Combining Statistical and Causal Mediation Analysis

Handbook of Research Methods in Social and Personality Psychology

Amanda Montoya

University of California, Los Angeles

This manuscript is submitted to be published in *The Handbook of Research Methods in Social and Personality Psychology*, Third Edition edited by Harry T Reis, Tessa West, and Charles M Judd, Cambridge University Press.

Combining Statistical and Causal Mediation Analysis

Handbook of Research Methods in Social and Personality Psychology

Introduction

Mediation analysis has become a staple of social and personality psychology research in recent decades. Understanding causal processes is important to theory development in all areas of psychology, but because many processes in social and personality psychology are difficult to directly manipulate or control, mediation analysis has become the tool of choice for testing questions about psychological processes. Mediation analysis is useful for testing questions about "how" or "why" effects occur. For example, increasing growth mindset can improve performance, but how exactly does this effect occur? One proposed mechanism is increased attention to mistakes (Schroder et al., 2017), where growth mindset increases attention to mistake which in turn increases performance.

Traditional implementation of mediation analysis in psychology has largely followed a process described as "statistical mediation analysis". A seemingly alternative approach called "causal mediation analysis" has developed in statistics, computer science, and epidemiology. While the specific labels of "statistical" and "causal" mediation analysis are relatively new, the distinction between these two methods has not been clearly described for psychological researchers. Each method provides a different types of inference: statistical mediation analysis is focused on conducting statistical inference—testing whether a quantity is non-zero at the population level, while causal mediation analysis is focused on causal inference (evaluating whether a sample value is an unbiased estimate of a causal effect at the population level).

In this chapter, I argue that mediation analysis practices in social and personality psychology could benefit from the integration of practices from statistical mediation analysis and causal mediation analysis. I briefly describe each method on its own, then provide recommendations for how to integrate practices from each method to simultaneously evaluate statistical inference and causal inference as part of a single

analysis. At the end of the chapter, I describe additional areas of recent development in mediation analysis that social and personality psychologists should also pursue to improve the quality of inference in their mediation analysis: latent variables and longitudinal models. Ultimately, this chapter is meant to be a kind introduction to causal inference in the context of mediation with very practical recommendations for how to implement these practices in research. Much of the details of these methods can be found in the cited work, and I encourage researchers interested in implementing these practices to familiarize themselves with the original work, acknowledging that much of the published literature in causal inference relies on a strong mathematical background.

"Statistical" mediation analysis

"Statistical" mediation analysis is an analytical process which typically involves estimating linear models using ordinary least squares (OLS) regression or structural equation models (SEMs). These statistical models provide an estimate of an indirect effect. The indirect effect is of greatest interest for testing a mediation hypothesis because it quantifies the effect of an independent variable on an outcome variable through a specified mediator variable (e.g., the effect of X on Y through M). Statistical inference is made to support a claim of "mediation" or "presence of an indirect effect" at the population level.

Statistical mediation analysis has become increasingly common in psychology and particularly social psychology over the last 40 years. Popularized through the seminal paper by Baron and Kenny (1986), mediation analysis has become a staple for researchers interested in evaluating processes. In this section I start with estimating the models using linear regression, then define total, direct, and indirect effect pathways, how to include covariates in the analysis, and inferential methods. Finally, I will conclude this section with some common critiques of statistical mediation analysis.

Estimating Equations with Linear Regression

Throughout this chapter, we will only consider a very basic model with a single mediator and a single outcome, but extensions beyond this simple model are possible,

popular, and sometimes preferred (Preacher & Hayes, 2004; Hayes, 2015). Consider the case where we have a single independent variable (X) which we believe has a causal effect on an outcome variable (Y) through a proposed mediator (M). The independent variable could be an experimental manipulation or an observed variable; however, as will be discussed later, experimental manipulations are preferred for aiding causal inference. In this chapter, we will focus on continuous outcomes and mediators, though it is possible to have categorical variables and non-normal distributions (Preacher, 2015; Iacobucci, 2012). This chapter focuses on studies where the X, M, and Y variables are observed once for each participant[1].

Two regression equations are used to represent the effects of interest in the mediation analysis. First is the model predicting the mediator using the independent variable:

$$M_i = a_0 + a_1 X_i + e_{M_i} \tag{1}$$

In this equation the dependent variable is the mediator (M). The model includes an intercept (a_0), which is the predicted value of the mediator when X is zero. The coefficient a_1 represents the predicted difference in M for two cases which differ by 1 unit on X. Lastly, the residual, e_{M_i}, represents the difference between the predicted value and the observed value on the outcome M for each case. We typically assume these residuals are normally distributed with a mean of zero and an unknown variance $\sigma^2_{e_M}$. Note that the subscript i is used to represent quantities which are specific to each case, and values with no i subscripts are parameters that are estimated in the model. When referring to estimates from specific models "hat" notation will be used (e.g., $\hat{a}_1$ is the sample estimate of the population parameter a_1).

The second regression equation is a model predicting the outcome variable using both the mediator and the independent variable:

[1]It is also possible for mediation analysis to be conducted in repeated measures (Montoya & Hayes, 2017) and longitudinal designs, as discussed at the end of this chapter.

$$Y_i = c'_0 + c'_1 X_i + b_1 M_i + e_{Y'_i} \quad (2)$$

In this equation the dependent variable is the outcome (Y). The model includes an intercept (c'_0), which is the predicted value of the outcome when X and M are both zero. The coefficient c'_1 represents the predicted difference in Y for two cases which differ by one unit on X but have the same score on M. Similarly, b_1 represents the predicted difference in Y for two cases which differ by one unit on M but have the same score on X. Finally, the residual, $e_{Y'_i}$, represents the difference between the predicted and observed value on the outcome for each case. This residual is typically assumed to be normally distributed with a mean of zero and an unknown variance $\sigma^2_{e_{Y'}}$.

These two equations can be estimated in any OLS regression software, or as part of a SEM program. The coefficient estimates from these models are then used to define the total, direct, and indirect effects which are important for mediation analysis.

Path Analysis and Indirect Effects

Equations 1 and 2 can be represented using a path diagram like Figure 1. This figure visually represents how these models estimate the potential pathways from X to Y. One pathway is the *direct effect*, c'_1. The direct effect does not go through the mediator, rather directly from X to Y. The interpretation of c'_1 is the expected difference on Y between two cases that differ by one unit on X and have the same score on Ms. Because the cases being considered do not vary on M, there is no way for this path to capture variability due to M. The second pathway is by tracing from X to M on the a_1-path, and then from M to Y on the b_1-path. This represents the *indirect effect*. The indirect effect is quantified by the product of a_1 and b_1 (i.e., $a_1 b_1$), and this product is the expected difference on Y for two cases that differ on X by one unit, solely due to the effect of X on M and the subsequent effect of M on Y. Consider an example where $a_1 = 2$ and $b_1 = 3$: A one unit difference on X leads to a 2-unit difference on M. The coefficient b_1 means that a one-unit difference on M leads to a 3 unit difference on Y. To calculate how much X affects Y through M we

take the product of a_1 and b_1 ($a_1 b_1 = 6$). The 1-unit difference on X results in a 2-unit difference on M which results in a 6-unit difference in Y.

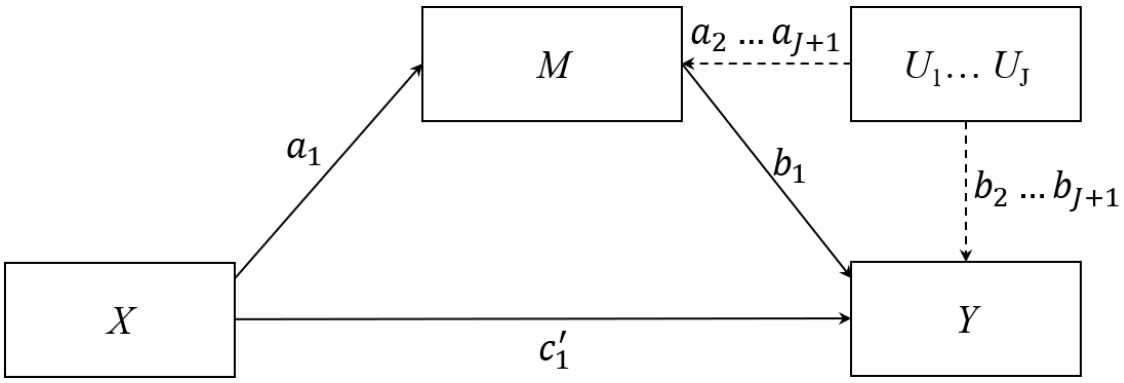

Figure 1. A path diagram describing a simple mediation model where X indirectly affects Y through M. If covariates U are included in the model, path diagram would also include paths denoted with dashed lines.

The indirect effect and direct effect sum together to a quantity called the *total effect* (i.e., $a_1 b_1 + c_1' = c_1$), where c_1 is the expected difference in Y between two cases that differ by one unit on X. The total effect can be estimated using a separate regression equation:

$$Y_i = c_0 + c_1 X_i + e_{Y_i} \tag{3}$$

The notation for this equation aligns with the previous ones. Notably, c_1 estimates the expected difference in Y between two cases who differ by one unit on X. This could be interpreted as the "main effect" of X on Y. While this quantity is not directly important for mediation analysis, it demonstrates how the main effect of X on Y is partitioned into two parts: the direct and indirect effect of X on Y.

Including Covariates

It is very common in linear regression analysis and statistical mediation analysis to include covariates as part of the models. There are multiple reasons why these covariates could be included: accounting for potential joint causes of some of the variables in the

model and improving statistical power. The three estimating equations for the mediation analysis can accommodate covariates as follows:

$$M_i = a_0 + a_1 X_i + \sum_{j=1}^{J} a_{1+j} U_{ji} + e_{M_i} \quad (4)$$

$$Y_i = c_0' + c_1' X_i + b_1 M_i + \sum_{j=1}^{J} b_{1+j} U_{ji} + e_{Y_i'} \quad (5)$$

$$Y_i = c_0 + c_1 X_i + \sum_{j=1}^{J} c_{1+j} U_{ji} + e_{Y_i} \quad (6)$$

In all three equations above, J covariates can be incorporated into the equation as the variables $U_{1i}...U_{Ji}$ with their coefficients $a_2...a_{J+1}$ in the Equation 4, $b_2...b_{J+1}$ in Equation 5, and $c_2...c_{J+1}$ in Equation 6. When the same covariates are included in all three equations the indirect and direct effect still sum to the total effect, but this is not true if different covariates are used in different equations. Including the covariates the interpretations of the primary coefficients in the indirect effect change. The interpretation of a_1 is now the predicted difference in M for two cases which differ by 1 unit on X but have the same score on all covariates $U_1...U_J$. Similarly, the b_1 coefficient is now interpreted as the predicted difference in Y for two cases which differ by one unit on M but have the same score on X and all covariates, $U_1...U_J$. The indirect effect is now the expected difference on Y for two cases that differ on X by one unit but are the same on all covariates, $U_1...U_J$, solely due to the effect of X on M and the subsequent effect of M on Y. This is another area where including different covariates for different equations could make it very difficult to interpret the indirect effect. Figure 1 incorporates the covariates by adding in the lines indicated with dashes to the previous path diagram.

Inference

If, at the population level, the indirect effect is zero, this means that M is not a mechanism through which X affects Y. Inference about indirect effects from a sample is needed to provide evidence for or against research questions about mechanisms. Some of

the earliest approaches to inference in mediation analysis were proposed in the 1980s (e.g., Sobel, 1982; Baron & Kenny, 1986) and new and different methods continued to develop over the subsequent decades (Shrout & Bolger, 2002; Selig & Preacher, 2008). In the early 2000s there were a host of methods available, but very limited information about which methods worked better than others. Monte Carlo simulation research in this area has explored this issue deeply and come to a very clear consensus on methods that work well, and those that don't. In this section I summarize the findings of this work and point to software resources which implement the methods that work.

Inferential Methods that Work. Based on a variety of simulation studies there are four methods which seem to perform very similarly in most circumstances: percentile bootstrap confidence intervals (CIs), Monte Carlo CIs, distribution of the product, and joint significance test. I will briefly summarize each method and discuss software implementation. Notably the differences among these methods are very slight compared to the difference between methods used in this section and those included in the next section.

Percentile bootstrap CIs. This is a non-parametric method where many bootstrap samples are generated by sampling from the original sample with replacement. Thousands of resamples, which are all the same size as the original sample, are created, and the indirect effect is estimated in each resample. Specific percentiles of the distribution define the bounds of the CI. For example, if we have 1,000 resamples and want a 95% CI, the lowest 2.5% corresponds to the 25th lowest resampled indirect effect and the highest 2.5% corresponds to the 25th highest resampled indirect effect. Typically, this CI is examined to see if it includes zero, and if it does not, the null hypothesis that the indirect effect is zero is rejected; however, these CIs also provide information above and beyond this dichotomous decision making, including general magnitude and uncertainty around the estimate of the indirect effect. Common tools available which provide percentile bootstrap CIs include the PROCESS macro for SPSS, SAS, and R (Hayes, 2022) and the `psych` package in R (Revelle, 2022).

