

***R* squared effect-size measures and overlap between direct and indirect effect in mediation analysis**

Peter de Heus

Published online: 19 August 2011

© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract In a recent article in this journal (Fairchild, MacKinnon, Taborga & Taylor, 2009), a method was described for computing the variance accounted for by the direct effect and the indirect effect in mediation analysis. However, application of this method leads to counterintuitive results, most notably that in some situations in which the direct effect is much stronger than the indirect effect, the latter appears to explain much more variance than the former. The explanation for this is that the Fairchild et al. method handles the strong interdependence of the direct and indirect effect in a way that assigns all overlap variance to the indirect effect. Two approaches for handling this overlap are discussed, but none of them is without disadvantages.

Keywords Effect-size · Mediation analysis

In the 25 years since the seminal article by Baron & Kenny (1986), mediation analysis has become an indispensable part of the statistical toolkit for researchers in the social sciences. Although this has given rise to an extensive literature (for an overview, see MacKinnon, 2008), measures of effect size in mediation analysis appear to have lagged behind. In a recent article, Fairchild et al. (2009) discussed the limitations of previous effect size measures, and proposed *R* squared effect-size measures as a viable alternative. In the present study, I will show that these *R* squared measures suffer from a serious problem, an

asymmetric treatment of the direct and indirect effect that can not be justified on normative grounds. I will indicate the source of this problem (assigning all overlap between direct and indirect effect to the indirect effect), and provide some solutions that may provide some help to researchers in deciding how to handle this interdependence in variance explained by the direct and indirect effect.

R squared measures for the direct and indirect effect

In order to keep the conceptual points of the present article as simple and clear as possible, I will limit myself to the three-variable case with one dependent variable (*Y*), one independent variable (*X*), and one mediator (*M*). Also for simplicity (because it removes standard deviations from all formulas), my description will be in terms of standardized regression weights (betas).

The total, the direct, and the indirect effect

The *total effect* (β_{tot}) of *X* on *Y* is given by β_{YX} , the regression weight in a regression analysis in which *Y* is predicted by *X* only. The *direct effect* (β_{dir}) of *X* on *Y* is given by $\beta_{YX.M}$, the beta of *X* in a regression analysis in which *Y* is predicted by both *X* and *M*. The *indirect effect* (β_{ind}) of *X* on *Y* via *M* is given by the product $\beta_{MX} \beta_{YM.X}$, in which β_{MX} is the beta in a regression in which *M* is predicted by *X* only, while $\beta_{YM.X}$ is the beta of *M* in a regression analysis in which *Y* is predicted by both *X* and *M* (Fig. 1).

Following these definitions and using standard equations from ordinary least squares regression analysis (e.g. Cohen, Cohen, West and Aiken, 2003), the total,

P. de Heus (✉)
Department of Psychology, University of Leiden,
Wassenaarseweg 52,
2333 AK Leiden, the Netherlands
e-mail: deheus1@fsw.leidenuniv.nl

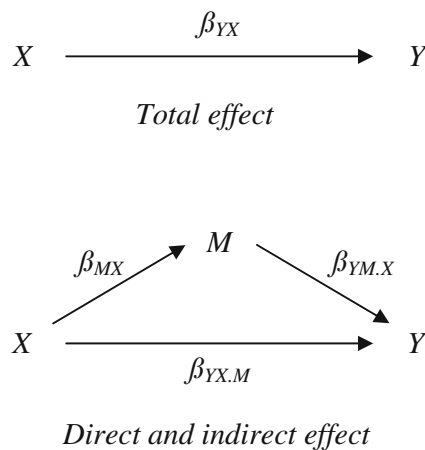


Fig. 1 The total, the direct, and the indirect effect

direct, and indirect effect can be computed from the intercorrelations of Y , X , and M as follows.

$$\beta_{tot} = \beta_{YX} = r_{YX} \quad (1a)$$

$$\beta_{dir} = \beta_{YX.M} = \frac{r_{YX} - r_{MX}r_{YM}}{1 - r_{MX}^2} \quad (1b)$$

$$\beta_{ind} = \beta_{MX}\beta_{YM.X} = \frac{r_{MX}(r_{YM} - r_{MX}r_{YX})}{1 - r_{MX}^2} \quad (1c)$$

As described in almost all texts and articles on mediation analysis (e.g. MacKinnon, 2008), and as can be verified by adding the right terms of Eqs. 1b and 1c, the total effect is the sum of the direct effect and the indirect effect.

