

Reflections on the Baron and Kenny model of statistical mediation

Antonio Pardo* and Marta Román

Universidad Autónoma de Madrid

Título: Reflexiones sobre el modelo de mediación estadística de Baron y Kenny.

Resumen: En los 25 años que han transcurrido desde que Baron y Kenny (1986) publicaran sus ideas acerca de cómo proceder para analizar e interpretar la mediación estadística, pocos trabajos han sido tan citados como éste y, probablemente, ningún otro ha influido de una forma tan decisiva sobre la forma en que los investigadores aplicados entienden y analizan la mediación en el ámbito de las ciencias sociales y de la salud. Pero la utilización masiva de un procedimiento no lo convierte, por sí sola, en una estrategia fiable o segura. De hecho, en estos 25 años no pocos investigadores han puesto de manifiesto las limitaciones de la propuesta de Baron y Kenny para demostrar mediación. El objetivo de este trabajo es (1) realizar una revisión de las limitaciones del método de Baron y Kenny prestando especial atención a la debilidad de la lógica confirmatoria en la que se basa y (2) ofrecer una demostración empírica de que, aplicando la estrategia de Baron y Kenny, los datos obtenidos en el marco de un mismo escenario teórico (sea o no de mediación) pueden ser compatibles tanto con la hipótesis de mediación como con la hipótesis de no mediación.

Palabras clave: Baron y Kenny; mediación estadística; efectos indirectos.

Abstract: In the 25 years since Baron and Kenny (1986) published their ideas on how to analyze and interpret statistical mediation, few works have been more cited, and perhaps, so decisively influenced the way applied researchers understand and analyze mediation in social and health sciences. However, the widespread use of a procedure does not necessarily make it a safe or reliable strategy. In fact, during these years, many researchers have pointed out the limitations of the procedure Baron and Kenny proposed for demonstrating mediation. The twofold aim of this paper is to (1) carry out a review of the limitations of the method by Baron and Kenny, with particular attention to the weakness in the confirmatory logic of the procedure, and (2) provide an empirical example that, in applying the method, data obtained from the same theoretical scenario (i.e., with or without the presence of mediation) can be compatible with both the mediation and no-mediation hypotheses.

Key words: Baron and Kenny; statistical mediation; indirect effects.

Introduction

The term *statistical mediation* or simply *mediation*, refers to a *causal chain* in which it is assumed that the effect of one or more independent variables is transmitted to one or more dependent variables through third variables. In the simplest case, the term *mediation* is used to indicate that the effect of an independent variable (X) is transmitted to a dependent variable (Y) through a third mediator variable (M). Therefore, statistical mediation refers to a causal sequence such as $X \rightarrow M \rightarrow Y$ (MacKinnon, Fairchild, & Fritz, 2007). A mediator variable is very useful to help understand the mechanism through which a cause (independent variable) produces an effect (dependent variable) (see Fairchild & MacKinnon, 2009).

Twenty-five years have gone by since the appearance of Baron and Kenny's (1986) influential work on how to proceed in order to identify mediator variables (M) in the relationship between two variables (X - Y). Using work by Judd and Kenny (1981) as a starting point, Baron and Kenny explain the meaning of statistical mediation and propose a simple method that, apparently, allows identifying mediator variables using the sequential adjustment from several linear regression models (see next section). During these twenty-five years, few works have been more cited than Baron and Kenny's (more than 18.000 mentions up until July 2012, according to the *Social Science Citation Index*), and perhaps, so decisively influenced the way applied researchers understand and analyze mediation in health and social sciences.

In a review of the studies published in the *Journal of Counseling Psychology* in 2001, Frazier, Tix, and Baron (2004) found a statistical mediation analysis in 10 out of 54 published articles (19%) during that year. In an informal review, Preacher and Hayes (2004) pointed out that 22% of the articles published in the *Journal of Applied Psychology* contained a mediation based analysis (most of them using Baron & Kenny's method). Iacobucci, Saldanha, and Deng (2007) informed that, between 1991 and 2004, approximately 25% of the published articles in the *Journal of Consumer Psychology* and the *Journal of Consumer Research* contained a mediation analysis. When searching for articles in the PsycInfo database that contained the word *mediation* in the title and had cited Baron and Kenny's (1986) article, Mackinnon et al. (2007) found 291 articles (98 from the field of social psychology, 70 from the field of clinical psychology and the remaining from other fields of psychology).

Baron and Kenny's method has not only been the most widely used method in the last years in order to demonstrate mediation in social and health sciences (see MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002; Wood, Goodman, Beckman, & Cook, 2008) but it is also very possibly the most used method still. For example, Barsevick, Dudley, and Beck (2006) have used Baron and Kenny's mediation method to demonstrate that the functional state of a person diagnosed with cancer acts as a mediator in the relationship between fatigue and symptoms of depression. Hofmann, Rauch, and Gawronski (2007) have identified imipramine as a mediator in the effect of cognitive-behavioral therapy in patients diagnosed with a panic disorder. Jiménez, Musitu, and Murgui (2008) have studied the mediating effect of self-esteem on the relationship between bad family functioning and drug consumption. Horcajo, Petty, and Briñol (2010) have studied the overall effect of the quality of an argument

*** Dirección para correspondencia [Correspondence address]:**

Antonio Pardo. Facultad de Psicología. Universidad Autónoma de Madrid. Cantoblanco. 28049 - Madrid (España).
E-mail: antonio.pardo@uam.es

and the status of the person that has proposed it on specific attitudes using the strength of the argument and the trust in the person as mediator variables.

The great popularity of Baron and Kenny's (1986) proposed strategy could lead one to think that it is the best way of demonstrating mediation (or, at least, a good way of doing it). But the massive use of a method doesn't guarantee it's a safe strategy. In fact, Baron and Kenny's proposed method contains important limitations.

