How and Why Criteria Defining Moderators and Mediators Differ Between the Baron & Kenny and MacArthur Approaches

Helena Chmura Kraemer,
Department of Psychiatry and Behavioral Sciences, Stanford University, CA

Michaela Kiernan,
Department of Medicine, Stanford University, CA

Marilyn Essex, and
Department of Psychiatry, University of Wisconsin, Madison, WI

David J. Kupfer
Department of Psychiatry, University of Pittsburgh, PA

Abstract

Objective—In recognition of the increasingly important role of moderators and mediators in clinical research, clear definitions are sought of the two terms to avoid inconsistent, ambiguous, and possibly misleading results across clinical research studies.

Design—The criteria used to define moderators and mediators proposed by the Baron & Kenny approach, which have been long used in social/behavioral research, are directly compared to the criteria proposed by the recent MacArthur approach, which modified the Baron & Kenny criteria.

Results—After clarifying the differences in criteria between approaches, the rationale for the modifications is clarified and the implications for the design and interpretation of future studies considered.

Conclusions—Researchers may find modifications introduced in the MacArthur approach more appropriate to their research objectives, particularly if their research might have a direct influence on decision making.

Keywords

moderators; mediators; risk factors; randomized clinical trials; effect size

---

In recent years, the importance of moderators and mediators has become increasingly apparent both in testing the efficacy of clinical interventions in randomized clinical trials (RCTs) (Kraemer, Frank, & Kupfer, 2006; Kraemer & Kupfer, 2006; Kraemer, Wilson, Fairburn, & Agras, 2002; Thompson & Higgins, 2005) and in observational research examining risk factors for disease (Kraemer, Stice, Kazdin, & Kupfer, 2001). In RCTs, because an effective intervention is not equally effective for all in any population, establishing *moderators* of intervention response—who responds and who does not— can prompt researchers to optimally target the intervention, and to seek better interventions for nonresponders. Establishing *mediators* of intervention response— how an intervention works— can prompt researchers either to strengthen, add, or remove certain intervention

---

components to make the intervention either more efficacious or more cost-effective (Kraemer et al., 2002). In risk research studies, it is evident that diseases of current scientific interest such as heart disease, cancer, depression, schizophrenia, and autism are not caused by a single risk factor such as a gene, organism, or environmental exposure, but by complex chains of risk factors. Identifying such chains, which may differ in different subgroups, requires more than compiling a list of risk factors for a particular disease; it requires determining how chains of risk factors might "work together" (Kraemer et al., 2001). Establishing *moderators* of risk factors—who is susceptible to exposure and who is not—improves our understanding of disease etiology and early diagnosis. Establishing *mediators* of risk factors can lead to the design of interventions to break the chains, i.e., for prevention or early detection of diseases.

Given the increasingly important role of moderators and mediators in clinical research, clear definitions of the two terms are essential to avoid inconsistent, ambiguous, and possibly misleading results. In social/behavioral research, the criteria for those definitions have long been those originally introduced by Baron & Kenny in 1986 (Baron & Kenny, 1986). Although there has been a proliferation of methods to implement these definitions (Collins, Graham, & Flaherty, 1998; James & Brett, 1984; Kenny, Kashy, & Bolger, 1998; MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002; Shrout & Bolger, 2002), the definitions themselves have hardly changed. Recently, modifications to those definitions have been introduced by a MacArthur Foundation Network subgroup (Kraemer et al., 2001, 2002), concerned with its clinical applications. In this paper, after clarifying the differences in criteria between approaches, the rationale for the modifications is explained and the implications for the design and interpretation of future studies are considered. Finally, changes made to the MacArthur approach since its introduction 5 years ago and future directions are discussed.

Terminology

Until about 20 years ago, the terms moderators and mediators were used more or less colloquially, not as scientific terms. However, in their landmark 1986 paper, Baron & Kenny proposed conceptual, strategic, and statistical definitions to specify and differentiate between the two scientific terms (Baron & Kenny, 1986). These conceptual definitions specified that a variable M is a *moderator* of the relationship between a target variable T and an outcome O in a particular population, if M explains *under what conditions* T is related to O. A variable M is a *mediator* of the relationship between T and O if M helps explains *how or why T* is related to O.

Although these conceptual definitions are succinct and clear, a systematic approach for applying these definitions in empirical research is necessary to determine whether the relationship between two variables $X1$ and X2 and an outcome O is one of moderation, mediation or neither, and whether $X1$ moderates (mediates) $X2$ or whether $X2$ moderates (mediates) $X1$.

