17

Mediator Relative Importance and Exploratory Mediator Analysis

Reality leaves a lot to the imagination

- JOHN LENNON

INTRODUCTION

In this chapter, I address two topics. First, given a set of mediators, we sometimes want to order them in terms of their relative importance as determinants of the outcome. The idea is that if limited resources or model complexity demand we focus on a smaller subset of mediators, then knowing the relative importance of mediators in shaping an outcome can help set program and analytic priorities. Second, I consider cases where one has measured so many plausible mediators that one must apply data reduction strategies to reduce them to a workable set for RET modeling. I do not consider judgments of relative importance for moderation, a topic I cover in Chapter XX.

Webster's dictionary defines importance as "a quality or aspect having great worth or significance; importance implies a value judgment of the superior worth or influence of something or someone." Importance judgments evaluate worth relative to a set of criteria, criteria that for many types of judgments vary by context, by constituency, and that are subject to disagreement. In Chapter 10, I discussed approaches to determining if an effect in an RET is meaningful. In addition to using focus groups with relevant constituencies to set standards, I described three strategies for exploring effect meaningfulness, (1) expert-based methods, (2) distribution-based methods, and (3) anchor-based methods. These strategies can, in principle, be extended to evaluate and order the degree of meaningfulness of each mediator for purposes of RET analysis. The present chapter, however, considers statistical criteria that researchers or program evaluators might use to help them make or lead discussions of relative importance.

RELATIVE IMPORTANCE OF OMNIBUS MEDIATION EFFECTS

It is not uncommon in the scientific literature for studies to order mediators in terms of their relative importance based on omnibus indirect effect sizes for each mediator. In traditional full information SEM (FISEM), mediation analysis encompasses multiple path coefficients that combine multiplicatively to reflect a mediator's implications for program effects on an outcome, namely the product of the coefficient from the treatment condition to the mediator ($p_{T\rightarrow M}$) multiplied by the coefficient from the mediator to the outcome ($p_{M\rightarrow Y}$). This multiplicative function is important. Consider two mediators and an outcome that all are measured on the same standardized metric (mean of 0 and a SD = 1). The path values and omnibus indirect (mediational) effect for the two mediators might be as follows:

	$p_{T\rightarrow M}$	$p_{M\rightarrow Y}$	Indirect effect
Mediator 1	0.00	0.50	0.00
Mediator 2	0.50	0.00	0.00

Each mediational chain suggests no mediation for the program effect on the outcome through the designated mediators because a link in the mediational chain is "broken"; a value of zero in one link negates the other link no matter what value it has. Although both of the above indirect effects suggest the mediators in question are unimportant, the zero indirect effect is qualitatively different in the two cases. For Mediator 1, the zero indirect effect is due to the program failing to create change in a relatively influential mediator; for Mediator 2, the zero indirect effect is due to the program targeting a mediator that has no causal influence on the outcome. How you as a program evaluator address the problem of a zero or weak indirect effect would differ in the two cases, as I have discussed in previous chapters. For Mediator 1, the problem lies not in the choice of the mediator per se but rather in the program's inability to change it. We need to revisit the program and figure out how to alter the program to bring about change in the mediator. For Mediator 2, the problem is not that the program fails to change the mediator but rather in the choice of the target mediator per se. The program assumed the mediator is relevant to the outcome but it is not. We need to "move on" from Mediator 2 and perhaps not focus program efforts on changing it. Note also that once I learn the values of the path coefficients for the two individual links for a given mediator, I have a reasonable sense of the strength of the omnibus indirect effect for a mediator. However, knowing only the result of the omnibus effect is non-diagnostic of what is happening at the per link level. This is why I like to work at a per link level; I get a fuller sense and understanding of the operative dynamics.

Andrew Hayes (2022), author of the PROCESS framework discussed in Chapter 9, argues against the analysis of individual links in mediational chains (other than noting the signs of the coefficients) and is critical of the joint significance test for mediation because of its focus on individual links. He also argues against evaluating link-specific effect sizes. He states that the analysis of omnibus product coefficient statistics as described in the previous section coupled with bootstrapping for purposes of statistical inference is the "correct" and "modern" way to pursue mediation analysis. I obviously disagree with him. How one approaches the analysis of models with mediational chains depends on the questions one seeks to answer. For purposes of program evaluation, a careful analysis of the individual links in mediational chains is informative. If one is confident that each link in a mediational chain is non-zero and meaningful, then, contrary to Hayes, it is not "incorrect" to conclude that mediation is present. If one of the links is clearly "broken," it also is not incorrect to conclude mediation is not operating. Hayes is correct that there are scenarios where bootstrapping outperforms the joint significance test in null hypothesis testing contexts, but there also are just as many if not more cases where the reverse is true (see Chapter 9). Further, one can use an array of statistical indices to gain perspectives on effect size for individual links in a mediational chain which are meaningful (see Chapter 10). Hayes criticizes but mischaracterizes the important work of Yzerbyt et al. (2018) on individual link analysis in mediation, asserting their approach is "outdated" and representative of "old habits dying hard." This is ironic because many methodologists would apply these terms to his PROCESS framework. I encourage readers to consult Yzerbyt et al. (2018) and to consider the material in the current and prior chapters to make their own conclusions. I believe that both omnibus mediation analysis and individual link analysis have their place, but one needs to tailor their use to the questions one seeks to answer.

Although I personally prefer individual link analyses in mediational chains to make decisions about mediator priorities, I review in this next section approaches that have been used in the literature to order mediators in terms of their relative importance using omnibus mediation effects, i.e., based on the full T→M→Y mediational chain where the various links within the chains are considered simultaneously. I then turn to the analysis of relative importance using the separate links, which is my preferred method for purposes of program evaluation.

Omnibus Indirect Effects Indexed by Percents

In the applied literature, a common strategy for assessing the omnibus relative importance of mediators is to obtain a numerical index of "percent explained variance" that is attributable to each mediator and then to order mediators as a function of this index. For example, Kelly et al. (2011) identified possible mediators of the effects of Alcohol Anonymous (AA) programs on drinking abstinence and the number of drinks per drinking day after completion of the program. They found the mediators as a whole accounted for about 50% of the variation in these outcomes. Based on a relative importance analysis of indirect effects for the individual mediators, Kelly et al. concluded that the most consistent pathways through which AA confers its recovery benefits is through mediators focused on the social networks of clients (e.g., the number of pro-drinking network members) and via abstinence self-efficacy. The index used by Kelly et al. divided the omnibus indirect effect for a mediator by the total indirect effect of all mediators. This equals the proportion of the total indirect effect accounted for by each mediator. Other researchers use a similar index, but the divisor is sum of the total effect of the mediators, not the total indirect effect of the mediators.

Let me illustrate the two indices. Consider a program with five presumed mediators designed to increase monthly retirement savings with each mediator reflecting a different program target (e.g., teaching budgeting strategies, teaching the importance of saving). The outcome is the number of dollars saved per month. Suppose the total effect of the program when comparing intervention and control individuals was to increase monthly savings on average by $95. The calculations for the two different proportion/percent indices are in Table 17.1.

Table 17.1. Relative Importance Using Omnibus Mediation Indices

	Mediator-Specific Indirect Effect	Percent of the Total Indirect Effect	Percent of the Total Program Effect
Mediator 1	30.00	(30/85)*100 = 35.29%	(30/95)*100 = 31.58%
Mediator 2	30.00	(30/85)*100 = 35.29%	(30/95)*100 = 31.58%
Mediator 3	0.00	(0/85)*100 = 0%	(0/95)*100 = 0%
Mediator 4	15.00	(15/85)*100 = 17.65%	(15/95)*100 = 15.79%
Mediator 5	10.00	(10/85)*100 = 11.76%	(10/95)*100 = 11.76%
	Σ = 85.00		

The indirect effect for a given mediator is shown in Column 1. It is the mean difference between the treatment and control groups in saved monthly dollars *as traced through that particular mediator*. It is the product of the unstandardized path coefficients $p_{T \rightarrow M}$ and $p_{M \rightarrow Y}$ (not shown in the table). Mediator 1 had an indirect effect of $30; it accounted for an increase of $30 savings per month in the treatment group relative to the control group. Mediator 4 had an indirect effect of $15; it accounted for an increase of $15 savings per month in the treatment group relative to the control group. The *total indirect effect* is the sum of the indirect effects for each mediator. It was $85. This means that the direct effect of the program on the outcome independent of the mediators was $95 - $85 = $10. Taken together, the total indirect effects plus the direct effect of the program sum to the total program effect of $95.

The middle column of Table 17.1 converts each indirect effect in the first column to the percent of the *total indirect effect* that it accounts for, i.e. the mediator specific indirect effect divided by $85. It is the index that Kelly et al. used. The last column uses a different divisor, namely the percent of the *total program effect* that it accounts for, or $95. One or the other of the resulting values in the last two columns are used to order the relative importance of the mediators. I refer to these indices as being "omnibus-based" because the entries in the last two columns are driven by the values in the first column in Table 17.1, which represent the product of $p_{T \rightarrow M}$ and $p_{M \rightarrow Y}$; hence they are omnibus effects. In this example, Mediators 1 and 2 are the most important in accounting for program effects, Mediator 3 is least important, and Mediators 4 and 5 are intermediate.