Some of these limitations have already been pointed out. Mckinnon et al. (2002) compared fourteen methods designed to contrast the mediation hypothesis and concluded that Baron and Kenny's method is less potent than others. Mallinckrodt, Abraham, Wei, and Russell (2006) arrived at the same conclusion when they compared Baron and Kenny's strategy with a method based on bootstrap techniques proposed by Shrout and Bolger (2002). James and Brett (1984), James, Mulaik, and Brett (2006) and many others (included in Kenny, 2008; Kenny, Kashy, & Bolger, 1998) have argued that there can be mediation and that studying it may make sense even when X and Y aren't related to each other, this is, even when there is no apparent relationship that can be mediated. Judd and Kenny (1981) and Frazier et al., (2004), amongst many, have warned against the limitations of linear regression analysis (and, in general, methods that don't control the measurement error) in the estimation of coefficients in a mediation model (Hoyle and Robinson, 2003, recommend using mediator variables whose reliability isn't under .90).

The great repercussion that Baron and Kenny's (1986) proposal has had and still has together with the ascertainment that it is an analytical strategy that holds important limitations has motivated us to carry out a study with two aims: (1) reviewing Baron and Kenny's mediation model from the point of view of its limitations in order to identify mediator variables (pointing out some of the ones that affect the essence of the argument proposed by Baron and Kenny) and (2) providing an empirical demonstration that, in applying the method, data obtained from the same theoretical scenario (i.e., with or without the presence of mediation) can lead to contradictory conclusions; specifically, we aim to demonstrate that a same theoretical scenario can generate data compatible with the mediation hypothesis as well as data compatible with the non-mediation hypothesis.

The Baron and Kenny proposal

As we have already pointed out, the term *mediation* indicates that the effect of an independent variable (X) is transmitted to a dependent variable (Y) through a third variable considered a mediator (M). The strategy proposed by Baron and Kenny to tackle with the study of mediation (1986; see also Judd & Kenny, 1981; Kenny, 2011; Kenny et al., 1998) consists of making a sequential verification of four conditions (reason why Baron & Kenny's proposal is known as the *four step model*):

1. Variables X and Y must be related, this is, coefficient c in Figure 1 must be different to zero in the expected direction. This condition is verified using a linear regression analysis of Y over X :

$$Y = i_1 + cX + e_1 \quad (1)$$

where i is the constant term, c is the regression coefficient that relates X to Y , and e are the random errors (this is, the part of Y that isn't explained by X), which are considered to be normally distributed, with constant variance and independent from each other. We represent the parameters with Latin letters instead of the more common Greek letters in order to respect the most commonly used notation in published papers on mediation (see, for example, Baron & Kenny, 1986; Kenny, 2011, o MacKinnon, 2008, pp. 48-49).

2. Variables X and M must be related, this is, coefficient a from Figure 1 must be different to zero. This condition is verified using a linear regression analysis of M over X :

$$M = i_2 + aX + e_2 \quad (2)$$

3. Variables M and Y must be related once the effect of X is controlled, this is, coefficient b from Figure 1 must be different to zero. This condition is verified using a linear regression analysis of Y over X and M :

$$Y = i_3 + aX + bM + e_3 \quad (3)$$

4. The relationship between X and Y must be significantly reduced when controlling the effect of M . This is, coefficient c' (direct effect in Figure 1) must be smaller than coefficient c (total effect in Figure 1). Baron and Kenny (1986, p.1176) explicitly point out that "the strongest mediation demonstration is when c' is zero".

Baron and Kenny conceive the incidental size reduction in coefficient c as a continuum: the larger that reduction is, the larger the degree of mediation. Therefore, when the reduction is maximum, this is, when coefficient c' is zero, there is evidence of the presence of only one mediator variable; on the other hand, if a reduction in the size of c' occurs without it reaching zero, there is evidence that more than one mediator variable is taking place. As a consequence of this, in Baron and Kenny's proposal there is a distinction between *total mediation* (all of the effect of X goes through M) and *partial mediation* (only part of the effect of X goes through M). The data are compatible with the total mediation hypothesis when the relationship between X and Y completely disappears when controlling M (this is, when coefficient c' is zero). The data are compatible with the partial mediation hypothesis when the relationship between X and Y is significantly reduced when controlling M but doesn't completely disappear (this is, when the absolute value of coefficient c' is smaller than c and, at the same time, greater than zero).

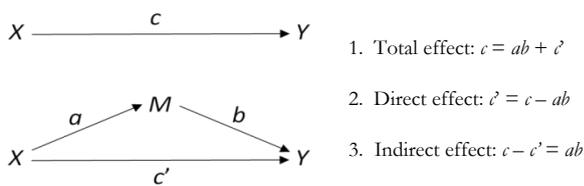


Figure 1. Details from Baron and Kenny's (1986) mediation model.

Limitations in Baron and Kenny's proposal

As we have already pointed out, the massive use of a method doesn't make it reliable or safe. In this section we will introduce some of the limitations in Baron and Kenny's (1986) method.

Our aim by doing this is not recommending researchers to abandon statistical mediation analysis, but to contribute to the awareness of the serious limitations this methodology has and to improve, in what is possible, the way in which they tackle with the study of mediation.

The role of the relationship between X and Y

The first of the four conditions from Baron and Kenny's (1986) proposal is that the independent variable X must be related to the dependent variable Y . Several experts (see, for example, Mathieu, & Taylor, 2006; Preacher & Hayes, 2004) think as do Baron and Kenny. This point of view is based on the idea that the main objective of mediation analysis is to contribute to the understanding of the relationship between two variables. In this sense, mediator variables are mechanisms which help clarify or understand the meaning or the nature of this relationship; if this relationship does not exist, there is nothing to mediate. It is from there that the first condition that must be satisfied, according to Baron and Kenny's proposal, is that "an effect that can be mediated must exist" (Kenny et al., 1998, p. 259).