Monte Carlo CIs. This is a semi-parametric method which simulates the sampling distribution for $\hat{a}_1$ and $\hat{b}_1$, then generates a sampling distribution for $\widehat{a_1 b_1}$. This method relies on the assumption that $\hat{a}_1$ and $\hat{b}_1$ both have normal sampling distributions, which is the case under the typical assumptions of linear regression (linearity, homoskedasticity, independence, and normality of residuals). A normal distribution with a mean of $\hat{a}_1$ and a standard deviation of $\hat{s}_{\hat{a}_1}$, where $\hat{s}_{\hat{a}_1}$ is the estimated standard error of $\hat{a}_1$, is used to generate many samples to represent the sampling distribution of $\hat{a}_1$. Similarly, a normal distribution is used to generate many samples to represent the sampling distribution of $\hat{b}_1$, using a mean of $\hat{b}_1$ and standard deviation of $\hat{s}_{\hat{b}_1}$. Pairs of samples from each distribution are multiplied together to give an estimated sampling distribution for $\widehat{a_1 b_1}$. From this point, the process for defining the bounds of the CI is the same as the percentile bootstrap. Tools such as the PROCESS macro for SPSS, SAS, and R (Hayes, 2022) and the `Rmediation` package (Tofighi & MacKinnon, 2011) generate this output. Notably, Monte Carlo CIs are the preferred method of inference for multilevel mediation models (Rockwood & Hayes, 2022), so researchers interested in this extension may want to familiarize themselves with this method in this simpler case.

Distribution of the product. This is a method for making inference about estimates which are the product of two quantities which have independent, normal distributions, as is the case for indirect effects (MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002). This method shares the same assumptions as the Monte Carlo CI, but uses mathematical functions rather than simulation to generate confidence intervals and p-values. While this method is not as highly adopted as the two previous methods, it has been shown to perform well in previous simulations studies (MacKinnon, Lockwood, & Williams, 2004; Williams & MacKinnon, 2008; MacKinnon et al., 2002). Software which implements this method is available in the `Rmediation` package for R (Tofighi & MacKinnon, 2011) and SAS (MacKinnon, Fritz, Williams, & Lockwood, 2007).

Joint Significance test. This is a method derived from the Causal Steps Method (described below), which relies on discrete hypothesis tests for the individual components of the indirect effect (a_1 and b_1). Null hypothesis significance tests are used for a_1 and b_1 separately, and if both tests are significant at a pre-specified α-level, then the null hypothesis that the indirect effect is zero is rejected. This method performs well with respect to making correct inferential decisions (Yzerbyt, Muller, Batailler, & Judd, 2018; Hayes & Scharkow, 2013), but does not generate a direct quantification of the indirect effect and does not provide a quantification of the uncertainty around the indirect effect, like the other methods described in this section. One advantage to this method is that special software is not needed, any program which can do OLS regression can be used to conduct this test.

Inferential Methods to Avoid. In this section, I discuss a few inferential methods which, based on simulation studies, are not recommended for use. The focus is more on understanding the criticisms of these methods, rather than the general procedures.

Bias-corrected (BC) and Bias-corrected and accelerated (BCA) bootstrap CI. The BC and BCA bootstrap CI are two methods which were proposed for use with indirect effects as they should be more general and flexible than the percentile bootstrap CI (Shrout & Bolger, 2002; MacKinnon et al., 2004). These methods both follow the same procedure as the percentile bootstrap CI for generating bootstrap samples; however, the methods they use for selecting the boundaries of the CI differ. The bias-corrected method makes an adjustment aimed to center the CI around the "true" indirect effect, assuming that there may be bias in the sample estimate and that this bias propagates to the bootstrap estimates. The acceleration in the BCA is another factor which aims to account for skewness in a transformed distribution. While early work on the BC and BCA methods touted their high power and ability to detect indirect effects which other methods could not (Hayes & Scharkow, 2013), the tides have turned recently as there is increased concern about managing type I error over maximizing power. Both these methods suffer from

elevated type I error rates, such that when a researcher sets their α-level at .05, they may be operating under a type I error rate more like .07 or .08 (MacKinnon et al., 2004; Biesanz, Falk, & Savalei, 2010). This issue has become more important with the increasing criticisms in social and personality psychology regarding replicability, and concerns that many published results could be type I errors (Earp & Trafimow, 2015). As such these methods are being increasingly abandoned (e.g., PROCESS V3 changed the default inferential method from BC to percentile bootstrap CI). There is existing work trying to understand why these methods do not perform well with indirect effects and propose alternatives, but current proposals continue to see a trade off between power and type I error and do not demonstrate a clear advantage over the percentile method (Chen & Fritz, 2021; Tibbe & Montoya, 2022).

Sobel test. The Sobel test (sometimes called a "delta method") operates under the assumption that the indirect effect has a normal sampling distribution, which has been demonstrated to be the case in very large samples (Sobel, 1982), but not in small samples (Stone & Sobel, 1990). Given this assumption, a standard error for the indirect effect is calculated and a Z-statistic is used to generate a p-value or CI. Sample sizes sufficiently large for the Sobel test to work well (i.e., 2,000+) are not common in social and personality psychology, which means this method may not perform well in application. Previous research has found that the Sobel test has low power and performs worse than other methods (MacKinnon, 2000; Biesanz et al., 2010; Hayes & Scharkow, 2013). Notably, Sobel tests continue to be a predominant inferential method for indirect effects in SEM applications, though there is no existing evidence to suggest that the Sobel test performs better in this context, other than the expectation that sample sizes are larger for SEM applications.

Causal Steps. The causal steps approach is perhaps the most well-known method for inference in mediation among social and personality psychologists. In their seminal paper, exploring the differences between mediation and moderation, Baron and Kenny

(1986) propose a 4 step process for identifying mediation:

1. Does X effect Y? Evaluated using a hypothesis test for c_1.
2. Does X effect M? Evaluated using a hypothesis test for a_1.
3. Does M effect Y? Evaluated using a hypothesis test for b_1.
4. Does X effect Y after controlling for M? Evaluated using a hypothesis test for c_1'.

Steps 1 - 3 are used to determine whether or not there is mediation, where if all paths are significant this supports a claim of mediation. Step 4 determines the type of mediation, where if c_1' is closer to 0 than c_1 and non-significant, this is called "complete" mediation, and if c_1' is significant it is "partial" mediation. From this process it is not clear what to do if the direct effect is further from zero than the total effect and whether that qualifies as mediation at all. Many papers have criticized this method (Hayes, 2009; Meule, 2019; Rucker, Preacher, Tormala, & Petty, 2011). The requirement that the total effect be significant is often criticized as overly conservative. If the indirect effect and direct effect are of opposite signs, the total effect may not be distinguishable from zero even if the indirect effect is. Additionally, Kenny and Judd (2014) demonstrate that when the total effect and indirect effect are equal, researchers can have greater power to detect the indirect effect than the total. The distinction between complete and partial mediation is also criticized as misaligned with the substantive claim. First, complete mediation is the stronger claim, but relies on accepting the null hypothesis. Second, even if the direct effect is zero, this does not preclude other mediators not in the model from carrying indirect effects from X to Y, and so even evidence that the direct effect is zero does not support the claim that all possible mechanisms have been identified. Simulation research has also shown that the causal steps method is very low in statistical power, and sensitive to factors unrelated to the indirect effect (MacKinnon et al., 2002). As such, many have abandoned this method, in particular Steps 1 and 4, which leaves Steps 2 and 3 which is the joint significance test. While Baron and Kenny (1986) has had a long staying power within social and personality psychology, recent years have seen this method decrease in

popularity based on these critcisms.

Critiques of Statistical Mediation Analysis in Psychology

Mediation analysis picked up at lightning speed in psychology research, especially social psychology. In 2018 - 2020 about 43% of articles published in the *Journal of Personality and Social Psychology* included at least one mediation analysis (Charlton, Montoya, Price, & Hilgard, 2021). But in recent years, there has been increasing push-back against mediation analysis, specifically the way it is commonly applied in psychological studies. Many of these are critiques of the contexts in which the methods are applied, and how they are used to make claims beyond the scope of the analysis.

Perhaps the most common criticism of mediation analysis is with respect to the type of data it is commonly applied to: completely cross-sectional data. Here I use this term to describe cases where X, M, and Y are observed variables measured all at the same time. By contrast, longitudinal data would involve X, M, and Y observed repeatedly over time. Another differentiation is experimental data, where X is randomly assigned. It is also possible to have experimental-longitudinal data where X is randomly assigned, then M and Y are measured repeatedly over time (discussed later). The issues with application of mediation analysis to completely cross-sectional data are two-fold: 1) mediation is an inherently causal process, and 2) mediation is an inherently longitudinal process. These two issues are linked and in reality do not constitute two separable issues, but differentiating these aids the discussion of the solutions to the problems.

Mediation is a hypothesis about causal effects, but statistical mediation analysis cannot support causal claims. Researchers trained in psychology are consistently reminded that "correlation is not causation", and yet when it comes to statistical mediation analysis it is not always apparent whether causal claims can be supported purely by the results of regression/correlational analyses, and if they cannot, what is the utility of mediation analysis? Under specific, yet often untestable assumptions (described more in-depth in the next section), mediation analysis of cross-sectional data can provide unbiased estimates of

indirect effects, which are causal effects. Yet it feels almost tautological to start with assumptions that align with causality, then to use the results to support causal claims. Even when X is randomly assigned, there is ambiguity in the causal order of M and Y such that experimental data still cannot provide unambiguous evidence of causal indirect effects (Bullock, Green, & Ha, 2010). This leaves researchers in a difficult spot, because they want to use mediation analysis to support causal hypotheses, but without making untestable assumptions, the conclusions from these analyses must be very vague. Given these criticisms of the lack of causal interpretations from "statistical" mediation analysis with cross-sectional or even experimental data, some researchers have turned to an emerging area called "causal" mediation analysis (Rohrer & Aslan, 2021). In the next section, I turn to this topic of causal mediation analysis.

"Causal" mediation analysis

The primary distinction between statistical and causal mediation analysis is that the latter is concerned with defining the assumptions under which the indirect effect is truly estimating a causal effect, a process typically called *identification*. While this requires defining the estimate, typically called Pure Natural and Total Natural Indirect Effects (PNIE and TNIE), the focus is primarily on defining the assumptions under which this estimate is an unbiased estimate of a causal effect in the population. The theoretical development of these methods occured in mathematics, statistics, and computer science, rather than applied fields like psychology, education, and marketing. This perhaps leads to a disconnect between the literature on causal mediation analysis, and those who are trying to apply these methods. A variety of research teams have outlined different sets of assumptions required to identify indirect effects as causal (Pearl, 2001; Imai, Keele, & Yamamoto, 2010; Robins & Greenland, 1992; Vanderweele & Robins, 2007), which may lead to some confusion among applied researchers on which set to use. Ultimately, any set is correct, and researchers can follow any approach which helps them to understand their analytic process more clearly. In this chapter, I will follow the general approach proposed

by Pearl (2001) which relies on the potential outcomes framework (Rubin, 2005).

An indirect effect is ultimately the combination of multiple causal pathways, and each pathway must be causally identified to get an unbiased estimate of the indirect effect. As such, in this section, I take a piecewise approach: first describing the potential outcomes framework and assumptions needed to identify a treatment effect, then addressing how this approach can be generalized to combined pathways such as indirect effects. Ultimately, for a treatment effect to be causally identified we must assume there are no unmeasured confounders of that effect ($X \rightarrow Y$). Similarly, within the mediation we must also assume there are no unmeasured confounders of any of the three effects involved in the mediation analysis ($X \rightarrow Y$, $X \rightarrow M$, $M \rightarrow Y$), but also some additional assumptions arise.

