Part 1

FUNDAMENTALS

1

Randomized Explanatory Trials

In God we trust. All others bring data

- W. EDWARDS DEMING

INTRODUCTION

Randomized controlled trials (RCTs) are considered by many to be the "gold standard" for evaluating program efficacy and effectiveness. An RCT randomly assigns people to receive one of several intervention forms, with one condition being a treatment group (which receives the program or intervention of interest) and the other condition being a standard of comparison or control group. Sometimes the comparison condition is a competing program, such as treatment-as-usual, with the goal being to compare the relative efficacy or effectiveness of the two programs. Major decision points include, among others, how to sample and randomly allocate individuals to condition, how and

when to measure the outcomes (e.g., at baseline, immediately after program participation ends, at a later follow-up), and how to analyze the data to make valid inferences about program effects, among others.

A thesis of the present book is that RCTs are, in many cases, inadequate gold standards for purposes of evaluating behavioral interventions. I propose instead what I call **randomized explanatory trials** (RETs). An RET is much like an RCT in that it is a program evaluation that uses random assignment to conditions. However, it differs from an RCT because it explicitly includes mediators and moderators of program effects. Given the presence of mediation and moderation, RETs require additional levels of theorizing, special methodological considerations, and alternative statistical analyses relative to traditional RCTs. The term "explanatory randomized trials*"* has been used by a small cadre of social and health scientists to contrast them with what they call **pragmatic randomized trials**. Some 45 years ago, Schwartz and Lellouch (1967) characterized an RET as an RCT designed to shed insights on the causal impact of a treatment on an outcome. By contrast, pragmatic trials are designed to compare the relative effectiveness of two or more treatments in practical conditions. Since their seminal paper, the terms explanatory and pragmatic trials have been used in diverse ways, but the essence of the distinction is a concern for understanding the causal mechanisms underlying the effect of a treatment under idealized experimental conditions on the one hand versus the comparative effects of treatments in practical settings on the other hand.

It is clear that since 1967, the concept of an RET has evolved considerably since its introduction. To be sure, the essence of a randomized explanatory trial remains that of understanding the causal mechanisms that account for the effects of treatments on outcomes but RETs have grown to include the concepts of mediation and moderation in conjunction with far more advanced conceptual, psychometric, and methodological tools than when Schwartz and Lellouch first coined the term. My adoption of the term updates and modernizes the notion of an RET to include a larger frame of newer scientific activities/constructs, not the least of which is the use of structural equation modeling to analyze randomized trials. I also take the RET concept a step further to argue that the dichotomization of trials into RETs versus pragmatic trials is counterproductive. Pragmatic trials often morph into RETs when we seek to understand why or why not interventions work in applied settings and to incorporate our understandings into intervention design and delivery. It no longer is enough to document the effectiveness rates of a treatment in an applied context or to demonstrate that one treatment works better than another treatment in an applied context. This is too narrow a perspective. Instead, we need to know *why* one treatment works better than another treatment in

applied contexts, *why* a treatment fails or succeeds in an applied context, and *for whom this is and is not the case*. In other words, pragmatic trials need to incorporate RET perspectives that elucidate mediators and moderators and that make use of modern scientific tools and methods to help us understand and improve our interventions.

The idea of conducting mediation and moderation analyses in a randomized trial certainly is not a new one. Agencies such as the National Institute of Health (NIH) and the Centers for Disease Control and Prevention (CDCP) have been pleading for years for researchers to adopt mechanism-based perspectives for program/intervention evaluation. The National Institute of Mental Health (NIMH) has invoked the concept of experimental therapeutics (i.e., identifying the mechanisms of complex behavior) in their Strategic Plan (NIMH, 2019), which highlights mediation at different levels of analysis. The concepts of personalized medicine, precision medicine, stratified medicine, and theranostics are medical approaches that segregate people into different groups and then tailors treatment protocols to patients based on the group to which they belong (Smith, 2012; Academy of Medical Sciences, 2015). This approach, at essence, represents a form of moderation analysis, with the guiding premise being that "one treatment protocol does not fit all."

There are, of course, many examples of RCTs with mediation and moderation dynamics incorporated into them. However, it is one thing to state at a global level that one should include mediation and moderation in an RCT; it is quite another to actually do so in a scientifically rigorous way. In this book, I address core issues for designing RETs to evaluate programs and behavioral interventions. My intent is to expand the approach that program evaluators bring to the task of program evaluation, from that of just determining program impact on outcomes to one where, in addition to outcome analysis, one seeks to gain an understanding of *why*, *for whom,* and *in what contexts* a program is effective. The idea is to provide guidance to program/intervention designers on how to improve their interventions to make them more effective. RETs help accomplish this.

Program evaluation can have different goals. One goal might be to determine if an existing program being used in a clinic or organization is effective and how one can improve its effectiveness. Another goal might be to evaluate a newly developed program that one is thinking of rolling out to clinics or organizations to determine if the program likely will be effective in those settings. A third goal might be for the evaluation study to advance scientific theory by explicating the active ingredients and boundary conditions of a type of program, such as evaluating cognitive behavior therapy principles or assumptions. All of these goals require us to attend to matters of mediation and moderation and to adopt a mechanistic perspective when pursuing program evaluation. Granted, there might be nuanced differences for designing the RET under the different goal structures, but inevitably, the general spirit of an RET applies in all cases.

In this chapter, I provide an overview of key theoretical, methodological, and analytic issues that typically must be considered when designing an RET. These include mediator and moderator mapping, the identification of confounders, recognizing and exploring non-linear functional forms, addressing possible reverse causality, dealing with measurement error, evolving logic models for temporal dynamics, and making decisions about sample sizes, among others. All of these topics are considered in depth in future chapters. The primary focus of this book is on the statistical analysis of RETs using structural equation modeling. However, theory, method, and statistics are intimately linked for RETs, so all three topics receive consideration.

RET FRAMING USING MEDIATION AND MODERATION

Figure 1.1 presents an influence diagram to illustrate core elements of an RET to evaluate a program aimed at preventing future drug use in adolescents. In this diagram, variables are indicated by rectangles and a causal relationship between two variables is indicated by a straight arrow. The arrow emanates from the presumed cause and points towards the presumed effect. These arrows are referred to as **causal paths**, which are assumed to vary in strength. I describe the circled *d* terms shortly. The influence diagram should be thought of at the level of theory, not data analysis; it is a formal heuristic for representing hypothesized causal relationships between variables in an RET.

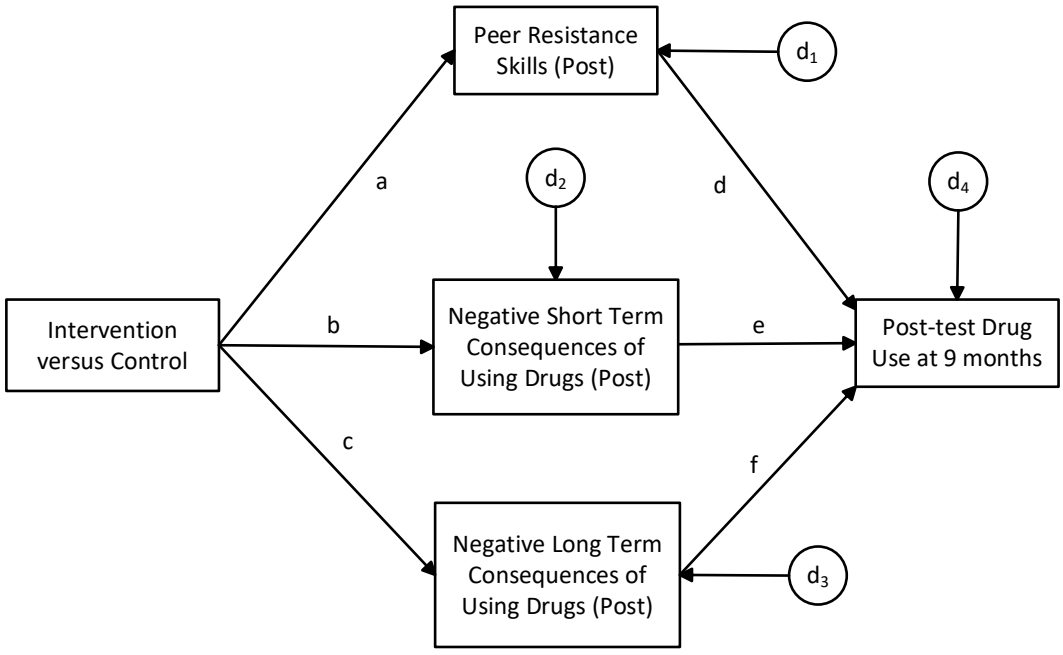


FIGURE 1.1. A Randomized explanatory trial

In Figure 1.1, there is an intervention versus control condition on the left most part of the figure, which, in this case, is represented as a two-valued variable (0 = person is in the control condition, 1 = person is in the treatment condition). The primary outcome that the program is designed to change is on the far right. Mediators are in the middle. A mediator is a mechanism through which the program is presumed to have its effect on the outcome. It explains *why* a program affects the outcome. For example, in Figure 1.1, the program is assumed to teach adolescents peer resistance skills (path *a*) and these skills, in turn, are presumed to affect future drug use (path *d*).

Linking the Program to Mediators

In an RET, a careful analysis of the program components is undertaken and the key mediators that each program component targets are identified. In Figure 1.1, the program has three components (1) a component to teach youth peer resistance skills, (2) a component to educate youth about negative short-term consequences of using drugs, and (3) a component to teach youth about negative long-term consequences of using drugs. Each mediator is measured after program completion as is youth drug use 9 months later.

In an RET, we expect the program will affect each mediator associated with a given program component, i.e., the effects signified by paths *a*, *b* and *c* all will be sizable. The magnitude of a causal effect is reflected by a **path coefficient**. In general, when the path coefficient equals zero, the path is presumed not to exist; the larger the absolute value of the path coefficient, the stronger the causal effect is presumed to be, everything else being equal. A common strategy used to determine the values of the path coefficients is linear regression, which regresses a measure of the presumed effect onto measures of the presumed causes. The path coefficient is the regression coefficient associated with a presumed cause in the regression equation. In Figure 1.1, if the path coefficient for *a*, *b* or *c* is found to be trivial in magnitude, then this suggests the program failed to change sufficiently a mediator that it was intended to change. If we have adequately mapped a mediator onto its corresponding program component(s), we learn from this analysis the specific component of the program that is culpable for a reduced effect on the outcome. A trivial coefficient is a signal, after the fact, that program designers need to revisit program activities used for that component because the ones used were not effective. Suppose in an RET that paths *a* and *c* are found to be reasonably sized but this is not the case for path *b*. This means program designers need to re-think the component aimed at impacting youth perceptions of the negative short-term consequences of using drugs. A powerful feature of RETs is that they help pinpoint where program components fail to bring about change in the targeted mediator and where change efforts were successful. This is invaluable information for improving the program.

