560     .0254    -.0852    -.0057     <- a3
treathrs    -.0177     .0103    -1.7220     .0870    -.0380     .0026
crave0       .1917     .0735     2.6100     .0099     .0467     .3368

Product terms key:

 int_1   mbrp     X   bdi0

**************************************************************************
Outcome: use4

Model Summary
       R      R-sq       MSE         F       df1       df2         p
   .7304     .5335     .2105   46.6070    4.0000  163.0000    .0000

Model
             coeff         se          t          p       LLCI       ULCI
constant    1.1298     .2150     5.2545     .0000     .7052     1.5544
crave2       .4810     .0402    11.9547     .0000     .4015     .5604     <- b1
mbrp         .0926     .0773     1.1979     .2327    -.0601     .2453     <- c'
treathrs    -.0199     .0056    -3.5720     .0005    -.0309    -.0089
crave0      -.0884     .0397    -2.2246     .0275    -.1668    -.0099

******************** DIRECT AND INDIRECT EFFECTS *************************
Direct effect of X on Y
   Effect         SE          t          p       LLCI       ULCI
    .0926     .0773     1.1979     .2327    -.0601     .2453     <- c'

Conditional indirect effect(s) of X on Y at values of the moderator(s):
                Values of W
Mediator          bdi0      Effect    Boot SE   BootLLCI   BootULCI
crave2         18.5082     -.1166     .0777     -.2955      .0228
crave2         25.2143     -.2631     .0859     -.4682     -.1237
crave2         31.9204     -.4097     .1347     -.7193     -.1834

<- (a1+a3W)b1 and 95% bootstrap confidence intervals

Values for quantitative moderators are the mean and plus/minus one SD from mean.
Values for dichotomous moderators are the two values of the moderator.

****************** INDEX OF MODERATED MEDIATION ************************

Mediator
           Index    SE(Boot)   BootLLCI   BootULCI
crave2     -.0219     .0102     -.0448     -.0050     <- Index of moderated mediation (a3b1) with 95% bootstrap confidence interval

****************** ANALYSIS NOTES AND WARNINGS ************************

Number of bootstrap samples for bias corrected bootstrap confidence intervals:
  10000