Potential Outcomes Framework

The potential outcomes framework is used to define causal effects stemming from the broader question "What does it mean for some variable to cause another variable?" In the potential outcomes framework, a causal effect is defined by multiple outcomes which could potentially follow some event (X). For example, consider a study examining the effect of career framing on students preference for majors in STEM (Siy et al., in press). In the study, career framing could be randomly assigned such that half of the participants get the "passion" framing ("follow your passions", $X = 1$) and half get the "resource" framing ("pursue money and stability", $X = 0$). Prior to receiving any framing cues, each participant has two potential outcomes: If they are assigned to the passions framing their outcome score will be $Y_i(X = 1)$, but if assigned to the resource framing, their outcome score will be $Y_i(X = 0)$. The subscript i here denotes that each individual may have their own potential outcome scores, and the number in parentheses denotes the conditions under which this potential outcome score is observed. The causal effect of career framing on participant i's outcome score is $Y_i(X = 1) - Y_i(X = 0)$. This individual level effect is impossible to observe because each participant is only ever assigned to one condition. Hypothetically, if we were able to observe both of these potential outcome scores for

everyone in the population, the average of this causal effect of career framing over all cases in the population would be the average treatment effect (ATE). For simple causal inference problems (i.e., estimating the effect of one variable on another), it is typical to try to estimate the ATE. With an experimental design, an unbiased estimate of the ATE is the difference between the group means. We can use the mean from one group to estimate the missing potential outcome scores for the other group, and vice versa. Using the means to estimate the missing potential outcomes is successful because individuals are randomly assigned to condition, so there should be no systematic difference between individuals in the treatment and control conditions other than the treatment. An excellent introduction to the application of the potential outcomes approach to estimating ATEs in simple experiments is Rubin (1974).

Imagine, an alternative study with the same data structure, but a different design: instead of randomly assigning participants to experience a specific framing, participants were asked which framing was more common in their lives. Much more caution would be needed in interpreting the difference between the group means as an estimate of a causal effect. The difference between the group means is not a good estimate of the causal effect of framing because it is unlikely that the mean from one group will be an appropriate estimate of the missing potential outcome scores for the other group. There could be systematic differences between individuals in the two groups other than the grouping. However, the potential outcomes framework can help us articulate under what assumptions we could identify the ATE even with this non-experimental data. One of the primary assumptions is with respect to confounders: variables which causally affect both the group and the outcome. This is not of concern in a randomized study, because condition is randomized, so nothing can affect it. In the non-randomized study of career framing, in order for the difference in the means to be an unbiased estimate of the ATE, we would need to assume that there are no variables that affect both X and Y (Rubin, 1974).

Confounding. Confounding is a very important concept for causal inference, because confounders can lead to additional covariance among group and outcome which is not due to the group's causal effect on the outcome. From a causal inference perspective this is troubling, because we want to know how much X causes Y, and confounding could lead to over or under-estimation of that effect. For example, what if most of the participants who indicated that the resource frame was more common were also older on average than those who indicated that the passions frame was more common? Additionally, there may be differences in the racial-ethnic make-up of the two groups, such that, for example, there were more Black, LatinX, and Asian participants in the resource group and more White participants in the passions group. If age and race also influence interest in STEM, then these are two examples of potential confounders. When confounders are not taken into account, it may look like there is a causal relationship between X and Y, when in fact this relationship is confounded.

A confounded effect of X and Y occurs when both variables share a common cause (e.g., U). When U is not controlled for in a model where X predicts Y there will be a relationship between X and Y, which could be confused for a causal effect of X on Y, unless U is included in the model as a covariate. Even if there is a causal relationship between X and Y, if the two variables also share a common cause U, then U needs to be included in the statistical model to provide an unbiased estimate of the causal effect of X on Y[2].

When confounding is suspected, controlling for U (whether this is one or multiple variables) will help provide a better estimate of the causal effect of X on Y. It is possible to account for potential confounders by controlling for them in a statistical model. To include confounders in our models we must measure them, and often assume they are measured without error. This would result in a different estimator for the ATE (e.g., an

[2]This is why X needs to be included in the model for Y when we estimate the b_1-path in mediation analysis.

adjusted mean difference) and require the assumption of no-omitted confounders. Using linear regression to account for confounding is very common; however, it is important to acknowledge that this assumes that the effects of the confounder on both the condition and outcome are linear and the confounder is measured without error. Alternative approaches for adjusting for confounders have been developed in the causal inference literature such as inverse probability weights (Thoemmes & Ong, 2015), propensity score matching (Rosenbaum & Rubin, 1984; Thoemmes & Kim, 2011), or doubly-robust estimation (Funk et al., 2011). However, these methods are beyond the scope of this article, but recommended for researchers interested in learning more about how to account for confounding using methods which are more robust to violations of linearity assumptions.

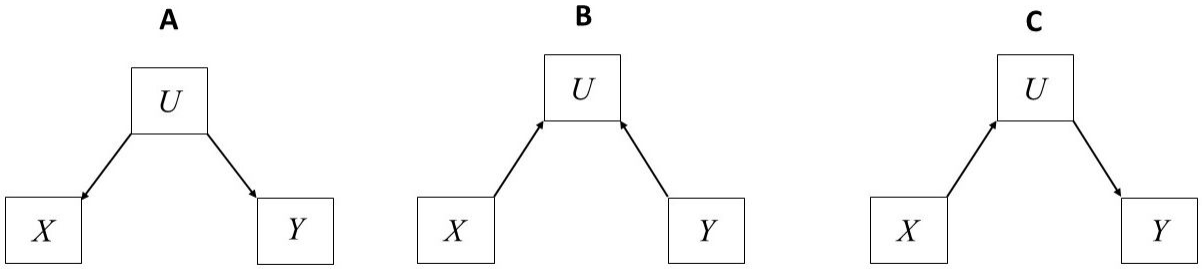

Figure 2. Examples of confounder, collider, and mediator relationships. A is an example where U is a confounder of the effect of X and Y. B is an example where U is a collider of X and Y. C is an example where U is a mediator of the effect of X on Y.

There are two kinds of variables which should not be controlled for when trying to estimate the causal effect of one variable on another: colliders and mediators. A collider is a variable which is caused by both X and Y (Figure 2 Panel B). The name suggests that the causal pathways from each variable "collide" at this variable. Conditioning on a collider can cause bias in the estimate of a causal effect, either suggesting there is an effect when one does not exist, or suggesting there is not an effect when one does exist. Similarly, controlling for mediators can bias the estimate of the causal effect (Figure 2 Panel C), because variability in the outcome which is due to the indirect effect will be misattributed.

Including covariates in a statistical model can be risky without thinking about the potential role that each variable plays from a causal perspective (Wysocki, Lawson, & Rhemtulla, 2022). Each covariate in the statistical model should be justified based on it's hypothesized role in the causal structure of the data.

Confounding is a core issue within the potential outcomes framework; however, the "no omitted confounders" assumption is not the only assumption needed to identify causal effects, even ATEs. In addition, SUTVA (Stable Unit Treatment Value Assumption) is needed, so in the next section I describe this assumption and it's implications. Notably, the following section will describe the particulars of causal mediation analysis which all build off of the approach to identifying ATEs, but now with a focus on indirect effects. Both confounding and SUTVA will come into play for these models as well.

Stable Unit Treatment Value Assumption (SUTVA). The potential outcomes framework relies on a specific assumption: the "Stable Unit Treatment Value Assumption" (SUTVA). There are two important implications of this assumption: no spillover and consistency. No spillover means that an individual's treatment response does not depend on the treatment of other cases. This assumption can be violated in cases with nested data. An example from social psychology might suggest that the causal impact of providing one member of a hiring committee anti-bias training, may depend on whether the other members of the committee were also assigned to training. Spillover is possible regardless of sample size, but can be affected by population size. Ensuring that the participants in the study are sampled from a large enough population can help reduce the likelihood that participants within the study affect each other. When studying small populations or groups of participants within some kind of cluster (e.g., households or classes), spillover is very important to consider.

Consistency is the idea that there are not multiple kinds of the treatment which are different from one another and would lead to different potential outcomes (Imbens & Rubin, 2015). This is particularly important when considering the variability of

experiences in the natural world. For example, in the observational version of the career framing study, you may have a participant who has only experienced the passions framing for their entire life. They would respond that the passions framing is most common for them, and have an associated response on the primary outcome. But, if their response on the outcome would be different than if they had generally had some exposure to both framing but experienced the passions framing more than the resource framing, this would be a violation of consistency. In both cases their level of the treatment is the same (passion framing) but if multiple kinds of the treatment (e.g., relative frequency of exposure to different framings) results in different outcomes, this is a violation of consistency. Similarly, this can occur in an experimental settings. Consider a mental health intervention where participants are randomly assigned to either get access a mindfulness app or a waitlist control, but otherwise are not provided additional information about what modules to complete within the app. If a participant might respond differently to the app depending on which modules they complete, this is a violation of consistency, because there is indeed not a single potential outcome which captures the participant's expected response if they use the app, because it depends on *how* they use the app. Consistency can be particularly troubling when participants engage in treatment to different degrees, but can in some ways be addressed by moderation analysis if the reason for different treatments can be measured and analysed.

Pure Natural and Total Natural Indirect Effects

Indirect effects are more complex than ATEs but can still be defined in the context of a difference in potential outcomes. The "what if" statement for ATEs is relatively simple: "What if the condition was different; would the outcome score be different too?" The statement for indirect effects is more complex. We must consider, "What would happen to the outcome if the mediator was at the level we would expect under the other condition?" A proposed mediator for the career framing example is drawing on your female role-congruent self (Siy et al., in press). For women in particular, being told to "follow your

passions" might cause them to draw more on their feminine identity, which would then decrease their interest in STEM fields (if, as research of stereotypes of STEM suggests, STEM fields tend to be perceived as incongruent with feminine identities). Considering the indirect effect of the "follow your passions" condition compared to the "resource driven" one, we consider the question "What if (all other things being equal) women drew on their role congruent selves just as much in the follow your passions frame as they do in the resource driven frame? What impact would that have on interest?" It is perhaps clear that this is a much more complex comparison of potential outcomes, compared to a effect of condition on the outcome, so in the following section I break this idea down more concretely.

To describe indirect effects in the potential outcomes framework requires nested potential scores. First, consider the potential mediator scores: $M_i(X = 0)$ under the control condition (e.g., "money and stability") and $M_i(X = 1)$ under the treatment condition (e.g., "follow your passions"). The outcome depends on the value of the mediator and the condition, so we have $Y_i(X, M)$ where X could be 0 or 1. M can take on it's two potential mediator values ($M_i(X = 0)$ or $M_i(X = 1)$), as these are the only two possible values within the study due to the two possible values of X.[3] So consider participants who are in the control condition ($X = 0$) we want to know the difference between their outcome if the mediator was the potential mediator score under the treatment compared to the potential mediator score under control:

$$Y_i(X = 0, M = M_i(X = 0)) - Y_i(X = 0, M = M_i(X = 1)) \tag{7}$$

This is called the Pure Natural Indirect Effect (PNIE). The term *pure* is used in causal inference when referring to cases in the control condition. Notice that this difference examines the difference in the outcome Y if we could change each participant's mediator value to the value under the treatment condition ($M_i(X = 1)$), but everything else about

[3]There is also a type of potential outcome which is called a "controlled" effect where M can take on any specified value, but these are not used for calculation of indirect effects.

them is consistent with the control condition ($X = 0$). For example, if we could change a person to draw on their female role congruent self just as much as they did when they were in the passions condition, but otherwise they experienced the resource condition, how different would the outcomes be? Then, to average this effect over all participants, we get an aggregate estimate of this PNIE.

When we consider participants under the treatment condition ($X = 1$), we want to know the difference between their outcome if the mediator was the potential mediator score under the treatment compared to under control:

$$Y_i(X = 1, M = M(X = 0)) - Y_i(X = 1, M = M(X = 1)) \tag{8}$$

This is called the Total Natural Indirect Effect (TNIE). The term *total* is used in causal inference to refer to differences in potential outcomes under the treatment condition ($X = 1$). This difference is very similar to the PNIE, in that it is examining the difference in potential outcomes if we change the mediator from the value in the control to the value in the treatment, but otherwise the person reacts as if they are in the treatment condition. An estimate of the aggregate of this effect across the whole population, provides the TNIE.