Linking Mediators to Outcomes

Also of interest in an RET is the strength of paths linking each mediator to the outcome (paths *d*, *e* and *f*). These paths represent assumptions that program designers made about the determinants of drug use when they designed the program, namely that such use is impacted independently by peer resistance skills, by the perceived negative short-term consequences of using drugs, and by the perceived negative long-term consequences of using drugs. If in an RET a given path in this portion of the model is found to be non-trivial in magnitude, then this is consistent with the assumptions made by program designers. If, however, the path is weak and non-significant, then this raises questions about the fundamental assumption of program designers for that path. For example, suppose path *f* is weak and non-significant. This means that although program designers thought that perceptions about the long-term negative consequences of using drugs would affect future drug use, the evidence from the RET does not support this assumption. Perhaps one should consider streamlining the program by dropping this component.

Note that this framing of RETs focuses analyses on the separate links in a mediational chain rather than an omnibus test of mediation per se, the latter of which is typical of most analyses of mediation. By focusing on the separate links, we gain more specific directives about how we can improve the program.

Disturbance Terms: Thinking about Variables Not in the Model

The influence diagram in Figure 1.1 also includes disturbance terms for the mediators; see d_1, d_2, and d_3. These terms reflect variables that are unmeasured and formally excluded from the RET model but that are thought to influence the mediator in question. For example, d_1 reflects all variables that are ignored by the investigator in the RET model that impact posttest peer resistance skills of individuals. The disturbance term d_2 reflects all variables ignored by the investigator that impact the posttest perceived short term consequences of using drugs. In Figure 1.1, the only specified determinant of the mediators is the treatment versus control condition, so the disturbance terms reflect everything else that influences each mediator besides this variable, respectively. Disturbance terms usually are characterized quantitatively using the concept of unexplained variance, i.e., the number association with them reflects the proportion of variation in the mediator that is unexplained by the variables in the RET model.

Inclusion of disturbance terms in the theory is important. They underscore the fact that in an RET, we not only must think about variables included in the RET model, but we also must think about variables that are not in the model and that we chose to ignore. You will see in future chapters how RET analytics incorporate disturbance terms into RET analysis and how important it is to do so. As a prelude, it turns out that we can use

the concept of disturbance terms to inform a concept known as **exceptions to the rule**. Consider path *a* in Figure 1.1. If this path is found to be statistically significant in the predicted direction, the general rule we formulate is that youth who participate in the prevention program have more peer resistance skills than youth who do not participate in the program. However, we know there invariably will be exceptions to this rule, i.e., some program participants will have lower peer resistance skills than some non-program participants. How many exceptions are there? If we randomly select a person from the treatment group and a person from the control group, how often will we find, contrary to the "rule," that the person in the control group shows more improvement than the person in the treatment group? Will this be true 5% of the time? 10%? 20%? As I show in later chapters, the larger the disturbance term, the greater the proportion of exceptions there will be. Program evaluations should document the extent to which such exceptions occur.

Also of interest in an RET is the disturbance term associated with the outcome, in this case, d_4 in Figure 1.1. Like program effects on mediators, this term can be used to document "exceptions to the rule" vis-a-vis the impact of mediators on the outcome. For example, in Figure 1.1. if path *d* is statistically significant and negative in value, we formulate the "rule" that people who score higher than others on peer resistance skills should also score lower than others on the outcome, namely future drug use. How many exceptions to this "rule" are there?

The disturbance term for the outcome serves another function. In SEM, the variance of the outcome disturbance term can be expressed so that it represents unexplained variance in the outcome relative to the mediation predictors. It is analogous to $1\text{-}R^2$ in a multiple regression analysis. If this value is large, then this suggests there may be many determinants of the outcome that are unaddressed by the program. It is a sign that perhaps program designers need to expand their program by adding new components to address these other sources of outcome variation. For example, suppose the variance of the outcome disturbance term in Figure 1.1 represents 80% unexplained variance in drug use. This tells me there possibly are other factors besides the three targeted mediators that significantly impact drug use in the sense that they account for substantial variation in it (though measurement error can inflate this estimate). Many of these factors will not be amenable to change (e.g., genetics, biological sex) but others might be. It is helpful if we can design the RET to gain perspectives on what these other factors might be.

Moderator Variables in RETs

Another feature of RETs is their focus on moderated relationships. Sometimes causal effects are stronger for some sub-groups than others. For example, for the RET in Figure 1.1, I might find that the intervention is effective at increasing peer resistance skills for

non-Latinx adolescents but not for Latinx adolescents; that is, the strength of path *b* changes as a function of ethnicity. Given this result, the feedback to program designers would be to revisit the program activities aimed at increasing peer resistance skills for Latinx adolescents because those activities are not working for these adolescents. Perhaps something different needs to be done for them versus non-Latinx adolescents.

Another possibility is that we might find that perceptions of the negative long-term consequences of using drugs impacts future drug use (path *e*) for adolescent girls but not for boys. In this case, the feedback to program designers might be that such perceptions do not seem relevant for boys and that perhaps for boys, the program activities aimed at increasing these perceptions should be dropped, thereby streamlining the program.

Whereas mediation seeks to answer the question of why an outcome effect occurs by identifying the underlying mechanisms for that effect, moderation seeks to identify the boundary conditions of effects and answers questions about under what conditions or for what subgroups an effect is stronger or weaker. These "conditions" might refer to different groups or subpopulations (e.g., for males but not females), different times (at time A but not time B), or different contexts (in situation A but not situation B). Such feedback also is useful for program designers and administrators because it provides perspectives on the generalizability of program effects as well as program "reach." As such, when designing an RET, we will want to think about providing feedback to program staff about program generalizability across key subgroups and settings. This is accomplished by incorporating moderator variables into the design.

Figure 1.2 presents an influence diagram illustrating moderator dynamics for the drug use prevention program. In this case, the program evaluation found that ethnicity moderated the impact of the program on peer resistance skills. The moderation is indicated by a straight arrow pointing to the causal path between the two variables rather than to a variable per se. This is because it is the causal relationship itself that varies as a function of the moderator variable. In this case, we found that the program is less effective at teaching peer resistance skills to Latinx adolescents than to non-Latinx adolescents (path *g*). Figure 1.2 also illustrates the case where biological sex moderates the impact of the mediator of perceived long-term negative consequences of drug use on future drug use (path *h*); the link is weaker for males than it is for females.

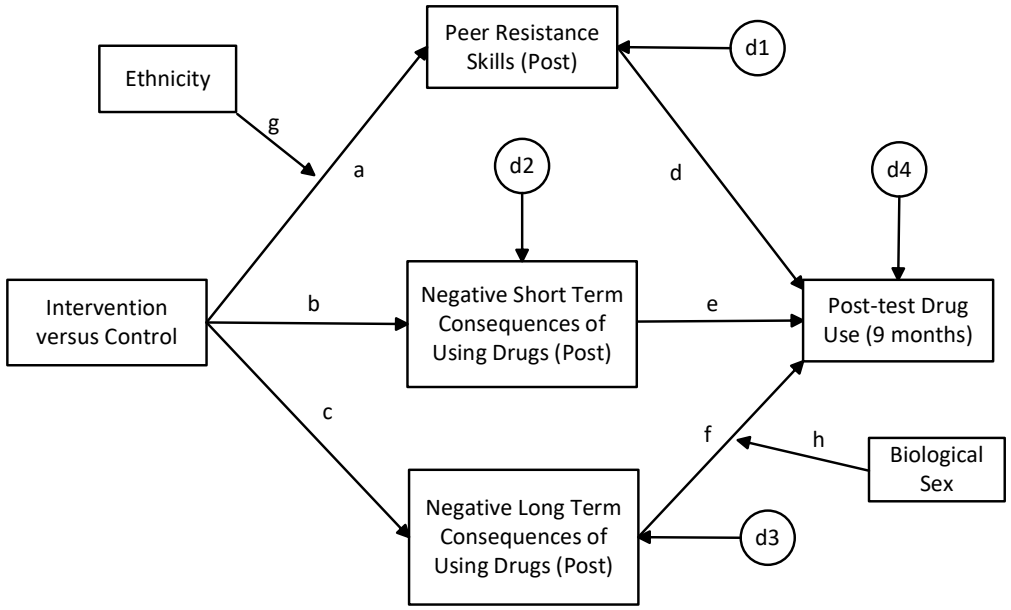


FIGURE 1.2. Model with moderated mediation

A unique feature of RETs is that they integrate mediation and moderation analyses. If a program affects a dependent variable differentially for one subgroup versus another, an RET can pinpoint where in the mediational system the subgroup difference occurs. As I discuss later, the corrective action program designers take to remove suboptimal response in one of the subgroups likely will differ depending on where in the system the moderation occurs. For example, the program revision might be different if the moderating effect of biological sex weakens path *c* for males (i.e., the program impacts the perceived long term negative consequences mediator for females but not males) versus path *f* (i.e., perceived long term negative consequences predict drug use for females but not males). In a traditional RCT that tests moderation, such information is not generated – we only learn that the program is more or less effective for one subgroup than another subgroup on the outcome. In an RET, we obtain information that identifies the locus of the subgroup effect within a mediation chain thereby providing perspectives on *why* subgroup differences on the outcome occur.