Level of confidence for all confidence intervals in output:
  95.00
```

References

Allen, B., Austin, K. E., Waldron, J. C., & Ollendick, T. H. (2015). Synchrony and desynchrony in the tripartite model of fear: Predicting treatment outcome in clinically phobic children. *Behavior Research and Therapy, 71*, 54–64.

Arch, J. J. (2014). Cognitive behavioral therapy and pharmacotherapy for anxiety: Treatment preferences and credibility among pregnant and non-pregnant women. *Behaviour Research and Therapy, 52*, 53–60.

Armstrong, T., Zald, D. H., & Olatunji, B. O. (2011). Attentional control in OCD and GAD: Specificity and associations with core cognitive symptoms. *Behavior Research and Therapy, 49*, 756–762.

Bailey, R., & Wells, A. (2016). Is metacognition a causal moderator of the relationship between catastrophic misinterpretation and health anxiety? A prospective study. *Behavior Research and Therapy, 78*, 43–50.

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.

Bauer, D. J., & Curran, P. J. (2005). Probing interactions in fixed and multilevel regression: Inferential and graphical techniques. *Multivariate Behavioral*

Research, 40, 373–400.
Beadman, M., Das, R. K., Freeman, T. P., Scragg, P., West, R., & Kamboj, S. K. (2015). A comparison of emotion regulation strategies in response to craving cognitions: Effects on smoking behaviour, craving, and affect in dependent smokers. *Behaviour Research and Therapy, 69*, 29–32.
Boden, M. T., John, O. P., Goldin, P. R., Werner, K., Heimberg, R. G., & Gross, J. J. (2012). The role of maladaptive beliefs in cognitive-behavioral therapy: Evidence from social anxiety disorder. *Behavior Research and Therapy, 50*, 287–291.
Bollen, K. A. (1989). *Structural equation modeling with latent variables*. New York: John Wiley & Sons.
Brandt, C. P., Bakhshaie, J., Zvolensky, M. J., Grover, K. W., & Gonzales, A. (2015). The examination of emotion dysregulation as a moderator of depression and HIV-relevant outcome relations among an HIV 1 sample. *Cognitive Behavioural Therapy, 44*, 9–20.
Breitborde, N. J. K., Srihari, V. H., Pollard, J. M., Addington, D. N., & Woods, S. W. (2010). Mediators and moderatorsin early intervention research. *Early Interventions in Psychiatry, 4*, 143–152.
Campbell, D. T., & Kenny, D. A. (1999). *A primer on regression artifacts*. New York: The Guilford Press.
Cerin, E., & MacKinnon, D. P. (2009). A commentary on current practice in mediating variable analysis in behavioral nutrition and physical activity. *Public Health Nutrition, 12*, 1182–1188.
Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). New York: Routledge.
Conklin, L. R., Cassiello-Robbins, C., Brake, C. A., Sauer-Zavala, S., Farchione, T. J., Ciraulo, D. A., et al. (2015). Relationships among adaptive and maladaptive emotion regulation strategies and psychopathology during the treatment of comorbid anxiety and alcohol use disorders. *Behaviour Research and Therapy, 73*, 124–130.
Darlington, R. B., & Hayes, A. F. (2017). *Regression analysis and linear models: Concepts, applications, and implementation*. New York: The Guilford Press.
Deacon, B., Kemp, J. J., Dixon, L. J., Sy, J. T., Farrell, N. R., et al. (2013). Maximizing the efficacy of interoceptive exposure by optimizing inhibitory learning: A randomized controlled trial. *Behaviour Research and Therapy, 51*, 588–596.
Dennis-Tiwary, T. A., Egan, L. J., Babkirk, S., & Denefrio, S. (2016). For whom the bell tolls: Neurocognitive individual differences in the acute stress-reduction effects of an attention bias modification game for anxiety. *Behavior Research and Therapy, 78*, 43–50.
Edwards, J. R., & Lambert, L. S. (2007). Methods for integrating moderation and mediation: A general analytical framework using moderated path analysis. *Psychological Methods, 12*, 1–22.
Emsley, R., Dunn, G., & White, I. R. (2010). Mediation and moderation of treatment effects in randomised controlled trials of complex interventions. *Statistical Methods in Medical Research, 19*, 237–270.
Fairchild, A. J., & MacKinnon, D. P. (2009). A general model for testing mediation and moderation effects. *Prevention Science, 10*, 87–99.
Fillo, J., Alfano, C. A., Paulus, D. J., Smits, J. A. J., Davis, M. L., Rosenfield, D., et al. (2016). Emotion dysregulation explains relations between sleep disturbance and smoking quit-related cognition and behavior. *Addictive Behaviors, 57*, 6–12.
Friedrich, R. J. (1982). In defense of multiplicative terms in multiple regression equations. *American Journal of Political Science, 26*, 797–833.
Fritz, M. S., Taylor, A. B., & MacKinnon, D. P. (2012). Explanation of two anomalous results in statistical mediation analysis. *Multivariate Behavioral Research, 47*, 61–87.
Gaume, J., Longabaugh, R., Magill, M., Bertholet, N., Gmel, G., & Daeppen, J.-B. (2016). Under what conditions? Therapist and client characteristics moderate the role of change talk in brief motivational intervention. *Journal of Consulting and Clinical Psychology, 84*, 211–220.
Gaynor, K., Ward, T., Garety, P., & Peters, E. (2013). The role of safety-seeking behaviours in maintaining threat appraisals in psychosis. *Behaviour Research and Therapy, 51*, 75–81.
Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs, 76*, 408–420.
Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York: The Guilford Press.