My own preference when working with omnibus indirect effects is to use the raw metrics in Column 1 rather than the percent-based indices in Columns 2 and 3. One problem with the percent-based indices is that if the total indirect effect or total program effects are small, then the calculated percents for a mediator can misleadingly create the impression the mediator indirect effect is more substantial than it is. For example, a mediator that accounts for 50% of a total program effect that increases retirement savings by only $5 per month is not meaningful despite the 50% figure associated with that mediator. Another problem is that if a mediator boomerangs to contribute adversely to the total effect for a positive outcome, its percent will be negative, which is nonsensical; and the negative percent will play havoc with the overall indices because they will no longer sum to 100%. Indeed, Kelly et al. found this to be the case in their study.

MacKinnon et al. (1995) found percent based omnibus indices tend to be unstable, showing considerable sample-to-sample variability across random samples from the same population. In their review of omnibus indirect effect size indices, Preacher and Kelly (2011) recommend against the use of the percent based indices. For these reasons as well as others provided by Preacher and Kelly (2011), my focus when characterizing relative importance using omnibus indices here will be on raw metric indirect effects as reflected by the first column of Table 17.1. Again, I do not recommend program evaluators use such omnibus indirect effects when evaluating mediator relative importance for purposes of program evaluation because they lack specificity about where mediational chains break down nor where they can be strengthened. However, some researchers may disagree with me and I recognize there may be contexts where such analyses can be informative. As such, I describe how to perform omnibus-based mediator analyses using the Mplus software.

A Numerical Example

In the remainder of this chapter I illustrate concepts using an example of a program to increase adherence to PrEP medication protocols. PrEP is a medication taken by high risk individuals for HIV infection that reduces their risk of contracting HIV. Adherence was measured during a six month period after program completion, with scores at the posttest ranging from 0 to 100 to reflect the percent of person adherence to the protocol. Perfect adherence across the sixth month period resulted in a score for the individual of 100. Ninety percent adherence across the six month period resulted in a score for the individual of 90. And so on. The program addressed six mediators, measures of which were obtained at baseline and just after program completion. The mediators were (1) social support for protocol adherence, (2) coping skills for dealing with PrEP side effects, (3) the perceived advantages of using PrEP, (4) perceptions of the risks of contracting HIV, (5) perceptions of the severity of HIV should it be contracted, and (6) depression. The first five mediators were assessed using multi-item scales with each item responded to on a -3 (strongly disagree) to +3 (strongly agree) metric. Each scale had a different number of items. Responses to the items were averaged to yield a total score. Even though the total score for each mediator ranged from -3 to +3, the measures probably should be considered to be measured on different metrics given they vary in the number of items feeding into the total score. Indeed, had I summed the item scores rather than averaged them, they unambiguously would be viewed as having different metrics. The depression measure ranged from 0 to 10. All mediators were positively associated with adherence except depression, which was negatively associated with adherence.

The RET causal model is shown in Figure 17.1, absent the baseline covariates to avoid clutter. I included correlated disturbances between social support and depression, the logic for which I consider later. All other mediators are assumed to be correlated by virtue of (a) the common effect of the treatment on them, and (b) the correlations between the baseline covariates that impact the mediators.

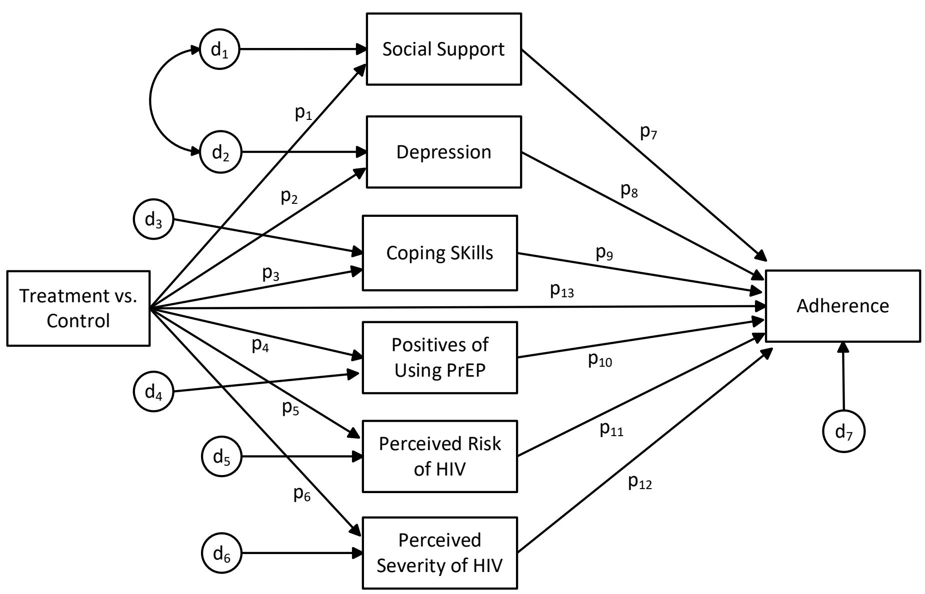

FIGURE 17.1. Adherence example

In reality, the model likely would be more complex by including causal relationships between some of the mediators. However, I use a simplified model here for purposes of pedagogy. The causal equations (with covariates included) implied by the model are:

$$\text{Support} = a_1 + p_1 \text{ Treat} + b_1 \text{ Support}_{\text{BASELINE}} + d_1 \quad [17.1]$$

$$\text{Depress} = a_2 + p_2 \text{ Treat} + b_2 \text{ Depress}_{\text{BASELINE}} + d_2 \quad [17.2]$$

$$\text{Coping} = a_3 + p_3 \text{ Treat} + b_3 \text{ Coping}_{\text{BASELINE}} + d_3 \quad [17.3]$$

$$\text{Positives} = a_4 + p_4 \text{ Treat} + b_4 \text{ Positives}_{\text{BASELINE}} + d_4 \quad [17.4]$$

$$\text{Risk-HIV} = a_5 + p_5 \text{ Treat} + b_5 \text{ Risk-HIV}_{\text{BASELINE}} + d_5 \quad [17.5]$$

$$\text{Severe-HIV} = a_6 + p_6 \text{ Treat} + b_6 \text{ Severe-HIV}_{\text{BASELINE}} + d_6 \quad [17.6]$$

$$\text{Adhere} = a_7 + p_7 \text{ Support} + p_8 \text{ Depress} + p_9 \text{ Coping} + p_{10} \text{ Positives} + p_{11} \text{ Risk-HIV} + p_{12} \text{ Severe-HIV} + p_{13} \text{ Treat} + d_7 \quad [17.7]$$

I use *p* in the equations to refer to path coefficients of interest and *b* to refer to coefficients for covariates.

Relative Importance of Omnibus Indirect Effects Using Raw Metrics

The standard Mplus syntax to evaluate the model is in Table 17.2. (I add to this syntax later to address mediator relative importance). I number each line for referencing but the line numbers are not part of Mplus syntax. You should be familiar with most of the syntax from previous chapters.

Table 17.2: Mplus Syntax for Adherence Example

```
TITLE: EXAMPLE CHAPTER 17 ;
DATA: FILE IS c:\mplus\ret\chap17M.txt ;
VARIABLE:
NAMES ARE
id support coping pos riskHIV sevHIV depress supportb
copingb posb riskHIVb sevHIVb depressb treat adhere  ;
USEVARIABLES ARE
support coping pos riskHIV sevHIV depress supportb
copingb posb riskHIVb sevHIVb depressb treat adhere  ;
MISSING ARE ALL(-9999) ;
ANALYSIS:
ESTIMATOR=MLR ;
MODEL:
!Specify equations
adhere on support depress coping pos riskHIV sevHIV treat (p7-p13) ;
support on treat supportb (p1 b1) ;
depress on treat depressb (p2 b2);
coping on treat copingb (p3 b3) ;
pos on treat posb (p4 b4) ;
riskHIV on treat riskHIVb (p5 b5);
sevHIV on treat sevHIVb (p6 b6);
```

```
!Specify correlated disturbances
support with depress ;
MODEL INDIRECT:
adhere IND treat ;
OUTPUT:
SAMP STDYX MOD(ALL 4) RESIDUAL TECH4 ;
```

The use of the *p* and *b* variable labels maps onto the labels in Figure 17.1. Because my initial focus is on the analysis of the relative importance of omnibus indirect effects, it probably is best to use percentile bootstrapping for the significance tests. However, doing so in this first analysis does not permit me to examine modification indices when evaluating global fit, so my initial run uses MLR to allow me to better evaluate model-data correspondence. Later, I shift to bootstrapping for purposes of sensitivity checks.

