Unpacking the "Black Box" of Programs and Policies: A Conceptual Overview of Mediation Analysis

Antonio Morgan-Lopez and Anupa Bir, RTI International

Policymakers and practitioners have a growing interest in answering questions beyond simply "does a program work?" Researchers are moving towards study designs to identify the mechanisms of action—known as mediators—that describe *how* their interventions work (Chen, 1990; MacKinnon, 2008). Without mediation analysis, it is difficult to specify how a program produces results or why it fails to do so. For example, an intervention can fail because it targeted the correct mechanisms but did not substantially change them. Or, the intervention may have targeted and changed mechanisms that did not lead to subsequent changes in the outcomes of interest. If one of the goals of evaluation is to build knowledge to improve future intervention attempts, these types of distinctions are important to make. In this brief, we discuss 1) key considerations in mediation analysis, 2) options for mediation design and analysis, and 3) some best practices to follow when designing and conducting a mediation analysis.

MEDIATION ANALYSIS: A MOTIVATING EXAMPLE

Evaluators who design mediation studies tend to be interested in two fundamental questions: 1) which risk factors predict the outcomes we wish to change, and 2) which policies and programs or services might best address those factors? Practically, it is also important to consider which of the hypothesized mechanisms can feasibly be changed through direct intervention.

For example, assume we are designing an intervention to improve students' neurocognitive functioning. A first step would be to design a conceptual model that identifies the risk conditions, or *mediators*, that may lead to poor neurocognitive functioning, such as child maltreatment (Briggs-Gowan, Carter, & Ford, 2012; Yoches, Beeber, Jones Harden, Malik, & Summers, 2011). An

Dr. Antonio Morgan-Lopez is a principal scientist in the Military and Family Risk Behavior Research Program at RTI International. He develops and applies advanced quantitative methods in researching the etiology, prevention, and treatment of substance abuse, alcoholism and other health outcomes. Dr. Anupa Bir directs the Center for Advanced Methods Development at RTI International. She leads evaluations of complex health and social policy interventions, including health care innovation, financial inclusion, and family strengthening during incarceration and reentry.

example mediation model is shown in Figure 1. First, we would establish the *conceptual theory* (labeled "β"), where we identify how the risk condition (child maltreatment) is linked to the outcome (neurocognitive functioning). Next, we must identify an *action theory* (labeled "α") to describe how the intervention is designed to change the risk condition (child maltreatment). For instance, we may look to reduce the risk of child maltreatment by strengthening parent-child bonding during the infant and toddler years (e.g., Dozier, Peloso, Lewis, Laurenceau, & Levine, 2008).

Figure 1 can also be conceived of as a path diagram, where estimates of the product of α and β can be assessed for whether it is significantly different from zero, constituting the indirect effect (also called the mediated effect); this captures whether the effect of the parent-child bonding intervention on increased neurocognitive functioning can be attributed to reductions in child maltreatment.

One purpose of this model is to identify a key variable—neurocognitive functioning—that we would expect to change as a result of intervening on child maltreatment. In Figure 1, neurocognitive functioning is the *proximal outcome* of interest that we expect to improve shortly after the intervention.