RETs Instead of RCTs

In sum, RETs are a strong form of program evaluation that go beyond simple outcome analysis in traditional RCTs. RETs force the program evaluator to think carefully about the specific program components and the potential mediating variables that each

component is intended to impact. RETs force us to think about whether and how the targeted mediators combine to impact study outcomes. RETs encourage us to ask questions about the generalizability of effects across subgroups/conditions and to integrate such questions with mediation dynamics. As I show in future chapters, RETs require specialized experimental designs and measurement strategies that allow us to gain perspectives on estimates of the presumed causal relationships specified in an RET model. Simple ANOVA, ANCOVA, or linear mixed models will not suffice. I argue instead that variants of structural equation modeling (SEM) are a preferred analytic tool and I build a case for this argument. Many researchers do not think of SEM as a tool for analyzing experiments. However, as I will show, the approach has much to offer for the analysis of RETs. RET designs in many contexts can identify program components that changed their targeted mediators and components that failed to do so. This provides specific directions to program designers for how to improve the program. RET designs also provide perspectives on assumptions about the importance of each mediator in influencing the outcome and thereby can identify program components that might be streamlined.

In my view, the "gold standard" for program evaluation should not be an RCT but instead it should be an RET. When we conduct program evaluations, it is too low a standard simply to determine whether the program has an impact on an outcome, which is the traditional focus of an RCT. Instead, we should demand our evaluations give us feedback on what program components are most effective and, if the program is not having much impact, understanding why this is the case. As well, our evaluations should provide perspectives on whether program effects are limited to certain subgroups at the expense of others. A well-crafted RET accomplishes these desiderata.

THE SCOPE OF EVALUATIONS AND THE CURRENT BOOK

There are many facets of program evaluation. Chen (2015) defines program evaluation as the process of gathering systematic data and contextual information about an intervention program to evaluate a program's planning, implementation, and/or effectiveness. Many excellent books have been written about program evaluation that emphasize one or more of these facets (e.g., Chen, 2015; Haggerty & Mrazek, 1994; Thomas. & Rothman, 2013). The current book is written more narrowly than these treatments and is designed to fill a significant gap in the program evaluation literature. Specifically, I focus on the planning, conduct, and analysis of RETs that use an infrequently applied analytic method to them, namely structural equation modeling (SEM). I articulate the unique strengths of SEM for evaluation designs that use mediation and/or moderation analysis could with

randomization to treatment and control groups and show how such approaches can provide essential feedback on how to improve programs.

These topics, taken together, receive scant attention in the evaluation literature. Importantly, the tools and perspectives I describe in this book extend well beyond traditional evaluations of existing programs in field settings. For example, a common enterprise in the social and health sciences is for scientists to develop new interventions or programs with the idea that if adopted, they will help mitigate significant social problems and/or improve the lives of large numbers of people. During such program development, scientists often conduct RETs on their programs to gain perspectives on the efficacy or effectiveness of them, also seeking insights into how to improve them or to advance theory surrounding the outcomes studied. The material covered in this book speaks to these scientists as well as more traditional program evaluators to help them design informative studies to attain their goals.

FACETS OF AN RET

In this section, I provide an overview of the major facets of an RET that you will need to consider as you plan and execute an RET for purposes of program evaluation. Each facet is elaborated in future chapters.

Facet 1: Mediator Mapping

A key step in designing an RET is what I call **mediator mapping**. This involves identifying the mediators that specific program activities target, presumably on the assumption that changes in them will bring about changes in the outcomes. In Figure 1.1, there were three distinct program components, each targeting a different construct. When designing programs, it is not uncommon for designers to develop an *a priori* logic model to describe program goals, activities, and foci (SAMHSA, 2012; Kellogg Foundation, 2004). Logic models can be quite detailed, often including manuals and field guides for program activities, timelines, required resources, methods for fidelity checks, and strategies for overcoming obstacles. Good logic models usually include descriptions of the relationships between program activities and desired outcomes, including the pathways through which the program activities produce the desired outcomes. They also sometimes describe contextual factors that might affect program effectiveness. I refer to such descriptions as **conceptual logic models**. Sometimes conceptual logic models are well articulated and other times they are implicit. The task of the program evaluator is to make the conceptual logic model of an intervention explicit. One way of doing so is by constructing an influence diagram, per Figure 1.1. An advantage of constructing an

influence diagram is that it forces one to elaborate facets of the underlying theory that might otherwise not be taken into account. If a program already has a well-articulated conceptual logic model, then it can be used by program evaluators to conduct mediator mapping. In the absence of such a model, program evaluators need to construct their own influence diagram, doing what the program designer should have done at the outset.

This step is important and designates a mindset to mediator identification that is sometimes downplayed, namely doing a deep dive into program contents to identify the specific mechanisms the program targets. The identified mechanisms/mediators then become part of the RET. As an example, I recently consulted in the conduct of an RET to evaluate a new web-based program tied to cognitive behavior therapy (CBT) to help people deal with having to live with chronic pain. I carefully examined the training manuals for staff as well as the materials used in the program and, together with the program designers and staff, identified eight program components that were addressed with clients, (1) self-distancing, (2) perspective broadening, (3) maladaptive worrying, (4) rumination, (5) perceived control, (6) expectancy bias, (7) context sensitivity, and (8) coping flexibility. These constructs then became seven of the mediators we focused on in the RET. We sought to determine in the RET (a) if the program did, in fact, change each of these mediators and (b) did each mediator have a reasonable impact on the outcomes of pain reduction and life satisfaction.

The Kitchen Sink Approach to Program Design

Some program designers do not focus their efforts on a few well-defined components when trying to impact an outcome. Rather, they seek to change many intervening constructs, making it difficult to separate out and evaluate those components that are primarily responsible for program success, often referred to as the **active ingredients** of the program. Consider the case of Positive Youth Development (PYD) programs to address adolescent problem behaviors. These broad-based programs seek to maximize child and youth development in different developmental domains (Catalano, Berglund, Ryan, Lonczak & Hawkins, 2004; Flay, 2002; Flay, Allred, & Ordway, 2001; Kirby, 1984). Catalano et al. (2004) specified 15 elements that are critical for PYD programs to address in order to be effective, including (1) promoting bonding, (2) fostering resilience, (3) promoting social competence, (4) promoting emotional competence, (5) promoting cognitive competence, (6) promoting behavioral competence, (7) promoting moral competence, (8) fostering self-determination, (9) fostering spirituality, (10) fostering self-efficacy, (11) fostering clear and positive identity, (12) fostering belief in the future, (13) providing recognition for positive behavior, (14) providing opportunities for prosocial involvement, and (15) fostering prosocial norms. Each of these variables potentially

represents a separate program component and each carry with it a set of possible mediators of program effects, usually on multiple outcomes (e.g., school performance, obesity, delinquency, bullying).

The presence of so many components, mediators and program outcomes raises the question of whether we truly expect a given PYD program to have strong, demonstrable effects on each of the 15 classes of mediators and whether we truly expect each of these mediators to have a strong, demonstrable effect on every measured outcome. Without a coherent, manageable theory organizing the linking of program components to specific mediators and, in turn, to specific outcomes, our efforts to construct cost-efficient, resource respectful, and effective positive youth development programs is hampered.

This problem does not just plague PYD program evaluations. For interventions aimed at adolescents that address the family context, for example, programs might address parental monitoring, parental supervision in specific domains (e.g., for school work, for friendship networks), parent-youth communication, shared activities between parent and child, parental discipline strategies, parenting styles (e.g., warmth, control), parental achievement orientations for their children, family stress, family cohesion, the quality of sibling relationships, and so on. The fact is, when interventions are mounted, there can be a tendency for program designers to bring to bear as many potentially relevant factors as possible in the hopes that a few of them "stick" and have an effect. The problem with this strategy is that when an outcome-only evaluation study is completed, we do not gain perspectives on the "active ingredients" of the program nor can we identify which program components are "not working" and why. In this book, I also consider traditional, big data, and machine learning methods for identifying the core mediators from a larger set of mediators that should be prioritized. Once identified, these prioritized mediators can be placed into an influence diagram and subjected to more detailed SEM analyses.

Balancing Mediator Specificity and Mediator Abstractness

Related to the above is the problem of relying on broad-based, abstract theoretical frameworks whose constructs can be conceptualized, dimensionalized, and related to one another in a multitude of ways as a basis for RET theoretical guidance. If one is designing a program to prevent unintended pregnancies for inner city, middle school Latinx youth, broad frameworks like Bronfenbrenner's (1979) ecological systems theory are not going to be of much assistance to a program designer who needs to formulate specific program activities for youth to engage in. Neither can knowing each of the 15 "core elements" of PYD programs identified by Catalano et al. (2004) be said to provide sufficient theoretical guidance for the evaluation of PYD programs because of the

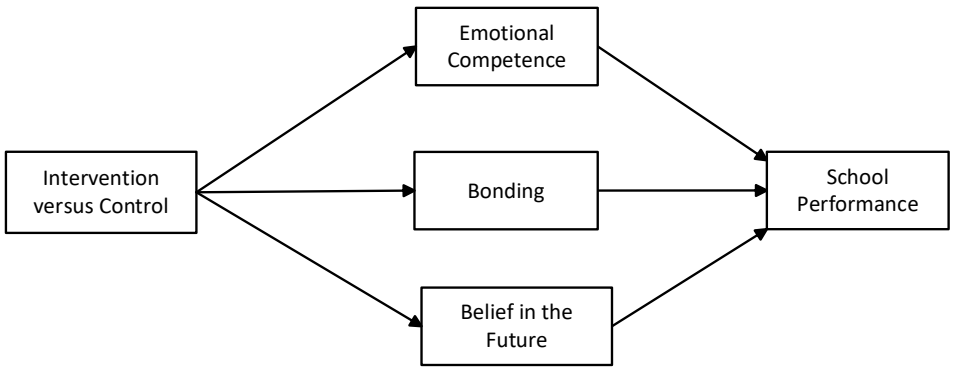
abstractness of the variable categories and the complex relationships that likely exist between the categories. For example, one of the 15 categories is emotional competence. What is "emotional competence" and what are the core dimensions of it that should be targeted by program activities? Catalano et al. (2004) state that emotional competence is the ability to identify and respond to feelings and emotional reactions in oneself and others. They cite Salovey and Mayer's (1989) five elements of emotional competence as a useful framework for this category, which includes (1) knowing one's emotions, (2) managing emotions, (3) motivating oneself, (4) recognizing emotions in others, and (5) handling relationships. This five-component elaboration is helpful because it articulates for program designers five skill sets on which to focus a program. However, each of the five elements still contains a level of abstractness that requires theoretical articulation for purposes of program design. For example, which emotions should a program focus on (e.g., anger, love, sadness, despair); what situational contexts in which those emotions might be experienced should a program focus on; do we need to address just the valence of the emotion or both the valence and intensity of emotions; and so on. Without more detailed theoretical guidance, program designers will use their own (sometimes misguided) intuition to evolve principles of change surrounding the five categories. Program evaluators need to strike a balance between being conceptually specific enough that program and intervention designers have a clear, evidence-based road map for how to re-structure a program based on an RET, but at the same time, giving program designers some flexibility to instantiate and adapt the program to their particular context.