$$\beta_{tot} = \beta_{dir} + \beta_{ind} \quad (2)$$

Effect size

Perhaps the most commonly used measure of *effect size* in mediation analysis is the *proportion mediated* (PM : Alwin and Hauser, 1975; MacKinnon and Dwyer, 1993), which simply gives the (direct or indirect) effect as a proportion of the total effect: $PM_{dir} = \beta_{dir} / \beta_{tot}$ and $PM_{ind} = \beta_{ind} / \beta_{tot}$. As demonstrated by simulation studies (MacKinnon, Warsi and Dwyer, 1995), the proportion mediated suffers from instability and bias in small samples, so it needs large samples ($N > 500$) to perform well. In the light of these limitations, investigating alternative effect size measures such as R squared measures seems to be a worthwhile effort (for a more complete overview of effect size measures in mediation, see Preacher and Kelley, 2011).

Given the previous definitions of the total, the direct, and the indirect effect, how much variance is explained by

each? The total amount of Y variance explained by X (i.e. the variance explained by the total effect), R_{tot}^2 , is equal to r_{YX}^2 , the squared correlation between Y and X .

Everything described until this point is generally accepted and completely uncontroversial. What's new in Fairchild et al. (2009) is a procedure to decompose this total amount of variance explained into two parts, one for the direct effect and one for the indirect effect (Fig. 2). According to these authors, R_{dir}^2 , the variance explained by the direct effect, corresponds to the Y variance that is explained by X but not by M , whereas R_{ind}^2 , the variance explained by the indirect effect, corresponds to the variance that is shared by Y , X and M together. Defined this way, the variance explained by the direct effect corresponds to the squared semipartial correlation $r_{Y(X.M)}^2$, and the variance explained by the indirect effect is the difference between total and direct effect variance explained. Of these three measures, the first two can only be positive or zero, but as noted by Fairchild et al. (2009), it is perfectly possible that the third, R_{ind}^2 , is negative.¹

$$R_{tot}^2 = r_{YX}^2 \quad (3a)$$

$$R_{dir}^2 = r_{Y(X.M)}^2 = \left(\frac{r_{YX} - r_{YM}r_{MX}}{\sqrt{1 - r_{MX}^2}} \right)^2 \quad (3b)$$

$$R_{ind}^2 = R_{tot}^2 - R_{dir}^2 = r_{YX}^2 - r_{Y(X.M)}^2 \quad (3c)$$

At first sight, this way of decomposing the total variance explained into a direct and an indirect part has very attractive features in comparison to proportion mediated. First, thinking in terms of variance explained is very common among social scientists (so common, actually, that if one uses the proportion mediated instead, one very often has to explain to one's colleagues that this proportion does *not* refer to variance explained). Second, the Fairchild et al. decomposition of variance explained allows for easy computation, using only well-known statistical concepts. Third, as shown by simulation studies, the proposed measures have nice statistical properties like stability and acceptable bias levels (Fairchild et al., 2009). Fourth, and

¹ Actually, Fairchild et al. (2009) used *commonality analysis* (Mood, 1969; Seibold and McPhee, 1979) for the decomposition of total effect variance. This has the advantage of allowing generalization to cases with two or more independent variables and/or mediators, but is not necessary for the more conceptual discussion of the single predictor - single mediator case to which the present article is limited, so I will not follow them in this regard.

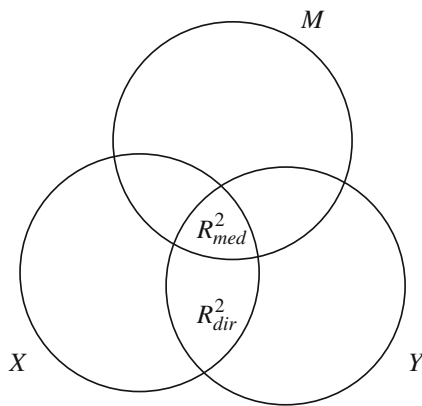


Fig. 2 Partitioning of total variance explained over direct, and indirect effect according to Fairchild et al. (2009)

perhaps most important, decomposing Y variance in this way appears to be intuitively clear. The conceptualization of variance explained by the direct effect is completely in accordance with how according to all textbooks (e.g. Cohen et al., 2003) we should compute the unique variance explained by a predictor in regression analysis. In addition, the idea that variance explained by the indirect effect is the variance shared by all three variables together (predictor, mediator, and dependent variable) also comes as a very natural one. If the part of the common variance of X and Y that is *not* shared by M refers to the direct effect, what could be more natural than assuming that the remaining part of the common variance of X and Y , the part that is also shared by M , represents the indirect effect?