Figure 1. Simplified Mediation Model
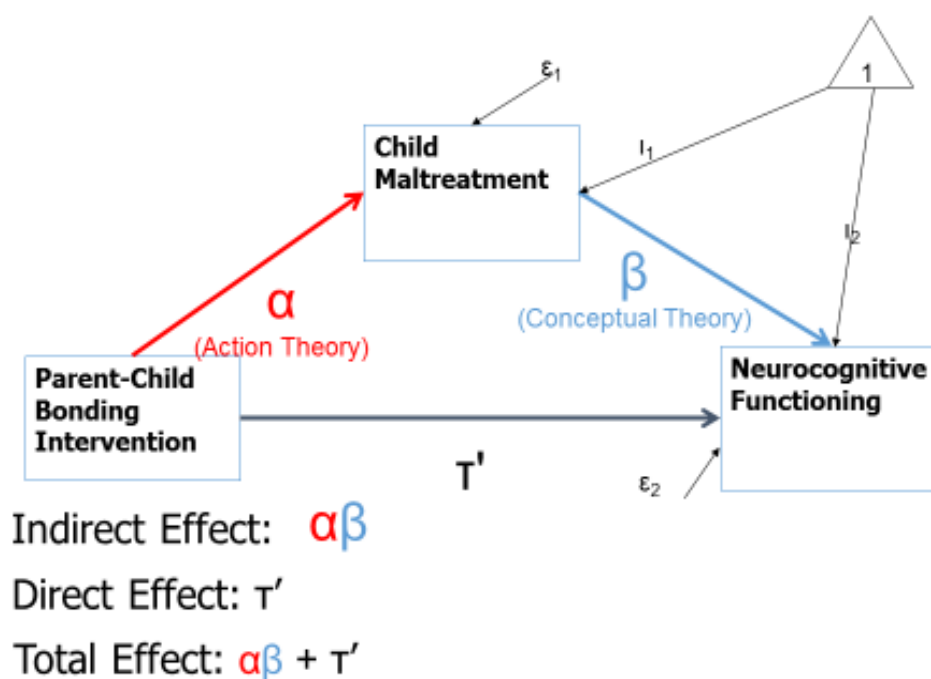
It may be that we are not only interested in neurocognitive functioning, but also its longer-term effects. For example, poor neurocognitive functioning has been demonstrated to cause later academic problems (Harden, 2015). However, we would expect first to see impacts on the proximal outcome, and impacts on the *distal outcome* (academic competence) further in the future.

OPTIONS FOR MEDIATION DESIGNS

Mediation analysis can be conducted in various study designs, either within the context of a program evaluation (e.g., a randomized controlled trial, single-group design, multiple-mediator design) or in observational studies that do not involve a program evaluation (e.g., cross-sectional designs). The type of design should be matched to the research question—conducting a mediation analysis with data from a weaker design (e.g., cross-sectional design) is likely to be beneficial in generative stages of research where potential intervention targets need to be identified, whereas mediation analyses in stronger designs are favored for examining underlying processes of effective interventions.

Randomized controlled trials. The gold-standard design for estimating program impacts on outcomes of interest is the randomized controlled trial (RCT). For example, we could randomly assign parents to receive or not receive classes designed to strengthen parent-child bonding and test if assignment to receive the classes leads to decreases in child maltreatment, and in turn, to improvements in neurocognitive functioning. However, the RCT may not be an option in all evaluation contexts (Chen, 1990), and when it comes to mediation, the RCT cannot address all questions about causality (Coffman & Zhong, 2012; MacKinnon, Taborga, & Morgan-Lopez, 2002).

The problem of establishing causality is further compounded because, even if individuals are randomly assigned to treatment groups, program conditions are rarely randomized and are often related to preexisting characteristics (e.g., demographics, baseline levels of the outcomes) that vary systematically across program sites. In other words, sites that may have better funding, infrastructure, stakeholder investment, or other potential site-level confounders may be predisposed to have more or higher quality programming. Many of these factors will co-vary with program content and impact the mediators and outcomes, leading to bias in both estimates of α and β (Coffman & Zhong, 2012; Imai, Keele, & Tingley, 2010).

Single-group designs. A nonrandomized mediation design frequently used in evaluation contexts is the single-group design. Here, all

participants receive the intervention, and data on mediators and outcomes are collected for at least two time points. With no comparison group, the source of variation in a single-group design is typically the intervention dosage each participant receives. A typical question is "does attending more sessions lead to greater change in the mediator?" Often, a problem with this design is that the intervention dosage is *self-selected*, meaning that participant characteristics may be related to whether they miss sessions of the intervention or drop out of the intervention entirely. Thus, although this two-time-point design provides stronger evidence of mediation than a cross-sectional study (see below), results may still be subject to selection bias. Special analytic frameworks designed to reduce selection bias can be applied after the fact to test the sensitivity of the findings (Coffman & Zhong, 2012).