However, many experts (see, for example, Collins, Graham, & Flagerty, 1998; James & Brett, 1984; James et al., 2006; Judd & Kenny, 2010; MacKinnon, 2009; MacKinnon, Krull, & Lockwood, 2000; MacKinnon et al., 2002; Shrout & Bolger, 2002; Zhao, Lynch, & Chen, 2010; etc.) have argued that the first condition of the sequence can be overlooked. From this point of view, mediation analysis could make sense even when no relationship between X and Y is observed. This absence of a relationship between X and Y in the mediation context can occur due to different reasons. For example, when applying a treatment it's possible to not find a relationship between the treatment and the dependent variable because there are non-identified suppressing or moderating variables that are altering that relationship (MacKinnon et al., 2000; Shrout & Bolger, 2002). It can also occur that different mediator variables produce opposite effects (MacKinnon et al., 2000). Take, for example, an intervention program designed to create awareness about the benefits of recycling trash among neighbors. It's a possibility that the

program (X) improves the neighbors' attitude (M), but that this isn't entirely reflected on the recycling behavior (Y) due to the troubles associated to recycling. The effect of a mediator variable (attitude toward recycling) could be partially neutralized by the effect of other variables (the trouble associated to recycling) and this could weaken the relationship between M and Y , and nullify the relationship between X and Y .

Adding to the idea that the proof of mediation doesn't require starting off from the relationship between X and Y , Shrout and Bolger (2002) argue that that is precisely what happens when the independent and dependent variables are separated by a long period of time, as is the case with longitudinal studies; the farther apart the independent and the dependent variables are in the causal chain, the less probable it is for the relationship between them to reach statistical significance (see also James et al., 2006).

On the other hand, the absence of statistical significance in the relationship between X and Y could be due to the use of designs which are not very potent, and not because no relationship exists. This is what Fritz and MacKinnon (2007) suggest is happening in a lot of cases.

Even Kenny (Kenny et al., 1998) admits, in a review of his work dating back to 1986, that the first of the four conditions (demonstrating that X and Y are related) could be overlooked in many cases. Therefore, even as logical as it may seem that to be able to talk about mediation there must be an effect to mediate, it doesn't seem necessary for that effect to reflect on the presence of a statistically significant relationship between X and Y .

Technical limitations

The fourth step in Baron and Kenny's (1986) method allows concluding that empirical evidence compatible with the mediation hypothesis when the direct effect (c') is less than the total effect (c), or what is equivalent, when the indirect effect (the product of ab) is different to zero. More specifically, Baron and Kenny indicate that the data are compatible with the total mediation hypothesis when the direct effect is null. As we've already pointed out, according to Baron and Kenny the strongest mediation demonstration occurs when c' is zero. But this way of *demonstrating* mediation has, in our opinion, important limitations.

In the first place, MacKinnon et al. (2002), after comparing fourteen strategies designed to identify the presence of mediator variables through the use of simulation, demonstrated that the strategy that consists of considering that mediation exists when coefficient c' is not significantly different to zero was the least potent one of all the ones compared. This strategy only reached acceptable potency with very large samples (more than 500 cases) as well as with indirect effects of the large size. In a pretty typical situation (a medium sized indirect effect), the potency of the procedure only reached a value of .28 with a sample of 100 cases and a value of .52 with a sample of 200 cases.

In the second place, the sequential adjustment of the three regression models implicated in Baron and Kenny's proposal (equations 1 to 3) is not a strong enough argument to conclude that the indirect effect is significantly different to zero. The fact that the total effect (c) is significantly different to zero while the direct effect (c') is not, doesn't imply that coefficients c and c' are different (see, for example, Preacher & Hayes, 2004). This problem is similar to the one we find when assessing the existing relationship between simple effects and the interaction effect in an analysis of variance; the fact that a simple effect is different to zero and another is not, doesn't imply that those simple effects are different from each other (see Pardo, Garrido, Ruiz, & San Martín, 2007). Zhao et al. (2010) point in this same direction when arguing that the strength of the mediation must be assessed starting from the size of the indirect effect (ab), and not from the absence of a direct effect (c'). Taking all this into consideration it is sensible to say that in order to determine if c and c' are different, knowing the individual statistical significance of coefficients c and c' is not enough, a comparison between them is necessary.

Therefore, the fourth step in Baron and Kenny's method requires applying a procedure that allows comparing coefficients c and c' in order to assess the statistical significance of the mediated effect. Out of the different available procedures, Baron and Kenny (1986; Kenny et al., 1998) have chosen and made popular the Sobel's test (Sobel, 1982). This test allows contrasting the null hypothesis by which the indirect effect is zero in the population (i.e., $ab = c - c' = 0$) dividing the product of the estimations of coefficients a and b by that product's standard error thus obtaining the statistical significance of that quotient using probabilities taken from the normal curve. However, this procedure has a weakness. Independently to how the standard error of the ab product is calculated (see Kenny et al., 1998; Sobel, 1982, 1986), MacKinnon, Warsi, & Dwyer (1995) have pointed out that the product of two normally distributed variables is not a normally distributed variable itself. And Bollen & Stine (1990) have demonstrated that the distribution of the ab product tends to be asymmetric (see also MacKinnon et al., 2002; MacKinnon, Lockwood, & Williams, 2004; Stone & Sobel, 1990). The consequence of this asymmetry is that Sobel's test is not as potent when working with small sample sizes.

In order to beat this problem, several authors (see for example Bollen & Stine, 1990; Shrout & Bolger, 2002; Preacher & Hayes, 2004, 2008) have proposed estimating the standard error of the ab product using resampling techniques (bootstrapping¹). The results obtained through these estimations seem to improve those obtained with Sobel's test,

however they still hold elevated Type I error rates when one of the coefficients (a or b) isn't different to zero (MacKinnon et al., 2004).

From all the simulation studies performed by MacKinnon et al. (2002) it can be deduced that the *joint significance test* (JST) is the strategy that offers the best balance between potency and Type I error rate control (this test holds the advantage that it's easy to apply: it simply consists of inspecting the regression results in order to verify the statistical significance of a and b ; if both coefficients are significantly different to zero, the conclusion is that indirect effect ab is also significantly different to zero), however, it being one of the best methods doesn't make it good enough. For example, the potency of the JST to detect a medium effect with a sample of fifty cases is .55; and if the effect is small, it doesn't reach .30 with a sample of two hundred cases. Therefore, no procedure that allows reliably contrasting the null hypothesis by which the indirect effect is zero seems to exist.