In a recent review of effect size measures in mediation, Preacher and Kelley (2011, p. 100) concluded that the Fairchild et al. measure "... has many of the characteristics of a good effect size measure: (a) It increases as the indirect effect approaches the total effect and so conveys information useful in judging practical importance; (b) it does not depend on sample size; and (c) it is possible to form a confidence interval for the population value." On the negative side, these authors noted that because of the possibility of negative values, the R squared measure is not technically a proportion of variance, which limits its usefulness. The final verdict was that "... it may have heuristic value in certain situations" (Preacher and Kelley, 2011, p. 100). In the next section, I will address a completely different problem with the R squared measure, which is not dependent on negative values (although it may be aggravated by them).

Problems

Problems started with a counterintuitive example, that proved to be a special case of a more general problem.

A counterintuitive example

Applying the R squared effect-size to an example in a course that I was teaching led to very counterintuitive results. In this course example ($N = 85$) about the effect of social support at work (X) on depression (Y) with active coping (M) as mediator, zero-order correlations were $r_{YX} = -.336$, $r_{MX} = .345$, and $r_{YM} = -.391$, leading to the following standardized estimates for the total, direct and indirect effects: $\beta_{tot} = -.336$; $\beta_{dir} = -.228$; $\beta_{ind} = -.108$. In terms of proportion mediated, the direct effect ($PM_{dir} = .679$) was more than twice as strong as the indirect effect ($PM_{ind} = .321$). Given such a large difference in favor of the direct effect, one would definitely expect that the direct effect would explain more variance of the dependent variable than the indirect effect. However, computing the R squared measures from Fairchild et al. (2009) led to $R_{dir}^2 = .046$ and $R_{ind}^2 = .067$, so in terms of variance explained the indirect effect appeared to be much *stronger* than the direct effect. This reversal of effect size from one effect being twice as large than the other in terms of proportion mediated, but much smaller in terms of R squared effect-size is by no means trivial. At the time I could only say to my students that I did not understand this discrepancy between proportion mediated and proportion of variance explained, but that I would try to find out. The present article is a direct consequence of that promise.

Asymmetry of direct and indirect effect

A first step in clarification is to be more specific about what we should expect from an R squared effect size measure. To my opinion, one should expect an R squared measure to be *symmetric* in the sense that a direct and indirect effect of the same magnitude (β_c) should lead to the same amount of variance explained. After all, if a one standard deviation gain in X leads to an average gain of β_c standard deviations in Y , there is no reason why it should make a difference for variance explained whether this gain is due to the direct or to the indirect effect. Since the change in Y as a function of X is identical in both cases, and since change in Y is all that matters (or should matter) for Y variance explained, one would expect amount of variance explained to be identical.

If this reasoning is correct, one would expect that exchanging the values for the direct and indirect effect would lead to a simple exchange of the R square change measures. In order to check this, the course example was modified. By keeping r_{YX} and r_{MX} at $-.336$ and $.345$ respectively, but changing r_{YM} to $-.697$, a reversal of the original values for the direct and indirect effect was accomplished with β_{dir} and β_{ind} equal to $-.108$ and $-.228$

respectively. However, the R squared measures in this modified example were *not* the reverse of those in the previous analysis (.046 and .067). Instead, the actual values were $R_{dir}^2 = .010$ and $R_{ind}^2 = .103$, so whereas in terms of proportion mediated the indirect effect is now about twice as strong as the direct effect, in terms of variance explained it is more than ten times as strong. In a last example, the direct and indirect effect were made identical. Keeping r_{YX} and r_{MX} at $-.336$ and $.345$ respectively, but changing r_{YM} to $-.545$, led to identical betas for the direct and indirect effect, $\beta_{dir} = \beta_{ind} = -.168$, but the obtained R squared change measures were $R_{dir}^2 = .025$ and $R_{ind}^2 = .088$. At least in these three examples, the general pattern seems to be that the R squared measures by Fairchild et al. (2009) are not symmetric, but strongly biased in favor of the indirect effect.

In hindsight, even without these examples we might have known that the R squared measures for the direct and indirect effect can not always be symmetric, because as mentioned by Fairchild et al. (2009), R_{dir}^2 will always be positive or zero, whereas R_{ind}^2 may also be negative. When R_{ind}^2 is negative, exchanging the betas for the direct and indirect effect can never lead to exchanging their R squared measures, because R_{dir}^2 can not become negative. The general question is how it is possible that identical direct and indirect effects, with identical effects on Y , can be completely different in terms of R squared effect size?