Two sets of criteria are used to establish whether any variable functions as a moderator or mediator: eligibility and analytic criteria. Eligibility criteria identify whether a variable is a candidate for consideration as a potential moderator (or mediator) based on temporal precedence and association. The ultimate goal of moderation/mediation analyses is to detect possible causal chains among variables leading to the outcome. The criteria required to claim causality have been a topic of more than two millennia of philosophical argument and continues today (Kenny, 1979; Pearl, 2000; Rothman & Greenland, 1998; Rubin, 2004). However, there is agreement that to establish that variable X causes the outcome O, it is necessary (not sufficient) to show (1) temporal precedence, i.e., that X precedes O in time,

and (2) association, i.e., that X is correlated with O. Consequently, these particular criteria are used to determine whether particular variable is eligible as a moderator or mediator. Analytical criteria are statistical criteria used to empirically demonstrate whether an eligible variable actually functions as a moderator (or mediator).

For convenience here and to use the mathematical model that underlies the Baron & Kenny approach, the target (T) variable is binary (e.g., intervention versus control in a RCT, presence/absence of a risk factor in an observational study), the moderator or mediator (M) is either ordinal (e.g., age) or binary (e.g., gender), and the outcome variable (O) is ordinal.

Criteria for Establishing Moderators and Mediators

The Baron & Kenny approach for establishing moderators and mediators is based on a linear model (Shrout & Bolger, 2002):

$$O = \beta_0 + \beta_1 T + \beta_2 M + \beta_3 TM + \varepsilon, \tag{1}$$

where the variance of the error term ε is independent of both T and M. Although the MacArthur approach is not limited to linear models (Kraemer et al., 2001), this linear model is here used for the MacArthur approach as well, so that the two approaches can be directly compared.

Although both approaches rely on the conceptual definitions described earlier, the approaches differ in their eligibility criteria for moderation and in their analytical criteria for mediation (summarized in Table 1). To establish that M moderates T in its relationship with O, the Baron & Kenny approach does not specify an eligibility criterion regarding the temporal precedence of M and T. The Baron & Kenny approach also does not specify an eligibility criterion regarding whether M and T are associated or not. Rather, it only suggests that "it is desirable" that M be independent of T (Baron & Kenney, 1986, p. 1174).

In contrast, the MacArthur approach stipulates that M must precede T and that M and T must be independent. Both approaches specify the same analytic criterion: that an interaction between M and T be demonstrated.

To establish that M *mediates* T in its relationship with O, the Baron & Kenny approach originally did not impose an eligibility criterion of temporal precedence (Baron & Kenny, 1986), but there is growing consensus that such a criterion be imposed: T precedes M. The original Baron & Kenny approach did, however, stipulate the necessity of demonstrating an association between T and M. The MacArthur approach imposes the same eligibility criteria for mediation. However, the approaches differ in the analytic criterion used in the linear models. The Baron & Kenny approach assumes that there is no interaction between T and M (and, thus, sets the interaction term to zero). The Baron & Kenny approach establishes mediation by demonstrating that the relationship of T with O, when both M and T are considered, differs from that when M is not considered. In contrast, the MacArthur approach includes the interaction between T and M in the model, and establishes mediation by the demonstrated presence of either a main effect of M or an interaction between T and M.

Rationale for the MacArthur Modifications

To understand why the approaches differ in eligibility criteria for moderation and analytical criteria for mediation, the rationale for the modifications made by the MacArthur approach is presented, and then each of the differences in criteria are considered separately.

Including the Interaction Between M and T in the Model for Mediation

The Baron & Kenny approach assumes that the interaction between *M* and *T* is zero in the population for mediation and, thus, does not include the interaction in the linear model. However, assuming the interaction is zero does not make it so. If there is a nonzero interaction effect in the population, omitting it from the model results in a remapping of that effect. Part of the effect may be remapped into the main effects, thus biasing them and perturbing the probabilities of Type I error. The remainder is remapped into the error, increasing the probability of Type II error. For these reasons, the MacArthur approach stipulates a modification to the Baron & Kenny approach: that the interaction between *M* and *T* always be included in the linear model. To do so, however, raises the problem of how to center the independent variables for clear interpretability of the results.

Centering in Linear Models

It has long been recognized in the statistical literature that the coding of variables in a linear model with an interaction term affects the estimation of the main effects (Aiken & West, 1991; Cohen, Cohen, West, & Aiken, 2003; Kraemer & Blasey, 2004). Thus, in contrast to the Baron & Kenny approach which did not need to explicitly address coding, the MacArthur approach stipulates that a binary *T* be coded $+1/2$ and $-1/2$ and that ordinal *M* be coded as deviations from a meaningful central value (Kraemer et al., 2002). Here *M* is centered at a value dictated by what is known at the time *T* is determined. If *M* temporally precedes *T*, *M* is centered at its mean. If *T* precedes *M* (i.e., *M* is an event coded 1 or 0, or *M* is a change occurring between *T* and *O*), *M* is centered at zero. The reason for this modification lies in the current focus on effect sizes, rather than on null hypothesis significance testing.