Affirming the consequent

The mediated relationship is an intrinsically causal relationship (see, for example, James et al., 2006; James, Mulaik & Brett, 1982). Baron and Kenny's (1986) mediation model is not only a *relationship* model, but a *causal relationship* model. Its promoters present it as such and it is as such as other authors understand and focus it (see, for example, Fraizer et al., 2004; MacKinnon, 2008; Mallinckrodt et al., 2006; Mathieu & Taylor, 2006; Spencer, Zanna, & Fong, 2005).

However, in non-experimental studies, in which mediation models are usually applied, the used designs' own characteristics don't allow demonstrating that the detected relationships are causal in nature (see, for example, Spencer et al., 2005; Stone-Romero & Rosopa, 2008). In absence of a real experiment, the best way available of demonstrating that a relationship is causal in nature is using a solid theory that is capable of generating specific predictions and a set of data that confirms those predictions. It is here where problems start to arise with mediation models in general and specifically with Baron and Kenny's proposal.

Baron and Kenny's argument adjusts to a clear logical order: *first goes the theory and then the data*. This is, first a mediation model that specifies the relationship between the involved variables must be formulated and only after this is done is when it makes sense to review the existing empirical relationships between the variables. If the mediation model is correct, reality should behave just as the model predicts; and when this occurs, it will be possible to conclude that the data support or are compatible with the mediation hypothesis. However, the argument has no return: the fact that the data behave as predicted by the mediation model doesn't imply the existence of mediation.

The reason for this asymmetry is clear. Baron and Kenny's argument is based on verification, not falsation; therefore, it is an argument which is non-conclusive on its own. The starting premise is that " p implies q ", this is, "if p , then

¹ Obtaining the distribution of a statistic (for example, the standard error of the ab product) through the use of bootstrapping is relatively simple. A size N sample is taken as if it were the reference population and from it, n sized samples with replacement are extracted. For each sample, a , b and the ab product are calculated. After extracting, say, a thousand samples, there are a thousand values from which the standard error of the ab product can be estimated.

q ”, p being the statement that “the mediation model is correct” (this is, the affirmation that variable M acts as a mediator in the relationship between X and Y) and q being “the behavior expected from the data when mediation really exists” (specifically, the indirect ab effect is different to zero and the total c effect completely or partially disappears). The statement “ p implies q ” can be expressed in the following way: “If the mediation model is correct, then the indirect ab effect must be different to zero and the total c effect must completely or partially disappear”.

However, using this premise as the starting point, there are only two possibilities: the data behave as the mediation model predicts, or that this doesn’t occur². When the data don’t behave as the mediation model predicts, it can be concluded (following a valid logical argument: *modus tollens*), that the mediation model isn’t correct (if it were, the data should behave according to the model): if p implies q and the q event doesn’t occur, the conclusion is that p hasn’t occurred, because if p had occurred, q would have necessarily occurred.

On the other hand, the fact that the data behave as the mediation model predicts, doesn’t allow stating that mediation exists, it only allows stating that the data are compatible with the mediation model and that, consequently, the proposed model offers a plausible (and maybe useful) explanation of the data; but this conclusion doesn’t exclude the existence of other models as plausible and useful as the one contrasted; if p implies q and it is verified that q is found, the conclusion drawn is that *maybe* p has occurred.

Indeed, due to the fact that the starting point of the argument is “ p implies q ” and not “ q implies p ”, the verification that q has occurred doesn’t guarantee the occurrence of p , since it is possible that q occurred because of different reasons to p . If the conclusion that mediation exists when the data behave as if mediation really did exist is to be drawn, then a logical error (a fallacy) known as *affirming the consequent* is committed, since it is possible that the data could behave that way due to different reasons than mediation. As James et al. (2006) point out, the fact that a specific model offers good adjustment to the data doesn’t imply that that model is correct or true; the possibility that other models (with other relationships and the same or other variables) offer as good an adjustment or better than the proposed model still exists.

These arguments are in contrast with some affirmations that, in our opinion, don’t help applied scientists to do things correctly. For example, Baron and Kenny’s statement that “a variable works as a mediator when the following conditions occur...” (1986, p. 1176) is not a correct statement; what happens is rather that “when a variable acts as a mediator, the following conditions are satisfied...”, which is very different. This confusion is pretty common in published studies on mediation. And Shrout and Bolger (2002) summarize it as follows: “if the data suggest that c takes a value different to

zero and that c' isn’t different to zero, then Kenny et al. (1998) will conclude that complete mediation exists”. Therefore, recognized mediation experts that usually do things correctly, tend to occasionally forget the limitations of a confirmatory argument when making statements.

Complete mediation or partial mediation

In order to state that confirmatory evidence of the presence of mediation has been found, the theory must be previous to the data: “the inferences on mediation are based, first and mainly, on theories... and, secondly, on statistical relationship evidence” (Mathieu & Taylor, 2006, p. 1032). The arguments found above allow justifying this statement.

A researcher that considers the complete mediation hypothesis and finds data compatible with that hypothesis can conclude that he or she has found confirmatory evidence of his or her hypothesis (we have already pointed out that this doesn’t mean that the hypothesis is true or correct, although it’s plausible and probably useful). The same can be said of a researcher that considers the partial mediation hypothesis and finds data compatible with this hypothesis.

However, what can a researcher conclude when the complete mediation hypothesis was considered but data compatible with the partial mediation hypothesis are found? Or a researcher that has considered the partial mediation hypothesis and finds compatible data with the complete mediation hypothesis? Can he or she conclude that confirmatory evidence of the hypothesis was found? Obviously not because the data are not compatible with the hypothesis.

And, if the starting hypothesis is a mediation one without specifying if it is partial or complete (this seems to be the recommended hypothesis in Baron and Kenny’s proposal, and is also common practice), after reviewing the data, can partial or complete mediation be the conclusion? Obviously not. In order to conclude that mediation exists when applying a verification-based strategy, the carriage (the data) cannot be placed in front of the horses (the hypothesis). And, however, in practice, the usual (and incorrect) way of assessing if complete or partial mediation exists consists of examining the data (specifically, the statistical significance of coefficient c') and making a decision on whether complete or partial mediation exists according to what the data say (see James et al., 2006).

