Mediation Analysis and Warranted Inferences in Media and Communication Research: Examining Research Design in Communication Journals From 1996 to 2017

Michael Chan¹, Panfeng Hu¹, and Macau K. F. Mak²

Abstract
The number of studies employing mediation analysis has increased exponentially in the past two decades. Focusing on research design, this study examines 387 articles in the *Journal of Communication*, *Human Communication Research*, *Communication Research*, *Journalism & Mass Communication Quarterly*, and *Media Psychology* between 1996 and 2017. Findings show that while most studies report statistically significant indirect effects, they are inadequate to make causal inferences. Authors also often infer that they uncovered the “true” mediator(s) while alternative models and mediators are rarely acknowledged. Future studies should pay more attention to the role of research design and its implications for making causal inferences.

Keywords
mediation analysis, indirect effect, research design, causal inference, content analysis

A common strategy for theorizing and testing causal mechanisms in quantitative communication science is the use of mediation analysis (Slater & Gleason, 2012). That is, the explication and examination of the intervening variables that partly or fully explain

¹The Chinese University of Hong Kong, Shatin, Hong Kong
²University of Wisconsin-Madison

Corresponding Author:
Michael Chan, School of Journalism and Communication, The Chinese University of Hong Kong, Humanities Building, New Asia College, Shatin, NT, Hong Kong.
Email: mcmchan@cuhk.edu.hk
the effects from the independent to dependent variables. Such endeavors are important because many media theories in the field propose indirect effects (Holbert & Stephenson, 2003). Mediation analysis is thus an important conceptual and statistical tool for researchers to understand “how and why media effects occur” (Valkenburg et al., 2016, p. 324), and it is through the accumulation of such understandings that we build our body of knowledge on the processes and effects of media and communication in everyday life.

In this article, we focus on the role of research design in mediation analysis and its importance in generating findings that are warranted. In other words, how confident can we be that the patterns of observations among the independent variable(s), mediator(s), and dependent variable(s) (i.e., the mediation model) reported in studies can be inferred to the real world? This has important applied as well as theoretical implications, such as informing the design of appropriate messages to discourage harmful behaviors to the self, reduce intergroup conflict, and engender positive well-being, and other contexts. Communication science has an abundance of theories (Neuman & Guggenheim, 2011; Walter et al., 2018) that provide the necessary conceptual frameworks and justifications for proposing hypothesized mediation among variables. Primers on statistics and associated software (e.g., Hayes, 2018) allow researchers to execute sophisticated mediation analyses with just a few clicks of the mouse. What is currently lacking in the literature is a systematic examination of the research designs used for mediation analysis in communication studies, and the extent to which inferences derived from the results of such designs are warranted. Methodologists have long noted that “it is the design, not the statistical method, that permits causal hypotheses to be adequately tested” (H. E. Bullock et al., 1994, p. 254). Research design is a fundamental link between theoretical reasoning and statistical analysis in the overall research process. Neighboring disciplines including political science (J. G. Bullock et al., 2011), social psychology (Fiedler et al., 2011), and the organizational and management sciences (Stone-Romero & Rosopa, 2007) have addressed its importance for mediation analysis and made several recommendations, but its role in mediation analysis has received little attention in communication. This article addresses the gap.

We first provide an overview of the general trends in mediation analysis by examining relevant articles in the field’s journals known for publishing theory-driven quantitative research: Journal of Communication (JoC), Communication Research (CR), Human Communication Research (HCR), Journalism and Mass Communication Quarterly (JMCQ), and Media Psychology (MP) from 1996 to 2017. Although hardly exhaustive or representative of all communication scholarship, the articles in these journals provide a good overview of the predominant research designs used in the field for mediation analysis. We discuss some of the important assumptions of mediation analysis that pose distinct challenges for research design and show that many, if not all, of the designs adopted in these articles are less than adequate to infer a “true” mediation process. Moreover, we demonstrate that few articles adequately reflect upon or address the limitations of
their research designs while making very confident causal inferences about their “significant” findings.

It is not the purpose of this article to dispute or criticize past research that had undergone rigorous peer review and made important contributions to the field. Rather, our aim is to raise awareness on current practices and propose recommendations that can improve the methodological rigor of mediation analysis. In doing so, inferences can be made with greater confidence and weaknesses can be acknowledged to guide subsequent research. This article thus informs researchers who are designing mediation studies as well reviewers of such work and educators who teach quantitative research methods classes.