The global fit indices suggested reasonable model-data correspondence. The chi square test statistic (df = 50) was 50.18, $p < 0.47$, the RMSEA was 0.003 with a 90% confidence interval of 0.00 to 0.029, the p value for close fit was 1.00, the CFI was 1.00, and the standardized RMR was 0.028. None of the z tests comparing predicted and observed covariances were statistically significant. I obtained four modification indices that were slightly greater than 4 but decided they were chance results because they did not make conceptual sense (I know for a fact that the modification indices results were chance because I generated the population data to conform to the tested model). I next re-ran the analysis using bootstrapping but added `MODEL CONSTRAINT` commands that allowed me to address relative importance of the omnibus indirect effects. Before explaining the `MODEL CONSTRAINT` commands, let me first characterize the results for the overall total effect of the program.

The control group mean for adherence at the posttest was 45.38 (SD = 18.95). This represents relatively low levels of adherence, less than 50%. The model estimated mean adherence difference between the treatment and control groups in the Mplus output was 11.28 ±2.79, $p < 0.05$ (taken from the section of the output labeled `TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS`). Although the result is statistically significant, it seems somewhat small in magnitude – the program only increased what was a fairly low mean adherence to begin with by about 11 percentage points. The Cohen's *d* was almost 0.60 but this does not change the fact that the program total effect needs to be better.

To evaluate the relative importance of the omnibus mediation effects for each of the 6 mediators, I need to document the effect of the treatment on the outcome through each mediator and then compare these omnibus indirect effects with one another. Although the mediators are measured on different metrics because they have a different number of items, this is not a concern because it turns out that the indirect effect always reflects the mean difference between the treatment and control groups on the outcome, i.e., the effect is scaled in terms of the metric of the Y mean difference between the treatment and control groups. Here are the `MODEL CONSTRAINT` commands I added after line 25 of the syntax in Table 17.2.

```
25a. MODEL CONSTRAINT:
25b. NEW (ind1 ind2 ind3 ind4 ind5 ind6 diff12 diff13 diff14 diff15 diff16
25c. diff23 diff24 diff25 diff26 diff34 diff35 diff36 diff45 diff46 diff56);
25d. ind1=p1*p7 ;
25e. ind2=p2*p8 ;
25f. ind3=p3*p9 ;
25g. ind4=p4*p10 ;
25h. ind5=p5*p11 ;
25i. ind6=p6*p12 ;
25j. diff12 = ind1-ind2 ;
25k. diff13 = ind1-ind3 ;
25l. diff14 = ind1-ind4 ;
25m. diff15 = ind1-ind5 ;
25n. diff16 = ind1-ind6 ;
25o. diff23 = ind2-ind3 ;
25p. diff24 = ind2-ind4 ;
25q. diff25 = ind2-ind5 ;
25r. diff26 = ind2-ind6 ;
25s. diff34 = ind3-ind4 ;
25t. diff35 = ind3-ind5 ;
25u. diff36 = ind3-ind6 ;
25v. diff45 = ind4-ind5 ;
25w. diff46 = ind4-ind6 ;
25x. diff56 = ind5-ind6 ;
```

Lines 25b to 25i define the 6 indirect effects by multiplying the path from T→M by the path from M→O. Lines 25j to 25x calculate all possible differences between the 6 indirect effects. To bootstrap the analysis, in Table 17.2 I change the estimator from MLR to ML on line 12 and add the text `BOOTSTRAP = 5000` to it, as follows:

```
ESTIMATOR = ML ; BOOTSRAP = 5000 ;
```

Then, on the output line (line 27), I change the confidence interval statement to read

```
CINTERVAL(BOOTSTRAP)
```

I also remove `MOD(ALL 4)` because it is not permitted with bootstrapping. The results are

```
MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

New/Additional Parameters
    IND1               0.890      0.704      1.264      0.206
    IND2               1.258      0.404      3.114      0.002
    IND3               4.281      0.691      6.196      0.000
    IND4               4.209      0.480      8.778      0.000
    IND5               0.008      0.046      0.177      0.859
    IND6              -0.015      0.242     -0.064      0.949
    DIFF12            -0.368      0.729     -0.505      0.613
    DIFF13            -3.391      0.972     -3.490      0.000
    DIFF14            -3.319      0.823     -4.034      0.000
    DIFF15             0.882      0.704      1.252      0.210
    DIFF16             0.905      0.744      1.217      0.224
    DIFF23            -3.023      0.808     -3.739      0.000
    DIFF24            -2.951      0.618     -4.776      0.000
    DIFF25             1.250      0.408      3.066      0.002
    DIFF26             1.274      0.462      2.760      0.006
    DIFF34             0.072      0.869      0.083      0.934
    DIFF35             4.273      0.690      6.195      0.000
    DIFF36             4.296      0.727      5.913      0.000
    DIFF45             4.201      0.485      8.669      0.000
    DIFF46             4.225      0.549      7.690      0.000
    DIFF56             0.024      0.246      0.096      0.924
```

It is helpful to discuss these results by organizing them into a table with the results for the individual links that comprise the separate mediation paths. The table appears in Table 17.3. Keep in mind that I cannot compare mediators on the values of the $p_{T\rightarrow M}$ coefficients or the $p_{M\rightarrow Y}$ coefficients per se because the mediators are measured on different metrics. However, I *can* compare mediators on their respective indirect effects (the second to last column of Table 17.3) because the effects have the same metric. I order the mediators in Table 17.3 in terms of their relative importance based on the size of their omnibus indirect effects in the second to last column of the table.

Table 17.3: Analysis of Omnibus Indirect Effects

Mediator	$p_{T\rightarrow M}$	$p_{M\rightarrow Y}$	Indirect Effect (IE)	Critical Ratio for IE
Coping	0.44*	9.72*	4.28[a]	6.20*
Positives of Using	0.85*	4.93*	4.21[a]	8.78*
Depression	-0.25*	-5.00*	1.26[b]	3.11*
Support	0.09	9.22*	0.89[b,c]	1.26

Risk of HIV	-0.09	-0.08	0.01[c]	0.18
Severity of HIV	0.62	-0.03	-0.02[c]	0.06

(In IE column, mediators with common superscripts are not significantly different from one another, $p > 0.05$)

As examples, the treatment is predicted to raise protocol adherence, on average, by 4.28 ±1.38 units through the coping skills pathway; it is predicted to raise protocol adherence, on average, by 4.21 ±0.96 units through the mediator that emphasizes the positives of using the medication. And so on. Based on the magnitude of the effects and the significance tests, two of the mediated effects (coping skills and positiveness of use) are strongest, two are intermediate (depression and support), and two are of lesser import (perceived risk of contracting HIV and perceived severity of HIV). I contrast these conclusions with conclusions derived from the analysis of individual links later.

Correlated Disturbances and Causal Effects Among Mediators

I consider in this section several cautions for the analysis of the relative importance of omnibus indirect effects. Recall that for the model in Figure 17.1, I included correlated disturbances between social support for adherence and depression. Without it, the model is unable to adequately account for the covariation between these variables. For example, if I test a model that omits it, the z test comparing predicted versus observed covariances for the two variables was 4.93, which is highly statistically significant, suggesting the model is not accounting for the correlation between the two variables very well. When I generated the population data from which the sample data were selected, I purposely introduced a correlation between the two disturbances in the population on the premise that there are unmeasured variables external to the model that have a positive impact on social support for medication adherence and a negative impact on depression, such as generalized social support. Thus, the correlated disturbance should be in the model.

Suppose I naively fit a model without the correlated disturbance. Not only would I observe the above z value for the test of the predicted and observed covariances, but I also would observe two substantial modification indices, one for the correlated disturbances and the other for the omitted causal path from social support for adherence to depression. Thus, after fitting the misspecified model, based on model diagnostics, I would be faced with a choice between two plausible mechanisms that can account for the localized ill fit, (1) I can correlate the disturbances or (2) I can introduce a causal impact of the support variable on depression. Adding either one of these mechanisms to the model would eliminate the misfit localized on the observed correlation between the two focal variables. I need to decide which of the mechanisms to specify in the revised model, letting theory be y guide. Note that if I choose the second mechanism, then the estimate of the value of the omnibus indirect effect of social support on adherence changes. In Figure 17.2a, which assumes correlated disturbances, the omnibus indirect effect (OIE) for the effect of the treatment through social support is

$$OIE = p1*p7 = .09*9.22 = 0.89$$

whereas in Figure 17.2b, where a causal effect of support on depression is posited, it is

$$OIE = p1*p7+p7a*p8 = .09*9.22 + -.23*-5.00 = 1.98$$

because support is assumed to influence adherence through two chains not just one.