Explanation and solutions

Direct and indirect effect are not independent

A look at the equations and illustrations in the Fairchild et al. (2009) article leaves one with the impression that total variance explained can unequivocally be divided over the direct and the indirect effect, as if those two effects are completely independent. As will be argued below, this is not true. The key to understanding the divergent results for proportion mediated versus variance explained measures and for the asymmetry of the direct and indirect effect, is that the direct and indirect effect are heavily interdependent.

Although the mediator M is crucial for *estimation* of the direct and indirect effect, once the relevant path coefficients have been estimated, both direct and indirect effect are a function of the independent variable X only. For computing variance explained, the situation is the same as if we had a single predictor with a single regression weight that for some reason can be split into two parts. For example, if we predict income (Y) from the number of hours worked (X) for a group of laborers with the same hourly wages and the same hourly bonus, we are formally in the same situation as

with the direct and indirect effect. Here too we have an overall effect that is just the sum of two separate effects (regression weights) for the same independent variable: $\beta_{tot} = \beta_{wages} + \beta_{bonus}$. If we try to answer the question how much variance of income is explained by wages and how much by bonus, formally our problem is exactly the same as when we try to predict how much of the variance of the total effect is due to the direct effect and how much to the indirect effect.

The fact that (once β_{dir} and β_{ind} have been estimated) the direct and indirect effect both are a function of the same predictor and nothing else, makes them heavily interdependent. In fact, they are perfectly correlated, because each individual's predicted gain due to the indirect effect will always be $\beta_{ind} / \beta_{dir}$ times his or her gain due to the direct effect. The consequences of this become clear when we write the total proportion of variance explained as a function of the direct and indirect effect:

$$R_{tot}^2 = \beta_{tot}^2 = (\beta_{dir} + \beta_{ind})^2 = \beta_{dir}^2 + \beta_{ind}^2 + 2\beta_{dir}\beta_{ind} \quad (4)$$

As can be seen in eq. 4, total variance explained is now divided into three parts, of which the first two can be unequivocally related to the direct (β_{dir}^2) or indirect (β_{ind}^2) effect, but the third ($2\beta_{dir}\beta_{ind}$) is a *joint part*, for which ascription to either direct or indirect effect is not straightforward. This joint part may be positive or negative, depending on whether β_{dir} and β_{ind} are of same or opposite sign.

Two ways of dividing variance over direct and indirect effect

Given these three variance components, there are at least two defensible ways to divide variance explained over the direct and indirect effect, but neither of them is without disadvantages.

Unique approach Perhaps the most “natural” solution is just taking β_{dir}^2 and β_{ind}^2 as variance explained for the direct and indirect effect respectively.

$$R_{dir(U)}^2 = \beta_{dir}^2 \quad (5a)$$

$$R_{ind(U)}^2 = \beta_{ind}^2 \quad (5b)$$

These squared betas refer to the proportion of variance explained by each effect if it were completely on its own (i.e. if the other effect were zero). This unique approach is roughly comparable with the Type III sum of squares approach of partitioning variance (also known as the unique or regression approach: Stevens, 2007) in ANOVA. In our unmodified example, this leads to

proportions of variance explained of $(-.228)^2 = .052$ and $(-.108)^2 = .012$ for the direct and indirect effect respectively, while $2 \times (-.228) \times (-.108) = .049$ is not uniquely explained.

An advantage of this unique approach is that it has a clear causal interpretation: The proportion of variance explained by the effect if the other effect did not exist or was blocked somehow. Furthermore, computing variances explained in this way will never lead to negative variances or to order reversals between proportion mediated and variance explained as in the examples, and more generally, will not be asymmetric in its treatment of the direct and the indirect effect. Disadvantage is that this way of computing variance completely ignores the joint part $2\beta_{dir}\beta_{ind}$, which may add or subtract a large amount of variance to or from the total effect.²

Hierarchical approach If we have reason to view one effect (which I will call the primary effect) as more fundamental or more important than the other (the secondary effect), we could use a hierarchical approach. For the primary effect we use the variance explained by the effect on its own as in the unique approach, and for the secondary effect we use the additional variance explained over and above what was explained by the primary effect.

$$R^2_{primary(H)} = \beta^2_{primary} \quad (6a)$$

$$R^2_{secondary(H)} = \beta^2_{secondary} + 2\beta_{primary}\beta_{secondary} \quad (6b)$$

Because the secondary effect gets the joint (overlap) variance, its variance explained may be negative when direct and indirect effect are of opposite sign. This hierarchical approach is roughly comparable with the Type I sum of squares approach (also called the hierarchical approach) of partitioning variance in ANOVA (Stevens, 2007). If we take the direct effect as primary in our unmodified example, we get a slightly different version of our original, counterintuitive results: .052 and .061 for the direct and indirect effect respectively. If we take the indirect effect as primary, we get very different values: .101 (direct) and .012 (indirect).