The mediators in a conceptual logic model are sometimes conceptually specific, such as in the case of the drug prevention program in Figure 1.1. Other times, the mediators are conceptually abstract, such as in the case of the 15 "essential elements" of PYD programs. Not only is the analysis of theoretical mediators in the latter case unwieldy, but the "advice" program designers take away from the analysis will be abstract (e.g., strengthen the program so that it more effectively addresses spirituality and emotional competence). The more specific we can be when mapping program components onto theoretically relevant mediators, the more concrete prescriptions for program improvement can be. Indeed, if the identified mediators are too abstract (e.g., spirituality), we may not even have a clear sense of the more specific facets of the mediator that are relevant. I have found a useful heuristic for thinking about the level of specificity of a mediator is to ask whether evaluating that mediator in an RET will yield meaningful and concrete "advice" to administrators and managers of the program. Having said that, I recognize that there often is a balancing act in terms of a general versus specific focus of mediators. I revisit this matter in future chapters.

Articulating Causal Relationships among Mediators

When pursuing mediator mapping in an RET, it is important to recognize there may be causal relationships among mediators. Consider the two models in Figure 1.3 that relate three mediators identified by Catalano et al. (emotional competence, bonding, belief in the future) to future school performance. The first model (Figure 1.3a) assumes that none of the mediators are causally related to one another. The second model (Figure 1.3b) assumes that emotional competence influences school performance directly but it also affects school performance indirectly by virtue of its influence on bonding. The model in Figure 1.3a ignores this latter source of influence and therefore would underestimate the overall effect of emotional competence on school performance, i.e., individuals who are emotionally competent are more likely to bond with others in their school and, in turn, to school more generally. Failure to take the causal relations between the mediators into account can result in program evaluators judging a given mediator as less important than it is. For example, by omitting the causal path from emotional competence to bonding, the importance of emotional competence is judged purely in terms of its independent effect on school performance, ignoring its centrality to increasing bonding which, in turn, affects school performance. Plausible causal relationships among mediators need to be included in one's conceptual logic model, as appropriate.

(a)



(b)

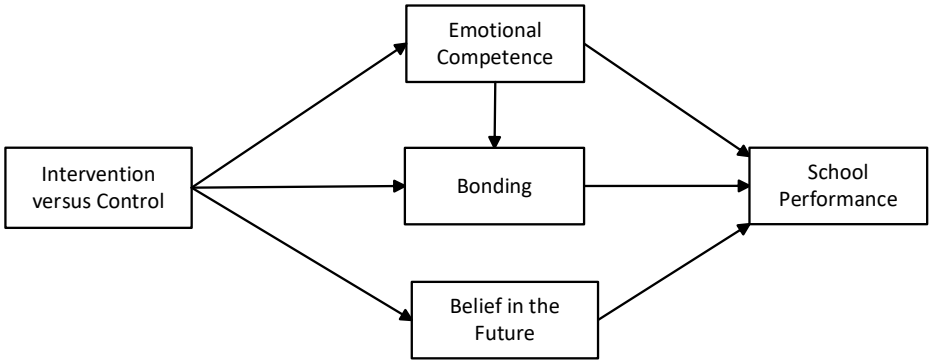


FIGURE 1.3. Two mediational models: (a) No causal influences between mediators, (b) Causal influences between mediators

Specifying the Functions by Which Mediators Impact Outcomes

A final facet of mediator mapping is to specify the likely functions that link each mediator to an outcome as well as how the mediators combine multivariately to impact the outcome. For the former, the relationship might be linear or non-linear. Depending on the assumed function, you might design the RET differently. As well, the conceptual logic models in Figures 1.1 and 1.3 presume the mediators combine additively to impact the outcome. It is, of course, possible that two or more of the mediators interact with one another to impact the outcome. This possibility also should be considered.

In sum, a major task for the design of an RET is to conduct mediator mapping, which includes (a) identifying the mediators that a program targets on the assumption that those mediators are relevant to the outcome, (b) identifying mediators that are not targeted by the program but that may be worth exploring for incorporation into the program in the future, (c) determining the appropriate level of specificity/abstractness of the mediators, (d) mapping the causal relationships among the mediators, and (e) specifying the likely functional forms that link the mediators to the outcome(s).

Differences between Traditional Mediation Analysis and Mediation Analysis in RETs

For purposes of program evaluation, mediation analysis can be divided into three segments. First, we want to determine if the program impacts the outcome(s) of interest. Second, we want to determine if the mediators a program targets are indeed relevant to

the outcome(s). Finally, we want to determine which of the mediators a program targets is successfully changed by the program. Stated another way, we pursue experimental designs and statistical analyses to answer three questions, (1) is there an overall effect of the program on the outcome? (2) are each of the targeted mediators indeed relevant to the outcome? and (3) does the program affect each of the targeted mediators?

As will be apparent in later chapters, the focus on these questions brings with it a somewhat different mindset than that of traditional mediation analysis. The traditional approach typically asks whether a given mediator, M, can account for some (or all) of the effects of a distal variable T (a treatment) on an outcome Y. This question is addressed by multiplying the estimated causal coefficient for the effect of T on M by the estimated causal coefficient for the effect of M on Y to yield an omnibus, single number that reflects the estimated effect of T on Y through the causal chain T→M→Y. For example, in Figure 1.1, it might be found that the coefficient for the effect of the program on drug use (quantified as the number of days a person has used drugs in the past 30 days) through the refusal skills mediator is -3.0, i.e., the treatment decreases drug use, on average, by 3 days through this chain.

In my opinion, such omnibus tests are not as informative as a careful analysis of the individual links in the mediational chain for purposes of program evaluation. The omnibus test ignores these individual links and provides perspectives on them only indirectly or partially. For example, if the omnibus test is non-zero and statistically significant, this suggests that each of the links in the mediational chain, namely T→M and M→Y, are non-zero. If the omnibus test is statistically non-significant, we know at least one of the links in the mediational chain is "broken." However, we do not know which one nor how many of the links are "broken." By analyzing the individual links in the chain, we can pinpoint where the "broken link" occurs so that we can then decide if the link is fixable. If the "broken link" is that the treatment does not meaningfully affect the mediator, perhaps program staff or we as scientists can figure out how to change the treatment so that it affects the mediator. However, is it even worth doing so if we also learn that the M→Y link is "broken"? If the M→Y link is "broken," what implications does this have for program revision? Should we abandon program activities that seek to change M in this case? Might it be possible to alter the program to strengthen the causal coefficient linking M to Y?

I find I can provide useful advice to program staff by analyzing and juxtaposing results for the individual links of mediational chains, with only marginal information gain added by focusing on the omnibus tests of mediation. As you will see in later chapters, by relegating omnibus mediation tests to the conceptual and substantive backyard, many of the statistical challenges of mediation analysis go with them. This is not to say scenarios

do not exist where omnibus tests are of interest. However, for purposes of providing feedback to program developers, a focus on the three core facets of evaluation (does the program meaningfully affect the outcome; are the target mediators relevant to the outcome; does the program meaningfully affect the targeted mediators) takes precedence.

Some scientists argue that omnibus indices of mediation are useful for identifying the most important mediators among a set of mediators. For example, I might find that the refusal skills mediation chain in Figure 1.1 reduces drug use, on average, by 3 days, that the short-term consequences chain reduces drug use, on average, by 2 days, and that the long-term consequences chain reduces drug use by 1 day, on average. As I will show in future chapters, reliance on these indices is not necessarily the best way to document the relative import of mediators in program evaluation because they confound (a) the effect of the program on the mediator with (b) the effect of the mediator on the outcome. I prefer to unconfound these effects when evaluating the role of mediators in producing program effects.

My general orientation to mediation analysis for program evaluation is to work from the ground up, one link at a time in each mediational chain and then integrating the knowledge about each link to make broader statements about mediation dynamics. The traditional approach to mediation analysis, by contrast, focuses on omnibus mediation dynamics by collapsing across all links in a mediational chain without digging deeper into what is happening at the level of individual links. In this book, I seek to bring a balance to the two approaches, but with a preference for individual link analyses.

Facet 2: Moderator Mapping

Another key facet of RET design is what I call **moderator mapping**. This involves identifying moderators of the T→M link in a mediational chain and/or moderators of the M→Y link. As noted, moderator mapping often takes the form of locating subgroups of the program population for whom the program does not work as well as other subgroups. Program revisions can then be made to remove the group disparity. A strength of RET analysis is that it can pinpoint where in the broader causal system the moderated effect occurs and provides direction to program revisions that might be considered.

Figure 1.4 illustrates the logic and highlights the advantage of using RET designs for moderator analysis. Suppose I test if a cognitive behavior therapy (CBT) program to address chronic pain affects patient subjective pain experiences relative to a non-CBT treatment as usual protocol (TAU). Both CBT and TAU have 10 sessions. One mediator addressed by CBT but not TAU is coping flexibility. Conventional wisdom is that a single coping strategy (e.g., positive reappraisal of pain) is universally effective in all contexts. Research suggests, however, that the effectiveness of a coping strategy depends

on the context in which it is used. In CBT, patients are taught to accurately read and adapt to contextual cues across situations and deploy a coping strategy that is the "best fit" for that context. This is referred to as coping flexibility.