Focus on Effect Size Rather Than Null Hypothesis Significance Testing

The Baron & Kenny approach focuses on testing a null hypothesis of random association in a linear model, consistent with approaches developed in an era when the primary emphasis of statistical analysis was significance testing of null hypotheses. Consistent with the growing recognition of the limitations of such testing and the current emphasis on meaningful effect sizes (Altman et al., 2001; Shrout, 1997; Thompson, 1999; Wilkinson & The Task Force on Statistical Inference, 1999), the MacArthur approach is based on the effect size of *T* on *O*. The effect size usually associated with the linear model is Cohen's δ, the standardized mean difference between the two groups defined by *T* (Cohen, 1988; Rosenthal, 1994), not the mean difference emphasized in the Baron & Kenny approach. If one has two groups (determined by *T*), where the means of the two groups differ, say, by 10 units, and the within group standard deviation is 100, there is virtually total overlap of the distributions of *O* in the two groups. However, if the within group standard deviation is 1, there is virtually no overlap. Thus the difference in means, ignoring variance, will mislead consideration of the size of effects, a fact long acknowledged in meta-analysis.

The effect size for this linear model is the standardized difference in means with centering, namely here:

$$\delta = \frac{\beta_1 + \beta_2(\mu_1 - \mu_2) + \beta_3(\mu_0 - C)}{\sqrt{V + .5[(\beta_2 + \beta_3/2)^2 - (\beta_2 - \beta_3/2)^2 \sigma_2^2]}} \tag{2}$$

where μ_1 and μ_2 are the means of centered *M* in the two groups defined by *T*, $\mu_0 = (\mu_1 + \mu_2)/2$, c is the centering value of *M*, σ_1^2, and σ_2^2 are the variances of centered *M* in those two groups and *V* is the variance of the error term assumed in the model to be the same in both groups defined by *T*.

Temporal Precedence of M Before T as an Eligibility Criterion for Moderation

Because the Baron & Kenny approach does not specify a criterion of temporal precedence for moderation, it is unclear which of two variables is the moderator and which one is being moderated. Consider an example: Suppose depressed participants in a RCT are randomized to one of two interventions as part of a study examining moderators and mediators of intervention. During the intervention period, some participants in each intervention experience a traumatic event (loss of job, death of loved one, etc.) that is *not* associated with random assignment of intervention. If participants in one intervention cope better with the traumatic event than participants in the other intervention, that might be reflected in an interaction effect between the intervention and the traumatic event on the outcome. The Baron & Kenny approach could conclude either that intervention moderates the event *or* that the event moderates intervention, given the absence of a temporal precedence criterion for intervention and the event. To avoid such situations, the MacArthur approach imposes the eligibility criterion of temporal precedence for moderation: a moderator *M* must precede *T*. Based on this criterion, the MacArthur approach would unambiguously conclude that here intervention moderates the event, since intervention here precedes the traumatic event, the intervention was not associated with the traumatic event, and the intervention predicts the way participants respond to a subsequent traumatic event.

In applying the Baron & Kenny approach, social/behavioral researchers often resolve such ambiguities by referring to their theoretical or conceptual model, specifically their 'a priori' assumptions. Thus, what one *assumes* about the position of *M* and *T* in a chain leading to *O*, determines whether a variable is labeled a moderator or as being moderated. Once a variable is potentially labeled as a moderator, this 'a priori' assumption also determines which variable is labeled as the moderator and which variable is labeled as being moderated. Consequently, multiple researchers analyzing the exact same data could draw contradictory conclusions about the roles of *M* and *T* simply because they made different assumptions about what the causal chains might be. This leads to nonreplicable results, a serious problem in all scientific research. Even more importantly, it may provoke confusion about the target of intervention in the context of prevention or treatment.

Absence of Association Between M and T as an Eligibility Criterion for Moderation

As can be seen in Equation (2) of the MacArthur approach, when *M* and *T* are not associated and *M* precedes *T*, then $\mu_1 = \mu_2$, $\mu_0 = c$, $\sigma_1^2 = \sigma_2^2$, and the effect of *T* on *O* in the total population does not depend on *M* (hence not a mediator). However, the effect of *T* on *O* in the subpopulation with $M = m$ depends on m when the interaction effect (β_3) is nonzero and is:

$$\delta = \frac{\beta_1 + \beta_3(m - \mu_0)}{\sqrt{V}}$$

where *V* is the common within group standard deviation. Thus, with the MacArthur approach, *M* plays a role in identifying sub-populations of the general population that have different relationships of *T* with *O*, consistent with the conceptual definition of a moderator.