Multiple mediator designs. While the example in Figure 1 is limited to one mediator, many program evaluations target and assess the impact of programs on several mediators simultaneously using multiple mediator designs (MacKinnon, 2008). Generally, there are two types of multiple mediator designs used in program evaluation: *all-or-nothing designs*, where participants receive programming related to *all* of the targeted mediators or do not receive any programming at all; and *dismantling designs*, where participants receive all, some, or none of the program content for each of the targeted mediators (MacKinnon, Taborga, et al., 2002). With an all-or-nothing design, even if the intervention is randomized, it is difficult to disentangle the effects of the intervention on the outcome through any specific mediator. In the dismantling design, it is possible to isolate the effects of each specific mediator.

Cross-sectional designs. When it is not possible to conduct an RCT, cross-sectional studies without random assignment can provide information on mediators that may be potential targets for intervention. These "generative" research studies use nonexperimental data from a single point in time to help researchers identify potential risk factors associated with problematic outcomes. These designs are weaker than those described above and cannot be considered causal for several reasons, including 1) it is impossible to control for all confounding variables that might account for associations between mediators and outcomes; 2) temporal precedence cannot be established; and 3)
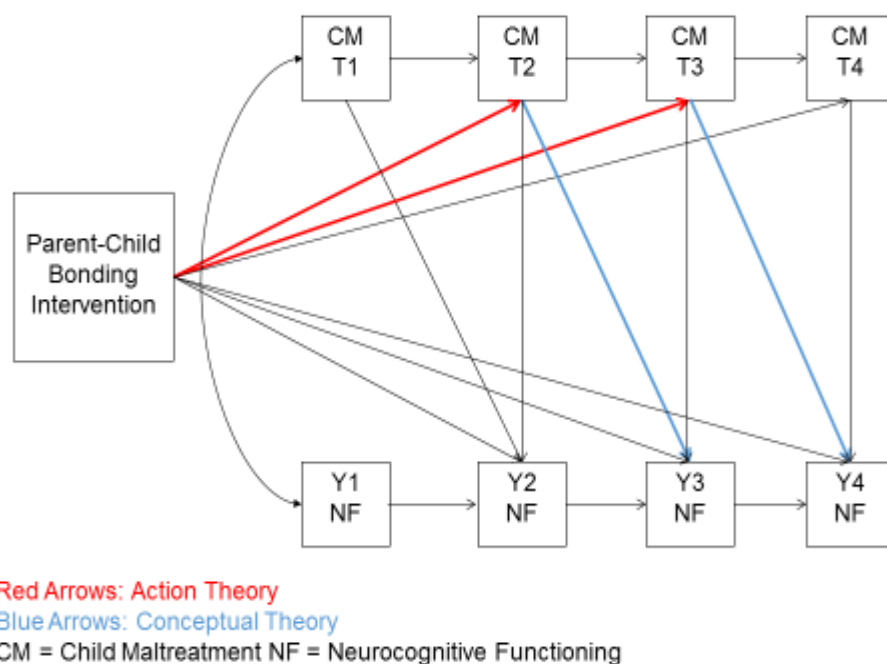
reciprocal causation cannot be ruled out (i.e., A may cause B or B may cause A). However, generative studies may suggest a relationship between a risk factor and an outcome exists (MacKinnon, Taborga, et al., 2002), and offer a starting point for understanding program mechanisms.

OPTIONS FOR MEDIATION ANALYSIS

The discussion of different types of designs for mediation goes hand-in-hand with the discussion of how to analyze and assess mediation. As noted in the previous sections, longitudinal data are very important for making causal inference from mediation designs, but longitudinal data also present another set of options and challenges for evaluation. The analysis options discussed here, along with modern causal analysis that is specific to mediation, can reduce some of the limitations of weaker designs in ways that were not previously possible. In turn, these approaches to mediation will give evaluators more options for making causal inferences regarding interventions and aid in making better and more informed policy decisions.