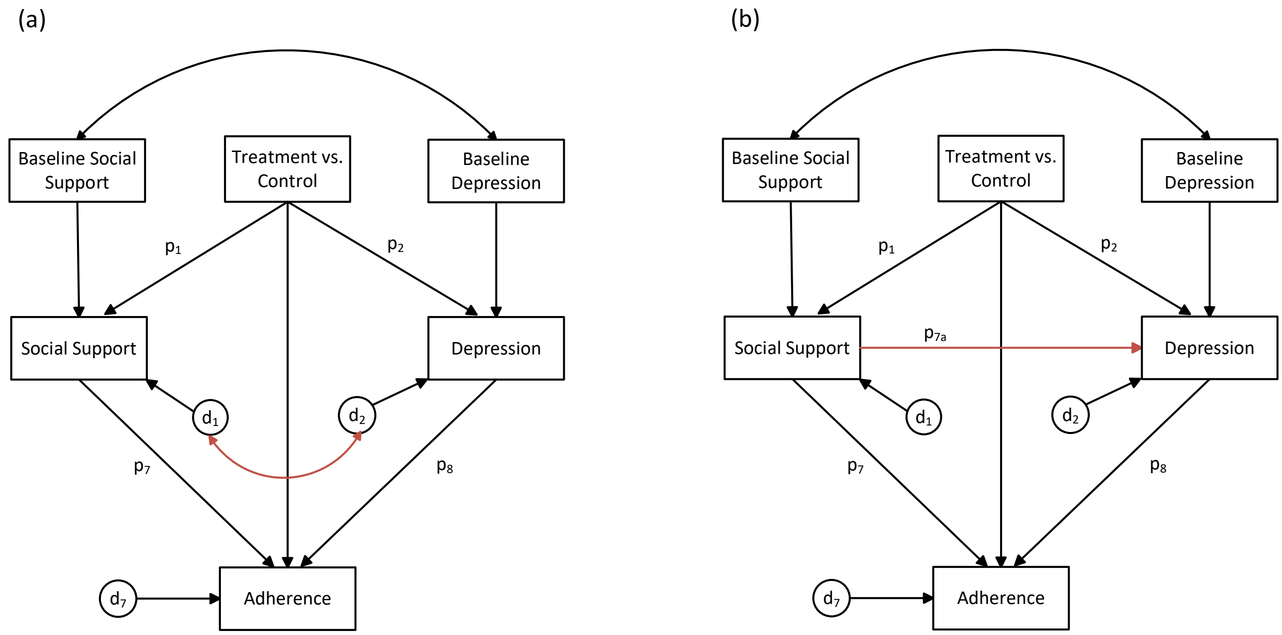

FIGURE 17.2. Adherence example with correlated disturbances versus causal links

This point is critical. How I choose to model the ill fit impacts the OIE, so I must give careful thought as to how to address that ill fit. I know in the present case that the correct choice is to allow correlated disturbances because I created the hypothetical population data based on this dynamic. However, I would not know this in practice. Upon reflection, I might even decide that *both* mechanisms are plausible and posit a model with both dynamics. Or, I might consider any or all of the dynamics in Figures 17.3a through

Figure 17.3g as better accounting for the ill fit for the covariance between depression and social support. In the absence of strong theory to guide me, I must recognize there are multiple models that might account for the data. This fact makes me approach my results on the mediation analyses with humility and tentativeness.

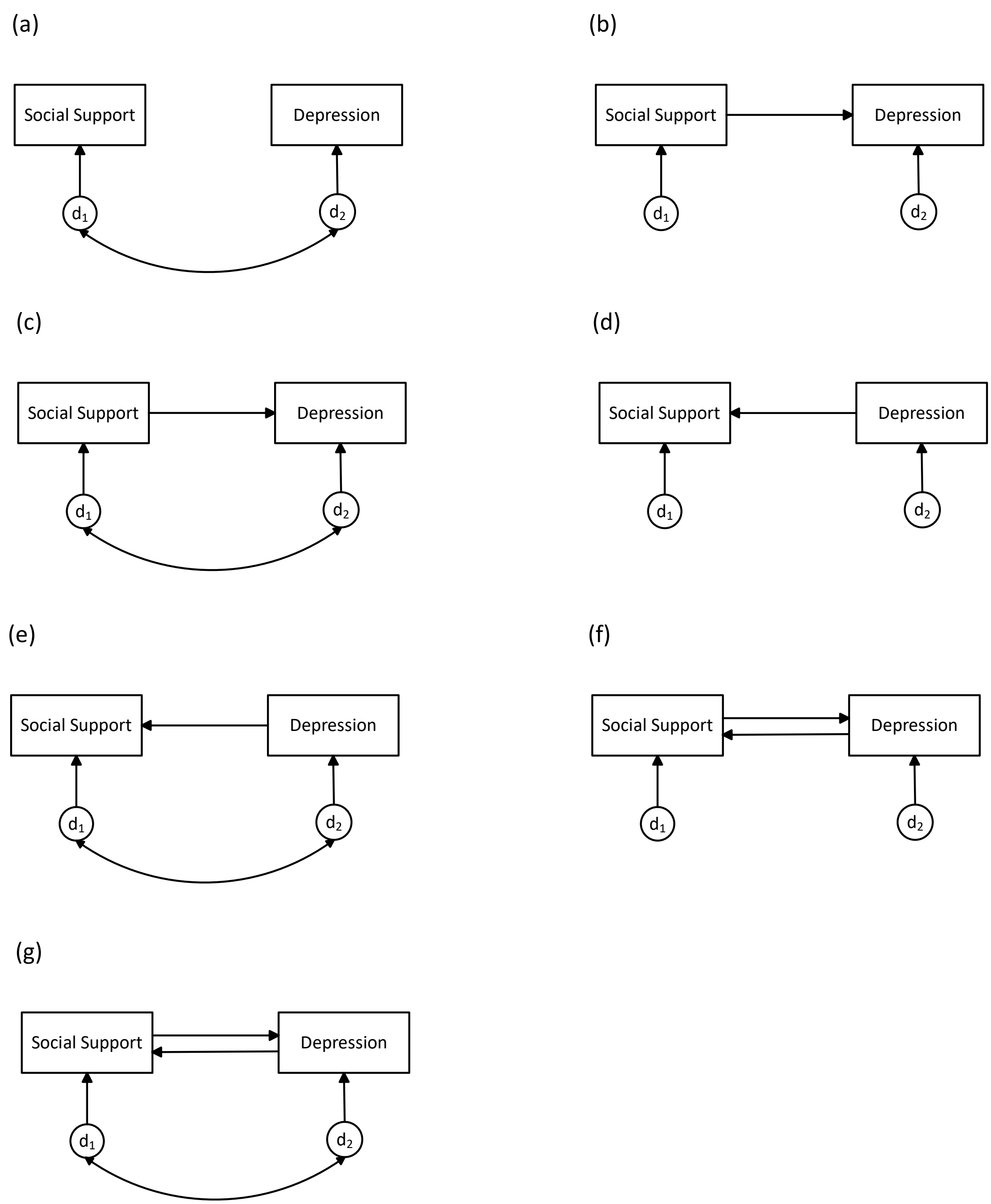

FIGURE 17.3. Alternative models of mediator relationships

TESTS OF RELATIVE IMPORTANCE OF MEDIATOR EFFECTS ON OUTCOMES

As stated, I believe that tests of omnibus indirect effects for purposes of program evaluation are limited because they confound or mix up conceptually distinct dynamics, namely (1) whether the program has meaningfully affected the mediator and (2) whether the mediator is meaningfully relevant to the outcome. In these next sections, I discuss how to index the relative importance of the effect of mediators on outcomes independent of the effects of the program on the mediator. Doing so prevents conflation of the effects.

When addressing the relative importance of mediators in influencing an outcome, the task is particularly challenging because mediators often are measured on different metrics. This means one cannot directly compare their respective path coefficients for M→Y. The magnitude of these values are metric dependent. In the numerical example, several mediators were indexed by multi-item scales in which each item was responded to on a -3 to +3 disagree-agree metric. I averaged item responses so that the total scores on a scale also varied on a -3 to +3 metric. As noted, this does not mean, however, that the mediators have been measured on the same metric. Because the total scores are based on a different number of items, a 1 unit change on one of the total scores can mean something distinct from a one unit change on the other. Suppose one mediator had 5 items and the other had 15 items. If I summed rather than averaged the item responses, the total scores would range from -15 to +15 and -45 to +45, respectively. Researchers would not treat the two total scores as being on the same metric. Averaging the item responses to form a total score does not change this dynamic. The technical definition of a metric in statistics is complex (Wilcox, 2021). I emphasize here that although total scores for multi-item mediators may share the same number system, it does not necessarily mean they have a common metric. They may or may not.

There is a large literature in linear regression (not SEM) on methods for evaluating the relative importance of predictors in a regression equation. I consider this literature here as applied to predicting outcomes from mediators. I first discuss coefficient statistical significance as a way of classifying predictors into those that are "important" and those that are "not important." I then consider a strategy called **best subset analysis** that uses either stepwise regression, lasso regression, all possible regressions, or generalized additive models. Next, I discuss the use of correlations, standardized regression coefficients, and semi-part correlations as indicators of predictor importance. Finally, I consider dominance analysis. I discuss the methods in the context of linear regression but then I consider how they can be applied in SEM contexts. From the PrEP numerical example, I focus on Equation 17.2 predicting adherence from the mediators and the treatment variable directly and use it to explore relative mediator importance. I repeat that equation here for convenience:

$$\text{Adhere} = a_7 + p_7\,\text{Support} + p_8\,\text{Depress} + p_9\,\text{Coping} + p_{10}\,\text{Positives} + p_{11}\,\text{Risk-HIV} + p_{12}\,\text{Severe-HIV} + p_{13}\,\text{Treat} + d_7$$

Using p Values as Indicators of Mediator Relative Importance

One strategy for identifying "important" versus "non-important" predictors in a regression equation uses significance tests of the coefficients and declares variables that fail to yield statistically significant regression coefficients as "unimportant" and those that yield statistically significant coefficients as "important." As straightforward as this seems, the approach can be problematic. The most obvious problem is that a study may have low statistical power leading it to declare one or more important predictors as statistically non-significant and hence, unimportant. The power problem is more insidious than many who use this approach realize. To illustrate, Maxwell (2000) notes that the typical correlation between variables in psychological research is about 0.30. If five predictors in a population are each correlated 0.30 with the criterion as well as 0.30 with each other, then the percent of unique explained variance in Y for each predictor will be 1.5% and each population regression coefficient will be non-zero and equal in value if the predictors all are measured on the same metric, i.e., all the predictors are equally important. The sample size needed to obtain statistical power of 0.80 for a significance test of a given path/regression coefficient in this case is about 420, which is well above typical sample sizes in many studies.