Logic of Mediation Analysis and Causal Inference

At a theoretical and conceptual level, a mediating (or intervening) variable constitutes part of a chain of relationships where an independent variable (X) is said to influence the mediator (M), which in turn influences the dependent variable (Y). On a statistical level, mediation analysis tests whether X has a statistically significant indirect effect on Y through M. Or if there are multiple mediators in the model, which specific indirect effects or pathways between X and Y are significant. Thus, mediation analysis is fundamentally concerned with explicating and testing causal relations. All introductory texts on quantitative research methods emphasize the three basic requirements to demonstrate causality between the independent and dependent variable: (a) observable association, (b) temporal order, and (c) non-spuriousness, such that the relationship is not explained by an unknown third variable. Causal inference can be implied when all three conditions are met. Another important requirement is a detailed understanding of the mechanism explaining why X affects Y, which requires knowledge of the literature and strong theoretical reasoning (Yanovitzky & Greene, 2009). As shown in Figure 1, adding a single mediator (M) extends the theoretical requirements to explain the $X \rightarrow M \rightarrow Y$ relationship. Moreover, X should precede M and M should precede Y and unknown third variables should not confound the associations between $X \rightarrow M$ and $M \rightarrow Y$. Adding even more mediators stretches the requirements even further. For example, Sotirovic and McLeod’s (2001) proposed causal chain of post materialism values $\rightarrow$ news consumption $\rightarrow$ discussion diversity $\rightarrow$ public affairs knowledge $\rightarrow$ political participation would require measuring the variables at five different time points to establish temporal order. Yet, the study itself uses cross-sectional data to test the model.

The use of cross-sectional designs for mediation is common in other fields (Kline, 2015) and reflects the reliance on theory and logical reasoning to justify proposed mediation processes and subsequent findings even though such research designs are often inadequate to attest to such causal claims. With no time precedence, correlational data that support an $X \rightarrow M \rightarrow Y$ mediation model can feasibly support a myriad of other models, such as $X \rightarrow Y \rightarrow M$ (“reflection”), $X \rightarrow M$ and $X \rightarrow Y$ (“common cause”), and $M \rightarrow Y$ (“alternative cause”) among others (Fiedler et al., 2018; Kline,
The possible combinations are increased exponentially if more mediators are added such that the probability that the combination of variables relating to each other in the theorized directions decreases exponentially (Saylors & Trafimow, 2020). So even if statistical analyses support the proposed mediation model, one cannot discount the possibility that the same data can fit other equally plausible models.

With these issues in mind, we begin this study by establishing the number of published articles that feature mediation analysis within the study timeframe (RQ1), the type of research design used to test mediation (RQ2) and the relative complexity of the mediation models (RQ3). As part of our interest in the general trends of mediation analysis we also examine the number of studies that appear in each article because multi-sample mediation studies are prevalent in fields such as social psychology (RQ4). It is also likely that advanced training in multivariate statistics has become more widespread among doctoral programs. As Slater and Gleason (2012) noted “it is likely that studies addressing mediation are becoming more common now that the software to conduct appropriate statistical tests is readily available and knowledge of how to conduct such analyses is likely increasing over time” (p. 233). Therefore, we examine the statistical approaches (RQ5) and software used (RQ6) to test mediation. We then focus specifically on the role of the mediator(s) and causal inferences. First, for experiments did authors manipulate the mediator to address possible M → Y confounds in addition to the independent variable to address X → Y confounds? (RQ7). Then, we examine the extent to which authors acknowledged the limitations of their research designs for testing mediation (RQ8), whether they tested alternative models (RQ9) and proposed alternative mediators that can explain the X→Y relationship (RQ10). Finally, we note how authors inferred their “causal” findings to the real world (RQ11).
Method

Sampling

Full texts articles were downloaded from two databases: (a) “Academic Search Ultimate: Communication Source” for JoC, HCR, and MP; and (b) “Communication Studies: A SAGE Full-Text Collection” for CR and JMCQ. The same search parameters were used for both databases: “mediation,” “mediating effect,” and “indirect effect” together with the time period: “January 1, 1996” to “December 31, 2017.” After deleting nonrelevant articles (e.g., review articles, articles where “mediation” appeared in the text, but had no relation to mediation analysis etc.) a total of 387 articles were included in the final study sample (CR = 157, JoC = 68, JMCQ = 58, HCR = 52, MP = 52).2

Coding Protocol and Intercoder Reliability

The first author of this study developed a coding protocol for the purpose of assigning values to the respective variables (N = 10) for subsequent analysis. Two additional coders (who are also coauthors of this study) were then trained to apply the protocol to mediation studies in articles that were not part of the study sample. Reliability checks and formal discussions among coders on the initial coding led to refinement of the protocol (see Table 1), which was then applied independently by the three coders to the study sample. Intercoder reliability checks for every variable of every case were conducted incrementally after every set of 50 articles using Krippendorff’s alpha (Hayes & Krippendorff, 2007).3

Any disagreement among the three coders was deliberated upon to reach a consensus before moving on to the next set of 50 articles. Agreement for manifest content (e.g., variables #3 and #4) was very high (e.g., $\alpha = .95$ to complete agreement) upon checking of the first set of 50 articles. There was initially less agreement for latent content such as #10 where reliability was $\alpha = .80$. However, it improved in the subsequent checks and high reliability was attained ($\alpha = .88$ to complete agreement). Again, coders conferred to reach consensus on any disagreements.