Interpretation of the Interaction Between M and T as an Analytical Criterion for Mediation

As seen in Equation (2) of the MacArthur approach, when *T* precedes *M* and when *M* and *T* are associated, then $\mu_1 \neq \mu_2$ and possibly $\mu_0 \neq c$ and $\sigma_1^2 \neq \sigma_2^2$. Then the effect of *T* on *O* may be explained at least partially by the differences between the means and variances of *M* in the two groups defined by *T* (each possibly the effect of *T* on *M*). This is the case if either the main effect of *M* (β_2) *or* the interaction effect of *M* and *T* on O (β_3) is nonzero. Thus,

when M and T are associated, the interaction between M and T is indicative of *mediation*, not moderation. M plays a specific role in determining the effect size δ and thus fulfills the conceptual definition of a mediator of T.

For instance, in a RCT, a mediator may itself be an outcome of the intervention (associated with T). The intervention may change the level of the mediator (inducing changes in the mean), perhaps changing some participants in the population more than others (inducing changes in the variance), and changing the impact a given level of M has on the outcome (inducing an interaction effect). In summary, an interaction between an independent M and T indicates moderation whereas an interaction between an associated M and T indicates mediation.

Not Drawing Causal Inferences From Observational Data

The role of causal assumptions and inferences in medical research is a major conceptual distinction between the two approaches. Using specific eligibility and analytic criteria to establish whether variables are moderators or mediators is an important step toward demonstrating a potential casual role of moderators and mediators in explanatory chains of variables, but never sufficient to establish causality.

In applying the Baron & Kenny approach, it is often assumed that the causal pathways are known and this assumption is then used to identify whether variables are moderators or mediators. For example, a moderator has been defined as "a third variable exhibiting statistical interaction by virtue of its being antecedent or intermediate *in the causal process* under study" (Last, 1995, p. 107). A mediator is defined as "a variable that occurs *in a causal pathway* from an independent to a dependent variable. It *causes* variation in the dependent variable, and itself is *caused* to vary by the independent variable" (Last, 1995, p. 87; emphasis added).

In the social/behavioral contexts where the Baron & Kenny approach arose, a causal relationship between a target variable T and outcome O is often assumed as part of a theoretical or conceptual model which in turn determines an analytic model (Baron & Kenny, 1986). Then, if the data do not contradict that assumption, causal inferences are drawn about T. As a result, causal inferences are often drawn from observational studies or from cross-sectional studies, and from nonsignificant goodness-of-fit tests in structural equation modeling.

Because the fundamental reason for seeking moderators and mediators is to gain insight into detecting the unknown causal processes involved, any approaches that *assume* that the causal process is already known prejudices the search. Indeed, if the causal process were known, there would be no need for the detection of moderators or mediators.

In clinical research, there has always been greater reluctance to accept causal inferences than in social/behavioral theoretical research. There is strong reason for this. For example, from the 1970s and early 1980s, it was known that premature ventricular contractions (PVCs) were a risk factor for sudden cardiac death, and it was widely believed that any prophylactic program against sudden cardiac death must involve the use of antiarrhythmic drugs to subdue PVCs (i.e., reduction in PVCs would causally mediate the effect of any intervention on cardiac death). This was so widely accepted that the Federal Drug Administration accepted PVC frequency as a surrogate outcome instead of mortality in RCTs testing the efficacy of certain drugs. On this basis, several drugs were approved. Only later, in 1987, was a RCT conducted to actually test whether these PVC suppressors were effective in reducing mortality. The trial was terminated early when the results clearly indicated that the drugs *increased* the risk of death. Before the final RCT was begun, it was estimated that

there were 57,000 prescriptions per month being filled in the United States, that some 200,000 patients in the United States were taking the drugs, and that ultimately the number of excess drug-related fatalities in the United States were thought to be numbered in the tens of thousands (Moore, 1995; Silverman, 1998). For reasons such as these, premature claims of causality must be carefully avoided in clinical research.

In the MacArthur approach, the criteria used to define moderators and mediators do not assume the causal pathways are known nor assume any necessarily causal role for moderators or mediators once identified. Instead, the criteria for establishing whether variables are moderators or mediators are based solely on temporal precedence, association, and the nature of the joint association of T and M with O. Establishing that variables are moderators or mediators indicates the existence of relationships between T, M, and O that *may* or *may not* be causal. Establishing that variables are moderators or mediators in initial studies does not lead to conclusions about causality; rather, it leads to *hypotheses* about the possible causal role for identified moderators or mediators to be tested in future studies specifically designed for that purpose.

During discussions comparing the two approaches among developers and users of both approaches[1], the Baron & Kenny approach was often described as being "theory driven," and the MacArthur approach as "data driven." In fact, both are theory-driven and both are data-driven. Choices are made in sampling, measurement, design, and analysis that reflect a theory comprising scientific knowledge, mathematical assumptions appropriate to the present study such as linearity or equal variances, and specification of any hypotheses to be tested for which there is sufficient rationale and justification. In both approaches, a study is designed and measures are collected based on the theory and the assumptions to test the hypotheses— hence "theory-driven." The data obtained in the study then "drive" inferences drawn to expand scientific knowledge— hence "data driven". The validity of conclusions is always conditional on the assumptions made. Thus, a major difference between the two approaches is the unwillingness by the MacArthur approach to *assume* that M is in the causal chain leading from T to O as a way to prove that M is a mediator of T in its relationship with O and thus *might* be in the causal chain leading from T to O.