Consider for a moment, the linear models assumed in the previous section in Equations 1 and 2. In this case the predicted potential mediator score for M in each condition would be:

$$M(X = 0) = a_0 \tag{9}$$

$$M(X = 1) = a_0 + a_1 \tag{10}$$

These potential mediator scores can then be plugged into Equation 2 to generate the predicted difference in potential outcome scores as defined by the PNIE:

$$\begin{aligned} &Y(X = 0, M = M_i(X = 0)) - Y(X = 0, M = M_i(X = 1)) \\ &= (c'_0 + c'_1 \times 0 + b_1 a_0) - (c'_0 + c'_1 \times 0 + b_1(a_0 + a_1)) = b_1 a_1 \end{aligned} \tag{11}$$

The derived PNIE is the same as the indirect effect from the statistical mediation section. Completing the same calculation process with the TNIE, the result is the same.

$$Y_i(X = 1, M = M_i(X = 0)) - Y_i(X = 1, M = M_i(X = 1)) =$$
$$(c'_0 + c'_1 \times 1 + b_1 a_0) - (c'_0 + c'_1 \times 1 + b_1(a_0 + a_1)) = b_1 a_1 \quad (12)$$

Why then does causal mediation analysis differentiate between the PNIE and the TNIE? In causal analysis the goal is typically to make as few assumptions as possible. The linear models most typically used in social and personality psychology for mediation analysis require many assumptions to generate an unbiased estimate of the indirect effect. In particular, these estimating equations assume that all the relationships are linear and errors are independently and identically distributed. Additionally, there is an assumption in Equation 2 that the relationship between M and Y is constant across values of X (i.e., there is no XM interaction). Because this assumption is very restrictive, it is typically not made for causal mediation analysis, instead a different linear model could be estimated:

$$Y_i = c'_0 + c'_1 X_i + b_1 M_i + c'_2 X_i M_i + e_{Y'_i} \quad (13)$$

Under this model, following the previous process we find that the PNIE and TNIE are different:

$$\begin{aligned}
PNIE &= Y_i(X = 0, M = M_i(X = 0)) - Y_i(X = 0, M = M_i(X = 1)) \\
&= (c'_0 + c'_1 0 + b_1 a_0 + c'_2 a_0 0) - (c'_0 + c'_1 \times 0 + b_1(a_0 + a_1) + c'_2(a_0 + a_1)0) \\
&= b_1 a_1 \\
TNIE &= Y_i(X = 1, M = M_i(X = 0)) - Y_i(X = 1, M = M_i(X = 1)) \\
&= (c'_0 + c'_1 \times 1 + b_1 a_0 + c'_2 a_0 1) - (c'_0 + c'_1 1 + b_1(a_0 + a_1) + c'_2(a_0 + a_1)1) \\
&= b_1 a_1 + c'_2(a_0 + a_1)
\end{aligned}$$

Thus, in causal mediation analysis even with a linear model, there is no single indirect effect, but rather two different indirect effects depending on the condition —

treatment or control. This equivalence is described more deeply in MacKinnon, Valente, and Gonzalez (2020), but the distinction is important for understanding the differences between statistical and causal mediation analysis. Because the philosophy in causal mediation analysis is to estimate indirect effects under the most general assumptions possible, a distinction between the two indirect effects is needed. If instead, researchers are comfortable assuming that there is no XM interaction (a testable assumption), then a single estimate is sufficient. Vo, Superchi, Boutron, and Vansteelandt (2020) recommend testing the interaction using a hypothesis test prior to conducting the mediation analysis to select the model for Y.[4]

Assumptions for Identification

Pearl (2001) outlined a set of four assumptions under which the PNIE and TNIE are identified:

1. No unmeasured confounders of the effect of X on Y conditioned on covariates
2. No unmeasured confounders of the effect of M on Y conditioned on covariates and X
3. No unmeasured confounders of the effect of X on M conditioned on covariates
4. No measured or unmeasured confounders of the effect of M on Y which are affected by X conditioned on covariates.

These first 3 assumptions are very similar, in that they assume that there are no unmeasured confounders (common causes) of the effect of one variable on another. None of

[4]It is worthwhile to note that causal mediation analysis typically does not use the linear model in Equation 2, but rather non-parametric estimation approaches like sequential g-estimation or inverse probability weighting are used to estimate the models (Valente, Pelham, Smyth, & MacKinnon, 2017). However, for the purposes of this chapter, I do not explore these options as they are not required to undertake causal mediation analysis, and the overwhelming adoption of linear models throughout psychology suggests that using these models to describe causal mediation analysis may help to elucidate the differences in approaches which are not due to estimation procedure. The future of mediation analysis in psychology could involve adoption of these more non-parametric approaches.

the paths in the mediation model can be confounded, otherwise there will be bias in the indirect effect. Examining Figure 3, this means that any variable of the form U_1, U_2, U_3 must be measured and included in the model with the correct functional form. The fourth assumption requires some additional explanation. Similar to the second assumption, Assumption 4 focuses on confounders of the M-Y relationship, but considers a specific kind of confounder of this relationship: one which is affected by X. This type of variable is represented by U_4 in Figure 3. Variables of this form are troublesome because if measured and included as a covariate in the model, part of the indirect effect ($X \rightarrow U_4 \rightarrow M \rightarrow Y$) will not be included in the indirect effect estimate, but if unmeasured it will bias the M-Y relationship. An alternative way to describe this kind of variable is an alternative mediator. While this chapter does not directly discuss models with multiple mediators, there are resources for handling such conditions within a causal mediation framework (Daniel, Stavola, Cousens, & Vansteelandt, 2015; VanderWeele & Chiba, 2014).

Assumptions under Random Assignment of X. In psychology research it is very common to run experiments where the causal variable of interest is randomly assigned to participants. Due to the ubiquity of this design, it is worth discussing the assumptions required to test mediation using this design. Researchers often associate the use of an experiment with valid causal inference, and as can be seen in Figure 3, where the dashed lines indicate paths removed by randomization on X: confounders like U_1 and U_3 are no longer of concern when X is randomized. But randomization of X does not rule out the potential for confounders like U_2 and U_4. So, while it is true that the ATE is identified under random assignment of X, it is not the case that the PNIE or TNIE is identified given random assignment alone. There are still untestable assumptions required to make causal inference about indirect effects when X is randomized. Under successful random assignment, Assumptions 1 and 3 are met but Assumptions 2 and 4 are unaddressed. Researchers will need to discuss why they believe these assumptions are met if causal inference is desired, and what the implications would be if these assumptions are not met.

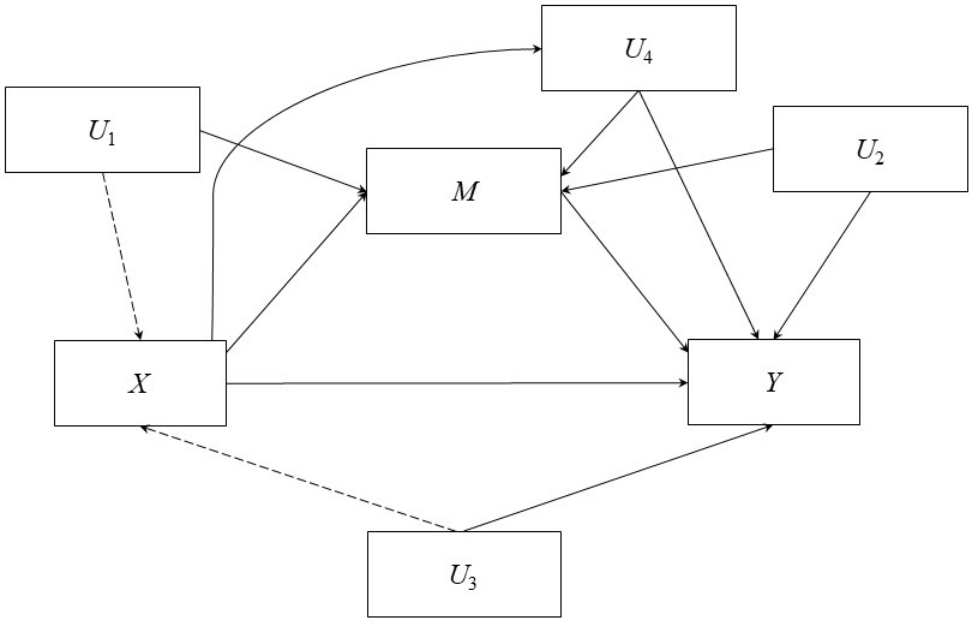

Figure 3. Example of confounders aligning with each of the four assumptions for identification of indirect effects. Dashed lines are paths which are removed by randomization of X.

One method for evaluating possible violations of these assumptions, is sensitivity analysis.

Sensitivity Analysis

Causal mediation analysis clearly outlines the assumptions required for the indirect effects to be identified, but all of these assumptions are untestable — it is impossible to know if they are true or not, even given population data. A method called sensitivity analysis aims to provide some information about whether the indirect effect can still be identified under a certain amount of confounding (Imai et al., 2010). While this is not a direct test of the assumption, it provides some information about the degree of assumption violation which would be needed to contradict the claims from the analysis.

Sensitivity analysis is particularly well developed for the case where X is randomized, and therefore the greatest concern is in confounding of the M-Y relationship. When there is a confounder of the M-Y relationship, this will induce a correlation between the residuals of the two outcomes (e_{M_i} and $e_{Y_i'}$). This is because there is some factor (e.g., U_2) which causes M and Y but is not included in the model, so the residuals will capture that shared explanation. The basic idea behind sensitivity analysis is to quantify the largest possible correlation between the residuals for which the indirect effect is still identified (Imai et al., 2010). This effect is typically expressed as a correlation coefficient. If, for example, even a small confounding effect (e.g., $\rho = .05$) would result in an unidentified indirect effect, it may not be feasible to assume that there are no confounders which individually or collectively have such a small effect. If instead the correlation is large (e.g., $\rho = .8$) it may be feasible to argue that missing such a large effect in the model is not plausible. However, especially for this second case, it is important to remember that it may not just be a single confounder, but rather a set of confounders (perhaps each individually inducing correlations of .3 or .4) which generate such an effect. Ultimately, sensitivity analysis is not designed to provide direct support for assumptions, but rather provide additional information for researchers and consumers to evaluate the plausibility of the assumptions made and the impact these assumptions have on the causal claims made about indirect effects.

Available software can implement sensitivity analyses for mediation analysis where X is randomly assigned. The `mediation` package in R can conduct sensitivity analysis when X is randomly assigned, and just the M-Y path needs to be evaluated for potential confounding (Tingley, Yamamoto, Hirose, Keele, & Imai, 2014). Currently, there are not tools for evaluating simultaneous confounding for multiple relationships in the mediation model, meaning that when completely cross-sectional data is used for mediation analysis it will be difficult to evaluate simultaneous violations of the assumptions. This issue adds to the growing concern about using completely cross-sectional data for evaluating mediation hypotheses.

Mediation Analysis in the Modern Age

Researchers planning on conducting mediation analysis should not feel the need to choose between "statistical" and "causal" mediation analyses. These two methods are complimentary and can be used in concert to generate research findings. In this section I outline how researchers might utilize tools from both frameworks when planning, conducting, and reporting mediation analysis.

Planning a Mediation Analysis

When research is intended to evaluate a mediation hypothesis, the design of the study can be leveraged to help meet some of the assumptions for causal interpretation. For example, a study where X is randomly assigned will help meet two of the four assumptions of mediation analysis. There are other study designs which improve causal inference in estimating indirect effects. For example, Bullock and Green (2021) suggest generating versions of the treatment which do and do not affect the mediators. If there could be a version of the treatment which is the same in all other respects but does not affect the mediator, and this could be compared to a version which does affect the mediator.[5] Another example of a design-based approach to mediation analysis comes from Imai, Tingley, and Yamamoto (2013) who propose a variety of designs which involve manipulation of the mediator (see also Valente et al., 2017). A core takeaway here is if the primary goal of a study is to evaluate a mediation hypothesis, a simple experimental design where X is randomized is not optimal (though is still preferable to a completely cross-sectional design). The above discussed designs should be used instead. However, researchers are frequently testing multiple hypotheses in a single study, and most of the above proposed designs will not be appropriate for evaluating hypotheses about, for

[5] Notice the link here between this approach, which generates observations of the outcome under treatment that does effect the mediator (i.e., $Y(X = 1, M = M(X = 1))$) and under treatment but where the mediator is the same as it would be under control (i.e., $Y(X = 1, M = M(X = 0))$). This method generates an estimate of the TNIE.

example, total effects, or interaction effects.

An important contribution of the causal mediation analysis approach is clearly articulating the assumptions required for the identification of indirect effects. Evaluating the feasibility of these assumptions in advance can ensure a stronger study. For example researchers should identify potential confounders of the relations in the model and make sure to measure these and evaluate how to include them in the model. One tool for identifying confounders is Directed Acyclic Graphs (DAGS; Pearl (1995), for an approachable introduction see Rohrer (2018)). Additionally, Assumption 4 means there cannot be measured or unmeasured variables which are affected by X and confound the M-Y relationship. Researchers can consider possible variables which might fit this description prior to conducting their research, and alter the manipulation to avoid impacting these variables.