While some variables are self-explanatory others require further explication. For #1, experiment is any study that presents one or more manipulated stimuli (X and/or M) to participants and the subsequent response (Y) is measured. Cross-sectional survey designs do not manipulate X nor M.4 Longitudinal designs involve data collection at more than one point in time, such as a two-wave panel study. Mixed designs may appear in multistudy articles, such as a cross-sectional survey in one study and experiment in another. For #2, a single mediator model refers to one mediator with at least one X and one Y. Parallel mediators refers to two or more parallel mediators between one or more X and Y. Serial mediators refers to a causal chain model with three or more paths, such as $X \rightarrow M_1 \rightarrow M_2 \rightarrow Y$. Complex refers to models that have characteristics of both parallel mediators and casual chains of three or more paths. Moderated mediation refers to a mediation model where one or more moderators influence the indirect effect(s) through one or more mediators. For #8, the term “alternative models”
Table 1. Final Protocol for Coding the Articles in the Study Sample.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Values</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. What research design is used to test mediation?</td>
<td>E = experiment S = surveys L = longitudinal M = mixed (for multistudy articles)</td>
<td>.97</td>
</tr>
<tr>
<td>2. What type of mediation model is tested?</td>
<td>Si = simple mediation P = parallel mediation Se = serial moderation C = complex mediation M = moderated mediation</td>
<td>.94</td>
</tr>
<tr>
<td>3. How many studies are in the article?</td>
<td>(Number)</td>
<td>.98</td>
</tr>
<tr>
<td>4. What statistical approach is used to test mediation?</td>
<td>R = linear regression S = structural equation modeling B = both regression and SEM O = other</td>
<td>.98</td>
</tr>
<tr>
<td>5. Is any statistical software for conducting mediation analysis mentioned?</td>
<td>N = no Y = yes + (record software name)</td>
<td>.96</td>
</tr>
<tr>
<td>6. For experimental studies, which variable is manipulated?</td>
<td>I = independent variable M = mediator B = both</td>
<td>.94</td>
</tr>
<tr>
<td>7. Are limitations of the research design addressed?</td>
<td>N = no Y = yes + (record verbatim)</td>
<td>.94</td>
</tr>
<tr>
<td>8. Are alternative mediation models tested?</td>
<td>N = no Y = yes</td>
<td>.98</td>
</tr>
<tr>
<td>9. Are alternative mediators proposed that can also explain the indirect effect?</td>
<td>N = no Y = yes Ys = yes. Specific mediators named</td>
<td>.94</td>
</tr>
<tr>
<td>10. Is the role of the mediator inferred in the past or present tense?</td>
<td>N = no inference stated NS = null mediation Y-now = present tense + (record verbatim) Y-past = past tense + (Record verbatim)</td>
<td>.92</td>
</tr>
</tbody>
</table>

Note. SEM = structural equation modeling. Krippendorff’s alpha for intercoder reliability was calculated for all articles after each set of 50 articles was coded. The reported figure above represents the average of the eight alpha values for each variable.

only refers to models that have different variable configurations compared with the hypothesized model. It does not include “respecified models” where the same mediation model is tested with the addition or removal of component paths, which is a common analytical step in structural equation modeling (SEM). For #10, we coded
Y-now” even in cases where the discussion section of the article used both present and past tense to infer the mediation findings.

**Results and Discussion**

**Growth and Trends of Mediation Analysis in the Field**

Figure 2 illustrates the exponential growth in the number of articles employing mediation analysis (RQ1). Another way of describing the growth is by examining the relative proportion of studies to total articles published. Thus, in 1999, 2% of articles in the journals included some form of mediation analysis, which rose to 7% in 2007 and then 22% in 2017. This shows that articles employing mediation analysis were not only being published in greater numbers, but also comprised of an increasing proportion of media and communication scholarship. The figure also shows that experiment and cross-sectional survey designs for testing mediation were equally prominent (RQ2). Experiments accounted for 44% of studies (N = 172), whereas surveys accounted for 43% (N = 167). Longitudinal studies accounted for 11% (N = 43) and mixed designs (i.e., articles with different designs across multiple studies) accounted for 1% (N = 5).

In terms of the kinds of mediation models examined in the articles (RQ3), Figure 3 suggests a trend over time for articles to examine more elaborate mediation models (e.g., multiple mediators and moderated mediation) rather than single mediator models. Of the 171 survey studies, 93 (54%) adopted parallel, serial, and complex mediation designs, which limited the plausibility of the models being the correct model because the variables were measured at the same time. The figure was 70 (44%) for experiments. Moreover, 343 articles featured a single study (89%), whereas 41 featured two studies (10%) and three featured three studies (1%) (RQ4).
Not surprisingly, multistudy articles were more common in experiments with 33 articles (19%) compared with three for survey studies (2%).

Figure 4 highlights the statistical approaches (RQ5) and software used (RQ6) over time and shows the pervasiveness of regression and SEM to test mediation. However, regression-based approaches became more prominent compared with SEM in the second decade. This might be attributable to the increased proliferation and use of user-friendly software macros that extended the capabilities of general-purpose statistical packages (e.g., SPSS) to run mediation analyses. The earlier macros were created to test specific mediation model configurations. For example, the INDIRECT macro only allowed the testing of simple and parallel mediation (i.e., multiple mediators between X and Y; Preacher & Hayes, 2008), whereas the MODMED macro was specific to testing moderated mediation models (Preacher et al., 2007). All the capabilities of these macros were eventually subsumed and expanded by PROCESS (Hayes, 2018), which was used extensively in articles published from 2015 to 2017. SEM approaches meanwhile generally require more academic training and dedicated software, which might explain the lower proportion of overall studies using the approach. After considering the general trends we now turn to the specific role of the mediator.