Cross-lagged panel mediation model. In models with repeated measurements of the mediator and outcome variables, the traditional analytic approach is a cross-lagged panel mediation model (MacKinnon, 2008; Selig & Preacher, 2009). The cross-lagged panel mediation model first assesses the intervention's effect on the mediator (time *t*), adjusting for the level of the mediator at the previous time point (time *t-1;* see Figure 2). Next, this model assesses the effect on the outcome attributed to the mediator at the previous time point (time *t-1*), adjusting for the outcome at time *t-1*. In addition, this second model assesses the remaining effect of the intervention on the outcome that is not attributed to the mediator. By controlling for previous time points, this model helps to minimize bias and eliminate spuriously inflated estimates of β. Figure 2 demonstrates what this model would look like using the previous example of a parent-child bonding intervention study that collected longitudinal data. When the study design does not include random assignment to intervention, a key concern is the possibility that adjusting for the mediator at time *t* may produce different results than looking at trajectories of the mediator over time (Lord, 1967).

Mediated latent growth model. Mediated latent growth models can tell us about trajectories over time (Cheong, MacKinnon, & Khoo, 2003). Cross-

Figure 2. Cross-lagged Panel Mediation

Red Arrows: Action Theory
Blue Arrows: Conceptual Theory
CM = Child Maltreatment　NF = Neurocognitive Functioning

lagged panel mediation models measure mediators and outcomes at a single time point and adjust for previous time points; in contrast, mediated latent growth models use a "slopes-as-outcomes" structure to assess latent trajectories of change in the mediator and outcome. A regression model is used to determine how much change in the mediator over time is due to the intervention (see Figure 3). While these models are useful in evaluation contexts, a primary limitation of mediated latent growth models is that it is impossible to pinpoint the specific time when the critical shift or shifts occur. More recent work has attempted to address this shortcoming (e.g., MacKinnon, 2008; Selig & Preacher, 2009).

Conditional mediation model. An intervention may not work in the same way across different populations. In the example in Figure 1, previous research suggests the intervention may work differently for boys and girls. In other words, child sex might be a *moderator* in the mediation model. Moderator variables may also be continuous in nature, such as family income level.

Understanding potential moderator variables is critical to refining and adapting policies and programs because interventions may need to adjust the mode of intervention delivery, the target mediators of the program, or both, to achieve desired program outcomes for a particular

population (Fairchild & MacKinnon, 2014; MacKinnon et al., 1991). Intervention modifications will depend on whether the moderator is at the individual or group level and whether the moderator variables are themselves malleable (e.g., baseline levels of risk vs. less-malleable factors such as sex).

Future directions. A key challenge in mediation analysis is causal inferences regarding mechanisms of action. These may be tenuous even in RCTs. In RCTs, the link between the program and mediator is considered a causal link because the program is randomized, but the relation between mediators and outcomes is not causal because the mediator itself is not randomized (MacKinnon, Taborga & Morgan-Lopez, 2002). Recent advances in causal mediation may improve causal inferences in nonrandomized evaluations. One such methodology is propensity scoring, which in mediation analysis involves changing the propensity to receive a continuously measured treatment into probability density values from two models that can be fit in ordinary least squares regression: a "numerator" model and a "denominator" model (Coffman & Zhong, 2012; Imai & van Dyk, 2004). Instrumental variable approaches are also emerging as a potential approach to address issues in causal inference in mediation analysis (Reardon, Unlu, Zhu, & Bloom, 2014).