Implications for the Design and Interpretation of Future Studies

The modifications made by the MacArthur approach to the eligibility and analytical criteria for defining moderators and mediators have important implications for the design and interpretation of future research studies. To see the potential source of the differences, let us consider a hypothetical example, in which it is possible to systematically compare which terms (moderator, mediator, or neither) the two approaches would apply for all possible empirical situations.

Suppose participants in a RCT are randomized to either intervention A, with probability π_1, or to intervention B (variable $X_1 = 1$ if A, 0 if B). After a month of initial intervention, participants either remain on the same intervention with probability π_2 or switch to the other intervention (variable $X_2 = 1$ if A, 0 if B). Participants are assessed on the outcome O a year after randomization. This example considers all possible empirical situations because all combinations of probabilities of π_1 and π_2—that vary from 0 to 1 and can differ from one another— can be systematically considered for the two time points. The variables are initially labeled X_1 and X_2 because the criteria for each approach will be used to establish

---

which variable is labeled T and which variable is labeled M in each of the possible empirical situations.

Suppose that the population means in the four groups were AA, AB, BA, and BB. In this situation, X_1 and X_2 are not associated if and only if $\pi_2 = .5$. If one followed the Baron & Kenny approach for mediation and fit the linear model without the interaction (thus assuming $\beta_3 = 0$), any interaction in the population would be remapped in part to the main effects of X_1 (β_1) and X_2 (β_2) to a degree that would be determined by π_1 and π_2, and the rest of the missing interaction effect would serve to exaggerate the error variance. In contrast, if one followed the MacArthur approach for mediation and fit the linear model in Table 1 with an interaction term, the model would always fit the four means exactly.

Here, where X_1 is known to temporally precede X_2, what conclusions about the relationship between X_1 and X_2 with O are warranted using the two approaches, given a large enough sample size? Is the relationship between X_1 and X_2 with O one of moderation, mediation both or neither? In general, when there is no association between X_1 and X_2 ($\pi_2 = .5$), neither approach would conclude the relationship is mediation. Similarly, when there is no interaction effect, neither approach would conclude the relationship is moderation. In the first three situations in Table 2, the conclusions are the same for both approaches: neither moderation nor mediation could be established. In the fourth situation, when there is an association between X_1 and X_2 ($\pi_2 \neq .5$), no interaction effect (and hence no moderation in either approach), and a main effect of X_2, both approaches will conclude that X_2 mediates X_1.

However, there are other situations where the conclusions about the relationship between X_1 and X_2 for the Baron & Kenny approach and the MacArthur approach would differ (Table 2). In the fifth and sixth situations, when there is no association between X_1 and X_2 ($\pi_2 = .5$) and there is an interaction effect (regardless of the main effect of X_2), the Baron & Kenny approach could conclude either that X_1 moderates X_2 or that X_2 moderates X_1, given the absence of a temporal precedence criterion. In contrast, the MacArthur approach would conclude only that X_1 moderates X_2. Finally, in the last two situations, when there is an association between X_1 and X_2 ($\pi_2 \neq .5$), and there is interaction effect, the Baron & Kenny approach could conclude that X_2 mediates X_1 and/or that X_1 moderates X_2 and/or that X_2 moderates X_1, given the absence of a criterion regarding association. In contrast, the MacArthur approach would conclude only that X_2 mediates X_1. Thus, although the two approaches will sometimes agree, the Baron & Kenny approach could apply the same term to quite different empirical situations and apply different terms to the same empirical situation whereas the MacArthur approach produces a *unique* and consistent conclusion about which term to apply for each empirical situation.

Because there are no eligibility criteria for moderation in the Baron & Kenny approach, the distinction between showing that M moderates T or that M mediates T in many cases, is whether the analyst chooses to include the interaction term or not. Different analysts, once again, may choose differently and reach different conclusions from the same data. In the MacArthur approach, there is an unambiguous distinction between a moderator and a mediator, and an unambiguous determination of which variable is the moderator and which variable is being moderated.

The MacArthur approach clarifies that the same general construct may underlie both a moderator and mediator of T and the same instrument may be used to measure that construct but the same *variable* cannot be both a moderator and a mediator of T. For example, suppose two variables are measured in a RCT: prerandomization level of social support and postrandomization change in social support. Here the general construct of social support per

se is the same and the measurement instruments used at pre- and postrandomization may also be the same. However, the prerandomization level of social support and the postrandomization change in social support are *not* the same variables and may not even be highly correlated. The prerandomization level of social support may or may not moderate the relationship between the intervention and the outcome. The postrandomization change in social support that may result during intervention may or may not mediate the relationship between the intervention and the outcome, but these are separate research questions. The temporal precedence criterion imposed by the MacArthur approach to establish whether the variables are potential moderators or mediators prevents confusing what are two distinct variables.