**Experiments Manipulate Independent Variables But Not the Mediators**

The experiment is the gold standard for demonstrating causality because a research design that randomly assigns people to a manipulated X and subsequently observing the Y can generally fulfill the three criteria of causality mentioned previously. In a simple mediation model, however, a manipulated X can only attest the causal influence of the X → M and X → Y effects. If the mediator is measured rather than manipulated, one cannot exclude the possibility that a third variable confounds the M → Y relationship (J. G. Bullock et al., 2010; Spencer et al., 2005). Such a measurement-of-mediation...
research design is common in the social sciences and the articles in this study showed that the communication field is no exception. All experiments in the sample adopted such a design and none appeared to manipulate the mediator (RQ7). The procedure used by Schmuck et al. (2017) in their laboratory study of anti-Muslim populist ad exposure on perceived hostility toward the mainstream population was typical of this approach. In brief, Muslim participants in the treatment condition were exposed to multiple right-wing anti-Muslim ads and those in the control condition were exposed to generic ads. After the treatments, all participants completed a questionnaire that measured among other things their levels of perceived discrimination, national identification, self-esteem, and hostility toward majority members in society. Mediation analyses supported their proposed model, such that “exposure to right-wing populist ads triggered the feeling of being discriminated against, which decreased individuals’ national identification and self-esteem. A decrease in national identification and self-esteem in turn resulted in higher hostility toward the majority population” (p. 622). These serial mediation pathways were tested with correlational data between the three mediators and the dependent variable. Thus, the model had the very same limitations as cross-sectional studies in its ability to draw causal inferences. It was equally plausible that the mediators could serve as dependent variables. The authors acknowledged this in the discussion (see Table 2), but such self-reflection in experimental studies is not common.

**Addressing Limitations of the Research Design**

To what extent did authors address the limitations of their research designs to make causal claims (RQ8)? A reading of the discussions and limitations sections in the
articles showed that they generally focused more on discussing “technical” limitations of their study, such as sampling, conceptualization, and operationalization issues related to study variables. Research design related specifically to mediation is acknowledged in 111 survey studies (67%), which is typically skimped over with an almost obligatory caveat sentence or paragraph that consisted one or several of the following elements: (a) The data were cross-sectional, so (b) causal claims could not be made on the direction of the effects, even though (c) the findings were based on strong theoretical foundations and/or past research. Therefore, (d) future longitudinal and/or experimental studies were needed to verify causality of the relationships found in the study. For experiments, 30 studies (17%) acknowledged the correlational nature of the $M \rightarrow Y$ relationship that inhibited the ability to draw concrete causal inferences. More specifically, 20 mentioned that the mediator and dependent variables were measured at the same time, eight cautioned that the mediator was not manipulated, and two studies gave both reasons. Table 2 provides several examples of how such limitations were phrased.

**Table 2. Example Statements of Limitations of Experiment Designs.**

<table>
<thead>
<tr>
<th>No experimental manipulation of mediator(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>“This claim, however, should be stated somewhat tentatively and confirmed with additional research because goal frequency and appropriateness were not directly manipulated” (Palomares, 2013, p. 92).</td>
</tr>
<tr>
<td>“Furthermore, we did not explicitly manipulate emotions in our design, and thus strictly we cannot make valid inferences about the role of emotions as a causal mechanism of framing effects” (Powell et al., 2015, p. 1012).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Correlational relationship between mediator(s) and dependent variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Third and most important, because we measured anticipated regret and intentions at the same time points, the mediational analyses are correlational and limit our certainty about the causal order” (van Koningsbruggen et al., 2016, p. 1039).</td>
</tr>
<tr>
<td>“Moreover, although we used an experimental approach to establish a causal relationship between political ad exposure and perceived discrimination, we measured all dependent variables at one point of time, which strictly speaking only allows correlational evidence for the mediators and dependent variables” (Schmuck et al., 2017, p. 626).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Both explanations</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Also, it should be noted that although one can make a strong causal inference about the direct effect of slant, a manipulated factor, the relationships between personal opinion and perceptions are correlational and must be interpreted with more caution” (Gunther, 1998, p. 498).</td>
</tr>
</tbody>
</table>

Note. Underline added for emphasis.

**Alternative Models and Mediators Are Rarely Acknowledged, Proposed or Tested**

A total of 36 survey studies (22%) attempted to strengthen their causal claims by testing alternative models against the data (RQ9). For example, Slater et al. (2007) framed
their article from the outset as a model comparison study and tested six different model configurations to examine the influence of sensation seeking and negative experience on risk judgments about alcohol via news exposure and attention to crime news. This approach was uncommon, however. More typical were confirmatory alternative mediation models that were tested and reported in the results or endnotes sections after the “significant” findings of the originally proposed model were reported. Morgan and Shanahan (2017), for example, proposed that heavy TV viewing was associated with increased authoritarianism tendencies, which in turn increased intention to vote for Trump in the U.S. election. After finding that the “data was consistent” with the proposed model they ran several alternative models with different configurations of X, M, and Y to show that the alternative mediation models did not fit the data as well, thus strengthening the argument for the proposed model. These methods may go some way in determining which models are consistent or not with the data, but they fundamentally cannot address the issue of causality.