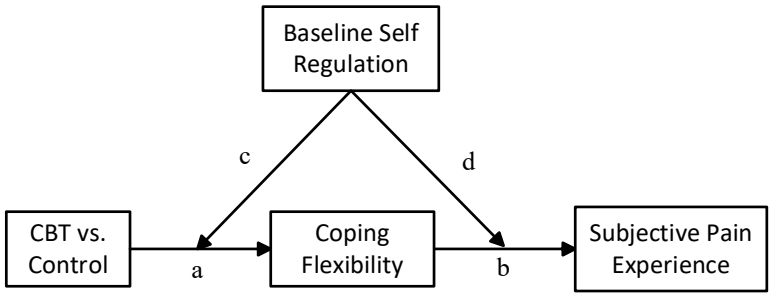


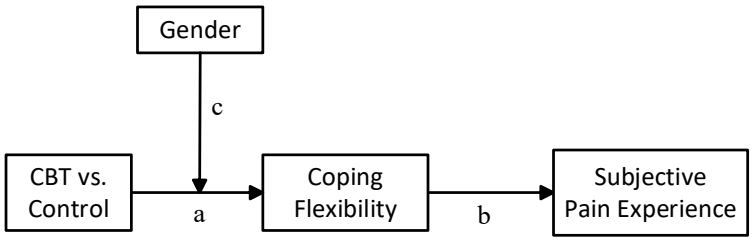
FIGURE 1.4. Moderated mediation

A potential moderator of CBT effects on subjective pain experience might be a person's ability for self-regulation as measured at baseline. As shown in Figure 1.4, there are two loci where the moderation might occur. First, low self-regulation might disrupt the effect of CBT treatment on the mediator (path *c*); individuals low in self-regulation may lack the cognitive resources to complete the tasks demanded by CBT and, as a result, the strength of path *a* is lower for low self-regulators than high self-regulators. Another possibility is that those with low self-regulation cannot translate the coping skills they learn in CBT to real life settings, hence low self-regulation disrupts path *b* via the moderation denoted by path *d*. A powerful feature of RET analysis is that we are able to identify where in the mediational chain the moderator disrupts treatment effects on the outcome, either via path *c*, path *d*, or both. As a remedy to the self-regulation effects, patients who are relatively high in self-regulation might receive traditional CBT per protocol. Patients who are low in self-regulation might still receive CBT but they might undergo an additional treatment module to address self-regulation deficits. Rather than trying to change the patient's stable self-regulation (which could be challenging), the program might use the additional module to teach patients self-management strategies that offset the detrimental effects of low self-regulation. The strategies would be tailored to the particular path in the mediational chain that low self-regulation disrupts, either *c* or *d*. This type of moderation analysis is often referred to as **moderated mediation**.

Another type of moderation analysis is called **mediated moderation.** Consider the same study but where the moderator is gender and its effects are concentrated in path *c* of Figure 1.5a, i.e., the program is more effective at changing coping flexibility for females

than it is for males. The question becomes why? Based on past research, I might speculate that females are more likely to attend all or most of the 10 therapy sessions than males, the result being reduced treatment exposure for males. In this case, CBT superiority cannot reveal itself for males because the dosage levels are in adequate. Figure 1.5b adds a mediator to the moderated relationship to reflect this dynamic.

(a)



(b)

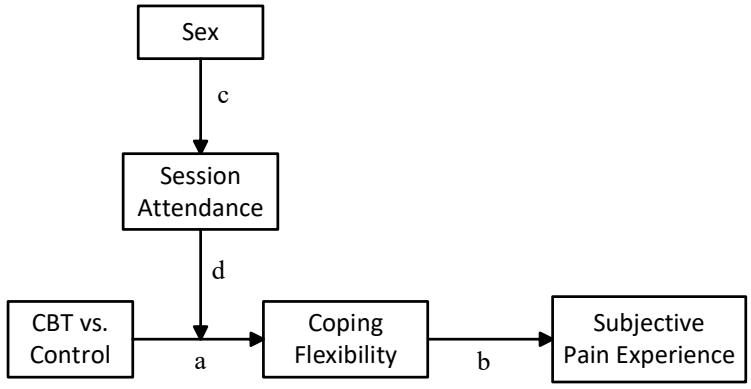


FIGURE 1.5. Mediated moderation

There is a third type of moderation effect in RETs that also are of substantive interest, known as **exposure-mediator moderation** or **treatment-mediator moderation**. The dynamic is shown in Figure 1.6. In this case, the effect of a mediator on an outcome (path *b*) is either stronger or weaker in the treatment group as compared to the control group. Specifically, the size of the M→Y coefficient varies by treatment condition. For example, a program to reduce drug use in adolescents may not increase mean levels of agreement with beliefs that using drugs leads to adverse health consequences (path *a* in Figure 1.6). However, the program may have the effect of making the adverse health

consequences more salient in working memory when adolescents encounter opportunities to use drugs, thereby strengthening the causal effect of such perceptions on drug use propensities (path *b*). Kraemer and Fairburn (2002) refer to such treatment-mediator interactions as mediation because they reflect a mechanism by which the treatment impacts the outcome. Other researchers, however, do not adopt this nomenclature and refer to the dynamic as moderation. Whatever the case, possible treatment-mediator moderation is important to consider when seeking to understand program effects.

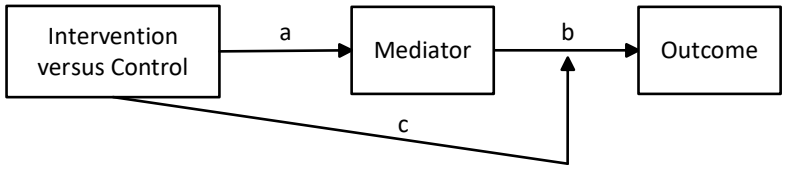


FIGURE 1.6. Treatment-mediator interaction

In sum, in addition to mediators, RETs encourage evaluators to articulate moderators within their conceptual logic models. Doing so allows one to identify the boundary conditions of program effects and the extent to which program effects generalize across subgroups and contexts. As the boundary conditions of mediational effects are identified and elaborated, corrective steps can be taken to increase the generalizability of the program. Whereas boundary conditions can be explored in traditional RCTs, doing so in an RET provides more information and insight because one can isolate where in the mediational chain and for which mediators the moderation occurs. Three types of moderation to be sensitive to are moderated mediation, mediated moderation, and treatment-mediator moderation.

Facet 3: Identifying Confounders

Another important facet of RET design is the identification of confounders that bias estimates of causal effects and that need to be controlled. The essence of a confounder is illustrated in the causal models in Figure 1.7. Figure 1.7a specifies a causal relationship between a mediator M and an outcome Y; M and Y are correlated for one and only one reason, namely, because M causes Y. In Figure 1.7b, M also causes Y, but there is an additional source contributing to the correlation between M and Y, namely the common influence of C on both M and Y. If we seek to estimate the strength of the causal effect of M on Y from the association between them, then we need to remove the inflating (or

deflating) effects of C on that association. In other words, we need to "control for" C. Potentially more disturbing is the case of Figure 1.7c, where C is the only source of the association between M and Y, i.e., there is no causal effect of M on Y. Researchers might erroneously infer a causal relationship between M and Y because M and Y are correlated, but the correlation is spurious.

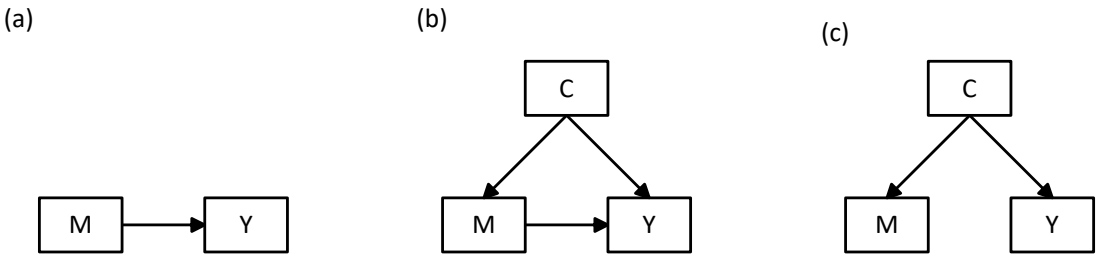


FIGURE 1.7. Different examples of confounders

Confounders are nuisance variables that are external to the focal causal system and that distort inferences about causal effects in that system. Confounders can contaminate RETs in multiple ways, some of which I illustrate in Figure 1.8. The key causal paths "in the system" are paths *a*, *b*, and *c*. Note that in this system there is a direct causal path from the treatment to the outcome independent of the mediator (path *c*). This path recognizes that the treatment can affect the outcome through mechanisms other than M, the measured mediator. In terms of confounders, consider first C1 and the mediator-outcome relationship (path *b*). C1 is a common cause of M and Y (see paths *f* and *g*) and as a result, its presence can bias estimates of the magnitude of the true causal effect between M and Y. The degree of bias is dependent on the strength of paths *f* and *g*; if these paths are weak, the degree of bias will be small. Interestingly, the presence of C1 cannot only bias path *b*, but it also can indirectly create bias of estimates of path *c*, a dynamic I discuss in future chapters.

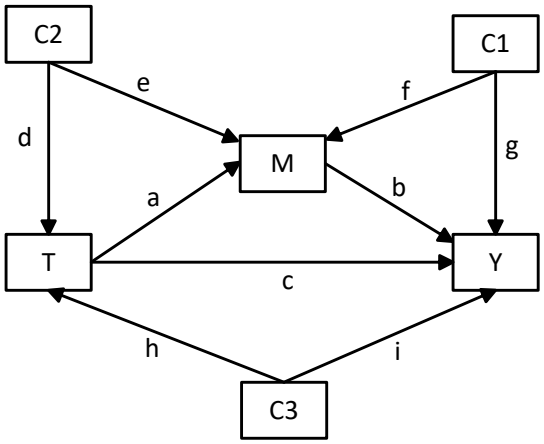


FIGURE 1.8. Three types of confounders

A second confounder in Figure 1.8 is C2. C2 is a common cause of T and M (see paths *d* and *e*). As a result, its presence can bias estimates of the causal effect of T on M, namely path *a*. The degree of bias again is dependent on the strength of paths *d* and *e;* if these two paths are weak, the degree of bias will be small. It may be surprising to some that path *d* can exist at all because variability in T is random by virtue of the use of random assignment. However, sometimes random assignment fails due to treatment dropouts, missing data (not everyone completing measures of M), and program-control contamination. If C2 impacts these sources of compromised random assignment and C2 also impacts M, then C2 will be a confounder and lead to bias in the estimate of path *a*.

A third confounder in Figure 1.8 is C3. C3 is a common cause of T and Y (see paths *h* and *i*) and as a result, its presence will bias estimates of the magnitude of the direct effect of T on Y, namely path *c*. This confounder also is tied to the successful implementation of random assignment. Indeed, if random assignment is successful in an RET, the primary threat of confounders lies in the mediator-outcome link because both mediators and outcomes are measured rather than manipulated.