Perhaps the most important implication of the MacArthur approach is the necessity of using longitudinal studies with at least two and usually three time points to establish moderators and mediators. The MacArthur approach cannot be applied in cross-sectional or retrospective studies as the Baron & Kenny approach often has.

For the MacArthur approach, two time points can only be used when considering constructs that are fixed markers. Fixed markers do not change over time, e.g., gender, ethnicity, year of birth, or genotype (Kraemer et al., 1997). It does not matter when such variables are measured, because they precede any event during the lifetime of the study participant. However, because fixed markers cannot change, only two questions can be asked about such variables: Do they moderate subsequent variables in their relationship with O? Are they mediated by subsequent variables in their relationship with O? For example, a genotype moderating the relationship of subsequently measured environmental variables with an outcome have been reported (Jaffee et al., 2005; Smeraldi et al., 1998). However, in absence of a fixed marker, three time points are needed to establish moderation (M moderates T with respect to O) or mediation (T is mediated by M with respect to O).

Because of this emphasis on timing, the MacArthur approach also helps to resolve problems such as "backward" or "reciprocal" causation. It may well be that X_1 measured at time 1 is mediated by Y_2, a change in Y between times 1 and 2, which in turn is mediated by X_3, a change in X between times 2 and 3, which in turn is mediated by Y_4, a change in Y between times 3 and 4 etc., forming a mediational chain (Collins et al., 1998). However, the criteria used in the MacArthur approach acknowledge that X_1 and X_3 are different variables, as are Y_2 and Y_4, even if the same instruments are used to measure X_1 and X_3 or to measure Y_2 and Y_4. If such a mediational chain were subsequently shown to be a causal chain, each cause would precede its effect. There is thus no "backward" or "reciprocal" causation in the MacArthur approach.

Finally, because of this emphasis on timing, no variable can be described as a moderator-mediator or a mediator-moderator with the MacArthur approach, a relationship sometimes claimed when applying the Baron & Kenny approach (see situations 7 and 8 in Table 2). In the MacArthur approach, a moderator must always precede and not be associated with that which it moderates, and a mediator must always follow and be associated with that which it mediates.

Role of Exploratory and Hypothesis-Testing Studies

In applications, both the Baron & Kenny and the MacArthur approaches can be used as a basis of hypothesis-testing using standard methods. To date, their primary value been in exploratory studies (Essex et al., 2006; Jaffee et al., 2005; Smeraldi et al., 1998). These exploratory studies are not designed to formally test hypotheses about the existence of particular moderators and/or mediators. Rather, they seek to generate strong hypotheses for future studies, either for experimental studies (e.g., RCTs) that test *a priori* hypotheses about

causal roles for moderators and/or mediators on an outcome (Kraemer, Frank, & Kupfer, 2006) or observational studies that test a priori hypotheses about the role (not necessarily causal) of particular moderators and/or mediators in predicting outcomes.

Evolution of the Macarthur Approach and Future Directions

Emphasis here has been on the rationale for the modifications by the MacArthur approach to the Baron & Kenny approach. Undoubtedly, the MacArthur approach has limitations that will emerge as it is applied, as has happened with the Baron & Kenny approach over the last 20 years. Some have already been identified.

When introducing the MacArthur approach (Kraemer et al., 2001), the criteria were defined, but without specification of methods to apply those criteria. It was rapidly found that in absence of specific methods, researchers found it difficult to apply the definitions consistently. Consequently, in considering the specific application to RCTs (Kraemer et al., 2002), the same linear model was used as used by the Baron & Kenny approach. What then proved problematic was the issue of centering of the independent variables. Because centering did not affect the interaction term, the problem was not in the definition of moderator, but in the definition of the mediator. The suggestion first presented in this paper of centering M at what is know about M at the time of T is one way to resolve that question to ensure the clear interpretability of the coefficients in the regression model.

An unsolved problem common to both approaches but more salient to the MacArthur approach is that of "proving" that two variables are not associated. For mediation, the Baron & Kenny approach has focused on whether the association is statistically significant ($p < .05$ or whatever significance level is set). However, this means that whether or not an association is claimed to exist depends largely on sample size, for the larger the sample size, the more likely one will see statistical significance. To date, in applying the MacArthur approach, a double criterion has been used: Association is shown to exist if $p < .05$ *and* the magnitude of an appropriate correlation coefficient exceed a threshold value, say .2. Both the selected significance level (.05) and the threshold value (.2) would best be based on consensus. Alternatively, a better tactic must be identified.