Similarly, 27 experimental studies (16%) tested alternative models and all were reported post hoc with the purpose to show that the original proposed model better fitted the data despite the correlational $M \rightarrow Y$ relationship. For example, de Graaf (2014) proposed that both self-referencing and identification mediated the effects of protagonist similarity on story consistency beliefs. The findings showed that only self-referencing mediated the effect. Then, to “rule out alternative explanations,” the author followed up with additional statistical tests for serial mediation (self-referencing $\rightarrow$ identification), which was not significant.

At best, even the most rigorous designs and appropriate statistical tests that support the proposed model cannot discount the possibility that other mediators not included in the study can exert influence in the $X \rightarrow Y$ effect. Causal inference can only be made that the proposed mediation model is one of many possible models. Again, following Fiedler et al. (2018) we examined whether the discussion section of the studies acknowledged this (RQ10). Only 28 survey studies of the total (17%) proposed alternative mediators that could also explain the $X \rightarrow Y$ relationship, whereas the number was 25 for experimental studies (15%) and three for longitudinal studies (7%). As shown in Table 3, the statements can range from very general acknowledgements that other unnamed mediators could account for the mechanisms, to the identification of specific mediators that are more helpful for future research.

Unwarranted Inferences?

Of the 387 studies in the sample, 382 reported one or more statistically significant mediation models. Yet, we had already established that the predominant research designs are often not up to the task of making strong causal inferences. Following the same procedure as Fiedler et al. (2018), we examined whether authors inferred their significant mediation findings in the past or present tense (RQ11). The implication is that explicit statements about the mediator(s) in the past tense confined causal inferences to within the study, whereas statements in the present tense extended causal inferences to a “generalizable law.” For example, Chia (2010) tested with cross-sectional
data the notion that different social influences (i.e., parents and friends) would mediate the effects of media advertising on young children’s sense of materialism. With significant SEM results, the author concluded in the discussion section that “there is a significant indirect effect of advertising on adolescents’ materialistic values” and that the “indirect effect is mediated by adolescents’ perception of the influence of advertising on friends, but not by adolescents’ perception of the influence of advertising on parents” (p. 414)” (italics added for emphasis). By referring the significant and insignificant mediation in the present tense, it is implied that, (a) the same pattern of relationships and mechanisms found in the study can be inferred to the real world, and (b) the significant mediator is the “true” mediator that explained the relationship.

Table 4 provides additional examples of how causal inferences were phrased following significant statistical tests of mediation. Of the 387 studies, five (1%) did not report significant findings, 18 (5%) did not make any explicit causal inferences derived from the mediator(s), 152 (39%) did so in the past tense, and 212 (55%) did so in the present tense.

General Discussion

Commenting on the general state of mediation analysis, Kline (2015) made the assessment that most studies used research designs “that are inadequate to establish mediation, so relatively little of the extant literature on mediation is actually worthwhile” (p. 210). On organizational research, Saylors and Trafimow (2020) pointed out that the field’s propensity for complex mediation models meant that “much of the knowledge generated in top journals is likely false” (p. 1). Our review of mediation analysis published in communication journals in the past two decades is consistent with the notion that extant research designs are often inadequate to draw strong causal inferences. Moreover, authors were generally eager to claim that their significant mediation
models can be generalized to the real world but paid less attention to acknowledging limitations of their research designs and the possibility of other plausible mediators that could also account for the proposed mechanisms.