A growing body of research on statistical power in mediation analysis suggests that tests of indirect effects tend to be low in statistical power. This is true for both statistical mediation analyses (Fritz & MacKinnon, 2007; Charlton et al., 2021; Vo et al., 2020) and causal mediation analysis (S. H. Liu, Ulbricht, Chrysanthopoulou, & Lapane, 2016). The interaction term in causal mediation might cause additional concern, because interactions typically have small effects and tests of interaction are also low in power (Aguinis, Beaty, Boik, & Pierce, 2005). Researchers should conduct power analyses to either select a sample size appropriate for their planned analyses or to evaluate what size of effects they are sufficiently powered to detect given their planned sample size, which may be limited based on other factors (e.g., limited available population, financial resources). Power analysis tools for statistical mediation analysis abound (Zhang & Yuan, 2018; Schoemann, Boulton, & Short, 2017; Zhang & Wang, 2013; Aberson, 2019; Kenny, 2017; X. Liu & Wang, 2019). Notably, all these tools assume no XM interaction, and so additional power analyses for the interaction might be completed separately, or the Shiny App from Qin (under revision) can be used for models which include the XM interaction.

Finally, with respect to planning, researchers intending to conduct a mediation analysis should consider preregistering their plan. Recent research on mediation analysis in psychology and marketing suggests there is potential evidence for questionable research practices (Götz, O'Boyle, Gonzalez-Mulé, Banks, & Bollmann, 2021; Charlton et al., 2021). Mediation analysis has been criticized throughout the field of psychology and beyond for its excessive flexibility. Mediation analysis has many potential researcher degrees of freedom, especially given the potential to swap around variables in their roles as outcomes, mediators, and covariates. Researchers can preregister using a service like the Open Science Framework or AsPredicted. Preregistration creates a time-stamped record of a research plan which is publicly accessible (though an embargo can be put on a preregistration while the study is underway). Many journals in social and personality psychology offer badges to papers with preregistered studies. Preregistrations of mediation analyses can include the planned sample size, α level, role of different variables in the analysis (e.g., condition, mediator, outcome, covariate), inferential method and any important specifications for that method (e.g., how many bootstraps, seed number), and plans for sensitivity analysis. For sensitivity analysis, researchers should provide some benchmarks for correlations which they would deem acceptable for evaluating potential confounding (e.g., we believe that correlations greater than .5 would suggest robustness to reasonable levels of confounding).

Overall, in planning a mediation analysis researchers can integrate elements of both causal and statistical mediation analysis by planning the design of their study to maximize causal support, select covariates to account for confounding, select a sample size which is appropriately powered for the study analyses (including detecting the XM interaction), and preregistering this plan.

Conducting a Mediation Analysis

Some practices which may be common among practicing social and personality researchers should be left at the wayside. Given strong hypotheses and preplanning of the analysis, researchers should avoid iterating through different variables in different roles

(e.g., mediators, outcomes, and covariates) unless this process will be explicitly described in the final product and described as exploratory. The recommended inferential methods include percentile bootstrap confidence intervals, Monte Carlo confidence intervals, and the distribution of the product method. Researchers should also pre-select their preferred inferential method and avoid trying multiple different inferential methods or rerunning methods which rely on random number generators (e.g., bootstrapping) as this may result in conflicting results and tempt the researcher to select the result most aligned with hypotheses.

In the remaining part of this section I focus on two elements of the mediation analysis process which are not currently common practice: evaluating the XM interaction and sensitivity analysis. Though this chapter does not focus much on software implementation, it is worthwhile to note that the PROCESS macro for SPSS, SAS, and R (Hayes, 2022) allows for estimation of the XM interaction, specification of covariates, and two of the recommended inferential methods (percentile bootstrap and Monte Carlo CI).

Testing the XM interaction. One practice that is not currently common among researchers doing mediation analysis is testing the XM interaction (Vo et al., 2020). Common practice in psychology is to not include interactions unless directly hypothesized (Yzerbyt, Muller, & Judd, 2004), causal mediation analysis proponents suggest including the XM interaction by default (Vo et al., 2020). Two alternative approaches could be adopted: testing prior to fitting the mediation model or using this test as a robustness check. Ultimately, more research is needed to establish best practices in evaluating the XM interaction.

It is generally recommended to evaluate whether including the XM interaction is important prior to conducting the mediation analysis. This could be done using a statistical test, effect size evaluation, or some other means. To test the XM interaction prior to fitting the mediation model, a researcher would fit Equation 13 and conduct a hypothesis test on the c_2' coefficient, to determine if there is a significant interaction. Based

on the test of the interaction, researchers can select the model for Y they would like to use. If the interaction is significant, it should be included in the model for Y and statistical inference for each indirect effect should be conducted separately. If it is not significant, it may not be needed in the model for Y and statistical inference can be made about the single indirect effect. This process of testing the interaction prior to analysis has not been extensively studied in simulation research, and may result in decision errors especially when power to detect the interaction is low. An alternative approach would be evaluating the effect size of the interaction (e.g., R^2 or η^2) and preregistering a minimum value for the effect size which would result in including the interaction in the model.

Rather than using information from the data to select a model, this decision could be made *a priori* by selecting the model which is believed to be most probable (interaction or no interaction). Then as a robustness check, the alternative model can be fit and reported. If the conclusions are dependent on the choice of including or excluding the interaction, this can be noted as a limitation. When using this robustness check approach, researchers should preregister their analysis plan, including the plan to include or exclude the interaction, as reporting both analyses (or only the one that resulted in a significant indirect effect) may be perceived as p-hacking.

Covariates and Sensitivity Analyses. Researchers should include covariates in their models which they believe could be potential confounders of the effects in the model. Whether including covariates or not, sensitivity analyses are useful for evaluating the extent of a violation of one of the no-confounder assumptions which would result in a lack of identification for the indirect effect. Sensitivity analyses should be conducted for any paths which, based on the experimental design, could be confounded. Even when X is randomly assigned, sensitivity analyses for the assumption of no unmeasured confounding of the $M - Y$ relationship (Assumption 2) are important to include. Many excellent resources are available which provide introductions to sensitivity analyses in mediation (VanderWeele, 2010; Smith & VanderWeele, 2019; Tingley et al., 2014; Imai et al., 2010;

Valente et al., 2017).

Reporting a Mediation Analysis

Psychology research papers that use mediation analysis should be mindful to report important information throughout the paper to provide readers with sufficient information to evaluate the study. In this section, I walk through suggestions for reporting mediation analyses. Much of this section aligns with Lee et al. (2021), who proposed the AGReMA statement (A Guideline for Reporting Mediation Analyses).

Introduction. Prior theoretical evidence which might support the proposed causal order of the variables in the mediation model should be presented in the introduction (Fiedler, Harris, & Schott, 2018). This is especially important because common designs in psychology cannot support all paths in a mediation analysis as causal. When X is randomized, researchers should develop a theoretical case for the causal order of M and Y. For completely cross-sectional designs, previous evidence for a causal relationship for all paths should be discussed. Any research suggesting potential confounders should be discussed in the introduction section as well. As a cautionary note, even if there is prior evidence of a causal effect between variables in a model, if that effect is not being controlled through some kind of manipulation, confounders can bias the estimate of the effect leading to incorrect conclusions.

Methods. The methods section should include information about the design of the study and the measures should be described in a way that the coefficients are interpretable. For example, if a summary score of a scale is used, researchers should clarify whether it is a sum, average, or something else. The methods section should also include the rationale for sample size planning (e.g., power analysis). If the sample size was selected based on other factors, such as limited resources, this should be explicitly reported.

The planned mediation analyses should be described in the methods or data analysis section, and include specific elements to ensure transparent and clear reporting of the analysis. A visual representation of the model (e.g., path diagram) should be included.

Diagrams should include all covariates, not just the X, M, and Y variables. The specific assumptions for causal inference should be described in this section, and not just reserved for the limitations section. Researchers can use the potential outcomes framework to describe these assumptions. In addition, the statistical analysis process should be clearly described, including estimation method (OLS, SEM, or something else), how covariates were included in the models, whether the XM interaction is tested prior to estimation and whether it was included in the model, what inferential method was used for the indirect effect, details specific to the inferential method (e.g., number of bootstraps), and information about the software used for the analysis (including version numbers for program and package). Additionally, details about the planned sensitivity analyses should be included in the methods section. Open data and code are being increasingly required by funding agencies (Kozlov, 2022) and journals (Nuijten et al., 2017). Code and data should be made openly available to ensure reproducibility.

Results. Researchers should report the estimates and inferences for the important paths in the mediation model in the results section. APA recommendations state it is not sufficient to report only p-values for effects, but coefficient estimates and confidence intervals should also be reported (American Psychological Association, 2020). For mediation analysis, the a and b paths need to be reported separately and together as the indirect effect. This is important to evaluate the alignment of the results with the hypothesized patterns of effects. For example, if a researcher hypothesized the impact of the condition on the mediator is positive ($a > 0$), and the impact of the mediator on the outcome is negative ($b < 0$), this would lead to a negative indirect effect. If instead the researcher found that the effect of the condition on the mediator was negative ($a < 0$), and the impact of the mediator on the outcome was positive ($b > 0$), this also results in a negative indirect effect, but is not necessarily supporting of the specific hypothesized pattern of results. Without transparent reporting of the individual paths, such an inconsistency would be difficult to identify.

Report the results of tests of the XM interaction, whether done prior to selecting a mediation model or as a robustness evaluation. If the XM interaction is included in the model, estimates of both indirect effects should be provided (PNIE and TNIE) with uncertainty estimates. If the XM interaction is not included in the model, the single indirect effect can be reported with an uncertainty estimate. Sensitivity analyses for each parameter in the model should be reported, in particular those paths which may be confounded due to design limitations.

Discussion. The discussion section is perhaps the most important with respect to balancing evidence from the analysis with limitations. As it is currently conducted in social and personality psychology today, broad causal claims are often made based on the results of a mediation analysis. The limitations and assumptions with respect to causal claims are only presented right at the end. In recent years, researchers have internalized the concern that causal inferences from mediation analysis are problematic. This has led to some unusual interpretations of mediation results which attempt to toe the line of causality. A few examples: "...[M] to mediate the association of [X] with [Y]" (Syropoulos, Lifshin, Greenberg, Horner, & Leidner, 2022) and "... [M] consistently predicts [Y], and [M] mediates the relationship between [X] and [Y]" (Martínez, van Prooijen, & Lange, 2022). These examples are nonsensical because an association/relationship is not causal, but mediation is causal. Researchers know they should not say "effect" because it sounds causal, but they do not realize that saying "mediation" is equally causal. Rather than avoiding this language altogether, especially when it is aligned with the theoretical model, researchers should discuss the limitations of their causal assumptions explicitly.

The limitations section of a paper reporting a mediation analysis should discuss each assumption for identifying the indirect effect, and discuss limitations of each assumption. For example, stating "We believe there are no common causes of drawing upon your female role-congruent self and interest in STEM" or "Factors such as mother's occupation could potentially impact the degree to which students draw on their female-role congruent self

and interest in STEM. Future studies should measure this variable and examine whether accounting for this potential confounder reduces or eliminates the association between our drawing upon your female role congruent self and interest in STEM." Statements such as these may also prompt reviewers to consider possible confounders which should be accounted for in the current study or future research more so than current statement such as "causality cannot be supported from the current data." Explicitly describing the assumptions within the context of the variables being examined allows other researchers to directly evaluate the plausibility of the causal claims.

The Future of Mediation Analysis

While this chapter has focused on the integration of what is typically referred to as "statistical" mediation analysis and "causal" mediation analysis, there are of course other major developments in mediation analysis in recent years which have yet to be widely adopted in social and personality psychology. In particular, all of the above discussion assumes that the variables are measured without error. This is not typically the case in psychology, and so measured variables are best seen as indicators of latent variables, where the hypotheses are specific to the latent variables. Mediation analysis with latent variables has been an area of great growth in recent years, but the integration of this literature with the causal inference literature is still very nascent (Gonzalez & Valente, 2022). In addition, this chapter only focused on the issue of causal inference, rather than the issue of measuring change over time, which would require longitudinal designs. Longitudinal mediation analysis, in its many forms, has seen much methodological development in the last 20 years, and many of the ideas from this chapter can be extended to but also complicated by longitudinal designs. These two primary areas suggest direction for the future of mediation analysis in psychology, and methodological researchers interested in contributing to this work might consider developments in these areas, particularly at the intersection of these areas with causal inference.

Mediation Analysis with Latent Variables

Many constructs important to social and personality psychology theory are not directly observable, they are latent (Flake, Pek, & Hehman, 2017; Ledgerwood & Shrout, 2011). Typically, when researchers are interested in latent constructs they use a scale with multiple items all designed to tap into this latent variable. For example, college students' interest STEM can be measured with questions like "How much do you intend to major in computer science" or "How likely are you to pursue a major in computer science?". We might ask the same questions about a variety of STEM fields (e.g., engineering, physics). The theoretical model behind these types of measures is that each student has some unobserveable interest in STEM (broadly), and that interest then causes them to respond the way they do on these questions. This process is modeled through a factor model, where a latent variable (η) causes each item response (Gorsuch, 1983). Models of just these measurement processes are frequently called factor analyses, and models with structural relationships (e.g., where one latent variable predicts another) are called SEMs.

