Journal of Experimental Social Psychology 75 (2018) 95–102

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

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

Unwarranted inferences from statistical mediation tests – An analysis of articles published in 2015☆

Klaus Fiedler*, Chris Harris, Malte Schott

University of Heidelberg, Germany

A B S T R A C T

Recent attempts to improve on the quality of psychological research focus on good practices required for statistical significance testing. The scrutiny of theoretical reasoning, though superordinate, is largely neglected, as exemplified here in a common misunderstanding of mediation analysis. Although a test of a mediation model X ➔ Z ➔ Y is conditional on the premise that the model applies, alternative mediators Z′, Z″, Z‴ etc. remain untested, and other causal models could underlie the correlation between X, Y, Z, researchers infer from a single significant mediation test that they have identified the true mediator. A literature search of all mediation analyses published in 2015 in Sciencedirect shows that the vast majority of studies neither consider alternative causal models nor alternative mediator candidates. Ignoring that mediation analysis is conditional on the truth of the focal mediation model, they pretend to have demonstrated that Z mediates the influence of X on Y. Recommendations are provided for how to overcome this dissatisfying state of affairs.

1. Introduction

A growing number of recent publications are driven by the laudable motive to improve the scrutiny of psychological science. How can we foster solid research findings that are reliable and replicable at the empirical level and well understood at the theoretical level? A glance at the pertinent literature shows that the suggested interventions and the implemented changes in the publication process focus on data sharing, good practices in documentation, and appropriate significance testing. Accordingly, the key to improved science seems to lie in stricter compliance rules for data management and in still more weight given to proper significance testing. There is a conspicuous paucity of discourse on strict theorizing and logic of science (Fiedler, 2017).

In this article, we emphasize the need for proper theorizing and the priority of theoretical reasoning as a major precondition of good science. Research design beats statistical testing, and theoretical reasoning beats research design. Even sophisticated statistical testing is worth nothing if the underlying research design is flawed. And the cleverest and most refined design is useless when applied to a logically inappropriate or undiagnostic hypothesis.

While there are many ways to substantiate this point, the present article concentrates on one issue, namely, the reliance on mediation analysis when drawing scientific inferences. Testing mediation models has become a gold standard for research submitted to prominent journals (Bullock, Green, & Ha, 2010; MacKinnon, Fairchild, & Fritz, 2007). It is supposed to enable rigorous process accounts of otherwise unexciting findings and to allow for causal inferences about what crucial factor mediates the influence of an independent on a dependent variable.

The present article is neither meant to deny the scientific potential of mediation analysis nor to criticize the pertinent statistical methods (cf. Hayes, 2009; MacKinnon, 2008). We are simply concerned with the scientific status of theoretical inferences informed by mediation analysis. Drawing on a universe of 102 articles (126 mediation analyses) solicited by the keyword “mediation analysis” in the internet platform Sciencedirect,[1] we demonstrate that the vast majority of theoretical inferences drawn from such mediation tests are logically unwarranted. Even when state-of-the art statistical procedures for mediation analysis are applied to well designed and carefully conducted experiments, often published in high-ranking journals, most theoretical inferences and practical take-home messages are misleading and logically incorrect.

☆ The work underlying the present article was supported by a grant provided by the Deutsche Forschungsgemeinschaft to the first author (FI 294/23-1).

* Corresponding author.
E-mail address: kf@psychologie.uni-heidelberg.de (K. Fiedler).

[1] The first author's experience as an Associate Editor of this journal, which is covered on the Science Direct platform, originally motivated this critical perspective on mediation analysis (cf. Fiedler, Schott, & Meiser, 2011).

https://doi.org/10.1016/j.jesp.2017.11.008
Received 31 January 2017; Received in revised form 25 September 2017; Accepted 16 November 2017

1.1. What inferences can(not) be informed by mediation tests

Let us illustrate the problem with a thought experiment. Imagine an epidemiologist has found that the emergence of a virus (X) is statistically related to the observation of a disease (Y). The epidemiologist holds a biologically well-founded theory that infection is transmitted by sexual contact (Z); sexual contact is the means by which the virus can infect other people. Indeed, when Z is entered as a third variable in a mediation test, a substantial part of the covariance shared by X and Y is explained by the model X ➔ Z ➔ Y. In this case, a statistical mediation test actually substantiates a causal model, which is reasonable on theoretical grounds.

However, now suppose that the correlations between X, Y, and Z are exactly the same but Z is fever (a symptom of Y) rather than sexual contact (a reasonable infection mechanism). We know, theoretically, that fever is not a mediator, but Sobel tests, regression analysis, or a bootstrapping algorithms do not have causal world knowledge; they are only sensitive to the tri-variate data array but not to the causal surplus meaning of a symptom (fever) versus an infection mechanism (sexual contact). Thus, when fed with the same correlation pattern, the significant test might be mistaken to imply that fever mediates the relation between virus and disease based on the unwarranted assumption that causality can be inferred inductively from a statistical test.

The example highlights the priority of theoretical over statistical reasoning. By the same token, a significant result for sexual contact as a third variable can be reframed theoretically as a moderator rather than a mediator effect. Then, the virus is transmitted only among sexually active people, but not among sexually abstinent people. Choosing between moderator (person groups) or mediator accounts (contagion mechanism) is an essentially theoretical problem that cannot be solved statistically (cf. Fiedler, Walther, Freytag, & Stryczek, 2002).