Maxwell (2000) reported a simulation study in which a multiple regression analysis was conducted using the above scenario, but with a sample size of only 100. Maxwell found that the most frequently occurring pattern of results, occurring 45% of the time, was the case where one predictor had a statistically significant regression coefficient but the other four predictors did not. The next most common pattern, occurring 32% of the time, was that two of the predictors had statistically significant regression coefficients, but three did not. Thus, in a situation where each of the five predictors is of equal import in the population, there was a high likelihood that only one or two of the predictors would exhibit statistical significance. The predictors that showed a statistically significant coefficient among the five predictors were essentially random. Such results should give theorists using smaller sample sizes pause about declaring a mediator "unimportant" if it receives a statistically non-significant regression coefficient.

Another shortcoming of the p value method is that importance is treated dichotomously; a predictor is either important or not. Sometimes we seek more nuanced importance judgments than this. The approach also equates statistical significance with practical significance, an equivalence that many researchers question (Amrhein, Greenland & McShane, 2019; Aguinis et al., 2010; Wasserstein, Schirm & Lazar, 2019). There are better approaches than one based exclusively on p values that can be used.

Best Subset Analysis as Indicators of Mediator Relative Importance

I next consider four methods of mediator subset analysis, (1) stepwise regression, (2) lasso regression, (3) all possible regressions, and (3) generalized additive models. These methods also seek to divide predictors into two groups, those that are "important" and those that are "not important," i.e., they "subset" the predictors into two classes. Each method has non-trivial flaws but it is important you are aware of them because they continue to appear in the literature, albeit somewhat inconsistently.

Stepwise Regression

In stepwise regression, the predictor variables represent a pool of potentially "important" mediators for the outcome. Stepwise algorithms enter variables sequentially into a regression equation for purposes of predicting the outcome but only "important" mediators are selected for inclusion in the equation on any given step if the analysis. Sequential inclusion is based on how much each mediator augments the squared multiple correlation relative to predictors/mediators already in the equation. One variant is a **forward approach** which first enters into the equation the mediator with the highest correlation with the outcome.[1] The next mediator added is based on the mediator which adds the most unique explained variance relative to the mediator(s) already in the equation *and* whose regression coefficient is statistically significant when it is added. The process of adding mediators continues until none of the remaining mediators in the predictor pool add statistically significant incremental explained variance to the equation. The mediators in the final equation are deemed "important" and those that do not make it into the final equation are deemed "unimportant."

A variant of the forward approach is the **stepwise approach** which uses the forward strategy but also tests for the statistical significance of *all* of the mediators in the equation at each step that a new mediator is added. If a previously entered mediator now yields a statistically non-significant coefficient with the addition of the new mediator, it is dropped from the equation. Yet another variant is the **backward elimination approach** which starts with an equation that includes all mediators and then at each step removes the mediator that accounts for the least incremental explained variance and whose regression coefficient is statistically non-significant. The process continues until at the

[1] If there are demographic or other covariates to control, you would force entry of those variables on the first step in the analysis and then conduct the stepwise analysis with those variables already in the equation.

final step, only statistically significant mediators remain in the equation.

Instead of using statistical significance as a criterion for stopping the process in forward, backward, or stepwise analyses, some researchers use an *a priori* specified proportion of unique explained variance (e.g., 0.05) as a standard. For example, the backward elimination approach continues until only mediators that each account for at least 5% unique explained variance are in the equation; or, the forward approach continues until none of the remaining mediators in the predictor pool add at least 5% unique explained variance to the existing equation.

Numerous objections to these approaches have been raised, including the misleading nature of the significance tests (Altman & Anderson, 1989), bias in the regression coefficients (Tibshirani, 1996), and a general failure of the approaches to accurately identify variables that are part of the true generating function for outcomes (see Derksen & Keselman, 1992; Mantel, 1970). To illustrate one problem, the first mediator that enters the equation in a stepwise or forward analysis is the mediator that has the highest zero order correlation with the outcome. At the second step, all remaining mediators are considered for inclusion relative to that first mediator. Only mediators that add significant unique explained variance relative to it are candidates for inclusion at the second step. Suppose M1 has a sample correlation of 0.30 with the criterion and M2 has a sample correlation of 0.29 and the two variables are correlated 0.70. No other mediator has a correlation as large with the outcome. M1 will enter the equation first, even though its correlation with the outcome is only larger by a miniscule amount compared to M2. M2 will not enter the equation at later steps because its explained variance in Y is largely redundant with M1 (as evidenced by its high correlation with M1). It is entirely possible that the correlation between M2 and Y is larger in the population than the correlation between M1 and Y and that the reversal of rank order of the correlations in the sample data reflects nothing but sampling error. Despite this, M1 is given theoretical priority and enters the equation first. In this case, a relatively small amount of sampling error alters the mediators that enter the equation not only at the first step but throughout the entire mediator selection process because the entry of the remaining mediators depends on the variables that are in the equation from previous steps. The results can be misleading.

Lasso Regression

Another subset-based approach is **lasso regression** (James, Witten, Hastie & Tibshirani, 2013).[2] Lasso regression is a constrained version of traditional OLS regression. It applies OLS to a prediction equation but with the constraint that the sum of the absolute value of

[2] Lasso is an acronym for least absolute shrinkage and selection operator.

the regression coefficients must be less than an a priori defined constant, *L*.[3] The maximum value *L* is allowed to have is the value that would result if one summed the absolute values of all the traditional OLS coefficients. The minimum value is 0. When *L* is set to its maximum value, the resulting equation is simply a standard OLS regression. Lower values of *L* tend to shrink the values of the coefficients toward 0, more so when small values of *L* are specified. This shrinkage results in biased coefficients but such bias, the logic goes, is compensated for by lower sample-to-sample variation in the coefficients, a trade off some are willing to make.

Lasso regression is executed in successive steps. At step one, all predictors/mediators are assigned coefficients of zero. At each successive step, one additional mediator is allowed to have a non-zero coefficient. By the final step, all mediators are "in the model" in the sense that they all are allowed to have non-zero coefficients. The final step is in essence the OLS model. One typically evaluates the quality of prediction at each step and stops the sequential process when further addition of mediators trivially improves model fit as indexed by a reduction of prediction error. When the assumptions of OLS are met, OLS will, by definition, yield the lowest prediction error in the sample. However, lasso regression has a penalty function for lack of parsimony and potential overfitting, so the equation with all mediators is not necessarily the best. The approach can be summarized as:

a. First, set all regression coefficients for mediators in the mediator pool to zero.

b. Next, find the mediator, M_k, most correlated with Y, and allow its coefficient, B_k to be non-zero.

c. Increase the value of the coefficient B_k by a small amount in the direction of the sign of M_k's correlation with Y, calculating the residual $e = Y_i – \hat{Y}_i$ for individuals for each increment in B_k. Stop increasing the value of B_k and move to the next step when some other mediator M_m with coefficient B_M has as much correlation with the e scores as does M_k. Before moving on, calculate an index of model fit/adequacy given the current set of mediators with non-zero coefficients in the model (I describe the index below).

d. Next, increase (B_k, B_m) in their joint least squares direction until some other predictor/mediator X_n has as much correlation with the residual based on the predicted Y from B_k and B_m. Continue this process, adding a mediator at each successive step, until the final model has all mediators in it.

[3] If the predictors are measured on different metrics, it usually is necessary to standardize them to apply lasso regression for L to be meaningfully used as a variable inclusion/elimination criterion.

e. Select the model at the step where further additions of mediators will not substantively improve the reduction of prediction error.

A popular index of model fit/adequacy or prediction error at a given step is called **Mallows' C** and is abbreviated by C_p. It is defined as
$C_p = 1/N\ [RSS + (2k)(s_e^2)]$

where RSS is the residual sum of squares for the model, *k* is the number of predictors and s_e^2 is the estimated residual variance. The lower the value of C_p, the better the model fit, everything else being equal. Essentially the value of C_p is a function of the amount of error variance plus a penalty function that increases as the number of mediators in the model increases (the term $(2k)(s_e^2)$). One typically chooses the model that has the lowest value of C_p. Lockhart et al. (2014) developed a significance test for variable entry at each step of lasso regression which is sometimes used to inform model selection, but it is controversial. Once a best fitting equation is selected, the mediators in it with non-zero coefficients are deemed "important."