In a mediation model it is possible for one or more of the variables involved to be latent (e.g., See Figure 4). In our career framing example, both interest in STEM (Y) and drawing on your female role-congruent self (M) were measured with multi-item questionnaires. The underlying latent variables are the variables of interest in the mediation model. It is not uncommon to undertake mediation analyses using summaries of these scales (e.g., a sum or mean of the items); however doing so can lead to statistical errors which could be overcome by using latent variable models. Using summary scores does not account for the possibility of measurement error (i.e., that the summary score does not perfectly reflect the latent variable). Not accounting for measurement error can bias the estimate of the indirect effect, but it can also increase uncertainty around the estimate (Hoyle & Kenny, 1999; Cole & Preacher, 2014). Mediation models in particular can behave in a complex manner when the mediator is latent, because the mediator is both a predictor and an outcome (Gonzalez & MacKinnon, 2021).

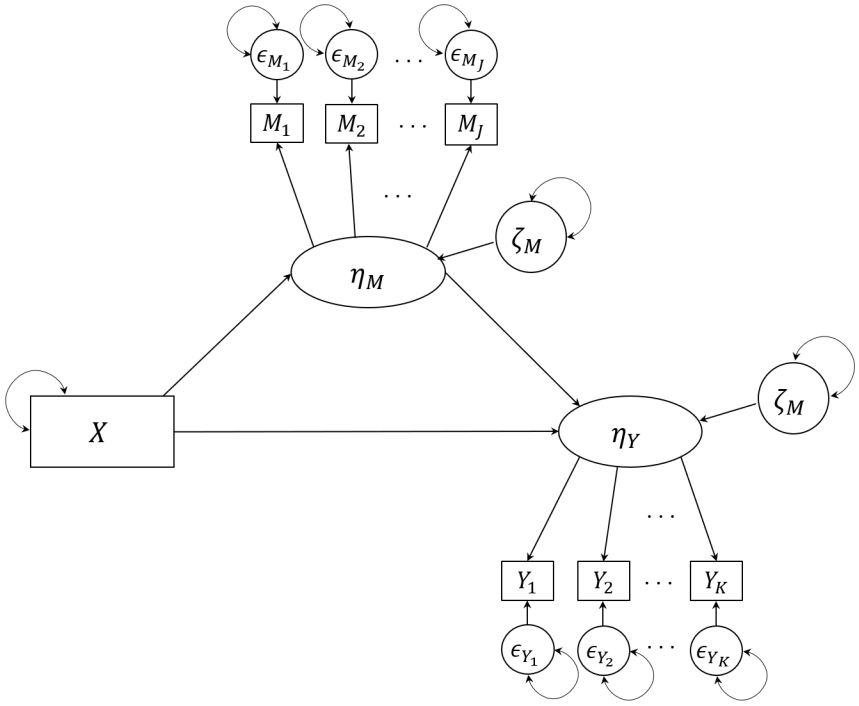

Figure 4. An example of a latent variable mediation model where X is observed (e.g., randomly assigned), and M and Y are latent. Latent variables (η) are represented in ovals, whereas observed variables are represented in rectangles. η_M has J observed indicators, and η_Y has K observed indicators. Residuals for latent variables (ζ) and residuals for observed variables (ϵ) are in circles. Curved arrows pointing at a variable represent estimated variance.

Researchers can account for latent variables by fitting mediation models in SEM programs (e.g., lavaan in R (Rosseel, 2012), Mplus (Muthèn & Muthèn, 1998 – 2011), and AMOS (?, ?)), where both the measurement model and structural model can be specified. This can allow for more optimal statistical performance of the analysis. SEMs may suffer from a trade off between precision and accuracy, such that the coefficients are estimated without bias, but the standard errors are larger (Ledgerwood & Shrout, 2011). Researchers may want to leverage latent variable modeling to increase power without increasing their sample sizes, but using latent variables will not always improve power, because of additional uncertainty. Calls to increase latent variable modeling in mediation in social and

personality psychology have largely gone unanswered (Ledgerwood & Shrout, 2011). There are a number of resources available for researchers interested in learning more on these topics (Muthèn & Asprouhov, 2015; G. W. Cheung & Lau, 2008; Falk & Biesanz, 2015; Iacobucci, Saldanha, & Deng, 2007). Future research will need to explore the intersection of latent variable models and causal inference for mediation.

Longitudinal Designs and Analysis

Earlier in this chapter, two primary criticisms of mediation analysis with completely cross-section data were introduced. The first criticism was that mediation is inherently causal, and statistical mediation analysis does not necessarily address concerns about causality. To address this concern, approaches from causal mediation analysis can be integrated into the analysis process. The second criticism was that mediation is an inherently longitudinal process, and cross-sectional data cannot speak to change over time. Recent work has consistently demonstrated that mediation analysis with cross-sectional data provides biased estimates of the longitudinal process as it unfolds (Maxwell & Cole, 2007; Maxwell, Cole, & Mitchell, 2011; Lindenberger, von Oertzen, Ghisletta, & Hertzog, 2011; O'Laughlin, Martin, & Ferrer, 2018; Shrout, 2011). The primary issue within this context is that the underlying ideas of mediation are that one variable causes another, but how much time is needed for an effect to take place is unclear. Most changes in important constructs (e.g., depression, motivation) take time. This is particularly problematic as researchers have demonstrated that under many circumstances mediation analysis with cross-sectional data provides essentially no information about longitudinal indirect effects (Maxwell & Cole, 2007; Maxwell et al., 2011; O'Laughlin et al., 2018). Given this major criticism of mediation analysis with cross-sectional data, many researchers will turn to longitudinal designs as a potential solution. See (Gordon & Thorson, in preparation) for a general introduction to modeling longitudinal data.

Addressing this concern involves changing the way that the data is collected to integrate multiple measurements over time, longitudinal data. Data for mediation can be

collected in multiple ways for a longitudinal design, including randomly assigning X then measuring M and Y repeatedly over time (experimental longitudinal designs). An alternative would be to examine how X, M, and Y change over time concurrently (completely longitudinal designs). Analytic approaches to estimating mediation will depend on the design, but the important takeaway is that methods do exist for estimating indirect effects in longitudinal data. In this section I provide a brief review of three approaches to longitudinal mediation analysis: a cross-lag panel approach, multilevel approach, latent growth curve approach.

Cross-lag Panel Approach. This cross-lag panel approach provides perhaps the simplest point of entry into longitudinal mediation. Consider a design where X is either measured or assigned at Time 1, and M and Y are measured at Time 1, 2, and 3. Given this design, a simple approach to mediation analysis would be to use the statistical mediation approach described earlier, but to estimate the indirect effect of X at Time 1 (X_1) on Y at Time 3 (Y_3) through M at Time 2 (M_2). In this case, M_1 and Y_1 should serve as covariates in the model, because they are potential confounders of the effect of X_1 on M_2 or Y_3 and M_2 on Y_3. It is less clear if Y_2 should be included as a covariate in the model for Y_3 (it should not be included in the model for M_2; See Figure 5 Panel A). Estimation of the indirect effect remains relatively simple (a_1b_1), and inference could be carried out using bootstrapping or any other method described in previous section. This model could be estimated using PROCESS or any other tool that can estimate indirect effects in cross-sectional data. Extending this model past three time points would require estimation in an SEM framework, but generalizes easily (Cole & Preacher, 2014; Selig & Preacher, 2009; Cole & Maxwell, 2003; Preacher, 2015; MacKinnon, 2008; Roth & MacKinnon, 2012; Card, 2012).

Multilevel Approach. Typically for a longitudinal design, people (Level 2) are measured at a variety of time-points (Level 1). In multilevel mediation, the model is described by the levels of each variable: X-M-Y. For example, an experimental

longitudinal design would be a 2-1-1, because the X variable is assigned to the person and does not vary across time, but M and Y are measured at each time point. Alternatively a completely longitudinal design would be a 1-1-1 (Krull & MacKinnon, 2001; Kenny, Korchmaros, & Bolger, 2003; Pituch, Whittaker, & Stapleton, 2005). Figure 5 Panel B shows a 2-1-1 design, because it lends itself best to causal inference, though still requires specific assumptions for causal inference (Bind, Vanderweele, Coull, & Schwartz, 2016). For those completely unfamiliar with multilevel models, I recommend Raudenbush and Bryk (2002) and Nezlek (2011) for introductions.

A particular advantage of multilevel modeling is that each participant can be allowed to have their own intercept and slope in the model; these are called random effects. In this way, mediation analysis can be complicated because the a-path, the b-path, or both can be random, meaning each participant can have their own indirect effect. Additionally, when both paths are random, they can covary, and this covariance must be taken into account when estimating and conducting inference on the indirect effect (Bauer, Preacher, & Gil, 2006). Bootstrapping approaches are complicated by the nesting of the data, so Monte Carlo confidence intervals are recommended for inference for indirect effects in multilevel mediation (Bauer et al., 2006; Rockwood & Hayes, 2022). Additionally, for 1-1-1 models indirect effects can occur at Level 1 and/or Level 2, called contextual effects (Zhang, Zyphur, & Preacher, 2009). The advantage of contextual effects is that they can provide insight into whether indirect effects can be attributed to differences across people as compared to changes within people across time.

Latent Growth Curve Approach. Latent growth curve approaches to mediation require SEMs, but are some of the most widely recommended methods for longitudinal mediation (M. W. L. Cheung, 2007). These models are very flexible, being able to account for linear and non-linear change, allow variability in change among participants, and accommodate unequal time intervals between measurement occasions. In this section, I focus on a design where X is randomly assigned at baseline, then measurements of M and

Y continue over time, but this analytic process can be expanded to involve longitudinal measurements of X as well.

In this latent growth curve model, M and Y will each have an intercept which can be thought of as the “starting" point of the curve over time (μ_0 for M and η_0 for Y), the level of the variable at the beginning of the study. The second parameter is the rate of growth (μ_1 for M and η_1 for Y). In a linear model this is considered a slope. This is the rate of change that we expect in our different variables over time. Latent growth curves allow for a random intercept and slope for both M and Y, and so similar to the multilevel mediation, participants are allowed to vary from each other in their model parameters. Both the intercept and slope could be affected by X, and this opens doors for a variety of indirect effects. For example, X could effect the starting point of M, which could affect the starting point of Y. Alternatively, X could affect the slope of M which could effect the slope of Y. Typically there are three possible indirect effects (intercept-intercept: a_0b_0, intercept-slope:a_0b_2, slope-slope:a_1b_1). Researchers hypothesizing mediation for these types of models, will need to clearly specify which of these indirect effects aligns with their hypothesis. Figure 5 Panel C provides a path diagram representation of the model with 4 time points.

Latent growth curve modeling approaches to mediation are very flexible, which can be an advantage or a disadvantage. Even when the X variable is manipulated it is difficult to know the causal ordering of the mediator and the outcome in these models, not unlike the cross-sectional mediation analyses. Could the intercept of the outcome affect the slope of the mediator? Researchers should think about the assumed causal order of the mediation models in all types of designs, but particularly when using latent growth curve modeling. In Figure 5 Panel C all potential indirect effects where a slope influences an intercept have been omitted, because this would likely violate assumed causal order. However, there may be cases where these effects are of interest and deemed causally plausible, especially because the intercept parameter does not necessarily need to be at the

first time point but could be an average or at a different time point. Causal inference in longitudinal designs can become increasingly difficult because of time-varying covariates. For more on causal inference in longitudinal mediation, see Bind et al. (2016). Sensitivity analysis for latent growth curve mediation analysis given both a randomized and observed X are available in Tofighi et al. (2018).

Summary

Mediation analysis within psychological science has come a long way. From Baron and Kenny (1986)'s seminal paper on the differentiation between mediation and moderation, researchers have moved to a much more sophisticated understanding of both mediation as a hypothesis, and mediation analysis as a statistical tool. Previous calls to social and personality psychologists to adjust their thinking about the inferential claims in mediation have been answered (Rucker et al., 2011). More recently, the limitations of statistical mediation analysis for making causal claims have become increasingly clear (Rohrer & Aslan, 2021). By integrating tools from both statistical and causal mediation analysis, researcher can increase the transparency about the assumptions made in these analyses. While many of these assumptions are untestable, researchers can explore the sensitivity of their results to these assumptions to temper their findings appropriately. By integrating common practices from causal mediation analysis, such as explicitly describing assumptions, evaluating the XM interaction, and sensitivity analyses, mediation analysis in social and personality psychology can improve. Integrating these practices into the planning, conduct, and reporting of mediation analyses is integral to overcoming the major criticisms of statistical mediation analysis as it is currently conducted in social and personality psychology. It is important to note that by integrating these practices into the mediation analysis process, researchers are by no means guaranteeing that they are now estimating causal effects, which is a common misunderstanding of "causal" mediation analysis. Instead they are making more transparent the limitations of the current design with respect to causality, and providing information about the sensitivity of the model to

potential confounding. Recent developments in latent variable and longitudinal models for mediation from a statistical point of view provide solutions to many criticisms of mediation analysis with cross-sectional data, but future research needs to integrate these developments with causal inference methods.