Finally, another problem common to both approaches but more salient to the Baron & Kenny approach is the dependence on linear models which often do not fit the research situation very well. There is nothing in the MacArthur approach that is specifically dependent on the linear model, but the linear model is both the most familiar to researchers and easiest to use. However, moving to a nonparametric approach is well underway. In such an approach, no assumptions would be made about the distribution of the outcome measures, linearity, or equal variances. However, such developments are yet in their infancy.

Conclusion

The catalyst for the MacArthur approach was concern about the ambiguities inherent in the Baron & Kenny approach, which are of great concern in research that underlies clinical decision-making. Such clinical decisions may include whether or not to recommend the use of an intervention (e.g., use hormone replacement therapy? get breast implants? prescribe a specific drug?), to recommend a reduction in risk factors (e.g., reduce fats? carbohydrates? proteins?), or to recommend the adoption of a protective factor (e.g., drink red wine to prevent heart disease?). Often there are serious consequences when such recommendations are wrong. Such recommendations, if inaccurate, can influence clinical decision-making before their inaccuracies can be detected and corrected by later research studies. As was apparent with PVC reducing drugs, the negative effects of incorrect inferences may not even be correctable for many to whom those clinical decisions were applied.

In emphasizing the concerns about the ramifications of erroneous inferences in clinical research, we do not suggest that those working in the social/behavioral sciences are protected against the consequences of inference errors associated with ambiguities. Today, many working in the social/behavioral sciences are working in areas with direct influence on policy decision making with consequences every bit as serious and immediate as those in clinical research. Such researchers may find the modifications designed to protect such decision-making introduced in the MacArthur approach more appropriate to their research objectives. Such a decision would require understanding what the similarities and differences between the two approaches are, the issue of focus in this discussion.

Acknowledgments

Glossary

β	Greek beta
σ	Greek sigma
δ	Greek delta
ε	Greek epsilon
μ	Greek mu

References

Aiken, LS.; West, SG. Multiple regression: Testing and interpreting interactions. Newbury Park, CA: Sage Publications; 1991.

Altman DG, Schulz KF, Hoher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: Explanation and elaboration. Annals of Internal Medicine. 2001; 134:663– 694. [PubMed: 11304107]

Baron RM, Kenny DA. The Moderator-Mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology. 1986; 51:1173–1182. [PubMed: 3806354]

Cohen, J. Statistical power analysis for the behavioral sciences. Hillsdale, NJ: Erlbaum; 1988.

Cohen, J.; Cohen, P.; West, S.; Aiken, L. Applied multiple regression/correlation analysis for the behavioral sciences. Hillsdale, NJ: Lawrence Erlbaum Associates; 2003.

Collins LM, Graham JW, Flaherty BP. An alternative framework for defining mediation. Multivariate Behavaioral Research. 1998; 33:295–312.

Essex MJ, Kraemer HC, Armstong JM, Boyce WT, Goldsmith HH, Klein MH, et al. Exploring risk factors for the emergence of children's mental health problems. Archives of General Psychiatry. 2006; 63:1246–1256. [PubMed: 17088505]

Jaffee SR, Caspi A, Moffitt TE, Dodge KA, Rutter M, et al. Nature X nurture: Genetic vulnerabilities interact with physical maltreatment to promote conduct problems. Development and Psychopathology. 2005; 17:67–84. [PubMed: 15971760]

James LR, Brett JM. Mediators, moderators, and tests for mediation. Journal of Applied Psychology. 1984; 69:307–321.

Kenny, DA. Correlation and causality. New York: Wiley; 1979.

Kenny, DA.; Kashy, DA.; Bolger, N. Data analysis in social psychology. In: Gilbert, D.; Fiske, ST.; Lindzey, G., editors. Handbook of social psychology. Vol. 1. New York: McGraw-Hill; 1998. p. 233-265.

Kraemer HC, Blasey C. Centring in regression analysis: A strategy to prevent errors in statistical inference. International Journal of Methods in Psychiatric Research. 2004; 13:141–151. [PubMed: 15297898]

Kraemer HC, Frank E, Kupfer DJ. Moderators of treatment outcomes: clinical, research, and policy importance. Journal of the American Medical Association. 2006; 296:1–4.

Kraemer HC, Kazdin AE, Offord DR, Kessler RC, Jensen PS, Kupfer DJ. Coming to terms with the terms of risk. Archives of General Psychiatry. 1997; 54:337–343. [PubMed: 9107150]

Kraemer HC, Kupfer DJ. Size of treatment effects and their importance to clinical research and practice. Biological Psychiatry. 2006; 59:990–996. [PubMed: 16368078]

Kraemer HC, Stice E, Kazdin A, Kupfer D. How do risk factors work together to produce an outcome? Mediators, moderators, independent, overlapping and proxy risk factors. The American Journal of Psychiatry. 2001; 158:848–856. [PubMed: 11384888]

Kraemer HC, Wilson GT, Fairburn CG, Agras WS. Mediators and moderators of treatment effects in randomized clinical trials. Archives of General Psychiatry. 2002; 59:877–883. [PubMed: 12365874]

Last, JM. A dictionary of epidemiology. New York: Oxford University Press; 1995.