Figure 3. Mediated Latent Growth Model
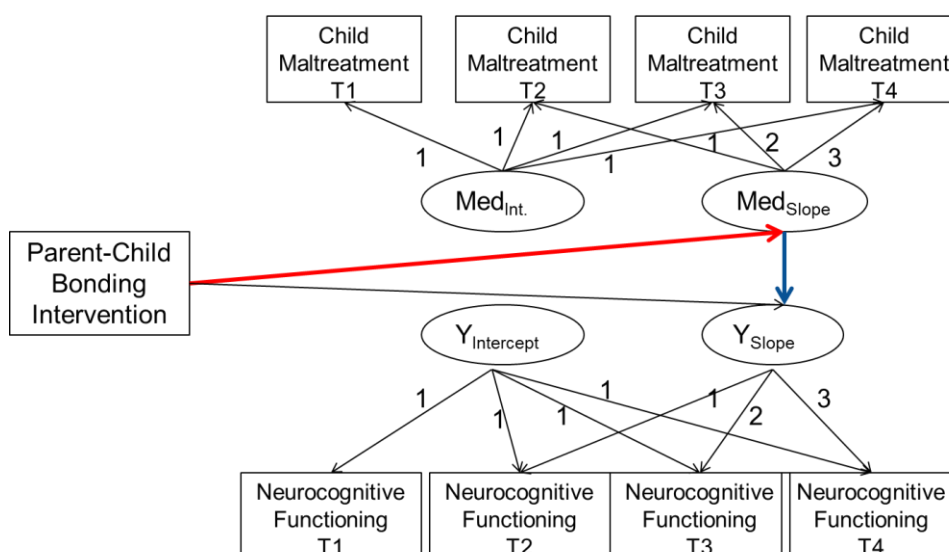
SOME BEST PRACTICES IN MEDIATION

❖ **When developing the intervention logic model, carefully consider which constructs should be measured.** This discussion should involve both methodologists and practitioners and a consideration of existing research and theory. Three elements should be measured: 1) the mechanisms of the action theory that indicate how the program or policy is expected to work (e.g., behaviors that constitute parent-child bonding like soothing, swaddling, cuddling); 2) the expected mediator itself; and 3) the proximal outcome. Ideally, data on all three types of variables would be collected at baseline and subsequent time points. Depending on the timeframe of the study, it may also be appropriate to measure the distal outcome.

❖ **The theories underlying the intervention's mechanisms of change need to be carefully thought out at the beginning of the study.** Having a clear understanding at the front end of the study will allow for concrete recommendations, in the event the mediation analysis reveals an action theory failure, a conceptual theory failure, or both.

❖ **Develop an analysis plan that appropriately accounts for the data structure.** Interventions are often delivered in a way that creates non-independence in the data, either because the same individuals provide multiple data points

over time, or because individuals are clustered within larger units (e.g., schools, classrooms, neighborhoods). Multilevel effects have received considerable attention (e.g., Zhang, Zyphur, & Preacher, 2009), and different intervention effects can be observed at the individual level or at the larger unit level using multilevel frameworks. Analysis strategies like mediated latent growth models are particularly well-suited to estimating multilevel effects. The plan should take into account non-normal mediators or outcomes such as binary data or count data (MacKinnon, 2008).

❖ **Conduct a power analysis for mediation analyses.** When estimating statistical power, the match between the assumptions that underlie the power analysis and the actual planned data analysis should be considered. In addition, the power analysis must be considered in terms of whether the study has a fixed sample size. If the sample size (and number of units, such as schools) cannot be changed, then the power must ask "What is the minimum effect size that can be detected at 80 percent power for a fixed sample size N and X number of schools?" In this approach, the effect sizes for paths α and β (Figure 1) must be changed together to find an optimal but realistic balance. This approach must also take into account the intraclass correlations (the amount of variation in the outcome due to repeated measures, the amount

of variation attributable to individuals being nested in larger units, or both). If the study is being powered based on fixed effect sizes (e.g., known effect sizes from the literature) and the sample sizes can vary, then the question becomes: "What is the fewest number of participants and schools we need to detect an effect size of X for achieving 80 percent power?" More often than not, the complexity of the proposed models for mediation designs requires statistical simulation models for estimating statistical power (e.g., Muthén & Muthén, 2002).