Our view is less somber than the scholars above based on both pragmatic and theoretical grounds. Practically, insistence on watertight research designs for mediation analysis will impede the progress of knowledge creation and dissemination. Some parts of the field may even grind to a halt. It is important to emphasize that even mediation studies with less than adequate designs can be sources of new ideas, perspectives, and propositions that serve as springboards for further exploration using more rigorous research designs. An exemplar is Shehata and Amnå’s (2019) application of the communication mediation model using five-wave panel data to demonstrate the causal dynamics over time among news use, political discussion, and political interest. Such a large undertaking and test of the model would most probably not be attempted if not for the body of previous theorizing and evidence derived from cross-sectional data across different cultural contexts (e.g., Chan et al., 2017; Shah et al., 2007). This is consistent with the notion that knowledge generation and theory building are cumulative (Neuman & Guggenheim, 2011) and that studies with inadequate research designs can still have value, especially if authors clearly acknowledge the limitations of their findings and make concrete proposals on what future studies should do to improve causal claims and inferences of their proposed mechanisms.
Given the methodological shortcomings of current studies, the interesting question arises as to why studies that use mediation analysis are appearing in greater numbers in the field’s journals? There are several possible interrelated and mutually reinforcing reasons. The first is the pervading indirect effects paradigm in the field that places emphasis on explicating mechanisms that underlie media effects (Neuman & Guggenheim, 2011; Valkenburg et al., 2016). This creates a mutually reinforcing cycle where journal editors and reviewers privilege “mechanisms” and “processes” that underlie theories while authors observe the propensity for elaborate models in the extant literature and conclude that this is what is required for publication. In social psychology these expectations are explicit as the following editorial statement exemplifies: “First, this journal has traditionally sought to publish articles that make a meaningful theoretical advance by linking empirical findings to underlying processes” (Smith, 2012, p. 1). Given the proximity of social psychology and quantitative communications science, it would not be surprising if this norm has filtered into our field, especially for research that uses experiments. Second, with free and easy-to-use software such as PROCESS (Hayes, 2018), the bar for testing and reporting sophisticated mediation models has been lowered exponentially. Anyone with a rudimentary understanding of regression statistics can run mediation analysis and iterations of different mediation models can be tested in a matter of seconds even though some researchers may not understand the logics that underlie the statistics. This may lead to more studies based on ad hoc data discovery and subsequently to a greater overall volume of studies examining indirect effects, which are then submitted and published in greater numbers in the field’s journals. Third, researchers are following norms in the field. If mediation studies based on cross-sectional surveys and experiments that only manipulate the independent variable are perceived to be sufficient for publication in the field’s “top” journals, then they can reasonably assume that such research designs are sufficient. This reduces the incentive to conduct multiple studies or adopt more rigorous designs that requires more time and resources. Fourth, it is possible that while doctoral programs in the past decade have emphasized more on statistical and software training, they may have neglected to some extent the role of research design for making warranted causal inferences.

**Recommendations**

In view of the above, we offer some suggestions and recommendations to improve the design and reporting of mediation analyses. Causal inferences in mediation depend not only on strong theoretical reasoning and appropriate statistical analyses, but also rigorous research design that can attest to the proposed mediating causal paths (Stone-Romero & Rosopa, 2007). Several recommendations have appeared in some form in previous reviews and critiques (e.g., J. G. Bullock et al., 2010; Fiedler et al., 2018; Kline, 2015; Spencer et al., 2005), but it is worth restating a few of them for the context of mediation analysis in media and communication research. We take J. G. Bullock et al.’s (2010) position that views “mediation analysis as a cumulative enterprise” (p. 550) and that the discovery and confirmation of mechanisms in communication
require a systematic series of studies rather than one standalone study. Moreover, it is highly unlikely that existing norms and practices of mediation analysis demonstrated in this study are going to change in the short term unless the threshold for publishing mediation studies in the field suddenly rises to a level that would force researchers to adopt more rigorous approaches. Therefore, we distinguish between recommending incremental improvements to current practices and adopting more rigorous research designs that can help the researcher make stronger causal claims.

**Improving Existing Practices**

Researchers need to temper claims of causal inference even when they demonstrate that their mediation models were found to be “statistically significant.” The predominant research designs in the current sample comprise cross-sectional surveys and experiments that only manipulate the independent variable. This means that all the studies have “effects” that are fully or partly correlational. Therefore, authors should acknowledge that the data can potentially account for other alternative models or mechanisms. As shown in Table 2 (i.e., correlational nature of data) and Table 4 (i.e., confining inferences to the study using the past tense), such practices are already being done to some extent though only in a few studies. Such acknowledgements should be more widespread in future studies when inadequate research designs are adopted because authors can only claim that the pattern of results is consistent with a proposed model based on prior theorizing and logical expectations. Nor can authors imply that their mediating mechanism is the only mechanism that intervenes between X and Y. Other possible mediators need to be acknowledged, such as those shown in Table 3 because such discussions serve important roles in setting the agenda and contributing ideas and insights on what subsequent studies should focus on when studying the same topic or attempting to replicate or expand on the mechanism. Proposing specific mediators would be even more helpful.

Continued use of longitudinal studies is recommended because X, M, and Y are assumed to have time precedence. For example, using a three-wave panel survey, van Oosten et al. (2015) demonstrated that young females viewing sexually provocative music videos by male artists in Time 1 was related to affective engagement in Time 2 (i.e., the video was arousing), which in turn was related to acceptance of female token resistance at Time 3 (i.e., perception that females say “no” to sexual advances when they actually mean “yes”). With three-wave panel designs it is still not possible to rule out other possible confounds. But the findings can be supplemented and strengthened with experiments. For example, in the van Oosten et al. study, participants can watch different videos (e.g., sexual vs. nonsexual) and then their arousal levels and acceptance of token resistance are measured afterwards. And then a second experiment can manipulate arousal levels and acceptance of token resistance (i.e., a double randomization design to be discussed later). Therefore, there is utility in testing the same proposed mediating mechanisms using different research designs. An example of this approach is Saleem et al.’s (2017) study, which used both survey and experiment to examine a mechanism where media depictions of Muslims as terrorists increased the
perception that they are aggressive, which in turn increased support for policies to restrict their civil rights and support for military action in Muslim countries. Despite the relative weaknesses of the designs to draw causal inferences they did to some degree strengthen the theoretical premise of the proposed relationships that could be the basis of later research. As noted earlier most articles in the field feature a single study. Multistudy articles are useful because proposed mechanisms can be replicated to other contexts. These include testing the same variables as the first experiment with another sample but extending some parameters in the second experiment, such as a second mediator (e.g., Arroyo et al., 2014); using samples from different countries or cultures (e.g., Men & Muralidharan, 2017), and examining if the proposed mediator is still present when different types of stimuli (e.g., film vs. television) are used (e.g., Bartsch & Schneider, 2014). Again, it should be noted that the use of multiple studies does not necessarily help to strengthen causal inferences, but they can be useful to accumulate evidence of possible causality, which can guide subsequent research.