Although the epidemiological example is clearly relevant to health psychology, it may be worthwhile illustrating the same point with a genuinely social psychological example: The same tri-variate covariance pattern allows for several theoretical interpretations of the role of Z relative to X and Y. For instance, the cognitive responses Z (pro or contra responses) to a persuasive communication in a thought-listing task are often interpreted as a mediator of the impact of a persuasive message X on a changing attitude Y. But Z may be conceived as another measure of the dependent variable, attitude change. Or, it may be framed as a moderator, restricting attitude change to those people who engage in active cognitive responses to the message content.

The viability of different mediation models can vary strongly on a-priori grounds. Thus, the encoding strategy applied to a persuasive message (e.g., trying to generate few or many arguments or counter-arguments; Tormala, Falces, Briñol, & Petty, 2007) logically affords a more viable candidate for a mediator variable than an enduring personality attribute (e.g., expert knowledge) that existed long before the persuasive attempt (as explained by Tate, 2015).

As a rule, statistical mediation tests are contingent on the validity of the mediation model (Waldmann, Hagmayer, & Blaisdell, 2006) and choosing an appropriate causal model is essentially a theoretical issue, not a statistical one. Therefore, if the causal model makes sense theoretically and logically, convincing and elucidating mediation analyses can be simple and straightforward. For a simple demonstration, take the finding that positive testing mediates the genesis of confirmation biases (Fiedler, Freytag, & Unkelbach, 2007). Most participants in a simulated classroom setting who were asked to test the hypothesis that girls are good in language whereas boys are good in science engaged in positive testing strategies. That is, they asked girls more questions in language classes and boys more questions in science. This difference in sample size was sufficient to subjectively confirm the hypothesis, even though boys and girls did not differ in the relative rate of correct responses in either discipline. The confirmation bias fully disappeared for the minority of participants who did not engage in positive testing search strategies. When the supposed mediator was manipulated experimentally, such that sample size was larger for boys in language and for girls in science, the resulting bias was reversed, thus ruling out a common gender stereotype as an alternative mediator.[2]

However, while a theoretically plausible top-down model can render mediation analysis convincing, bottom-up inferences from statistically significant ad-hoc tests are logically flawed. It is a category mistake to infer from a significant mediation test that "Z mediated the influence of X on Y". Just as it is inappropriate to infer the truth of H_1 from a significant result, or its falsehood from non-significance (Trafimow, 2003), it is particularly wrong to infer the causal status of Z from a significant test result of a mediation model X ➔ Z ➔ Y. Such a model test can happen to be significant for many other reasons than Z being the true mediator.

1.2. Two sources of uncertainty

On the one hand, it is a truism that for every correlation between two variables it is possible to find alternative accounts in terms of several third variables, which can never be identified and controlled exhaustively. Fever comes along with other physiological symptoms (e.g., weakness of the immune system, inflammation) or behavioral correlates (mood states; risk-taking strategies). In persuasion, too, the number of pro and contra responses to arguments is but one possible mediator; cognitive responses come along with experienced fluency, pragmatic inferences, self-perception and demand effects etc. Because it is never possible to include the entirety of all potential mediator candidates (Z, Z′, Z″, … etc.) in a regression model and to decide which one (or two, or three) is the true mediator, it is impossible to identify causes in a statistical bottom-up inference.

On the other hand, given only three variables, X, Y, and Z, of which one (i.e., X) is bound to be exogenous and is therefore never affected by the other two, there is still a variety of 12 different models that might describe the tri-variate causal structure (cf. Fig. 1).[3] The mediation model X ➔ Z ➔ Y is only one, and often not even the most plausible, of all these causal models. For example, fever, or cognitive responses to persuasion, might be consequences rather than mediating conditions of the disease or attitude change, respectively, reflecting reverse-mediation (X ➔ Y ➔ Z, denoted "reflection" in Fig. 1), which is hard to separate statistically from mediation proper (Lemmer & Gollwitzer, 2017; Thoemmes, 2015). An exhaustive bottom-up analysis aiming to identify the true mediator statistically would have to rely on diagnostic tests of mediation against countless other candidates and many alternative causal models. The number of different models increases dramatically when more than one mediator candidate is considered or when bi-directional or non-directional relations are allowed.

All this is by no means novel. Many methodologically-minded researchers and statisticians would pretend that it is actually common sense, asseverating that all serious scientists understand that there is always room for alternative mediators, and that alternative models exist. However, the reality of current behavioral research as it is published in peer-reviewed journals does not justify this disclaimer. In fact, the twofold problem of alternative mediators and alternative causal models is sorely neglected. As documented below, researchers rarely examine more than a single mediator variable, and they virtually never test other causal models than the standard mediation model. Nevertheless, they routinely and confidently infer from significant mediation tests that the arbitrarily chosen variable Z does mediate an effect, and they infer from non-significant tests that Z does not mediate an effect. We further observe that currently there is more of a tendency among statistical experts to facilitate complex analyses by developing

[2] In fact, no standard mediation test was required to substantiate the mediating role of positive testing.

[3] One might argue that mediation analysis is confined to those (six) models, in which X does affect Y. However, whether this condition is met is hardly known beforehand. None of the 12 models can be ignored theoretically.