Lasso regression is often used in the big data and machine learning literatures for purely prediction purposes to identify prediction equations for future events. However, like forward, stepwise, and backward elimination approaches, it sometimes is used to identify a subset of "important" predictors that one then theorizes about, builds substantive knowledge about, and ultimately models in broader multivariate contexts. I find it to be of limited use outside of prediction contexts, as elaborated below.

Related approaches to lasso regression include methods known as **least angle regression**, **elastic net regression**, and **ridge regression**. Use of these methods as well as lasso regression for variable reduction purposes and subset analysis is discussed in depth by Efron et al., (2004), Zou and Trevor (2005), and James et al. (2013).

Generalized Additive Models

In Chapter 16, I described the generalized additive model (GAM) as a method for modeling linear and non-linear relationships between mediators and outcomes. I assume here that you are familiar with my treatment of GAMs in that chapter. GAMs can be used for mediator subset selection in the same spirit as lasso regression when the number of mediators is not too large. The advantage of using GAMs instead of lasso regression is GAMs are sensitive to both linear and non-linear relationships between mediators and outcomes. The *mgcv* package in R offers two variable selection methods, one called the **shrinkage method** and the other called the **double penalty method**. Marra and Wood (2011) found that the double penalty method tends to perform better than the shrinkage method and they recommend it (it is the default selection method in the *mgcv* package). The double penalty method adds a second penalty to the traditional penalty matrix for wiggliness in GAMs, but the second penalty only affects functions in the null space targeting the overall linear portion of the smooth. The two penalties operate in ways that the influence for a given mediator in the prediction equation can be reduced to zero, much like lasso regression. For mathematical details, see Marra and Wood (2011). The double penalty approach can be computationally demanding because each smooth requires two smoothness parameters to be estimated. For an example of subset selection using GAMs, watch the video for the *generalized additive model* program on my website under the Programs tab.

As an alternative to these methods, some researchers use a backward method for predictor removal in GAMs. This method proceeds in the same way as the backward regression method described above but mediators are sequentially dropped based on their p-values in the GAM analysis. The approach begins by including all predictors from the predictor pool in the equation and then drops the mediator with the highest non-significant p-value. This elimination process continues sequentially with a refitted model at each step until all terms that remain in the equation are statistically significant. A weakness of this approach is that the p-values of the coefficients are only approximate, making it somewhat questionable to rely on them. Some researchers use AICs instead of p values when using the backward approach, evaluating changes in AIC as each backward step is executed. The backward process stops when dropping a predictor leads to a non-trivial change in the AIC.

All Possible Regressions

The final subset strategy I discuss is called **all possible regressions**. This approach examines prediction error for equations that comprise all possible combinations of predictors/mediators. For example, the predictor pool might consist of variables M1, M2 and M3. I would apply OLS regression to an equation that predicts the outcome Y from just M1, a second equation that predicts Y from just M2, a third equation using just M3, a fourth equation using both M1 and M2 as predictors, a fifth equation using both M1 and M3 as predictors, a sixth equation using both M2 and M3 as predictors, and a seventh equation with predictors M1, M2 and M3. The mediators in the equation that produces the lowest amount of prediction error or that has the best model fit is the subset of "important" mediators. In general, prediction error decreases as one adds predictors to a regression equation. Like lasso regression, the all possible regressions strategy compensates for this by using indices of prediction error that include penalty functions for larger numbers of predictors, such as Mallow's C or the AIC or BIC. As such, it is not necessarily the case that equations with the most mediators will be "best."

A disadvantage of all possible regressions is that it can be computationally intense as the number of mediators increases, so it is not well suited to scenarios where there are many mediators. The number of possible models is 2^k-1, where k is the number of predictors. With 10 mediators there are 1,023 possible models. This is a lot of models.

Concluding Comments on Best Subset Analysis

In general, I think best subset strategies are limited for ordering mediators on their relative importance. As seen later, I make use of some of them to help reduce a large number of plausible mediators to a workable number for purposes of more detailed SEM analyses, but as more fine-grained approaches to the identification of mediator relative importance, they have shortcomings. One primary shortcoming is that they only yield dichotomous characterizations of importance. Also, when fully automated, they equate importance with prediction rather than causation and they fail to take into account theoretical coherence. Predictive accuracy is defined in different and sometimes seemingly arbitrary ways across the methods. Judd and McClelland (1989) capture the essence of reservations expressed by many researchers when they state "it seems unwise to let an automatic selection algorithm determine the questions we do and do not ask about our data" (p. 204). You will encounter these methods in the mediation literature so it is important to know about them. However, as methods for determining mediator importance, there are better approaches.

Correlations, Standardized Coefficients, and Part Correlations as Indicators of Mediator Relative Importance

Two commonly used indices of mediator relative importance are the squared zero order correlation of the mediator with the outcome and the standardized regression coefficient for a mediator when it is embedded in a larger regression equation with the other mediators. I discuss each index in turn as well as the use of part correlations.

Squared Zero Order Correlations

If the mediator and the outcome are linearly related, normally distributed, and free of confounds, then the squared bivariate correlation coefficient between them is often interpreted as the proportion of variation in the outcome that is "explained by" or "accounted for" by the mediator. Mediators with larger squared correlations are said to be more important than those with smaller squared correlations. The aforementioned assumptions, however, are not trivial, especially the assumption that the relationship between the mediator and the outcome is confound free. Given there are almost always confounds between mediators and outcomes, the squared zero order correlation is not a good candidate as an index of the relative importance of mediators in RETs. The presence of such confounds makes them difficult to interpret. The squared correlation also ignores redundant explained variance with other mediators.

Standardized Regression Coefficients and Part Correlations

Another common practice is to use standardized regression coefficients as indicators of the relative importance of predictors/mediators in determining outcomes (see Chapter 10). With complete standardization, the metric of each predictor and the outcome is converted to a standard score purportedly to place all measures on a common metric. Predictors with larger absolute standardized regression coefficients when the outcome is regressed onto the mediators are then said to be more important than those with smaller standardized regression coefficients.

One problem with this strategy is that the assertion that standardization places predictors on a common metric is questionable (Willitt, Singer & Martin, 1998; Judd, McClelland, & Ryan, 2009; Blanton & Jaccard, 2006). For example, does a one standard deviation increase in social mobility really mean the same thing and have the same implications for age of onset of diabetes as does a one standard deviation increase in weight if both predictors have standardized regression coefficients of 0.35? Willitt et al. (1998) answer this question with a resounding "no." In their analysis of the relative importance of familial rule setting and maternal education as determinants of adolescent delinquent behavior, Willit et al. frame the issue as follows:

> *"Is a one standard deviation difference in rule setting the same as a one standard deviation difference in a variable like maternal education? The answer to this question depends upon the sample homogeneity with respect to these variables which in turn depends, in part, on researchers' decisions about target populations and sampling strategies. Yet, standardization effectively eliminates information about homogeneity from consideration, creating the false illusion that coefficients can be directly compared"* (p. 412).

A hypothetical but somewhat "tongue-in-cheek" example adapted from King (1986) makes evident the difficulties with the standardized coefficient approach. Consider the adage "an apple a day keeps the doctor away." Suppose I analyze the number of times people see a doctor per year as a function of the number of apples and oranges they eat per week and obtain the following unstandardized regression equation:

Number of visits = 3.0 + -1.0 Apples + -.25 Oranges.

Based on this equation, an apple a week decreases the number of annual visits to a doctor by 1, whereas an orange a week decreases the number of visits, on average, by 0.25. Based on this, I conclude that it takes only one apple to accomplish what four oranges accomplish. Now suppose the SD for apples eaten in a week is 0.50, for oranges it is 1.0, and for the number of per year doctor visits, the SD is 3.0. Given this information, the above equation can be expressed using standardized coefficients as follows:

Standardized number of visits = 0.0 + -0.17 Apples + -.08 Oranges.

It now appears that one apple has the same effect as two oranges (-.17 is double the size of -.08). Which answer is correct? Does one apple equal the effect of four oranges or does it equal the effect of two oranges? It turns out that the results using standard scores refers not to apples and oranges per se but rather to *scaled* apples and oranges vis-à-vis their standard deviations. The standardized result is thus driven not only by the causal effect of apples and oranges on visits to the doctor but also by their respective standard deviations, which are part of that scaling process. Given this, there is a mixture of multiple dynamics that make it difficult to interpret the coefficients in a straightforward way. Many methodologists are uncomfortable with this fact because they feel that standard deviations are sensitive to features of study design that have little to do with variable importance, such as who you include in your sample and variables you control for (see Greenland, Schlesselman & Criqui, 1986). The recommendation is therefore not to compare standardized regression coefficients for making inferences of relative causal importance.