Hayes, A. F. (2015). An index and test of linear moderated mediation. *Multivariate Behavioral Research, 50*, 1–22.
Hayes, A. F. (2016). *Partial, Conditional, and Moderated Mediation: Quantification, Inference, and Interpretation*. Manuscript submitted for publication.
Hayes, A. F., & Matthes, J. (2009). Computational procedures for probing interactions in OLS and logistic regression. *Behavior Research Methods, 41*, 924–936.
Hayes, A. F., & Montoya, A. K. (2016). A tutorial on testing, visualizing, and probing an interaction involving a multicategorical variable in linear regression analysis. *Communication Methods and Measures* (in press).
Hayes, A. F., & Preacher, K. J. (2013). Conditional process modeling: Using structural equation modeling to examine contingent causal processes. In G. R. Hancock, & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed., pp. 217–264). Greenwich, CT: Information Age.
Hayes, A. F., & Preacher, K. J. (2014). Statistical mediation analysis with a multicategorical independent variable. *British Journal of Mathematical and Statistical Psychology, 67*, 451–460.
Hayes, A. F., & Scharkow, M. (2013). The relative trustworthiness of inferential tests of the indirect effect in statistical mediation analysis: Does method really matter? *Psychological Science, 24*, 1918–1927.
Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods, 15*, 309–334.
Irwin, J. R., & McClelland, G. H. (2001). Misleading heuristics and moderated multiple regression models. *Journal of Marketing Research, 38*, 100–109.
Johnson, P. O., & Fey, L. C. (1950). The Johnson-Neyman technique, its theory, and application. *Psychometrika, 15*, 349–367.
Johnson, P. O., & Neyman, J. (1936). Tests of certain linear hypotheses and their application to some educational problems. *Statistical Research Memoirs, 1*, 57–93.
Kelley, K., & Preacher, K. J. (2012). On effect size. *Psychological Methods, 17*, 137–152.
Keng, S.-L., Seah, S. T. H., Tong, E. M. W., & Smoski, M. (2016). Effects of a brief mindful acceptance induction on implicit dysfunctional attitudes and concordance between implicit and explicit dysfunctional attitudes. *Behaviour Research and Therapy, 83*, 1–10.
Kenny, D. A., & Judd, C. M. (2014). Power anomalies in testing mediation. *Psychological Science, 25*, 334–339.
Kraemer, H. C. (2016). Messages for clinicians: Moderators and mediators of treatment outcome in randomized clinical trials. *American Journal of Psychiatry, 137*, 672–679.
Kraemer, H. C., Kiernan, M., Essex, M., & Kupfer, D. J. (2008). How and why criteria defining moderators and mediators differ between the Baron & Kenny and MacArthur approaches. *Health Psychology, 27*, S101–S108.
Kraemer, H. C., Wilson, G. T., Fairburn, C. G., & Agras, W. S. (2002). Mediators and moderators of treatment effects in randomized clinical trials. *Archives of General Psychiatry, 59*, 877–883.
Kromrey, J. D., & Foster-Johnson, L. (1998). Mean centering in moderated multiple regression: Much ado about nothing. *Educational and Psychological Measurement, 58*, 42–68.
Kuckertz, J. M., Amir, N., Boffa, J. W., Warren, C. K., Rindt, S. E. M., Norman, S., et al. (2014). The effectiveness of an attention bias modification program as an adjunctive treatment for post-traumatic stress disorder. *Behavioural Research and Therapy, 63*, 25–35.
Lebowitz, E. R., Shic, F., Campbell, D., Basile, K., & Silverman, W. K. (2015). Anxiety sensitivity moderates behavioral avoidance in anxious youth. *Behaviour Research and Therapy, 74*, 11–17.
MacKinnon, D. P. (2000). Contrasts in multiple mediator models. In J. S. Rose, L. Chassin, C. C. Presson, & S. J. Sherman (Eds.), *Multivariate applications in substance use research* (pp. 141–160). Mahwah, NJ: Lawrence Erlbaum Associates.
Magill, M. (2011). Moderators and mediators in social work research: Toward a more ecologically valid evidence base for practitioners. *Journal of Social Work, 11*, 387–401.
Manuel, A., & Wade, T. D. (2013). Emotion regulation in broadly defined anorexia nervosa: Association with negative affective memory bias. *Behaviour Research and Therapy, 51*, 417–425.
McLean, C. P., Yeh, R., Rosen, D., & Foa, E. B. (2015). Changes in negative cognitions mediate PTSD symptom reductions during client-centered therapy and prolonged exposure for adolescents. *Behaviour Research and Therapy, 68*, 64–69.
Montoya, A. K. (2016). *Extending the Johnson-Neyman technique to categorical independent variables: Mathematical derivations and computational tools* (Master's thesis). The Ohio State University.
Morgan, A. J., MacKinnon, A. J., & Jorm, A. F. (2013). Behavior change through automated e-mails: Mediation analysis of self-help strategy use for depressive symptoms. *Behaviour Research and Therapy, 51*, 57–62.
Muller, D., Judd, C. M., & Yzerbyt, V. Y. (2005). When mediation and moderated and moderation is mediated. *Journal of Personality and Social Psychology, 89*, 852–863.
Muthén, B., & Asparouhov, T. (2015). Causal effects in mediation modeling: An introduction with applications to latent variables. *Structural Equation Modeling, 22*, 12–23.
Newby, J. M., Williams, A. D., & Andrews, G. (2014). Reductions in negative repetitive thinking and metacognitive beliefs during transdiagnostic internet cognitive behavioral therapy (iCBT) for mixed anxiety and depression. *Behaviour Research and Therapy, 59*, 53–60.