The main advantages of the hierarchical approach are that all variance explained is neatly and uniquely divided over the direct and the indirect effect, and that both effects have a clear interpretation, namely variance explained by the effect if it were on its own (primary), and additional variance explained over and above what was explained by the primary effect (secondary). However, these advantages

only apply when we have good reasons for deciding which effect should be treated as primary, and which as secondary. An example of such good reasons is if one of the effects is the intended one and stronger than the other effect, which is supposed to represent only relatively minor side effects. One might think that the direct effect is more suitable for the role of primary effect, but this is not necessarily so. For example, if a psychotherapy is intended to reduce psychological complaints by encouraging self-efficacy, it makes sense to treat the indirect effect (from therapy via self-efficacy to psychological complaints) as primary, and the direct effect (which in addition to a possible “real” direct effect also represents all indirect effects for which the possible mediators have not been measured) as secondary. Furthermore, as will be argued in next section, when the direct and indirect effect are of opposite sign, it makes sense to treat the effect with the largest absolute size as primary. The main disadvantage of the hierarchical approach is that often there are no decisive theoretical or statistical arguments for deciding which effect is the primary one.

Negative variance explained

An often noted problem in texts on regression analysis (e.g. Cohen et al., 2003) in regard to graphical displays of variance explained like Fig. 1, is that the value of the overlap area of all three variables can be negative, whereas both area and variance are squared entities, which should always be equal or larger than zero. Although this possibility of negative variance/area is not a problem for regression analysis, since the reasons for it are well-understood, it ruins the area-represents-variance metaphor that gives these kinds of figures their heuristic value.

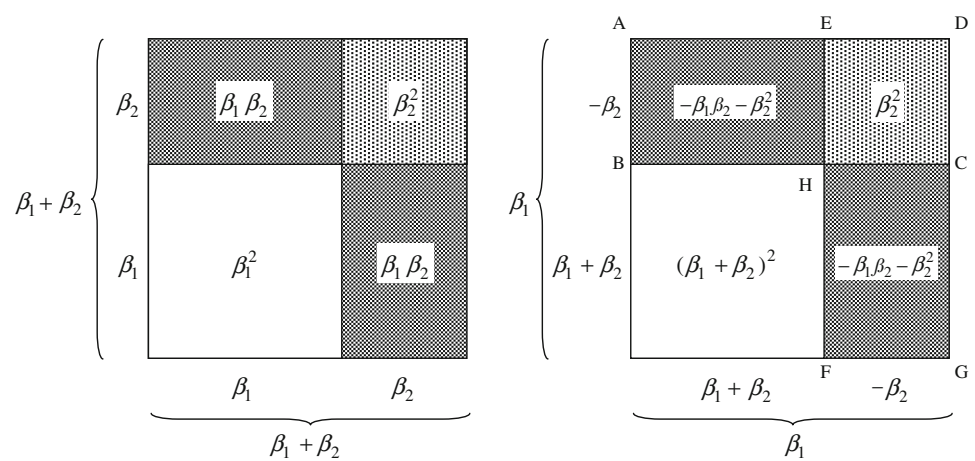
Here I will present a different graphical approach, that may provide a more helpful illustration of negative variance explained, using squares instead of circles and taking the signs of the two betas on which the total effect is based into account (Fig. 3). In what follows, the two betas will just be called β_1 and β_2 , because for the present purpose it does not matter which one represents the direct or the indirect effect. There are two interesting cases here, dependent on whether the two betas have the same sign (case 1) or different signs (case 2).³ Only in case 2 will we be confronted with negative variance explained.

Case 1 (same signs) Because total variance explained is equal to $\beta^2_{tot} = (\beta_1 + \beta_2)^2$, it is given by the area of the large square with sides $\beta_1 + \beta_2$ in Fig. 3a. The two squares

² The joint part $2\beta_{dir}\beta_{ind}$ can take values between $-(\beta^2_{dir} + \beta^2_{ind})$ (if $\beta_{dir} = -\beta_{ind}$), and $\beta^2_{dir} + \beta^2_{ind}$ (if $\beta_{dir} = \beta_{ind}$).

³ Of course, there is also the possibility that one or both betas are zero, but about this I have nothing of interest to say.