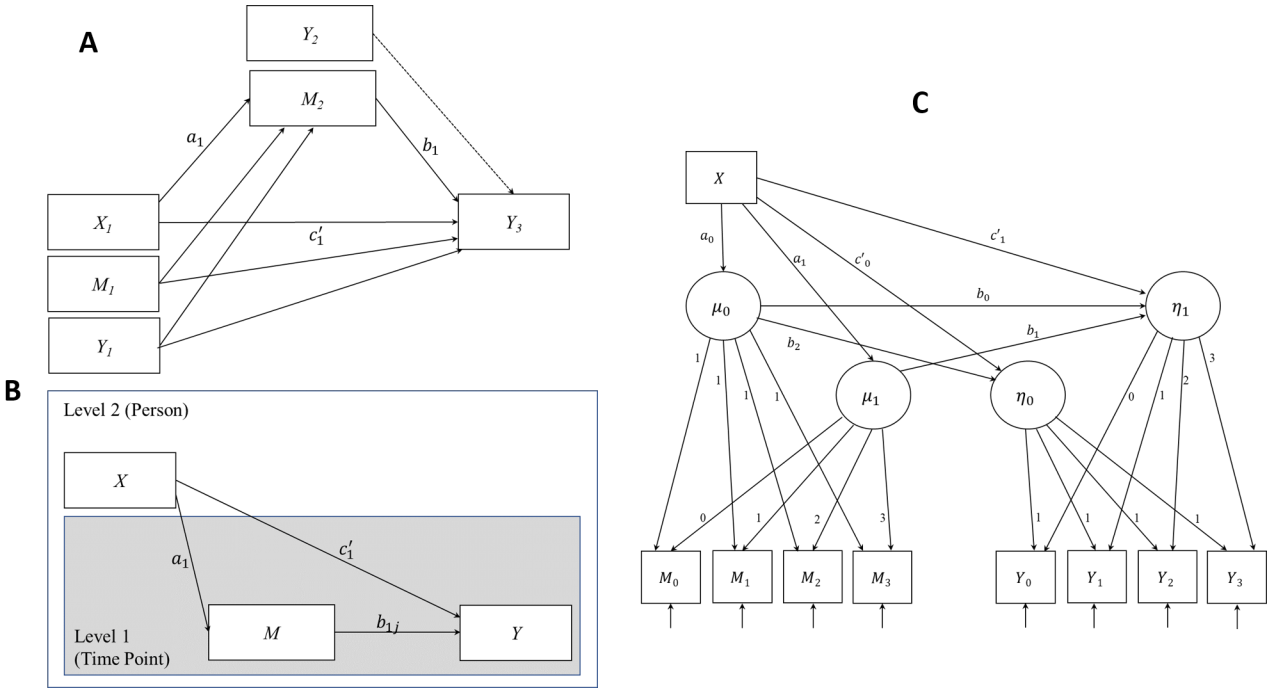

Figure 5. Visualizations of three longitudinal mediation models. Panel A) Cross-lag Panel Approach. Treatment (X) affects the mediator at Time 2, controlling for the mediator and outcome at Time 1. The mediator at Time 2 affects the outcome at Time 3 controlling for the mediator and outcome at Time 1 and the outcome at Time 2. Panel B) Multilevel Approach. The notation b_{1j} indicates that the coefficient b_1 is allowed to vary across person's j. Panel C) Latent Growth Curve Approach. Indicators of the mediator load onto a latent intercept (μ_0) and a latent slope (μ_1), while indicators of the outcome load onto a latent intercept (η_0) and a latent slope (η_1). Treatment X affects all latent variables, the mediator intercept (μ_0) affects both the intercept and slope for the outcome (η_0 and η_1), the slope for the mediator (μ_1) affects the slope for the outcome (η_1).

References

Aberson, C. L. (2019). *Applied power analysis for the behavioral sciences* (2nd ed.). New York: Routledge.

Aguinis, H., Beaty, J. C., Boik, R. J., & Pierce, C. A. (2005). Effect size and power in assessing moderating effects of categorical variables using multiple regression: A 30-year review. *Journal of Applied Psychology*, *90*(1), 94-107.

American Psychological Association. (2020). *Publication manual of the American Psychological Association 2020: the official guide to APA style* (7th ed.). American Psychological Association.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182. doi: 10.1037//0022-3514.51.6.1173

Bauer, D. J., Preacher, K. J., & Gil, K. M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: New procedures and recommendation. *Psychological Methods*, *11*(2), 142–163. doi: 10.1037/1082-989X.11.2.142

Biesanz, J. C., Falk, C. F., & Savalei, V. (2010). Assessing mediational models: Testing and interval estimation for indirect effects. *Multivariate Behavioral Research*, *45*, 661–701. doi: 10.1080/00273171.2010.498292

Bind, M.-A., Vanderweele, T., Coull, B. A., & Schwartz, J. D. (2016). Causal mediation analysis for longitudinal data with exogenous exposure. *Biostatistics*, *17*(1), 122–134. doi: 10.1093/biostatistics/kxv029

Bullock, J. G., Green, D. P., & Ha, S. E. (2010). Yes, but what's the mechanism? (don't expect an easy answer). *Journal of Personality and Social Psychology*, *98*, 550–558. doi: 10.1037/a0018933

Bullock, J. G., & Green, G. P. (2021). The failings of conventional mediation analysis and

a design-based alternative. *Advances in Methods and Practices in Psychological Science*, *4*(4), 1–18. doi: 10.1177/25152459211047227

Card, N. A. (2012). Multilevel mediational analysis in the study of daily lives. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (p. 479-494). New York: Guilford.

Charlton, A., Montoya, A. K., Price, J., & Hilgard, J. (2021). Noise in the process: an assessment of the evidential value of mediation effects in marketing journals. doi: 10.31234/osf.io/ck2r5

Chen, D., & Fritz, M. S. (2021). Comparing alternative corrections for bias in the bias-corrected bootstrap test of mediation. *Evaluation & the Health Professions*, *44*(4), 416–427. doi: 10.1177/01632787211024356

Cheung, G. W., & Lau, R. S. (2008). Testing mediation and suppression effects of latent variables: Bootstrapping with structural equation models. *Organizational Research Methods*, *11*(2), 296–325. doi: 10.1177/1094428107300343

Cheung, M. W. L. (2007). Comparison of approaches to constructing confidence intervals for mediating effects using structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(2), 227–446. doi: 10.1080/10705510709336745

Cole, D. A., & Maxwell, S. E. (2003). Testing mediational models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology*, *112*(4), 558–577. doi: 10.1111/acer.12126

Cole, D. A., & Preacher, K. J. (2014). Manifest variable path analysis: Potentially serious and misleading consequences due to uncorrected measurement error. *Psychological Methods*, *19*(2), 300–315. doi: 10.1037/a0033805

Daniel, R. M., Stavola, B. L. D., Cousens, S. N., & Vansteelandt, S. (2015). Causal mediation with multiple mediators. *Biometrics*, *71*(1), 1–15. doi: 10.1111/biom.12248

Earp, B. D., & Trafimow, D. (2015). Replication, falsification, and the crisis of confidence

in social psychology. *Frontiers in Psychology*, *6*, 1–11. doi: 10.3389/fpsyg.2015.00621

Falk, C. F., & Biesanz, J. C. (2015). Inference and interval estimation methods for indirect effects with latent variable models. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*, 24–38.

Fiedler, K., Harris, C., & Schott, M. (2018). Unwarranted inferences from statistical mediation tests – an analysis of articles published in 2015. *Journal of Experimental Psychology*, *75*. doi: 10.1016/j.jesp.2017.11.008

Flake, J. K., Pek, J., & Hehman, E. (2017). Construct validation in social and personality research: Current practice and recommendations. *Social Psychological and Personality Science*, *8*(4), 370–378. doi: 10.1177/1948550617693063

Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science*, *18*(3), 233–239. doi: 10.1111/j.1467-9280.2007.01882.x

Funk, M. J., Westreich, D., Wiesen, C., Sturmer, T., Brookhart, M. A., & Davidson, M. (2011). Doubly robust estimation of causal effects. *American Journal of Epidemiology*, *173*(7), 761–767. doi: 10.1093/aje/kwq439

Gonzalez, O., & MacKinnon, D. P. (2021). The measurement of the mediator and its influence on statistical mediation conclusions. *Psychological Methods*, *26*(1), 1–17. doi: 10.1037/met0000263

Gonzalez, O., & Valente, M. J. (2022). Accommodating a latent XM interaction in statistical mediation analysis. *Multivariate Behavioral Research*. doi: 10.1080/00273171.2022.2119928

Gordon, A., & Thorson, K. (in preparation). Longitudinal data: Design decisions and analytic strategies. In H. T. Reis, T. West, & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology.*

Gorsuch, R. L. (1983). *Factor Analysis*. Hillsdake, NJ: Lawrence Erlbaum Associates.

Götz, M., O'Boyle, E. H., Gonzalez-Mulé, E., Banks, G. C., & Bollmann, S. S. (2021). The

"Goldilocks" zone: (Too) many confidence intervals in tests of mediation just exclude zero. *Psychological Bulletin*, *147*(1), 95–114. doi: 10.1037/bul0000315

Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs*, *76*(4), 408–420. doi: 10.1080/03637750903310360

Hayes, A. F. (2015). An index and test of linear moderated mediation. *Multivariate Behavioral Research*, *50*(1), 1 – 22. doi: 10.1080/00273171.2014.962683

Hayes, A. F. (2022). *Introduction to mediation, moderation, and conditional process analysis* (3rd ed.). New York, NY: Guilford Press.

Hayes, A. F., & Scharkow, M. (2013). The relative trustworthiness of inferential tests of the indirect effect in statistical mediation analysis: Does method really matter? *Psychological Science*, *24* , 1918–1927. doi: 10.1177/0956797613480187

Hoyle, R. H., & Kenny, D. A. (1999). Sample size, reliability, and tests of statistical mediation. In *Statistical strategies for small sample research* (R. H. Hoyle ed., pp. 195–222). Thousand Oaks, CA: Sage.

Iacobucci, D. (2012). Mediation analysis and categorical variables: The final frontier. *Journal of Consumer Psychology*, *22*(4), 582–594. doi: 10.1016/j.jcps.2012.03.006

Iacobucci, D., Saldanha, N., & Deng, J. X. (2007). A meditation on mediation: Evidence that structural equations models perform better than regressions. *Journal of Consumer Psychology*, *17*(2), 140–154. doi: 10.1016/j.jcps.2012.03.006

Imai, K., Keele, K., & Yamamoto, T. (2010). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science*, *25*(1), 51–71. doi: 10.1214/10-STS321

Imai, K., Tingley, D., & Yamamoto, T. (2013). Experimental designs for identifying causal mechanisms. *Journal of the Royal Statistical Society*, *176*(1), 5–51. doi: 10.1111/j.1467-985X.2012.01032.x

Imbens, G. W., & Rubin, D. (2015). *Causal inference for statistics, social, and biomedical*

sciences: An introduction. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139025751

Kenny, D. A. (2017). *An interactive tool for the estimation of power in tests of mediation [computer software].* Retrieved from `https://davidakenny.shinyapps.io/MedPower/`

Kenny, D. A., & Judd, C. M. (2014). Power anomalies in testing mediation. *Psychological Science*, *25*(2), 334 – 339.

Kenny, D. A., Korchmaros, J. D., & Bolger, N. (2003). Lower-level mediation in multilevel models. *Psychological Methods*, *8*(2), 115–128.

Kozlov, M. (2022). NIH issues a seismis mandate: share data publicly. *Nature*, *602*, 558 - 559. doi: 10.1038/d41586-022-00402-1

Krull, J. L., & MacKinnon, D. P. (2001). Multilevel modeling of individual and gorup level mediated effects. *Multivariate Behavioral Research*, *36*, 249 – 277.

Ledgerwood, A., & Shrout, P. E. (2011). The trade-off between accuracy and precision in latent variable models of mediation processes. *Journal of Personality and Social Psychology*, *101*, 1174-1188.