As another example, suppose in a two group RET that a treatment relative to a control condition has a substantial effect on one mediator, M1, but not on a second mediator, M2. Suppose that both M1 and M2 are measured on the same metric, that they have equal variances at baseline, and that, in the abstract, each has the same causal effect on the outcome, i.e., the population unstandardized coefficient $\beta_{M1\rightarrow Y}$ equals $\beta_{M2\rightarrow Y}$. If I collapse across the treatment and control conditions, the posttest variability of M1 will be larger than the posttest variability for M2 because the program created a degree of separation between the treatment and control groups for M1 but failed to do so for M2. As a result of this effect on the variability of M1 but not M2, the two mediators have different standardized regression coefficients when Y is regressed onto them even though the unstandardized coefficients are identical. The differential standardized coefficients are an artifact of program effects on variability rather than causal importance.

Finally, when ordering mediators in terms of their relative importance using standardized regression coefficients, it seems wise to take into account sampling error by virtue of executing significance tests of those differences and by evaluating confidence or credible intervals of them. However, this is rarely done (see Cohen et al., 2003 for statistical methods for making such comparisons).

Standardized regression coefficients have other properties that are bothersome for mediator relative importance analysis (see Bring, 1994, 1996; Darlington, 1968; Grömping, 2007a,b for elaboration). There are better approaches you can use. Both partial and semi-partial correlations as indicators of relative importance are subject to the same artifacts as standardized regression coefficients, so they too are not good measures of relative importance (Grömping, 2007a,b, 2015). In addition, they focus only on unique explained variance of a predictor, which is limiting, as I discuss below.

Dominance Analysis and Mediator Relative Importance

A final approach to evaluating mediator relative importance is **dominance analysis**. Dominance analysis has its roots in methods that allocate shares of the explained variance of a linear model to each predictor, with the sum of the shares equaling the overall amount of explained variance, usually the squared multiple correlation. Approaches to such decompositions other than dominance analysis include (a) a method developed by Hoffman (1960; see also Pratt, 1987) that works with the sum of the product of the standardized coefficient multiplied by the correlation coefficient for each predictor, and (b) a principal components based relative weight method developed by Johnson (2000; Johnson & LeBreton, 2004). The latter method has been found to have non-trivial shortcomings (see Thomas, Zumbo, Kwan & Schweitzer, 2014) so I do not consider it further. The Hoffman/Pratt method is viewed by most statisticians as less satisfactory than dominance analysis (Grömping, 2007a,b, 2015; *cf* Menard, 2007), hence my primary focus here is on dominance analysis.

Budescu (1993) and Johnson and LeBreton (2004) argue that importance metrics for a predictor, X, in a multi-predictor regression analysis should simultaneously take into account the following criteria; (a) the contribution of X to Y when X is the only predictor, (b) the contribution of X to Y over and above all other predictors (unique explained variance), and (c) the contributions of X to Y considering different subsets of the other predictors. Dominance analysis uses all three criteria. Importance indices such as the standardized regression coefficient, the semi-part correlation, and t ratios of coefficients do not. Grömping (2015) lists a dozen different desiderata for importance indices, most of which are satisfied by dominance analysis. Of all the relative importance methods typically discussed in the literature on variable importance in regression analysis, dominance analysis usually is considered to be the method of choice, although it does have limitations (see Grömping, 2007a,b; 2015).

Dominance analysis uses an index called **general dominance** that is based on the average increase in R^2 for all subset models of equal size that include the predictor in question relative to models that do not include it. The index reflects the average unique explained variance contribution of X to the outcome across all possible subsets of predictors. It uses much more information and in more sophisticated ways than the relative importance indices discussed prior to this. A useful feature of general dominance indices is that across predictors they sum to equal the total R^2. Dominance analysis maps closely onto an approach by Lindeman, Merenda and Gold (1980), so you will sometimes see it referred to as the LMG method. As an example, with k=3 predictors, (A, B, and C), one can calculate the increase in R square that A yields relative to a model with no other predictors in it, that A yields over and above B, that A yields over and above C, and that A yields over and above B and C together. The increments are averaged within common subsets of the number of predictors (e.g., for the case where k = 1, then where k = 2, and then where k = 3) and then these averaged values are, in turn, averaged for the predictor in question to yield its general dominance index. This process is repeated for each predictor. For the 4 predictor case, (A, B, C, and D), one calculates the increase in R^2 that A yields over and above a model with no predictors, that A yields over and above B, that A yields over and above C, that A yields over and above D, that A yields over and above B and C together, that A yields over and above B and D together, that A yields over and above C and D together and that A yields over and above B, C, and D. Obviously, computations become intense for large numbers of predictors. For details and well explained numerical examples, see Azen and Budescu (2003).

I provide a program on my website called *Dominance analysis* that allows you to apply the method to a linear regression equation. I applied the program to the adherence example and the results are shown in Table 17.4 (details of the analysis and features of the program are provided in the video for the program on my webpage). Table 17.4 also includes results from the analysis of the omnibus mediational effects originally reported in Table 17.3 for comparative purposes. The first column in Table 17.4 is the traditional dominance index. The entries in the column sum to the value of R^2. The total R^2 was 0.71. The second column provides normalized dominance indices that sum to 100% and represent the percentage of the total explained variance that the mediator accounts for. For example, the mediator for coping skills is apportioned 23% units of explained variance of adherence (column 1) and this value represents 32% of the total explained variance by all of the mediators combined or R^2 (column 2). The program bootstraps pairwise comparisons between the dominance indices, the results of which are summarized using superscripts in column 1. In the last two columns, I present the indirect effects from Table 17.3, with the next to last column being the raw omnibus indirect effects and the last column being normalized omnibus indirect effects that sum to 100.

Table 17.4: Dominance Analysis

Mediator	Dominance M→Y	N-Dominance M→Y	Indirect Effect T→M→Y	N-Indirect Effect T→M→Y
Coping	0.226[a]	32.03	4.28[a]	40.26
Positives of Using	0.071	10.00	4.21[a]	39.61
Depression	0.119	16.79	1.26[b]	11.85
Support	0.289[a]	40.84	0.89[b,c]	8.37
Risk of HIV	<0.01[b]	0.24	0.01[c]	<0.10
Severity of HIV	<0.01[b]	0.10	-0.02[c]	<0.-.20

(Note: In columns with superscripts, mediators with common superscripts are not statistically significantly different from one another, $p > 0.05$. Comparisons are only valid within a column, not across columns. N in title signifies normalized result.)

There are striking differences between the results of Table 17.3 that relied on omnibus indirect effects and those of Table 17.4. First, in the dominance analysis, the mediator for social support tends to dominate the other mediators yet it is a relatively weak mediator when analyzed vis-a-vis the classic omnibus indirect effect method. The reason for this is because social support is a solid predictor of adherence resulting in its strong showing in the dominance analysis, but the program failed to meaningfully change it (see Table 17.3) so the T→M link in the mediational chain T→M→Y is "broken." Is the support mediator relatively important? The dominance analysis says "yes" because support is a good predictor of adherence; the omnibus indirect effect analysis says "less so" but it does not give us a clue as to why. We only learn why when we analyze the individual links in the chain and see that the T→M link for support is the source of weaknesses in the omnibus effect. To be sure, support does *not* formally mediate the effect of the program on adherence because of the broken T→M link. But it seems to be an important malleable determinant of adherence vis-à-vis the dominance analysis and program designers simply need to do a better job of influencing it. Of course, if it is determined that support is too difficult to change, then the program focus on it might be dropped.

Dominance analysis can be applied in SEM contexts using limited information SEM; one simply applies the *Dominance analysis* program on my website to the linear equation of interest in the larger system of equations defined by the model. Dominance analysis also can be applied to full information SEM with latent variables. I provide a program on my webpage called *Dominance analysis II* that permits such applications.

For a discussion of how to use dominance analysis as a theory generation heuristic, see Jaccard and Jacoby (2020). To extend dominance analysis to the analysis of binary outcomes, consider the approaches of Azen and Traxel (2009), Tonidandel and LeBreton (2009), or use a modified linear probability model. However, all such extensions are somewhat controversial. To apply dominance analysis to models with interactions and for ways of introducing covariates, see Grömping (2006). To apply dominance analysis to hierarchical linear models (HLM), see Luo and Azen (2012). For a Bayesian approach to dominance analysis, see Korpon (2014) and Wang (2016). For extensions of dominance analysis to multiple outcome scenarios, see Azen and Budescu (2006) and Huo and Budescu (2009). An R package that applies many of these variants is the *dominanceanalysis* package. For a discussion of the effects of measurement error on dominance analysis, see Braun, Converse and Oswald (2019). For a discussion of suppressor variables in dominance analysis, see Azen and Budescu, (2003).

Dominance analysis must be used with caution. It inherently relies on standardized rather than unstandardized statistics which, as noted, can be misleading (King, 1986; Achen, 1990). As applied to mediation analysis, it assumes the linear model is correctly specified and reflective of the true causal dynamics surrounding the mediators. It is challenging to apply when there are causal relationships among mediators in the target equation, although it can accommodate correlated disturbances among mediators because these correlations are absorbed into the correlations between predictors.