Confounders can operate in ways other than those characterized in Figure 1.8 and I elaborate these ways in Chapter 2. The key point here is that when designing RETs, it is important to be cognizant of the possible presence of confounders, to identify those confounders, and then to measure them, as feasible, so that they can be controlled and dealt with statistically or, alternatively, through design considerations.

Some methodologists argue that despite one's best efforts, there almost always will be confounders that remain unmeasured and uncontrolled and that undermine causal inferences. In my opinion, the onus is on such critics to specify what these confounders

are; it is not reasonable to invoke the possibility of confounders in the abstract and then dismiss the results of an RET as being due to confounding. Critics should specify what the offending confounders are. Having said that, you, as a program evaluator, should always be nervous about the possible presence of unmeasured confounders in an RET.

The presence of unmeasured confounders will often manifest itself in correlated disturbances. This is illustrated in Figure 1.9 for the M-Y relationship. In influence diagrams, non-causal relationships (i.e., correlations) are indicated by curved double-headed arrows. Such an arrow has been drawn between d_1 and d_2 in Figure 1.9. Suppose that a variable, C, is a common cause of both the mediator and the outcome but that it is unmeasured and not specified by the researcher. The disturbance term d_1 reflects all variables other than the treatment, T, that impact the mediator. C implicitly resides in this disturbance term. The disturbance term d_2 reflects all variables other than T and M that impact the outcome. C also implicitly resides in this disturbance term. If C resides in both disturbance terms, we would expect the two terms to be correlated because, after all, they have a common element, C. This dynamic yields the correlated error shown in Figure 1.9. Recognition of this property is important because formally modeling correlated disturbances is one way that a researcher can deal with unmeasured confounders, a strategy I discuss in Chapters 6 and 11.

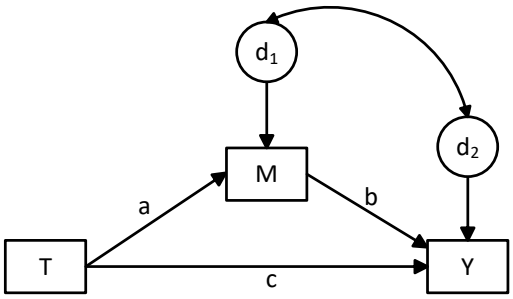


FIGURE 1.9. Model with correlated disturbances

In the mediation literature, you will often encounter what is known as the assumption of **sequential ignorability**. Definitions of it vary but it essentially refers to the assumption that net the formal control of (measured) confounders, there is no meaningful unmeasured confounding of the treatment-mediator, treatment-outcome, and mediator-outcome relationships. This assumption is typically necessary for accurate estimation of causal links, a point I return to in future chapters (see Chapter 10).

In sum, when designing an RET for purposes of program evaluation, the problem of confounders needs to be addressed. Confounders can distort causal inferences in the RET system and potentially lead program evaluators astray. In the presence of confounders, one can make faulty conclusions about the effects of the treatment on mediators and/or the effects of the mediators on outcomes. The most typical strategy for dealing with confounders is to generate a list of plausible confounders, prioritize them in terms of their importance, measure as many of the most important confounders as feasible, and then control for them, as described in later chapters.

Facet 4: Addressing Measurement Error

RETs typically rely on measurement of mediators, outcomes, and covariates. These measures may be subject to measurement error, both random and systematic. The presence of measurement error can bias parameter estimates and undermine causal inference. For example, if we want to determine the degree of association between a mediator and an outcome and if both measures have a reliability of 0.67 (meaning a third of the variability in each measure is random noise), then we likely are going to significantly underestimate the true association between the constructs. If a measure of IQ has 30% random noise in it, it is like we go to a random number table and select a fairly sizeable random number to add or subtract to a person's true score; the lower the reliability of the measure, the larger the random number we pick, to the point that, in this case, 30% of the variation in the IQ scores is just random. Needless to say, the more dominated an IQ measure is by such random error, the worse it is going to be at, say, forecasting GPA. When we conduct a statistical analysis that seeks to control a confounder by including a measure of that confounder as a covariate in a regression analysis, if the measure is contaminated by random error, then we may not adequately control for the confound because our measure is contaminated by random noise. For example, to the extent measures of socioeconomic status are subject to random error, then analyses that seek to control SES statistically may not adequately do so.

Systematic measurement error is not random and can bias parameter estimates in causal modeling. For example, socially desirable response tendencies is an individual difference variable that reflects a propensity to want to create good impressions on others. Individuals who are high on this trait are more likely to systematically underreport such things as drug use, unprotected sex, and depressive symptoms and to over report income, life satisfaction, and accomplishments. If social desirability response tendency is a common cause of, say, self-reports of both depression (a mediator) and alcohol use (an outcome), it can inflate the true association between the two variables.

When designing an RET, we want to use measures that are relatively free of both

random and systematic measurement error. In later chapters, I discuss strategies for adjusting for measurement error in the context of causal modeling.

Facet 5: Addressing Temporal Dynamics

Causal theory generally assumes that a cause temporally precedes an effect. The time it takes for a cause to translate into the effect can vary. Sometimes change is virtually instantaneous (milliseconds) while other times it is lengthy. Suppose a treatment to reduce child depression targets the parents of the depressed child and teaches parents more effective parenting strategies to reduce child depression. The effect of the newly acquired parenting skills on child depression will not be instantaneous. It will take time for the parents to apply them, for the child to notice a difference, and for the relationship between the parent and child to change to a positive enough state that the child starts to become less depressed. Suppose it takes a minimum of 2 months for the intervention to have its effect on the outcome. If an RET evaluates the effects of the intervention one month after treatment, the effect of the treatment will be underestimated; if the researcher had waited one more month to assess the outcome, a different conclusion of program effectiveness would have resulted. The same might be true for the time interval between assessments of the mediator and the outcome; it takes time for changes in the mediator to translate into changes in the outcome.

On the other hand, sometimes change in a mediator or an outcome can be nearly instantaneous. I might develop an intervention to change a person's intention to support policies favorable to the environment and, as part of the intervention, present convincing arguments for doing so. Immediately after hearing the arguments, people in the program revise their intentions to support "green" policies. If instead of an immediate assessment of intentions I instead wait 2 months to measure it, per the above research with parenting and depression, I am essentially studying not just the initial impact of the program but rather its immediate effect plus decay in those effects over a two-month period. However, perhaps I am interested in both immediate and decay dynamics. RET designers need to make educated judgments about such temporal dynamics (Cole & Maxwell, 2003; Maxwell & Cole, 2007; Maxwell, Cole & Mitchell, 2011) and incorporate those into their design decisions for RETs.

Facet 6: Addressing Reciprocal Causality

Most RETs are longitudinal in structure, assessing constructs at a baseline and then post-treatment. Some designs also obtain assessments mid-treatment and still others obtain measures at follow-up, say 6 months, or 12 months after the intervention has ended. In some cases, mediators and outcomes are measured at the same point in time. In such

cases, it is possible that the association between the mediator and the outcome reflects not just the effect of M on Y, but, as well, the effect of Y on M. If a program targets parent anxiety (M) with the idea that doing so will reduce child anxiety (Y), one must recognize that child anxiety also can impact parent anxiety. Adjustments need to be made to accommodate this reciprocal causality dynamic if one wants to obtain an accurate estimate of the effect of parent anxiety on child anxiety. This also is true when estimating causal relationships among mediators that might have a reciprocal causal dynamic.

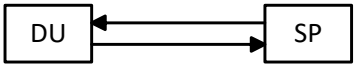
Strictly speaking, there can never be simultaneous reciprocal causation because there always must be a time interval, no matter how infinitesimally small, between the cause and the effect. If we map the true causal dynamics within a time frame for a reciprocal causal relationship between M and Y, the dynamic might appear as follows:

$$M_{t1} \rightarrow Y_{t2} \rightarrow M_{t3} \rightarrow Y_{t4}$$

where M_{t1} is the mediator at time 1, Y_{t2} is the outcome at time 2, M_{t3} is the mediator at time 3, and Y_{t4} is the outcome at time 4. As an example, consider adolescent performance in school as measured by grade-point average (the outcome) and adolescent drug use (the mediator). It is likely that school performance is adversely affected by drug use by interfering with students' completion of their homework and affecting their ability to concentrate on tests. At the same time, performing poorly in school likely puts adolescents at risk for drug use, as their interests drift away from doing well in school and as they spend more time with deviant peers rather than studying. A causal chain that describes this dynamic is

$$DU_{t1} \rightarrow SP_{t2} \rightarrow DU_{t3} \rightarrow SP_{t4}$$

where DU represents drug use at time *t*, SP represents school performance at time *t*, and the numerical subscript attached to *t* represents later time points as the numbers increase in value. If one is unable to assess these processes at this fine-grained level, and if these processes have already played themselves out when the assessments of drug use and school performance are made cross-sectionally, the resulting causal representation that captures what has transpired is this:



This diagram essentially reflects a summary of the sequential dynamics. The cross-sectional association we observe between drug use and school performance reflects both

the prior causal impact of drug use on school performance and the prior causal impact of school performance on drug use. If we were to only model a causal influence for the impact of drug use (the mediator) on school performance (the outcome), we would erroneously "give credit" for the association between the two variables as being entirely due to the effect of drug use on school performance when, technically, it deserves only some of the credit. Our estimate of the causal coefficient will be biased.

When designing an RET and if assessments of mediators and outcomes are obtained at the same time points, we need to consider the possibility of reciprocal causality between the mediator and the outcome. In later chapters, I discuss strategies for addressing this matter.

Facet 7: Making Sample Size Decisions

Yet another consideration when designing an RET is the determination of the sample size to use. Most researchers think about sample size in terms of maximizing statistical power, but RETs require a broader perspective given their multivariate character. In addition to statistical power, we must take into account covariance matrix stability, asymptotic theory, margins of error, and statistical estimation strategies. I consider sample size decisions for RETs in Chapter 28.

THE BIG PICTURE

Randomized explanatory trials generally are superior to traditional randomized controlled trials for evaluating behavioral interventions. RETs go beyond simply documenting whether a program impacts one or more outcomes. RETs provide feedback on what program components are effective and, if a program is not having sufficient impact, RETs provide perspectives why this is the case. A well-crafted RET will help us pinpoint which mediators/mechanisms are important to address in a program as well as how effective the program is in addressing those mediators/mechanisms. An RET also provides insights into the boundary conditions of program effectiveness and the groups for whom or circumstances under which the program works best. Program designers and administrators can then revise their programs based on RET feedback to try to make them more generalizable and widely applicable.