❖ **To make statistical inferences about mediation, use an approach that estimates the sampling distribution of the product** $\alpha\beta$. Current state-of-the-art approaches involve some way of estimating the sampling distribution of the mediation effect by estimating the confidence interval based on the moments of the distribution of the product (e.g., MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002), simulating the distribution of the product directly (i.e., parametric bootstrap; MacKinnon, Lockwood, & Williams, 2004; Selig & Preacher, 2009), or simulating the distribution of the product indirectly through resampling the data (i.e., nonparametric bootstrap; MacKinnon, Fritz, Williams, & Lockwood, 2007). There are several user-friendly software options to estimate confidence intervals for mediation effects available as stand-alone packages in SAS, SPSS, and R (e.g., Tofighi & MacKinnon, 2011); mediation effect modules within existing SEM packages (e.g., Muthén & Muthén, 1998–2012); and in web-based Java applets (e.g., quantpsy.org).

Note: This brief is based on a presentation from OPRE's 2014 meeting, *What works, under what circumstances, and how? Methods for unpacking the 'black box' of programs and policies.* The presentation is available at http://www.opremethodsmeeting.org/2014presentations.html.

REFERENCES

Briggs-Gowan, M., Carter, A., & Ford, J. (2012). Parsing the effects of violence exposure in early childhood: Modeling developmental pathways. *Journal of Pediatric Psychology, 37,* 11–22.

Chen, H.-T. (1990). *Theory-driven evaluations.* Newbury Park, CA: Sage.

Cheong, J.-W., MacKinnon, D. P., & Khoo, S.-T. (2003). Investigation of mediational processes using parallel process latent growth curve modeling. *Structural Equation Modeling, 10*, 238-262.

Coffman, D. L., & Zhong, W. (2012). Assessing mediation using marginal structural models in the presence of confounding and moderation. *Psychological Methods, 17*, 342–664.

Dozier, M., Peloso, E., Lewis, E., Laurenceau, J., & Levine, S. (2008). Effects of an attachment-based intervention on the cortisol production of infants and toddlers in foster care. *Development and Psychopathology, 20,* 845–859.

Fairchild, A. J., & MacKinnon, D. P. (2014). Using mediation and moderation to enhance prevention research. In Z. Sloboda & H. Petras (Eds.), *Defining prevention science* (pp. 537–555). New York, NY: Springer.

Harden, B. J. (2015). Services for families of infants and toddlers experiencing trauma: A research-to-practice brief. Brief prepared for the Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods, 15,* 309–334.

Imai, K., & van Dyk, D. A. (2004). Causal inference with general treatment regimens: Generalizing the propensity score. *Journal of the American Statistical Association, 99*, 854–866.

Lord, F. M. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin, 68*, 304–305.

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

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*, 384–389.

MacKinnon, D. P., Johnson, C. A., Pentz, M. A., Dwyer, J. H., Hansen, W. B., Flay, B. R., & Wang, E. Y. I. (1991). Mediating mechanisms in a school-based drug prevention program: First year effects of the Midwestern Prevention Project. *Health Psychology, 10,* 164–172.

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

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

MacKinnon, D. P., Taborga, M. P., & Morgan-Lopez, A. A. (2002). Mediation designs for tobacco prevention research. *Drug and Alcohol Dependency, 68,* S69–S83.

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus User's Guide* (7th Ed.). Los Angeles, CA: Muthén & Muthén.

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling, 9,* 599–620.

Reardon, S. F., Unlu, F., Zhu, P., & Bloom, H. S. (2014). Bias and bias correction in multisite instrumental variable analysis of heterogeneous mediator effects. *Journal of Educational and Behavioral Statistics, 39,* 53–86.

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

Tofighi, D., & MacKinnon, D. P. (2011). RMediation: An R package for mediation analysis confidence intervals. *Behavior Research Methods, 43*, 692–700.

Yoches, M., Beeber, L., Jones Harden, B., Malik, N., & Summers, S. (2011). Children's trauma exposure. In S. Summers & R. Chazan-Cohen (Eds.), *Understanding early childhood mental health: A practical guide for professionals* (pp. 79–98). Baltimore, MD: Brookes.

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.