When proceeding in this manner an essentially confirmatory method is becoming exploratory (LeBreton, Wu, & Bing, 2009). Specifically, an argument of the kind “if p , then q ” is becoming an argument of the kind³ “if p_1 , then q_1 ; if p_2 , then q_2 ”. When using the existence of q_1 or q_2 in order to decide if the correct premise is p_1 or p_2 a fallacious argument is being utilized (the affirmation of the consequent). The consequence of this way of arguing is that distinguishing when a direct effect is no longer significant because indeed variable

² In this context, as in many others, *statistical significance* is used in order to decide if the data behave or not in a specific manner. Therefore, it must not be overlooked that the arguments used are probabilistic in nature.

³ With p_1 = “complete mediation hypothesis”, p_2 = “partial mediation hypothesis”; q_1 = “coefficient c' is zero and different to c ”; q_2 = “coefficient c' is different to zero and to d ”.

M is a mediator variable in the X and Y relationship or because the relationship between X and Y is simply spurious is made impossible.

Mediation without mediation

Baron and Kenny's (1986) strategy to evidence mediation doesn't only suffer from logic problems. In this section we aim to demonstrate that Baron and Kenny's strategy is not a safe strategy to contrast the mediation hypothesis. More specifically, we aim to demonstrate that *the data can behave as if mediation exists when in reality it doesn't and as if it doesn't exist when in reality it does*.

We have included fictitious data compatible with different scenarios in the Appendix. The data corresponding to the variable trio with subscript 1 (this is, X_1, M_1, Y_1) are compatible with the *complete mediation* scenario (CM); the data corresponding to the variable trio with subscript 2 are compatible with the *partial mediation* hypothesis (PM); and the data corresponding to the variable trio with subscript 3 are compatible with a *non-mediation* or *mediation absence* scenario (NM).

Table 1 shows the value of Pearson's correlation coefficient between the variables of each trio. This coefficients are sample values whose degree of similarity with the corresponding parameter (population correlation coefficient) can be obtained using the information offered in Table 2. This table contains percentiles 2.5 and 97.5 from the sample distributions of Pearson's correlation coefficient for $n = 50$ and for two different coefficient values: .60 and .70. These percentiles are approximate; they're the empirical values obtained when simulating a thousand samples. In each one of these one thousand samples, three normally distributed variables (X, M and Y) were simulated with a theoretical correlation of .60 for the XY pair and .70 for the XM and MY pairs (Kenny et al., 1998, suggest that when searching for mediator variables, the XM relationship should be comparable to the size of the MY relationship). Percentiles 2.5 and 97.5 shown in Table 2 are the values between which 95% of the simulated coefficients are located. Therefore, those two percentiles indicate the limits between which the values of Pearson's correlation coefficients are expected to be found with a 95% level of confidence, when calculated using variables measured in random fifty-case samples extracted from a multivariate normal distribution where the XY pair correlates at a value of .60 and the XM and MY pairs correlate at a value of .70.

In the three cases (CM, PM and NM) we start from the same supposition: the real relationship (the population value of the correlation coefficient) is .60 for the XY pair and .70 for the XM and MY pairs. We could've chosen different scenarios, but all of them would've made us arrive at the same conclusion as this one. This is, other sample sizes and other population correlation coefficients would've led us to the same results. Smaller correlations (which are more realistic in health and social sciences) would have more disperse sample distributions and would make finding the effect we

aim to demonstrate easier. That is the reason why we've chosen relatively high correlation levels.

Therefore, the results we offer here with the specific values we have chosen to use (.60 for the XY pair and .70 for the XM and MY pairs) are an example of the results we could find with other values. And, more importantly, all of these results could correspond, indistinctively, to a *true mediation* scenario or to a *spurious relationship* scenario due to third variables that are not being taken into consideration. This is, the chosen population correlations (.60 for the XY pair and .70 for the XM and MY pairs) can be found in a scenario in which variable M truly acts as a mediator in the XY relationship, as well as in a scenario in which the XY relationship is spurious (due to, for example, X as well as Y depending on M).

Table 1. Pearson correlation coefficients between the variables of each trio from the Appendix table ($p < .0001$ in all cases).

		M_1	Y_1
1. Complete mediation			
	X_1	.759	.652
	M_1		.753
2. Partial mediation		M_2	Y_2
	X_2	.732	.702
	M_2		.692
3. No mediation		M_3	Y_3
	X_3	.748	.757
	M_3		.655

Table 2. Approximate percentiles (based on a thousand samples) from the sample distribution of the Pearson correlation coefficient.

Population correlation	Empirical percentiles	
	2.5	97.5
.60	.366	.781
.70	.498	.846

Complete mediation

Let's begin with the first of the three cases. The obtained coefficients are .652, .759 and .753 (see Table 1). The percentiles in Table 2 show that these sample values are perfectly compatible with the population values that we assume have generated them (.60 for the XY pair and .70 for the XM and MY pairs), this is, the three sample values are found between percentiles 2.5 and 97.5 of their corresponding sample distributions: coefficient .652 is found between values .366 and .781 (interval limits corresponding to coefficient .60) and coefficients .759 and .753 are found between values .498 and .846 (interval limits corresponding to coefficient .70). Therefore, in a scenario in which the true relationship between variables was .60 for XY and .70 for XM and MY , it could be probable that correlation coefficients calculated using samples of fifty cases would be .652, .759 and .753 respectively.

The results in Table 3 show that this relationship pattern is compatible with the presence of complete mediation: (1) regression coefficients a , b and c are significantly different to zero and (2) coefficient c' is significantly smaller than coefficient c (this is, the indirect ab effect is significantly different to zero) and isn't significantly different to zero.