Designing a high quality RET has unique theoretical, methodological, and statistical challenges. I have previewed some of these challenges in this chapter. The remainder of this book flushes out RET design and analysis in more depth. The key tasks for designing an effective RET outlined in this chapter can be summarized as follows:

Mediator mapping: The first task when designing an RET is to conduct a careful

mapping of target mediators onto program components and activities. This mapping takes the form of a conceptual logic model that often can be summarized by an influence diagram. The identified mediators should be relatively specific and reasonably tied to program components so that concrete prescriptions for program improvement can be made after the RET is completed. Abstract mediators often are less informative, but not completely so. When mapping mediators onto the program components and outcomes, you should consider possible causal relationships among the mediators. Ultimately, the RET will be able to tell us (a) which mechanisms/mediators are affected by the program and which are not (thereby flagging program components in need of revision), and (b) which mechanisms/mediators are most important in influencing the outcome and which ones are not, perhaps leading us to streamline the program by eliminating components that address mechanisms thought to be relevant but which are only minimally so.

Moderator mapping: Once the mediators are specified, you should expand the conceptual logic model by considering plausible moderators of the various mediational effects. These moderators specify potential boundary conditions of the program and allow one to evaluate generalizability of program effects across subgroups and/or contexts. Such feedback can be useful for program designers. When thinking about moderators, also consider the possibility of moderator effects of one mediator on another mediator in its effect on outcomes as well as moderated mediation, mediated moderation and treatment-mediator moderation.

Identification of confounders: Confounders are nuisance variables that cause us to make erroneous causal inferences. Confounders can operate for any causal effect in the causal system of interest. Make a list of all plausible confounders, prioritize them in terms of importance to control, and then measure them in the RET, as feasible, so they can be statistically controlled or, alternatively, control them via design considerations.

Addressing measurement error: Do everything you can to use measures that are valid and free of random and systematic measurement error. Adopt analytic strategies that can accommodate fallible measures without biasing parameter estimates, as feasible.

Integrating temporal dynamics into one's conceptual logic model: As you elaborate your target causal system, for each causal link in the system, think about the time interval it takes for the cause to translate into an effect. Specification of these intervals impact the design of your RET in terms of when to make assessments.

Integrating reciprocal causality into your conceptual logic model, as appropriate: Consider the possibility of bidirectional causal effects for every causal relationship in the target causal system. Integrate plausible ones into your conceptual logic model, as appropriate. This task is important because it ultimately will affect how you design your RET, as I discuss in future chapters.

Determining an appropriate sample size: Based on considerations of statistical power, covariance matrix stability, asymptotic theory, margins of error, and likely statistical estimation strategies, determine the sample size needed for your RET.

EXPERIMENTAL THERAPEUTICS

Experimental therapeutics is an evidence-based paradigm for evaluation that has gained traction at the National Institute of Health (NIH). Experimental therapeutics are grounded in clinical trials that focus on the "targets" (mechanisms of action) of an outcome rather than just the outcome itself. The goal of the trial is to "identify or verify" a target (i.e., a mediator) and to determine the ability of the intervention to affect that target and, in turn, the outcome. RETs are at the heart of experimental therapeutics. Consider these quotes from two different directors of the National Institute of Mental Health:

> *Discoveries in neuroscience and behavioral science can suggest malleable targets (potential mediators) for novel intervention strategies. Evaluating the relationship between changes in these targets or mediators and changes in symptoms allows us to fine-tune our understanding of mental illness, and helps us prioritize the most promising interventions for further investment. Consideration of these factors enables research aimed at refining therapies to increase potency and efficiency.* (Gordon, 2017)

> *"Target engagement" refers to verification that the intervention has had the predicted effect on the target. Targets may be molecular, cellular, circuit, behavioral, or interpersonal, commensurate with the intervention. Once target engagement is demonstrated, measures of target engagement are then related to clinical outcomes to test the hypothesis that the target is relevant to the clinical problem under study…. Even negative results can be informative because if a proposed target is engaged, but there is no effect on a relevant endpoint, we can rule out that target. For example, antidepressant effects have variously been proposed to involve changes in serotonin neurotransmission, hippocampal cell birth, and changes in stress hormones, among many other effects. By ruling out some targets and focusing on those involved in the biology of the disorder, we can direct treatment development much more efficiently.* (Insel, 2012, 2013)

Much of the enthusiasm for experimental therapeutics at NIMH has focused on biological processes and neural mechanisms as mediators of the effects of drug therapies and/or behavioral therapies on clinical disorders. In this sense, one can think of mediators and moderators in an RET as occurring at different levels of analysis, ranging from the macro-level of contexts to a more micro-level of mental processes to an even more micro-level of biological and neural mechanisms. Some RETs focus their analysis on a single level while others address multiple levels of analysis.

As an example of the latter, I recently consulted on the evaluation of a cognitive-behavioral-therapy (CBT) program to reduce hoarding behavior in people with debilitating hoarding disorder. The mediators the research team were interested in focused on brain activity measured using fMRI as participants performed hoarding related tasks on an iPad. The essence of the RET design is shown in Figure 1.10. The treatment condition was participation in the CBT program versus participation in a wait-list control. The research team was interested in (a) identifying the effects of CBT on different brain regions, and (b) mapping the effects of those brain regions on the outcome of hoarding symptomology. The RET design provided the research team with feedback as to which brain region mediators were related to hoarding symptoms and which brain region "mediators" were irrelevant (paths *d* to *f* in Figure 10). It also provided the research team with feedback on which brain region mediators CBT was affecting and which mediators it was not affecting (paths *a* to *c* in Figure 10). As I explained colloquially to the team when presenting the results of the study, the RET tells us which brain region "buttons" the CBT treatment is pushing and whether those "buttons" are the right or the wrong buttons to be pushing. Based on the results, I challenged them to consider how to alter their CBT protocol so that it would better affect the brain regions that seemed to matter but that currently were not be affected by CBT. A challenge in this particular project was that the researchers identified some 200 potentially relevant brain regions (mediators); I had to use specialized data reduction methods to reduce the number of plausible mediators to the 5 or so most critical brain regions that impacted hoarding symptoms. I discuss such reduction approaches in Chapter 17.

MIXED METHODS RETs

My discussion to this point has focused on classic quantitative-based RETs, but it is possible to supplement many RETs with qualitative methods so as to approach them from a mixed-methods framework. Plano-Clark (2010) analyzed mixed-methods research funded by 25 different agencies in the National Institute of Health and found that the most common type was research that included both a randomized trial component and a qualitative component tied to the trial. The qualitative portion of such trials studied

individuals who were randomized to one of two (or more) conditions, usually a treatment versus a control group. Individuals in each group were interviewed about their experiences and lives before the trial began, during the trial proper, and after the trial was completed. In some studies, direct observations of the study participants were undertaken in participants' homes or some other setting.

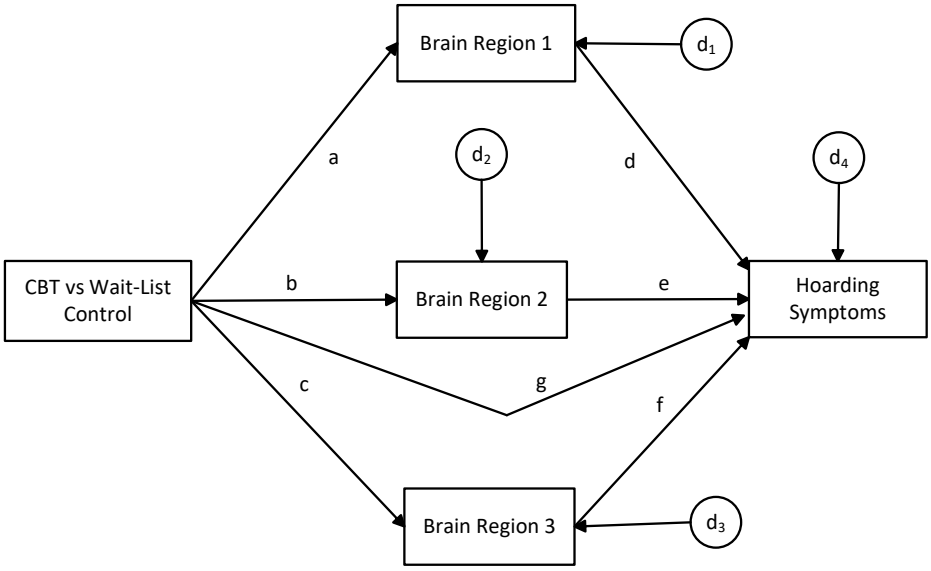


FIGURE 1.10. RET with brain activation mediators

Creswell and Plano-Clark (2018) describe three types of designs that can be used when mixing qualitative and quantitative research, (1) an **explanatory sequential design** in which the quantitative study is conducted first, then a follow-up qualitative study to supplement, enrich, and qualify the results from the quantitative study; (2) an **exploratory sequential design** in which the qualitative research is conducted first, which then informs the quantitative study; and (3) a **convergent parallel design** in which the qualitative and quantitative data are collected concurrently to compare results. You might consider using a mixed methods approach for your RET to enrichen program evaluation.

RETs AS THOUGHT EXPERIMENTS

I have consulted for numerous agencies to help them evaluate their programs. When I raise the prospects of conducting an RCT or an RET, agency administrators and staff sometimes object due to cost considerations and a reluctance to deny or delay treatment

to control individuals. When faced with scenarios where an RET is not feasible because of the above or other considerations, I still conduct an RET but it instead might take the form of a "thought experiment" conducted with different agency constituencies. The approach sacrifices scientific rigor, but it is cost effective and manageable for many agencies and invariably can yield useful information.

I begin by working with program management and staff to evolve a conceptual logic model organized by RET concepts. This requires mediator mapping for the current intervention/program, moderator mapping, consideration of confounders, talking through temporal dynamics, relationship functional forms, and elaboration of plausible reciprocal causal relationships. The logic model is evolved using focus groups and in-depth interviews. Once a tentative conceptual logic model is created, I repeat the process but now using focus groups and in-depth interviews with different constituencies, namely (a) providers who implement the program/intervention, and (b) clients who have participated in the program. Based on this input, I finalize an integrated conceptual logic model that details a thorough mapping of mediators and moderators as detailed by different constituencies and that represents a good theoretical accounting of the theoretical bases of the target program.