MacKinnon DP, Lockwood CM, Hoffman JM, West SG, Sheets V. A comparison of methods to test mediation and other intervening variable effects. Psychological Methods. 2002; 7:83–104. [PubMed: 11928892]

Moore, TJ. Deadly medicine: Why tens of thousands of heart patients died in America's worst drug disaster. New York: Simon & Schuster; 1995.

Pearl, J. Causality: Models, reasoning, and inference. Cambridge: Cambridge University Press; 2000.

Rosenthal, R. Parametric measures of effect size. In: Cooper, H.; Hedges, LV., editors. The handbook of research synthesis. New York: Russell Sage Foundation; 1994. p. 231-244.

Rothman, KJ.; Greenland, S. Modern epidemiology. Philadelphia: Lippincott Williams & Wilkins; 1998.

Rubin DB. Teaching statistical inference for causal effects in experiments and observational studies. Journal of Educational and Behavioral Statistics. 2004; 29:343–367.

Shrout PE. Should significance tests be banned? Introduction to a special section exploring the pros and cons. Psychological Science. 1997; 8:1–2.

Shrout PE, Bolger N. Mediation in experimental and non-experimental studies: New procedures and recommendations. Psychological Methods. 2002; 7:422–445. [PubMed: 12530702]

Silverman, WA. Where's the evidence? Debates in modern medicine. Oxford: Oxford University Press; 1998.

Smeraldi E, Zanardi R, Benedetti F, DiBella D, Perez J, Catalano M. Polymorphism within the promoter of the serotonin transporter gene and antidepressant efficacy of fluvoxamine. Molecular Psychiatry. 1998; 3:508–511. [PubMed: 9857976]

Thompson B. Journal editorial policies regarding statistical significance tests: Heat is to fire as p is to importance. Educational Psychology Review. 1999; 11:157–169.

Thompson SG, Higgins JP. Can meta-analysis help target interventions at individuals most likely to benefit? Lancet. 2005; 365:341–346. [PubMed: 15664231]

Wilkinson L. The Task Force on Statistical Inference. Statistical methods in psychology journals: Guidelines and explanations. American Psychologist. 1999; 54:594–604.

Table 1

Eligibility and Analytic Criteria Used to Establish Variables as Moderators and Mediators for the Baron & Kenny Approach and the MacArthur Approach

	Baron & Kenny approach		MacArthur approach	
	M moderates T	M mediates T	M moderates T	M mediates T
Eligibility criteria				
Temporal precedence of either T or M	[Not specified]	T before M	M before T	T before M
Association between T and M	[Not specified]	$\gamma_1 \neq 0$	$\gamma_1 = 0$	$\gamma_1 \neq 0$
Analytical criteria	$\beta_3 \neq 0$	Assume $\beta_3 = 0$; $\gamma_1\beta_2 \neq 0$	$\beta_3 \neq 0$	Either $\beta_2 \neq 0$ or $\beta_3 \neq 0$

Note. Both use the linear models; $O = \beta_0 + \beta_1 T + \beta_2 M + \beta_3 MT + \varepsilon$; and $M = \gamma_0 + \gamma_1 T + \varepsilon^*$, where ε and ε^* are independent of T and M.

Table 2

Conclusions About the Relationship Between X_1 and X_2 for the Baron & Kenny Approach and MacArthur Approach in a Hypothetical Example

	Eligibility criteria		Analytic criteria		Conclusions	
	X_1 precedes X_2	Association between X_1 and X_2	Interaction between X_1 and X_2	Main effect of X_2	Baron & Kenny approach	MacArthur approach
1	[Yes]	No	No	No	Neither	Neither
2	[Yes]	No	No	Yes	Neither	Neither
3	[Yes]	Yes	No	No	Neither	Neither
4	Yes	Yes	No	Yes	X_2 mediates X_1	X_2 mediates X_1
5	[Yes]	No	Yes	No	X_1 moderates X_2 X_2 moderates X_1	X_1 moderates X_2
6	[Yes]	No	Yes	Yes	X_1 moderates X_2 X_2 moderates X_1	X_1 moderates X_2
7	[Yes]	Yes	Yes	No	X_2 mediates X_1 X_1 moderates X_2 X_2 moderates X_1	X_2 mediates X_1
8	[Yes]	Yes	Yes	Yes	X_2 mediates X_1 X_1 moderates X_2 X_2 moderates X_1	X_2 mediates X_1

Note. [Yes] Is an issue only for the MacArthur approach for moderation. Here subjects are randomized to intervention A and B (X_1 = 1 if A, 0 if B), and after a month either remain on the same intervention or are switched to the other intervention (X_2 = 1 if A, 0 if B).