According to Baron and Kenny's (1986) argument, a researcher that had found these data would have found *confirmatory empirical evidence for the complete mediation hypothesis*. In reality, since coefficient c' isn't significantly different to zero, we could assume its population value to be zero and this would be "the strongest mediation demonstration" (Baron & Kenny, 1986, p. 1176). However, this is not relevant; what is relevant is that the found data corroborate (are compatible with) the complete mediation hypothesis.

Table 3. Regression coefficients (standard errors) and Sobel's test corresponding to the scenarios shown in Table 2.

	<i>Complete mediation</i>	<i>Partial mediation</i>	<i>No mediation</i>
<i>Coefficient a</i>	.765 (.095) $p < .0005$.724 (.097) $p < .0005$.671 (.086) $p < .0005$
<i>Coefficient b</i>	.590 (.140) $p < .0005$.421 (.155) $p = .009$.211 (.146) $p = .156$
<i>Coefficient c'</i>	.185 (.141) $p = .197$.456 (.154) $p = .005$.568 (.132) $p < .0005$
<i>Coefficient c</i>	.636 (.107) $p < .0005$.760 (.111) $p < .0005$.710 (.088) $p < .0005$
<i>Sobel's test</i>	$Z = 3.73$ $p = .0002$	$Z = 2.55$ $p = .011$	$Z = 1.42$ $p = .156$

Partial mediation

Let's see what happens in the second case. Although the empirical correlation coefficients have changed (.702, .732 and .692; see Table 1), the percentiles in Table 2 show that these coefficients are still compatible with the population values that we assume have generated them (the three sample values are located between percentiles 2.5 and 97.5 of their respective sample distributions). And the results in Table 3 indicate that this relationship pattern is compatible with the presence of *partial mediation*: (1) regression coefficients a , b and c are significantly different to zero and (2) coefficient c' is significantly smaller than coefficient c (this is, the indirect ab effect is significantly different to zero) and significantly different to zero. Therefore, the same starting point (.60 for XY and .70 for XM and MY) that generated *confirmatory empirical evidence (compatible with) the complete mediation hypothesis* has now generated *confirmatory empirical evidence (compatible with) the partial mediation hypothesis*; and the reason for this discrepancy is only due to small sample variations that are located within what is probable.

No mediation

Finally, let's see what happens in the third case. The empirical correlation coefficient values have slightly changed

(.757, .748 and .655; see Table 1) and the percentiles shown in Table 2 indicate that those coefficients are still compatible with the population values we assume have generated them (.60 for XY and .70 for XM and MY). The results shown in Table 3 indicate that the found relationship pattern is compatible with the *absence of mediation*: since coefficient b is not significantly different to zero, it's not possible to conclude that variable M is a mediator variable (the indirect ab effect doesn't reach statistical significance either). Therefore, the same starting point (.60 for XY and .70 for XM and MY) that two sections back led to finding confirmatory empirical evidence of complete mediation and that one section back led to finding confirmatory empirical evidence of partial mediation, has now led to the conclusion that there is no empirical mediation evidence; and the reason for this discrepancy is found, once again, in small sample variations that are within what is probable.

The fact that we are interested in highlighting here is that a same population scenario (.60 for XY and .70 for XM and MY) can generate compatible data with different hypothesis: complete mediation, partial mediation and mediation absence. And what is more alarming is that this can occur if the population values reflect true mediation or if they only reflect a spurious relationship, because the proposed correlation coefficients can be easily found in contexts in which variable M is a mediator variable as well as in contexts in which variable M is a confounding variable.

Conclusions

Kenny and his collaborators' (Baron & Kenny, 1986; Judd & Kenny, 1981; Kenny et al., 1998) proposal is, no doubt, the dominant strategy to analyze statistical mediation within health and social sciences (MacKinnon et al., 2007). In spite of this, not few experts have argued that it is a strategy that holds many limitations.

First, it seems that the mediation demonstration doesn't require the satisfaction of the first of the four conditions from Baron and Kenny's method (significant relationship between X and Y). As we have already argued, under certain circumstances (presence of mediator variables with opposite effects, presence of suppressing variables, temporal distance between X and Y), a mediator variable may be exercising its effect even when no significant relationship between X and Y is found. To consider that a complete effect doesn't need to occur in order for the study of mediation to make sense holds some advantages. On one hand, those who consider that the total effect must be significant won't tackle with the study of mediation in situations in which doing so might be interesting and useful. On the other hand, the fact that searching for mediator variables in absence of an initial total effect makes sense, obligates increasing the effort to justify the hypothesis that are formulated about the relationships that are expected to be found; and strengthening the premises of a confirmatory argument is taking steps in the right direction. Even Kenny (2008) has ended up admitting that the

relationship between X and Y must not necessarily be present for the search of mediator variables to make sense.

Second, Baron and Kenny's (1986) proposal as well as the improvements it has received in the aspects relating to how to assess the statistical significance of the indirect effect suffer from some technical limitations: they lack potency when assessing the indirect effect and, when the null hypothesis which states that the indirect effect is zero is true, the Type I error rate is too high (MacKinnon et al., 2002, 2004; Shrout & Bolger, 2002).

Third, since Baron and Kenny's (1986) argument is of the confirmatory kind, it's important to not forget that the premises (hypothesis) must go before their consequents (the data). When mediation exists, it's expected that the data will behave in a specific way. But the fact that the data behave in a specific way doesn't mean mediation exists: there's always the possibility that the data behavior is due to other reasons different to mediation. Therefore, although in a mediation study it's legitimate to conclude that *empirical evidence compatible* with the mediation hypothesis was found, it's not legitimate to state that the existence of such mediation has been in fact *proven*.

Fourth, since theory must go before data, it's not acceptable to consider a hypothesis open to mediation (without specifying if it's partial or complete mediation) and to then wait and see what the data say in order to decide if the conclusion of total or partial mediation is possible depending on the degree of reduction of the total effect. This way of reasoning is not compatible with a confirmatory argument as proposed by Baron and Kenny (if p , then q). A confirmatory argument requires a clear starting point. In this sense, we agree with James et al. (2006) when stating that the starting hypothesis must specify complete or partial mediation.