Also, while some studies acknowledged limitations in research design, it would be helpful for authors to better elaborate how future studies can improve causal inferences rather than use generic statements along the lines of “longitudinal studies are required to support causal claims” or that “the mediator and dependent variable are correlational so mediation has to be interpreted with caution.” A meaningful suggestion does not have to be lengthy, as demonstrated by van Koningsbruggen et al.’s (2016) succinct recommendation to improve the research design of their study: “Therefore, it is crucial that future studies replicating our results establish a temporal order between anticipated regret [M] and intentions [Y] in order to more precisely follow the causal chain approach assumed when testing mediation” (p. 1039). Notation is added for emphasis.

**Strengthening Future Practices**

Because mediation analysis is concerned about causality the experiment is the most appropriate method to test proposed mechanisms. In particular, scholars in other fields have recommended experiments that manipulate both X and M. Known as manipulation-of-mediator designs they can provide stronger evidence of a causal M → Y effect whereas measurement of mediation designs can only demonstrate association. One such design that would strengthen causal inferences in many experimental studies in this sample would be the double randomization design (Pirlott & MacKinnon, 2016), also known as the experimental-causal-chain design (Spencer et al., 2005) and two randomized experiments design (Stone-Romero & Rosopa, 2007). This design typically comprises two experiments. The first is a standard measurement of mediation design where X is manipulated and M and Y are measured, then the second experiment manipulates M and Y is measured. When the findings of both experiments are combined, stronger inferences can be made that M mediates the X → Y effect. As almost all the experiments reported in this study have essentially completed the first experiment, it would be a small step to conduct a follow-up second experiment to manipulate M and measure Y. Recall Saleem et al.’s (2017) experiment where perception of
Muslims as aggressive (M) and support for policies to restrict their civil rights (Y) were measured and are therefore correlational. A follow-up experiment can manipulate perceptions of aggressiveness and measure support for policies. If van Koningsbruggen et al. (2016) were to follow their own recommendation mentioned earlier, they could conduct a second experiment where anticipated regret (M) is manipulated and then intention (Y) measured.

More substantive examples can be found in neighboring fields such as psychology. One of them is Bélanger et al.’s (2019) four-study article that examined among other things how high levels of obsessive passion (X) led to moral disengagement (M), which then engendered willingness to engage in violent behaviors for a political or social cause (Y). The first two studies were cross-sectional, and they provided initial support for statistical mediation among the variables. The latter studies followed the double randomization design. In the third study the authors manipulated X by inducing an obsessive or harmonious mind-set through a writing task. The previous mediation analyses were replicated, which provided causal evidence for X → M and X → Y. In the fourth study they manipulated M by inducing moral disengagement through a writing task. Together, the two studies provided causal evidence for all components of the proposed mechanism.

While the double randomization design is appropriate for simple mediation models, mediation in communication research in recent years is characterized by more complex models. Recall the study of Schmuck et al. (2017) which proposed a causal chain from right-wing ad exposure (X), to perceived discrimination (M₁) → religious/national identification (M₂/M₃), self-esteem (M₄), and hostility toward mainstream society (Y). Systematically applying the logic of the double randomization design to this causal chain would require more than two experiments. One way to address this would be to reference in the literature review past experimental studies that have demonstrated causal evidence for individual components of the proposed causal chain, such as the effects of national identification on self-esteem (Martinot et al., 2016) and in-group identification on out-group derogation (McGregor et al., 2008). In doing so, one can focus on the underexamined and untested parts of the proposed mediation model.

Another manipulation-of-mediator design is the concurrent double randomization design (Pirlott & MacKinnon, 2016) where both X and M are manipulated in a single experiment. Again, such a design is more prominent in psychology. One is Pogge and Smith’s (2020) study of individuals’ perceptions of political polarization in the U.S. (X) on the perceived brokenness of the electoral system (M), which then increases ones’ desire for a third-party candidate (Y). Like Bélanger et al., they ran multiple studies to first establish causal evidence for X → Y, X → M, and M → Y. They then conducted a final study where X and M were both manipulated. To be consistent with the temporal order, participants first received the X stimulus (i.e., three short scenarios that primed low, medium, and high polarization) and then the M stimulus (i.e., two short scenarios that primed broken and not-broken electoral system). The advantage of this design compared with the double randomization design is that one can concurrently observe the effects of X and M on Y. As Pogge and Smith ultimately found: “the effect of perceived brokenness did not entirely subsume the effect of perceived polarization”
(p. 11), thus providing causal evidence that X and M concurrently cause Y. A shortcoming of the concurrent double randomization design is that it cannot directly demonstrate the $X \rightarrow M$ effect. Therefore, this design may work better as a follow-up experiment or in research situations where the $M \rightarrow Y$ effect is the primary focus and $X \rightarrow M$ is already a given based on prior theory or previous experimental evidence.