Fig. 3 An alternative picture of the partitioning of total variance explained over the direct and indirect effect (and their overlap), without negative area



Case 1. β_1 and β_2 have same sign (here both positive)

Case 2. β_1 and β_2 have different signs (here β_1 positive and $|\beta_1| > |\beta_2|$)

in the lower left and upper right of the large square, with areas β_1^2 and β_2^2 respectively, give the variance explained for effect 1 and 2 for the (here clearly hypothetical) situation in which the other effect were zero. The two rectangles in the upper left and lower right within the large square, both with area $\beta_1\beta_2$, together represent the overlap part $2\beta_1\beta_2$, which we can assign to neither effect (unique approach), or to the secondary effect only (hierarchical approach).

Case 2 (different signs) If direct and indirect effect have opposite signs, the effect with largest absolute value will always explain more variance than the total effect. This is illustrated in Fig. 3b, which looks very much like Fig. 3a, but now the large square has area β_1^2 and refers to the variance explained by the largest of the two effects. The variance explained by the total effect is represented by the smaller square with area $(\beta_1 + \beta_2)^2$ in the lower left within the large square. Here a hierarchical interpretation with the largest effect as primary is the most natural way to describe what happens to variance explained. It goes like this. The proportion of variance explained by effect 1 on its own (i.e. if β_2 were zero) would be β_1^2 (the large square), but because effect 2 makes β_{tot} smaller than β_1 , total variance explained is represented by the smaller white square within the large square. One way of describing how to get the area of this small square from the area of the large square is that we have to subtract from the area of the large square (β_1^2) the areas of the two rectangles ABCD and DEFG (both $-\beta_1\beta_2$), but because the square CDEH is part of both rectangles, its area (β_2^2) is subtracted twice, so to compensate for that, we have to add this area once. This leads to the subtraction $\beta_1^2 - (-2\beta_1\beta_2 - \beta_2^2) = \beta_1^2 + 2\beta_1\beta_2 + \beta_2^2$, which neatly equals β_{tot}^2 . In this graphical interpretation of variance explained, there is never negative area, but only *subtraction of one positive area from another*.

Incidentally, this interpretation only works well if we take the largest effect as primary in our hierarchical approach. Algebraically, it makes no difference whether we start from β_1^2 or from β_2^2 , but geometrically, I do not see an intuitively revealing picture that clarifies how to get from the small square with area β_2^2 to the larger square with area $(\beta_1 + \beta_2)^2$. Putting into words how exactly to get from so much variance explained by one effect to even more variance explained by a total effect that is in the opposite direction, is not easy either. Therefore, if the direct and indirect effect are of opposite sign, the most intuitively appealing approach is hierarchical, with the largest effect as primary.

Comparison with the Fairchild et al. approach

The method presented by Fairchild et al. (2009) can be seen as a slightly modified version of the hierarchical approach in the present article. Their R squared measure for the direct effect, the squared semipartial correlation (Eq. 3b), is closely related to the squared beta for the direct effect (which is used in the hierarchical approach when the direct effect is taken as primary), because it follows from Eqs. (1b) and (3b) that:

$$r_{Y(X.M)}^2 = (1 - r_{XM}^2)\beta_{dir}^2 \quad (7)$$

The R squared measure for the indirect effect gets all that remains of the total variance explained after removing direct effect variance. This means that in the Fairchild et al. approach the direct effect is taken as primary and the indirect effect as secondary. As a consequence, the indirect effect always gets the joint part $2\beta_{dir}\beta_{ind}$, so it will be favored in comparison to the direct effect when this joint part is positive, which is when both effects have the same

sign. When the joint part is negative, the indirect effect will be disfavored, possibly leading to negative variance explained. These are the reasons why in our examples (in which both effects always had the same sign) the direct effect was much lower than expected. It also explains why the indirect effect, but not the direct effect can become negative.

Modifying the Fairchild et al. approach

As argued previously, the problem in the Fairchild et al. approach is not in the direct effect measure (the squared semipartial correlation), which taken on its own makes perfect sense, but in the asymmetry between direct and indirect effect measures. Therefore, instead of the previous decomposition of equation (4) into different parts, an alternative approach is retaining the Fairchild et al. direct effect measure, but changing the indirect effect measure in a way that makes it symmetrical with the direct effect measure.