All these limitations affect the trust that can be set upon the decisions made when using Kenny's and his collaborators' strategy in order to prove mediation. But the limitation that is probably most alarming is the one we have named *mediation without mediation*. We don't exactly know up to what point we can trust an argument that allows reaching contradictory conclusions starting from the same theoretical scenario, but we are inclined to think that the trust it deserves is scarce. We have demonstrated that small variations in the

data (variations that are perfectly acceptable due to random sampling) can change a mediation conclusion into a non-mediation one, and the other way around. This being when the theoretical scenario is a mediation one as well as when it's not.

As a consequence of all this, our recommendation to those interested in applying the strategy proposed by Baron and Kenny (1986) is that they shouldn't overlook the serious limitations this methodology holds when reaching reliable conclusions.

Several experts, including Kenny and his collaborators, agree when recommending the use of structural equation models (SEM) in order to soften the problems that derive from the sequential adjustment in linear regression models (Baron & Kenny, 1986; Holmbeck, 1997; Hoyle & Kenny, 1999; Judd & Kenny, 1981; Kenny et al., 1998; see also Hoyle & Smith, 1994; Iacobucci et al., 2007). The logic of the analysis doesn't change just because of the use of one type of model or another, but, generally, SEM have some advantages: they allow controlling the measurement error, they offer information on the complete model adjustment degree and they are more flexible than linear regression models (they allow incorporating more than one independent variable, one mediator variable and one dependent variable into the analysis; they allow including variables that act as cause of mediator variables; they can incorporate repeated measures; etc.). However, the goodness of SEM can't make us forget that adjusting regression models and adjusting structural equation models is only useful to evaluate alternative models and to distinguish between the models that offer plausible data explanations and those that don't; the model adjustment isn't useful for identifying *true* models.

In Stone-Romero and Rosopa's opinion (2004, p. 250; see also Spencer et al., 2005), "the mediation model contrasts based on non-experimental studies have little or no capacity to make valid mediation inferences".

All these observations, together with some problems that haven't been pointed out here (see, for example, Ato & Vallejo, 2011, p.554), question the fact that the model proposed by Kenny and his collaborators (Baron & Kenny, 1986; Judd & Kenny, 1981; Kenny et al., 1998) is an appropriate strategy when analyzing statistical mediation.

References

- Ato, M., & Vallejo, G. (2011). Los efectos de terceras variables en la investigación psicológica. *Anales de Psicología*, 27, 550-561.
- 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.
- Barsevick, A. M., Dudley, W. N., & Beck, S. L. (2006). Cancer-related fatigue, depressive symptoms, and functional status. *Nursing research*, 55, 366-372.
- Bollen, K. A., & Stine, R. (1990). Direct and indirect effects: Classical and bootstrap estimates of variability. *Sociological Methodology*, 20, 115-140.
- Collins, L. M., Graham, J. W., & Flaherty, B. P. (1998). An alternative framework for defining mediation. *Multivariate Behavioral Research*, 33, 295-312.
- Fairchild, A. J., & MacKinnon, D. P. (2009). A general model for testing mediation and moderation effects. *Prevention Science*, 10, 87-99.
- Frazier, P. A., Tix, A. P., & Barron, K. E. (2004). Testing moderator and mediator effects in counseling psychology research. *Journal of Counseling Psychology*, 51, 115-134.
- Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science*, 18, 233-239.
- Hofmann, W., Rauch, W., & Gawronski, B. (2007). And deplete us not into temptation: Automatic attitudes, dietary restraint, and self-regulatory resources as determinants of eating behavior. *Journal of Experimental Social Psychology*, 43, 497-504.
- Holmbeck, G. N. (1997). Toward terminological, conceptual and statistical clarity in the study of mediators and moderators: examples from the child-clinical and pediatric psychology literatures. *Journal of Consulting and Clinical Psychology*, 65, 599-610.