**Conclusion**

The number of studies using mediation analysis in the field has risen exponentially in the field of communication, but no study has yet examined the pivotal role of research design in mediation analysis. An analysis of 387 articles found that many studies of mediation are based on inadequate designs that can lead to unwarranted causal inferences. Rather than dismiss these studies we should acknowledge their value to accumulate evidence of plausible causal mechanisms and provide theoretical and methodological insights for subsequent research. At the same time, we encourage authors to better acknowledge research design limitations, temper claims of causal inference, and provide concrete suggestions on how future studies can improve upon or verify the causal claims of their mediation models.

This study also exemplifies the challenges facing researchers employing mediation analyses. Sophisticated narratives and elaborate models are appealing. But, the greater number of mediators in the model and the more elaborate the mechanisms examined, the more difficult it becomes to make causal inferences about the proposed model. Implicit norms and expectations in the field may be partly responsible for this growing trend, but we should not lose sight that findings from simpler and more parsimonious research designs for mediation may ultimately be of greater benefit and value because stronger causal inferences can be made. This study makes several recommendations such as the use of manipulation-of-mediator designs. Applying them, however, poses another set of challenges. Practically, more time and resources must be expended to conduct multiple studies. Professionally, the demands of career advancement mean that researchers are enmeshed in what Mellado et al. (2020) call the “continuous cycle of publish and perish.” One practical consequence is that a researcher may view three single-study articles on different topics to be ultimately more valuable to their career than a single three-study article that rigorously tests and affirms a causal mechanism. A recent call for the field to adopt “open science” practices may partly address this tension by incentivizing replication studies (Dienlin et al., 2020). This may encourage the use of more rigorous research designs to attest causal mechanisms that were based on solid theoretical grounds but were examined with less than robust research designs. This is especially relevant to the field of communication as most articles analyzed in this study were based on single studies.

We acknowledge that this study focused on research design for mediation and not statistical techniques for mediation, such as sensitivity analysis (Imai et al., 2010). Nor did we consider mediation studies on communication-related phenomena outside the field, such as research on video games appearing in multidisciplinary journals (e.g., Gabbiadini et al., 2016) and trends toward the use of psycho-physiological measures
as mediating variables (e.g., Appel et al., 2019). Moreover, the increasing number of moderated mediation studies appearing in the field (e.g., Chan, 2018; Wojcieszak & Garrett, 2018) means that researchers need to carefully consider the role of the moderator from a theoretical and research design perspective (Holbert & Park, 2020). These can be the focus of future reviews on mediation analysis.

To conclude, one may think that good theory that is paired with sophisticated statistical analysis is sufficient to generate valid causal inferences on the mechanisms underlying media and/or communication effects. This review serves to remind researchers, reviewers, and educators in the field that elaborate statistical techniques for testing mediation cannot overcome the flaws of inadequate research design.

**Declaration of Conflicting Interests**
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding**
The author(s) received no financial support for the research, authorship, and/or publication of this article.

**ORCID iDs**
Michael Chan https://orcid.org/0000-0001-9911-593X
Panfeng Hu https://orcid.org/0000-0002-7820-2214
Macau K. F. Mak https://orcid.org/0000-0003-0819-1107

**Notes**
1. A full discussion of the history and background of mediation analysis is beyond the scope of this article. In this regard, Mathieu et al. (2008) provide an informative historical account of the origins and development of mediation analysis.
3. We started with *Journalism & Mass Communication Quarterly* from the oldest to newest articles, then repeated the same procedure for *Journal of Communication, Human Communication Research, Media Psychology*, and then *Communication Research*.
4. Survey experiments were coded as “experiment.”
5. Tables summarizing the percentage breakdown of the results are available from the authors on request.
6. The four most popular structural equation modeling (SEM) packages in the sample are LISREL, EQS, AMOS, and Mplus.
7. It should be noted that JMCQ switched from Chicago Style to APA Style in 2015. We did not observe any noticeable changes in the use of tenses before and after the switch.

**References**


Yanovitzky, I., & Greene, K. (2009). Quantitative methods and causal inference in media effects research. In R. L. Nabi & M. B. Oliver (Eds.), *The SAGE handbook of media processes and effects* (pp. 35–52). SAGE.

**Author Biographies**

**Michael Chan** (Chinese University of Hong Kong [CUHK]) is an associate professor at the School of Journalism & Communication, CUHK. His research examines the role of digital communications in political and social life. His works have appeared in *New Media & Society*, *Political Communication*, and *Journal of Computer-Mediated Communication*, among others.
Panfeng Hu (MA, Beijing Foreign Studies University) is a PhD student at the School of Journalism & Communication, Chinese University of Hong Kong. Her research areas include political communication, digital media, and feminism studies.

Macau K. F. Mak (MPhil, CUHK) is a PhD student at the School of Journalism and Mass Communication, University of Wisconsin-Madison. His research interests include political communication, new media, and media effects.