This alternative approach goes in two steps. First, the Y variance that is uniquely explained by M is given by the squared semipartial correlation $r_{Y(M.X)}^2$. However, not all this variance can be attributed to the indirect effect, because M is only partially explained by X . Second, multiplying $r_{Y(M.X)}^2$, the proportion of Y variance uniquely explained by M , with r_{MX}^2 , the proportion of M variance explained by X , gives the proportion of Y variance that is uniquely explained by X via M .

$$R_{ind}^2 = r_{Y(M.X)}^2 r_{MX}^2 \quad (8)$$

As a product of two squared numbers, R_{ind}^2 will never be negative. And because $r_{MX}^2 = \beta_{MX}^2$, and because it follows from exchanging X and M in eq. 7 that $r_{Y(M.X)}^2 = (1 - r_{MX}^2) \beta_{YM.X}^2$, eq. 8 can be rewritten as follows.

$$R_{ind}^2 = (1 - r_{MX}^2) \beta_{YM.X}^2 \beta_{MX}^2 = (1 - r_{MX}^2) \beta_{ind}^2 \quad (9)$$

Comparing eqs. 7 and 9 reveals that now the direct and indirect effect are treated in a completely symmetrical way. For both effects, the unique proportion of variance explained is computed by multiplying the squared beta with $(1 - r_{MX}^2)$, so identical betas for the direct and indirect effect lead to identical R squared measures of variance explained.

As in the approach based on squared betas, there is a joint part of variance explained, which now equals $2\beta_{dir}\beta_{ind} + r_{MX}^2(\beta_{dir}^2 + \beta_{ind}^2)$. In comparison to the squared beta approach, this joint part is more positive if the two effects are of the same sign, and less negative if they are of opposite signs (unless $r_{MX} = 0$, in which case it makes no difference). As before, we can use the unique or the hierarchical approach for assigning the three components

of variance explained to the direct and the indirect effect, with the same advantages and disadvantages as discussed before. The difference is that in the adjusted Fairchild et al. (2009) approach, the unique contribution of an effect is corrected for the other effect *as it is* in the present data, whereas in the approach based on squared betas the unique contribution of each effect is computed for the more hypothetical situation if the other effect would be zero.

After having derived this measure, I found out that although not described in Fairchild et al. (2009), a closely related measure was mentioned in MacKinnon (2008, p. 84, equation 4.6) as one of three measures requiring more development. The only difference is that instead of a semipartial correlation with X partialled out from M only, as in the present article, MacKinnon (2008) used a *partial correlation* of Y and M with X partialled out from *both* Y and M , leading to $R_{ind}^2 = r_{YM.X}^2 r_{MX}^2$. The effect on interpretation is that we are addressing the same variance in the two measures, but as a *proportion of a different total* in each, namely all Y variance (semipartial) versus Y variance not shared by X (partial). Both measures are equally valid as long as we are clear about from which total we take a proportion, but to my opinion, using the semipartial is preferable over the partial, because it is easier to understand, measures the direct and indirect effect as proportions of the same total instead of different totals, and is completely symmetric in its treatment of the direct and indirect effect.

Discussion and conclusions

Limitations

Three possible limitations of the present study should be discussed. First, one might wonder whether the restriction of discussing R squared effect sizes only in relation to *standardized* effects (betas instead of b 's) does limit the generality of the results. It does not, because R squared effect size measures are standardized themselves, completely based on correlations, which are covariances between standardized variables. We go from standardized to unstandardized effects by multiplying the betas for the total, direct, and indirect effect all with the same constant ($SD(X) / SD(Y)$), and apart from that, nothing changes.

Second, nothing was assumed or said about the *distributions* of X , M , and Y . A minor point here is that differences between the distributions of the three variables (e.g. when X is strongly skewed, but Y is not) may limit the maximum size of correlations, and therefore of R squared effect size measures. Apart from this minor point, it can be argued that for the purpose of the present article,

distributional assumptions were not necessary, because no attempt was made to estimate stability and bias of the effect size measures.

This absence of simulation studies in order to establish stability and bias of the R_{ind}^2 measure of eq. (8) is the third and most serious limitation of the present study. At present, the only information I can give comes from MacKinnon (2008, p. 84), who for his closely related measure (identical except for the use of partial instead of semipartial correlations) reported that in an unpublished masters thesis (Taborga 2000) minimal bias, even in relatively small samples, was found. Whether this generalizes to the present measure is a matter for future research.

Conclusion and recommendations

The most important conclusion of the present study is that in terms of variance explained there is strong overlap between the direct and indirect effect. In order to handle and quantify this overlap, a method has been presented to decompose variance explained into three parts (unique direct, unique indirect, and joint part). Because of this strong overlap, dividing variance explained over the direct and indirect effect is only possible if we make some choice about what to do with the joint part.