- Horcajo, J., Petty, R. E., & Briñol, P. (2010). The effects of majority versus minority source status on persuasion: A self-validation analysis. *Journal of Personality and Social Psychology, 99*, 498-512.
- Hoyle, R. H., & Kenny, D. A. (1999). Sample size, reliability, and tests of statistical mediation. In R. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 195-222). Thousand Oaks, CA: Sage.
- Hoyle, R. H., & Robinson, J. I. (2003). Mediated and moderated effects in social psychological research: Measurement, design, and analysis issues. In C. Sansone, C. Morf, & A. T. Panter (Eds.), *Handbook of methods in social psychology* (pp. 213-233). Thousand Oaks, CA: Sage Publications.
- Hoyle, R. H., & Smith, G. T. (1994). Formulating clinical research hypotheses as structural models: A conceptual overview. *Journal of Consulting and Clinical Psychology, 62*, 429-440.
- Iacobucci, D., Saldanha, N., & Deng, X. (2007). A meditation on mediation: Evidence that structural equations models perform better than regressions. *Journal of Consumer Psychology, 17*, 139-53.
- James, L. R., & Brett, J. M. (1984). Mediators, moderators and tests for mediation. *Journal of Applied Psychology, 69*, 307-321.
- James, L. R., Mulaik, S. A., & Brett, J. M. (1982). *Causal analysis: Assumptions, models and data*. Beverly Hills, CA: Sage.
- James, L. R., Mulaik, S. A., & Brett, J. M. (2006). A tale of two methods. *Organizational Research Methods, 9*, 233-244.
- Jiménez, T. I., Musitu, G., & Murgui, S. (2008). Funcionamiento familiar y consumo de sustancias en adolescentes: el rol mediador de la autoestima. *International Journal of Clinical and Health Psychology, 8*, 139-151.
- Judd, C. M., & Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review, 5*, 602-619.
- Judd, C. M., & Kenny, D. A. (2010). Data analysis in social psychology: Recent and recurring issues. In D. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (5th ed., Vol. 1, pp. 115-139). New York: Wiley.
- Kenny, D. A. (2008). Reflections on mediation. *Organizational Research Methods, 11*, 353-358.
- Kenny, D. A. (2011). *Mediation*. En: <http://davidakenny.net/cm/mediate.htm>.
- Kenny, D. A., Kashy, D., & Bolger, N. (1998). Data analysis in social psychology. In D. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed., Vol. 1, pp. 233-265). New York: McGraw-Hill.
- LeBreton, J. M., Wu, J., & Bing, M. N. (2009). The truth(s) on testing for mediation in the social and organizational sciences. In Ch. L. Lance, & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity and fable in organizational and social sciences* (pp. 107-147). New York: Routledge.
- MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. Mahwah, NJ: Erlbaum.
- MacKinnon, D. P. (2009). Current directions in mediation analysis. *Current Directions in Psychological Science, 18*, 16-20.
- MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology, 58*, 593-614.
- MacKinnon, D. P., Krull, J. L., & Lockwood, C. (2000). Mediation, confounding, and suppression: Different names for the same effect. *Prevention Science, 1*, 173-181.
- MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods, 7*, 83-104.
- MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: distribution of the product and resampling methods. *Multivariate Behavioral Research, 39*, 99-128.
- MacKinnon, D. P., Warsi, G., & Dwyer, J. H. (1995). A simulation study of mediated effect measures. *Multivariate Behavioral Research, 30*, 41-62.
- Mallinckrodt, B., Abraham, W.T., Wei, M., & Russell, D. W. (2006). Advances in testing the statistical significance of mediation effects. *Journal of Counseling Psychology, 53*, 372-378.
- Mathieu, J. E., & Taylor, S. R. (2006). Clarifying conditions and decision points for mediational type inferences in organizational behavior. *Journal of Organizational Behavior, 27*, 1031-1056.
- Pardo, A., Garrido, J., Ruiz, M. A., & San Martín, R. (2007). La interacción entre factores en el análisis de varianza: errores de interpretación. *Psicothema, 19*, 343-349.
- 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.
- Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and non-experimental studies: New procedures and recommendations. *Psychological Methods, 7*, 422-445.
- Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), *Sociological methodology, 1982* (pp. 290-312). Washington, DC: American Sociological Association.
- Sobel, M. E. (1986). Some new results on indirect effects and their standard errors in covariance structure models. *Sociological Methodology, 13*, 290-312.
- Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: Why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology, 89*, 845-851.
- Stone, C. A., & Sobel, M. E. (1990). The robustness of estimates of total indirect effects in covariance structure models estimated by maximum likelihood. *Psychometrika, 55*, 337-352.
- Stone-Romero, E. F., & Rosopa, P. (2004). Inference problems with hierarchical multiple regression-based tests of mediating effects. In J. Martocchio (Ed.), *Research in personnel and human resources management* (Vol. 23, pp. 249-290). Greenwich, CT: Elsevier.
- Stone-Romero, E. F., & Rosopa, P. (2008). The relative validity of inferences about mediation as a function of research design characteristics. *Organizational Research Methods, 11*, 326-352.
- Wood, R. E., Goodman, J. S., Beckmann, N., & Cook, A. (2008). Mediation testing in management research: A review and proposals. *Organizational Research Methods, 11*, 270-295.
- Zhao, X., Lynch, J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research, 37*, 197-210.

(Artículo recibido: 2-11-2011, revisado: 20-07-2012, aceptado: 30-7-2012)

Appendix. Datasets compatible with different mediation hypotheses: (1) subscript 1: *complete mediation* hypothesis; (2) subscript 2: *partial mediation* hypothesis; (3) subscript 3: *non-mediation* hypothesis.

<i>N° case</i>	X_1	M_1	Y_1	X_2	M_2	Y_2	X_3	M_3	Y_3
1	41	40	32	40	41	49	32	41	36
2	35	36	45	36	35	45	33	38	40
3	48	40	54	40	48	54	34	39	40
4	35	45	50	45	35	36	35	35	40
5	54	49	41	49	54	59	36	36	45
6	48	53	45	53	48	45	36	37	45
7	61	57	54	57	61	64	36	38	36
8	54	62	54	62	48	45	37	39	36
9	67	66	68	66	67	68	38	39	31
10	61	62	68	62	61	68	39	41	42
11	41	32	36	32	41	36	40	45	36
12	35	36	41	36	35	40	40	46	40
13	48	40	41	40	48	36	40	48	40
14	41	45	45	45	41	45	41	48	54
15	54	49	50	49	54	40	42	51	54
16	48	53	63	53	48	63	44	35	36
17	41	53	54	53	41	54	45	35	36
18	54	62	50	62	48	49	45	40	45
19	67	66	63	66	67	63	45	41	45
20	61	49	50	49	61	49	45	42	45
21	41	40	32	40	41	31	46	60	40
22	35	36	45	36	35	30	47	61	49
23	48	40	54	40	48	54	48	58	40
24	35	45	36	45	35	36	48	54	49
25	54	49	41	49	54	54	49	54	49
26	48	53	45	53	48	45	49	53	54
27	61	57	54	57	61	54	50	58	59
28	54	62	45	62	54	35	51	61	49
29	67	66	68	66	67	68	52	41	54
30	61	62	68	62	61	68	52	41	54
31	41	36	36	36	41	36	53	42	54
32	35	36	41	36	35	30	53	46	45
33	48	40	41	40	48	40	53	47	45
34	41	45	45	45	41	45	54	48	63
35	54	49	50	49	54	49	55	49	63
36	48	53	63	53	48	63	56	49	63
37	41	53	54	53	41	54	56	57	54
38	54	62	50	62	54	49	57	61	54
39	67	66	63	66	67	63	60	48	45
40	61	45	50	45	61	49	61	50	49
41	41	36	36	36	41	36	62	54	45
42	35	36	41	36	35	40	62	54	49
43	48	40	41	40	48	40	63	54	49
44	41	45	45	45	41	45	64	56	68
45	54	49	50	49	54	49	65	61	68
46	48	53	63	53	48	63	65	63	63
47	41	53	54	53	41	54	66	63	60
48	54	62	50	62	54	49	66	65	63
49	67	66	63	66	67	63	67	66	62
50	61	45	50	45	61	49	69	67	68