Lee, H., Cashin, A. G., Lamb, S. E., Hopewell, S., Vansteelandt, S., VanderWeele, T. J., ... McAuley, J. H. (2021). A guideline for reporting mediation analyses of randomized trials and observational studies: The agrema statement. *JAMA*, *326*, 1045–1056. doi: 10.1001/jama.2021.14075

Lindenberger, U., von Oertzen, T., Ghisletta, P., & Hertzog, C. (2011). Cross-sectional age variance extraction: What's change got to do with it? *Psychology and Aging*, *26*(1), 34–47. doi: 10.1037/a0020525

Liu, S. H., Ulbricht, C. M., Chrysanthopoulou, S. A., & Lapane, K. L. (2016). Implementation and reporting of causal mediation analysis in 2015: a systematic review in epidemiological studies. *BMC Research Notes*, *9*(354). doi: 10.1186/s13104-016-2163-7

Liu, X., & Wang, L. (2019). Sample size planning for detecting mediation effects: A power analysis procedure considering uncertainty in effect size estimates. *Multivariate Behavioral Research*, *54*(6), 822 – 839.

MacKinnon, D. P. (2000). Contrasts in multiple mediator models. In J. Rose, L. Chassin, C. C. Presson, & S. J. Sherman (Eds.), *Multivariate applications in substance use research: New methods for new questions* (pp. 141 – 160). Mahwah, NJ: Erlbaum.

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York: Lawrence Erlbaum Associates.

MacKinnon, D. P., Fritz, M. S., Williams, J., & Lockwood, C. M. (2007). Distribution of the product confidence limits for the indirect effect: Program PRODCLIN. *Behavior Research Methods*, *39*(3), 384 — 389. doi: 10.3758/bf03193007

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*, *7*(1), 83-104.

MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, *39*, 99 – 128.

MacKinnon, D. P., Valente, M. J., & Gonzalez, O. (2020). The correspondence between causal and traditional mediation analysis: The link is the mediator by treatment interaction. *Prevention Science*, *21*, 147–157. doi: 10.1007/s11121-019-01076-4

Martínez, C. A., van Prooijen, J., & Lange, P. A. M. V. (2022). A threat-based hate model: How symbolic and realistic threats underlie hate and aggression. *Journal of Experimental Social Psychology*, *103*, 1-13. doi: 10.1016/j.jesp.2022.104393

Maxwell, S. E., & Cole, D. A. (2007). Bias in cross-sectional analyses of longitudinal mediation. *Psychological Methods*, *12*(1), 23–44. doi: 10.1037/1082-989X.12.1.23

Maxwell, S. E., Cole, D. A., & Mitchell, M. A. (2011). Bias in cross-sectional analyses of longitudinal mediation: partial and complete mediation under an autoregressive

model. *Psychological Methods*, *12*, 23–44.

Meule, A. (2019). Contemporary understanding of mediation testing. *Meta-Psychology*, *3*. doi: 10.15626/MP.2018.870

Montoya, A. K., & Hayes, A. F. (2017). Two-condition within-participant statistical mediation analysis: A path-analytic framework. *Psychological Methods*, *22*(1), 6–27. doi: 10.1037/met0000086

Muthèn, B., & Asprouhov, T. (2015). Causal effects in mediation modeling: An introduction with applications to latent variables. *Structural Equation Modeling*, *22*(1), 12–23. doi: 10.1080/10705511.2014.935843

Muthèn, L. K., & Muthèn, B. O. (1998 – 2011). *Mplus user's guide* (Sixth ed.). Los Angeles, CA: Muthèn & Muthèn.

Nezlek, J. B. (2011). *Multilevel modeling for social and personality psychology*. Los Angeles, CA: Sage.

Nuijten, M. B., Borghuis, J., Veldkamp, C. L. S., Dominguez-Alvarez, L., van Assen, M. A. L. M., & Wicherts, J. M. (2017). Journal data sharing policies and statistical reporting inconsistencies in psychology. *Collabra: Psychology*(1), 31. doi: 10.1525/collabra.102

O'Laughlin, K. D., Martin, M. J., & Ferrer, E. (2018). Cross-sectional analysis of longitudinal mediation processes. *Multivariate Behavioral Research*, *53*(3), 375–402. doi: 10.1080/00273171.2018.1454822

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, *82*(4), 669–710. doi: 10.1093/biomet/82.4.669

Pearl, J. (2001). *Direct and indirect effects*. San Francisco, CA: Morgak Kaufman.

Pituch, K. A., Whittaker, T. A., & Stapleton, L. M. (2005). A comparison of methods to test for mediation in multisite experiments. *Multivariate Behavioral Research*, *40*, 1–23. doi: $10.1207/s15327906mbr4001_1$

Preacher, K. J. (2015). Advances in mediation analysis: A survey and synthesis of new

developments. *Annual Review of Psychology*, *66*, 825–852. doi: 10.1146/annurev-psych-010814-015258

Preacher, K. J., & Hayes, A. F. (2004). Spss and sas procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods*, *36*, 717–731. doi: 10.3758/BF03206553

Qin, X. (under revision). Sample size and power calculations for causal mediation analysis. Retrieved from `https://xuqin.shinyapps.io/CausalMediationPowerAnalysis/`

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Sage.

Revelle, W. (2022). psych: Procedures for psychological, psychometric, and personality research [Computer software manual]. Evanston, Illinois. Retrieved from `https://CRAN.R-project.org/package=psych` (R package version 2.2.5)

Robins, J. M., & Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, *3*(2), 143–155. doi: 10.1097/00001648-199203000-00013

Rockwood, N. J., & Hayes, A. F. (2022). Multilevel mediation analysis. In *Multilevel modeling methods with introductory and advanced applications* (To appear in A. A. O'Connell, D. B. McCoach, B. Bell (Eds.) ed., pp. –). Charlotte, NC: Information Age Publishing.

Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science*, *1*(1), 27-42. doi: 10.1177/2515245917745629

Rohrer, J. M., & Aslan, R. C. (2021). Precise answers to vague questions: Issues with interactions. *Advances in Methods and Practices in Psychology*, *4*(2), 1–19. doi: 10.1177/25152459211007368

Rosenbaum, P. R., & Rubin, D. B. (1984). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55. doi:

10.1093/biomet/70.1.41

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. doi: 10.18637/jss.v048.i02

Roth, D. L., & MacKinnon, D. P. (2012). Mediation analysis with longitudinal data. In J. T. Newsom, R. N. Jones, & S. M. Hofer (Eds.), *Longitudinal data analysis: A practical guide for researchers in aging, health, and social sciences* (pp. 181–216). New York: Routledge.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688–701.

Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of American Statistical Association*, *100*(469), 322–331. doi: 10.1198/016214504000001880

Rucker, D. D., Preacher, K. J., Tormala, Z. L., & Petty, R. E. (2011). Mediation analysis in social psychology: Current practices and new recommendations. *Social and Personality Psychology Compass*, *5*, 359–371.

Schoemann, A. M., Boulton, A. J., & Short, S. D. (2017). Determining power and sample size for simple and complex mediation models. *Social Psychological and Personality Science*, *8*(4), 379 — 386. doi: 10.1177/1948550617715068

Schroder, H. S., Fisher, M. E., Yanli, L., Lo, S. L., Danovitch, J. H., & Moser, J. S. (2017). Neural evidence for enhanced attention to mistakes among school-aged children with a growth mindset. *Developmental Cognitive Neuroscience*, *24*(1), 42–50. doi: 10.1016/j.dcn.2017.01.004.

Selig, J. P., & Preacher, K. J. (2008). Monte carlo method for assessing mediation: An interactive tool for creating confidence intervals for indirect effects [computer software].

Selig, J. P., & Preacher, K. J. (2009). Mediation models for longitudinal data in developmental research. *Research in Human Development*, *6*(2–3), 144–164.

Shrout, P. E. (2011). Commentary: Mediation analysis, causal process, and cross-sectional data. *Multivariate Behavioral Research*, *46*(5), 852–860. doi: 10.1080/00273171.2011.606718

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*, *7*, 422–445.

Siy, J. O., Germano, A., Vianna, L., Azpeitia, J., Yan, S., Montoya, A. K., & Cheryan, S. (in press). Does the follow-your-passions ideology cause greater academic and occupational gender disparities than other cultural ideologies? *Journal of Personality and Social Psychology*.

Smith, L. H., & VanderWeele, T. J. (2019). Mediational E-values: Approximate sensitivity analysis for unmeasured mediator-outcome confounding. *Epidemiology*, *30*(6), 835–837. doi: 10.1097/EDE.0000000000001064

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology*, *13*, 290–312. doi: 10.2307/270723

Stone, C. A., & Sobel, M. E. (1990). The robustness of estimates of total indirect effects in covariance structure models estimated by maximum likelihood. *Psychometrika*, *55*(2), 337–352. doi: 10.1007/BF02295291

Syropoulos, S., Lifshin, U., Greenberg, J., Horner, D. E., & Leidner, B. (2022). Bigotry and the human–animal divide: (dis)belief in human evolution and bigoted attitudes across different cultures. *Journal of Personality and Social Psychology*, *123*(6), 1264 - 1292. doi: 10.1037/pspi0000391

Thoemmes, F. J., & Kim, E. S. (2011). A systematic review of propensity score methods in the social sciences. *Multivariate Behavioral Research*, *1*, 90–118. doi: 10.1080/00273171.2011.540475

Thoemmes, F. J., & Ong, A. D. (2015). A primer on inverse probability of treatment weighting and marginal structural models. *Emerging Adulthood*, *4*(1), 40–59. doi: 10.1177/2167696815621645

Tibbe, T. D., & Montoya, A. K. (2022). Correcting the bias correction for the bootstrap confidence interval in mediation analysis. *Frontiers in Psychology*, *13*. doi: 10.3389/fpsyg.2022.810258

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). Mediation: R package for causal mediation analysis. *Journal of Statistical Software*, *59*(5), 1–38. doi: 10.18637/jss.v059.i05

Tofighi, D., Hsiao, Y.-Y., Kruger, E. S., MacKinnon, D. P., Horn, M. L. V., & Witkiewitz, K. (2018). Sensitivity analysis of the no-omitted confounder assumption in latent growth curve mediation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*(1), 94–109. doi: 10.1080/10705511.2018.1506925

Tofighi, D., & MacKinnon, D. P. (2011). Rmediation: A R package for mediation analysis confidence intervals. *Behavior Research Methods*, *43*, 692–700. doi: 10.3758/s13428-011-0076-x

Valente, M. J., Pelham, W. E., Smyth, H., & MacKinnon, D. P. (2017). Confounding in statistical mediation analysis: What it is and how to address it. *Journal of Counseling Psychology*, *64*(6), 659–671. doi: 10.1037/cou0000242

VanderWeele, T. J. (2010). Direct and indirect effect for neighborhood-based clustered and longitudinal data. *Sociological Methods & Research*, *38*(4), 515 – 544.

VanderWeele, T. J., & Chiba, Y. (2014). Sensitivity analysis for direct and indirect effects in the presence of exposure-induced mediator-outcome confounders. *Epidemiology, biostatistics and public health*, *11*(2), e9027. doi: 10.2427/9027

Vanderweele, T. J., & Robins, J. M. (2007). Four types of effect modification: A classification based on directed acyclic graphs. *Epidemiology*, *18*(5), 561–568. doi: 10.1097/EDE.0b013e318127181b

Vo, T., Superchi, C., Boutron, I., & Vansteelandt, S. (2020). The conduct and reporting of mediation analysis in recently published randomized controlled trials: results from a methodological systematic review. *Journal of Clinical Epidemiology*, *117*, 78–88.

Williams, J., & MacKinnon, D. P. (2008). Resampling and distribution of the product methods for testing indirect effects in complex models. *Structural Equation Modeling*, *15*, 23 – 51.

Wysocki, A. C., Lawson, K. M., & Rhemtulla, M. (2022). Statistical control requires causal justification. *Advances in Methods and Practice in Psychological Science*, *5*(2). doi: 10.1177/25152459221095823

Yzerbyt, V. Y., Muller, D., Batailler, C., & Judd, C. M. (2018). New recommendations for testing indirect effects in mediational models: The need to report and test component paths. *Journal of Personality and Social Psychology: Attitudes and Social Cognition*, *115*(6), 929–943. doi: 10.1037/pspa0000132

Yzerbyt, V. Y., Muller, D., & Judd, C. M. (2004). Adjusting researchers' approach to adjustment: On the use of covariates when testing interactions. *Journal of experimental social psychology*, *40*, 424–431. doi: 10.1016/j.jesp.2003.10.001

Zhang, Z., & Wang, L. (2013). Methods for mediation analysis with missing data. *Psychometrika*, *78*(1), 154 — 184.

Zhang, Z., & Yuan, K. H. (2018). *Practical statistical power analysis using webpower and R*. Granger, IN: ISDSA Press.

Zhang, Z., Zyphur, M. J., & Preacher, K. J. (2009). Testing multilevel mediation using hierarchical linear models: Problems and solutions. *Organizational Research Methods*, *12*, 695–719.