Next, I conduct a new round of focus groups and interviews with each constituency in which I present to them a "thought experiment" that essentially describes an RET. I ask each constituency to discuss what they think would be the likely results in such an experiment for each link in the model based on their personal experiences with the program. We discuss whether the program will likely affect each proposed mediator, discuss why this is the case, and share ideas about ways to improve the program so that its impact on a given mediator can be strengthened. We also review the relevance of each mediator to program outcomes and discuss whether each mediator/mechanism is, in fact, central to program success. We rank order the likely importance of the mediators in determining outcomes. As we discuss these matters, I raise issues of confounders, temporal dynamics, functional forms, and reciprocal causality. We then discuss the moderators in the model and discuss issues of subgroup and context generalizability. In essence, we conduct an RET as a collective but do so in the form of a thought experiment with independent input about likely RET results from key constituencies. This is a "poor person's" variant of an RET, but I have found it yields a wealth of useful information.

FACTORIAL RETs AND DISMANTLING DESIGNS

Some researchers evaluate multi-component interventions using factorial designs to gain perspectives on the relative contributions of intervention components. The recent

multiphase optimization strategy (MOST) by Collins (2018) advocates for such approaches. For an intervention that consists of three components, C1, C2 and C3, each targeting a different mediator, one might construct a factorial RET consisting of 8 groups that manipulate all combinations of the presence or absence of C1, C2, and C3, respectively, per the factorial structure in Table 1.1. Using this design, one can then evaluate main effect influences of each component on outcomes as well as interaction effects of component combinations on outcomes.

Table 1. 1: Factorial RET

	C1 Present		C1 Absent	
	C2 Present	C2 Absent	C2 Present	C2 Absent
C3 Present	Group 1	Group 3	Group 5	Group 7
C3 Absent	Group 2	Group 4	Group 6	Group 8

I view such approaches as a specialized form of **multi-treatment RETs**. In a multi-treatment RET, rather than just comparing a treatment and a control group, the researcher compares multiple treatments to each other and a control group, such as cognitive behavioral therapy for attention deficit disorder versus a medication for attention deficit disorder versus a control condition. In a factorial RET, there also are multiple treatment groups (per Table 1.1) but they are strategically defined in a way that allow you to tease out cross-component influences of multi-component interventions and possible interaction effects between those components. For the example in Table 1.1, suppose that component C1 seeks to change one mediator, that component C2 seeks to change a second mediator, and that component 3 seeks to change a third mediator. Figure 1.11 presents the relevant RET influence diagram for the factorial RET, which is no different in structure from Figure 1.1 where I first introduced RETs but had just two program conditions, a treatment group and a control group. The difference is that the program condition in Figure 1.11 consists of 8 groups instead of two and would be represented by seven dummy variables rather than a single dummy variable during data analysis. The seven dummy variables would be defined to isolate different contrasts among the various program types so as to conform to questions answerable in factorial designs. I discuss how to conduct these analyses in Chapter 28.

Despite the factorial structuring of treatments, the fundamental nature of a factorial RET remains the same as a standard RET; you measure the mediators of each component and you carefully analyze the individual links in each mediational chain. If a program

component fails to impact the outcome, the question becomes why? Was it because the component failed to change the mediator it was supposed to change? Was it because the mediator was not causally related to the outcome, contrary to expectations? Or was it because of both dynamics? RETs integrated with MOST designs provide perspectives on such matters. Outcome only RCTS do not.

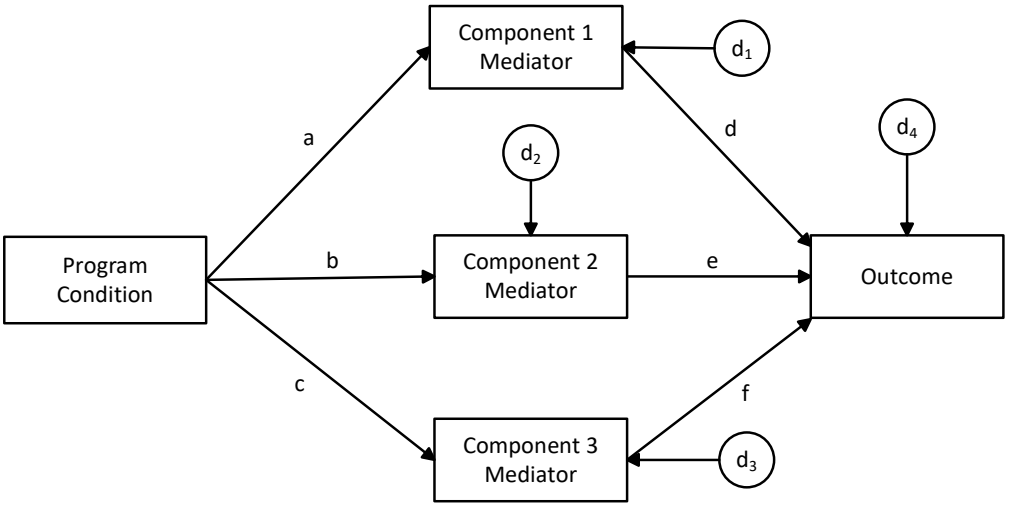


FIGURE 1.11. Multi-treatment RETs

Related to factorial RETs are **dismantling designs** (Papa & Follette, 2015). Dismantling designs seek to identify if all components of a multicomponent intervention are necessary. Researchers typically compare a group of participants exposed to an intervention consisting of only a subset of the intervention components with a group of participants who are exposed to the complete intervention to determine if dropping components affects efficacy. If the two groups yield comparable outcomes, then the components deleted to create the streamlined intervention are deemed as unimportant. The design and analysis of dismantling studies can be complex and I also address them in Chapter 28. Unfortunately, dismantling studies often adopt an outcome-only orientation, that is, they do not address both mediators and outcomes. Rather than focus just on outcomes, I argue that dismantling studies can benefit from the inclusion of mediators and moderators in their design so as to provide richer insights into the mechanisms through which each program component impacts (or fails to impact) outcomes (see Silverman et al., 2021, for an example). Dismantling designs should

embrace RET rather than RCT perspectives. For example, does a given component "fail" to affect an outcome because (a) the component failed to change the mediator it was intended to affect, and/or (b) because the mediator it targeted turned out to be only weakly related to the outcome, contrary to what was assumed? RETs provide perspectives on this matter. RCTs without mediation analysis do not.

RETs AND OTHER FACETS OF PROGRAM DESIGN/EVALUATION

An RET, of course, addresses only one facet of program evaluation, namely component efficacy/effectiveness. Program designers usually have many goals in the course of program evaluation (Chen, 2014; Royse, Thyer & Padgett, 2015). For example, one agency goal might be to make a program cost-effective. Another goal might be to make the program achieve a specified minimum level of effectiveness. Yet another goal might be to make a program briefer. RETs are only one facet of program evaluations.

Also important to keep in mind is that the foci of "program components" can be diverse. For example, components can focus on naturally occurring determinants of the outcome, they can address treatment adherence on the part of program participants, or they can address provider or staff fidelity of program implementation. Our concern in this book is primarily with naturally occurring determinants of outcomes, but any of the aforementioned targets can be framed and addressed in an RET. I address this in more detail in chapters 3 and 13.

CONCLUDING COMMENTS

The term randomized explanatory trial was first coined by Schwartz and Lellouch (1967) and juxtaposed against what they called pragmatic trials (PTs). Schwartz and Lellouch characterized an RET as a randomized controlled trial designed to shed insights on the causal impact of a treatment component on an outcome. By contrast, they argued that pragmatic trials are designed to compare the relative effectiveness of two or more treatments in practical contexts. Since their seminal paper, the terms explanatory and pragmatic trials have been used in diverse ways in the scientific literature, but the essence of the distinction has focused on a concern for understanding the causal mechanisms underlying the effect of a treatment on an outcome (RET) under idealized experimental conditions on the one hand versus the comparative effects of treatments in practical settings (PT) on the other hand. Flash forward some fifty years later and it is clear that the concept of an RET has evolved considerably since its introduction in 1967. To be sure, the essence of a randomized explanatory trial remains that of understanding the causal mechanisms that account for the effects of treatments on outcomes but RETs have

evolved to include the core concepts of mediation and moderation in conjunction with far more advanced conceptual, psychometric, and methodological tools than when Schwartz and Lellouch first coined the term. This book, in part, serves as an updating of the RET concept. More specifically, I contend that the dichotomization of trials into RETs versus PTs is counterproductive. To be sure, I fully recognize the unique influences on outcomes that are relevant in practical, real-life contexts. However, PTs morph or blend into RETs when we seek to understand the nature of those unique influences and to formally address them in our interventions and their roll-outs in applied contexts. It no longer is enough to document the effectiveness rates of a treatment in an applied context or to demonstrate that one treatment works better than another treatment in an applied context. This is too narrow a perspective. Instead, we need to know *why* one treatment works better than another treatment in applied contexts, *why* a treatment fails or succeeds in an applied context, and *for whom* this is and is not the case. In other words, PTs need to incorporate RET perspectives that elucidate mediators and moderators and that make use of modern scientific tools and methods to help us understand and improve upon their results. The focus should not only be on testing whether interventions work, but also on understanding why they work (or do not work) and whether intervention effects operate through the presumed mechanisms in applied contexts, whether those mechanisms are basic or unique to a given type of applied context.

Randomized explanatory trials have much to offer relative to traditional randomized controlled trials. However, implementing RETs raises theoretical, methodological, and analytic challenges, some of which have been introduced in this chapter. In the remainder of this book, I elaborate these challenges and lay the foundation for the conduct of informative RETs, emphasizing data analytic issues. The remainder of the first section of the book considers conceptual (Chapter 2), psychometric (Chapter 3), methodological (Chapter 4), and statistical (Chapters 5 through 8) foundations for RETs. This material sets the stage for Part 2 of the book, which constitutes the heart of RET analytic approaches. Chapters 9 through 17 address issues relevant to mediation analysis in RETs and Chapters 18 through 24 address moderation analysis in RETs. Chapter 25 discusses clustered RET designs and Chapter 26 addresses issues of treatment dropouts and missing data. Chapter 27 considers sample size issues and Chapter 28 addresses multi-treatment designs including factorial RETs.