Pictet, A., Jermann, F., & Ceschi, G. (2016). When less could be more: Investigating the effects of a brief internet-based imagery cognitive bias modification intervention in depression. *Behaviour Research and Therapy, 84*, 45–51.
Pitts, B. L., & Safer, M. A. (2016). Retrospective appraisals mediate the effects of combat experiences on PTS and depression symptoms in U.S. Army medics. *Journal of Traumatic Stress, 29*, 65–71.
Poon, J., Turpyn, C. C., Hansen, A., Jacangelo, J., & Chaplin, T. (2016). Adolescent substance use & Psychopathology: Interactive effects of cortisol reactivity and emotion regulation. *Cognitive Therapy and Research, 40*, 368–380.
Preacher, K. J. (2015). Advances in mediation analysis: A survey and synthesis of new developments. *Annual Review of Psychology, 66*, 825–852.
Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics, 31*, 437–448.
Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, and Computers, 36*, 717–731.
Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*, 879–891.
Preacher, K. J., & Kelley, K. (2011). Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. *Psychological*

Methods, 16, 93–115.
Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Assessing moderated mediation hypotheses: Theory, method, and prescription. *Multivariate Behavioral Research, 42*, 185–227.
Preacher, K. J., & Selig, J. P. (2012). Advantages of Monte Carlo confidence intervals for indirect effects. *Communication Methods and Measures, 6*, 77–98.
Rucker, D. D., Preacher, K. J., Tormala, Z. L., & Petty, R. E. (2011). Mediation analysis in social psychology: Current practices and new recommendations. *Personality and Social Psychology Compass, 5/6*, 359–371.
Sandin, B., Sanchez-Arribas, C., Chorot, P., & Valiente, R. M. (2015). Anxiety sensitivity, catastrophic misinterpretations and panic self-efficacy in the prediction of panic disorder severity: Towards a tripartite cognitive model of panic disorder. *Behaviour Research and Therapy, 67*, 30–40.
Seehuus, M., Clifton, J., & Rellini, A. H. (2015). The role of family environment and multiple forms of childhood abuse in the shaping of sexual function and satisfaction in women. *Archives of Sexual Behavior, 44*, 1595–1608.
Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods, 7*, 422–445.
Smith, P., Scott, R., Eshkevari, E., Jatta, F., Leigh, E., Harris, V., et al. (2015). Computerized CBT for depressed adolescents: Randomized controlled trial. *Behaviour Research and Therapy, 73*, 104–110.
Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology, 13*, 290–312.
Spiller, S. A., Fitzsimons, G. J., Lynch, J. G., & McClelland, G. H. (2013). Spotlights, floodlights, and the magic number zero: Simple effects tests in moderated regression. *Journal of Marketing Research, 50*, 277–288.
Spinhoven, P., Penninx, B. W., Krempeniou, A., van Hemert, A. M., & Elzinga, B. (2015). Trait rumination predicts onset of post-traumatic stress disorder through trauma-related cognitive appraisals: A 4-year longitudinal study. *Behavior Research and Therapy, 74*, 60–71.
Staring, A. B. P., van den Berg, D. P. G., Cath, D. C., Schoorl, M., Engelhard, I. M., & Korrelboom, C. W. (2016). Self-esteem treatment in anxiety: A randomized controlled crossover trial of eye movement desensitization and reprocessing (EMDR) versus competitive memory training (COMET) in patients with anxiety disorders. *Behaviour Research and Therapy, 82*, 11–20.
Tofighi, D., & MacKinnon, D. P. (2009). Rmediation: An R package for mediation analysis confidence intervals. *Behavior Research Methods, 43*, 692–700.
Torres, L., & Taknint, J. T. (2015). Ethnic microaggressions, traumatic stress symptoms, and Latino depression: A moderated mediational model. *Journal of Counseling, 62*, 393–401.
Trompetter, H. R., Bohlmeijer, E. T., Fox, J.-P., & Schreurs, K. M. G. (2015). Psychological flexibility and catastrophizing as associated change mechanisms during online acceptance & commitment therapy for chronic pain. *Behaviour Research and Therapy, 74*, 50–59.
Valeri, L., & VanderWeele, T. J. (2013). Mediation analysis allowing for exposure–mediator interactions and causal interpretation: Theoretical assumptions and implementation with SAS and SPSS macros. *Psychological Methods, 18*, 137–150.
Vincent, N., & Walsh, K. (2013). Hyperarousal, sleep scheduling, and time awake in bed as mediators of outcome in computerized cognitive-behavioral therapy (cCBT) for insomnia. *Behavioural Research and Therapy, 51*, 161–166.
Wan, Z., & Fan, X. (2015). Monotonicity of effect sizes: Questioning kappa-squared as mediation effect size measure. *Psychological Methods, 20*, 193–203.
Williams, J., & MacKinnon, D. P. (2008). Resampling and distribution of the product methods for testing indirect effects in complex models. *Structural Equation Modeling, 15*, 23–51.
Witkiewitz, K., & Bowen, S. (2010). Depression, craving, and substance use following a randomized trial of mindfulness-based relapse prevention. *Journal of Consulting and Clinical Psychology, 78*, 362–374.
Yuan, Y., & MacKinnon, D. P. (2009). Bayesian mediation analysis. *Psychological Methods, 14*, 301–322.
Zhao, X., Lynch, J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research, 37*, 197–206.