In a way, there is nothing new here. Handling such overlap is a routine matter in ANOVA for unbalanced designs and regression analysis with correlated predictors, and the unique and hierarchical approaches suggested in the present article are closely parallel to the most commonly used ways of handling overlap in regression and ANOVA. However, due to the interconnectedness of the direct and indirect effect, the amount of overlap is much larger than what we usually encounter in regression and ANOVA, so different choices of how to handle overlap may lead to radically different effect sizes (as was illustrated by the examples in the present study). Unfortunately, choices of how to handle such overlap are always arbitrary to some extent.

Is it possible to give some useful advice to the applied researcher? Due to the absence of knowledge of possible bias of the R_{ind}^2 effect size measure and the impossibility to eliminate all arbitrariness from the decision about what to do with the overlap part of variance explained, very specific guidelines are impossible. However, some general recommendations for using R squared effect size measures can be given.

1. Generally, the approach based on squared semipartial correlations should be preferred above the approach based on squared betas. In the present article, squared betas were useful for explaining the interconnectedness of the direct and indirect effect, but as measures of unique

variance explained they are too hypothetical for almost all research situations in the social sciences.⁴ Measures based on squared semipartial correlations are much more descriptive of the actual data.

2. If there are good reasons to indicate one effect as primary, or if the two effects are of opposite sign, use the hierarchical approach. In many situations, it is very natural to ask for the unique contribution of the primary effect, and then for the additional contribution of the secondary effect. If we have good reasons to choose the direct effect as the primary one, the original Fairchild et al. approach still is the thing to do.

3. If there are no good reasons for a primary versus secondary effect distinction, use the unique approach. The price to be paid is that a lot of overlap variance remains unexplained, but this is not too different from what we routinely accept when doing ANOVA or multiple regression analysis.

4. Whatever we do, in terms of R^2 and variance explained, there will always be substantial overlap between our effects and some arbitrariness in our handling of this overlap. If we accept this, we can make our decisions as described in the three points above. Alternatively, we might conclude that R^2 measures are not the best possible way to describe effect size in mediation, and consider other effect size measures, that possibly suffer less from overlap between effects. We may reconsider proportion mediated as a measure of effect size, or we could simply use β_{ind} (eq. 1c, discussed as the *completely standardized indirect effect* by Preacher and Kelley (2011)), or we could try one of the other measures discussed by these authors. At present, I see no measure which is satisfactory under all circumstances, so the quest for the perfect measure of effect size in mediation should go on.

⁴ In principle, squared betas can be useful if we want to predict how much variance would be explained by an effect if the other effect were blocked somehow. It is possible to imagine research situations in which this question has some use. For example, some experimental manipulation (X) may have both a direct effect on general mood (Y) and an indirect effect by inducing fear (M), which in turn influences general mood. Now if we could modify the experimental manipulation in a way that eliminates its effect on fear, without changing its direct effect on mood in any way, the squared beta of the direct effect in our original experiment would predict the proportion of variance explained in that new situation. However, even in this artificial situation, computing variance explained for a new experiment with the modified manipulation would be strongly preferable over assuming its value via squared betas, because it is an empirical question whether the direct and indirect effect of the modified manipulation will behave as expected by the investigator.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Alwin, D. F., & Hauser, R. M. (1975). The decomposition of effects in path analysis. *American Sociological Review*, *40*, 37–47.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.
- Fairchild, A. J., MacKinnon, D. P., Taborga, M. P., & Taylor, A. B. (2009). R^2 effect-size measures for mediation analysis. *Behavior Research Methods*, *41*, 486–498.
- MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York: Erlbaum.
- MacKinnon, D. P., & Dwyer, J. H. (1993). Estimating mediated effects in prevention studies. *Evaluation Review*, *17*, 144–158.
- MacKinnon, D. P., Warsi, G., & Dwyer, J. H. (1995). A simulation study of mediated effect measures. *Multivariate Behavioral Research*, *30*, 41–62.
- Mood, A. M. (1969). Macro-analysis of the American educational system. *Operations Research*, *17*, 770–784.
- Preacher, K. J., & Kelley, K. (2011). Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. *Psychological Methods*, *16*, 93–115.
- Seibold, D. R., & McPhee, R. D. (1979). Commonality analysis: A method for decomposing explained variance in multiple regression analyses. *Human Communication Research*, *5*, 355–365.
- Stevens, J. P. (2007). *Intermediate statistics. A modern approach* (3rd ed.). New York: Erlbaum.
- Taborga, M. P. (2000). *Effect size in mediation models*. Arizona State University, Tempe: Unpublished master's thesis.