Concluding Comments on Mediator Relative Importance

Methods for evaluating the relative importance of mediators in determining or predicting an outcome have received considerable attention in the social science literature. I have considered methods based on p values, subset analysis (including stepwise regression, lasso regression, generalized additive models, and all possible regressions), zero order correlations, part correlations, standardized regression coefficients, and dominance analyses. Each approach has strengths and weaknesses. In my opinion, there is no simple, statistical algorithm that can mindlessly be applied to determine the relative importance of a mediator compared to other mediators. Addressing such questions requires careful integration of substantive, theoretical, and statistical information, per Chapter 10. Of the methods described above, dominance analysis is in my opinion probably the most useful when evaluating the relative strengths of the $M \rightarrow Y$ link, though it is not perfect.

There are other approaches to assessing predictor relative importance that I have not addressed here. Ritter, Jewell and Hubbard (2014) have written an R package for importance analysis called multiPIM that uses estimating equation methodology (van der Laan & Robins, 2003) or targeted maximum likelihood (van der Laan & Rubin, 2006; van der Laan & Rose 2011) in conjunction with machine learning algorithms. Random forest approaches also are popular for analyzing predictor importance in prediction contexts (e.g., Liaw & Wiener, 2002; Kuhn 2008; Kuhn et al., 2011). An index of predictor importance often used in the literature on Bayesian additive regression trees (BART) is based on how frequently the predictor appears in the different regression trees during model estimation, also known as **usage indices** (see Sparapani, Spanbauer & McCulloch, 2021). These indices are somewhat limited in scope (Gelman, 2017), although some researchers find them to be helpful in certain BART contexts. Several non-parametric approaches to importance analysis have been proposed, as have neural network frameworks (Gevrey, Dimopoulos & Lek, 2020). In the final analysis, however, I lean towards the use of dominance analysis coupled with strong theory.

RELATIVE IMPORTANCE OF TREATMENT IMPACT ON THE MEDIATORS

Some mediators are harder to change than others. An intervention may not impact a mediator either because the program is poorly constructed, because it fails to adequately address the determinants of the mediator, or because the mediator is simply difficult to change and doing so is beyond program capacities. Some researchers seek to order the target mediators from those that were most effectively changed by the intervention to those that were least effectively changed by the intervention with the idea that if priorities must be set, mediators that are more effectively changed have the highest priority.

Quantifying mediator changes as a function of a treatment for comparative purposes can be challenging because the mediators often are measured on different metrics. One strategy for continuous mediators is to convert the absolute mean difference between the intervention and control conditions for mediators to a Cohen's *d* and then order the absolute *d* statistics across the different mediators. Larger absolute *d* values indicate mediators that were changed more. This approach has the weakness I discussed above for standardized regression coefficients – each *d* uses a different standardizer (i.e., a different standard deviation) and it is not known if the standardizer for one mediator, namely a standard deviation, has the same meaning and implications as that for another mediator.

A second strategy is to focus on the absolute value of the critical ratio for the significance test of the effect of the treatment condition on each mediator. Relative magnitude is then indexed for each mediator by dividing its absolute critical ratio by the sum of the absolute critical ratios across the mediators. The resulting index ranges from 0 to 1.00 and reflects the proportion of the critical ratio sum that the mediator accounts for. This method essentially uses as the standardizer the standard error of the mean difference between the treatment and control groups rather than the within group standard deviation of Cohen's *d* because the critical ratio is the mean difference divided by this standard error. Like Cohen's *d*, it suffers from the problem of whether a one standard error unit difference has the same meaning and implications across mediators.

For the case where all the mediators are binary, one can order the absolute proportion differences between the intervention and control groups for each mediator across the mediators. If the mediators are a mixture of continuous and binary predictors, then this strategy is not viable.

Some researchers use the aforementioned critical ratio approach when there is a mixture of binary and continuous mediators or they use the *d* approach by converting the proportion differences to analogs of Cohen's *d* using formulas from the meta-analysis literature and then order the *d*s across all the mediators. Hasselblad and Hedges (1995) suggest the following conversion formula for converting proportion differences to Cohen's *d* (see also Sanchez-Meca, Marin-Martinez, & Chacon-Moscoso, 2003; Whitehead, 2002):

$$d = [\ln(\text{Odds}_1 / \text{Odd}_2)][\sqrt{3} / \pi^2]$$

where $\text{Odds}_1 = p_1/(1-p_1)$, $\text{Odds}_2 = p_2/(1-p_2)$, ln = the natural log, and π = the mathematical constant for pi (which is approximately 3.1416).

For the PrEP numerical example, I calculated the covariate adjusted Cohen *d* value for each mediator by dividing the coefficient for the effect of the treatment condition on the mediator reported in the `MODEL RESULTS` section of the Mplus output by the square root of the unstandardized residual variance for the mediator reported in the `MODEL RESULTS` section in the sub-section labeled `Residual Variances`.[4] Here are the results organized in descending order as a function of their *d* values and their critical ratios:

Mediator	$p_{T\rightarrow M}$	Disturbance Var	Absolute d	Critical Ratio
Positives of Using	0.853	0.567	1.13	12.67
Severity of HIV	0.615	0.727	0.72	8.05
Coping	0.440	0.609	0.56	6.24
Depression	-0.251	0.760	0.28	3.21
Support	0.096	0.750	0.11	1.26
Risk of HIV	-0.094	0.754	0.11	1.20

[4] Each *d* is based on a different covariate that is extracted from the disturbance term (namely the baseline measure of the target mediator) which means we must assume that a covariate adjusted SD for one mediator has the same meaning and implications as a covariate adjusted SD for another mediator.

If one treats the *d* values and critical ratios as valid indicators of relative change, the program was most effective at changing the perceived positives of using PrEP followed by the perceived severity of HIV if it were to be contracted. Interestingly, the greatest amount of change was brought about on a mediator that had one of the lower dominance indices (the perceived positives of using PrEP), per Table 17.4.

In sum, no index of relative change is perfect. One can get a rough sense of the ordering of mediators in terms of the intervention's success in changing the mediator by using either Cohen *d* statistics or the critical ratios for the mediator significance tests. However, these methods make non-trivial assumptions about the equivalence of standardizers across mediators and must be used with caution. Other metric-free indices of change might be worth developing.

CONCLUDING COMMENTS ON RELATIVE IMPORTANCE

My own orientation to relative importance analysis is that the question of which mediators are most important probably is not all that interesting for purposes of program evaluation. Rather, I want to know (a) if a given mediator, M, has a meaningful effect on the outcome in its own right and (b) if the program has a meaningful influence on M in its own right. This information is apparent when I analyze the individual links in the T→M→Y chain. I tend to agree with King (1986) that seldom is a deeper understanding gained by hypothesizing a winner in a race of mediators; and to make such questions even more challenging as evidenced by my discussion of the various methods for determining relative importance, the declaration of a winner often depends on the vantage point (i.e., the method of analysis) one uses. Having said that, situations can arise where you want to order the relative importance of mediators either in terms of their presumed effect on Y or in terms of the extent to which the intervention brings about change in them. This might occur if resources are limited and you can only address a subset of the mediators in your program moving forward or if you must make modeling decisions that carve up a complex model for purposes of piecewise or LISEM based analyses. Analysis of mediator relative importance may help you make these decisions. The approaches discussed above are candidates for use in such scenarios, although their limitations must be kept in mind. I think the healthiest approach to the matter probably is to analyze relative importance from multiple perspectives, perhaps using both omnibus indirect effects as well as dominance analyses and one or more of the other methods I discussed. By doing so, you should be able to get a good handle on issues of relative importance.

WHEN THE NUMBER OF MEDIATORS IS LARGE: DATA REDUCTION

Scenarios occur for some RETs where there are a large number of potential program mediators, so many so that they simply are not amenable to analysis using SEM. In Positive Youth Development (PYD) programs, for example, there are a key set of elements that are seen as critical for a PYD intervention to be effective (Catalano et al., 2004), 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 in an RET and each is a possible mediator of program effects on outcome behaviors. Further complicating matters is that within a given variable category, there often are multiple plausible mediators. For the category emotional competence, which is defined as the ability to identify and respond to feelings and emotional reactions in oneself and others, Catalano et al. (2004) specify five elements that are key (1) knowing one's emotions, (2) managing emotions, (3) motivating oneself, (4) recognizing emotions in others, and (5) handling relationships. This five-component framework is somewhat abstract so that when articulated at a more specific level, we might have even more mediators within the category. The analysis for the entire program can quickly become unmanageable as the number of mediators multiply.

When interventions are mounted, some program designers try to change as many potentially relevant factors as possible in the hopes that a few of them "stick" and have an effect on the outcome. Faced with a small sample size in an RET to evaluate the program and a large number of potential mediators many of which probably are not serving as true active ingredients, we have what is known as the **curse of dimensionality**, i.e., small N and large *k*, where *k* is the number of potentially relevant variables. (Another term used to describe the curse is **high dimensionality**). High dimensionality can wreak statistical havoc, in which case we need to pursue some form of data reduction to make analyses manageable.

Factor Analysis and Principal Components Analysis

One form of data reduction is to use factor analysis, principal components analysis, or to define latent variables that underlie subsets of measures of the mediators so that the subset mediators become indicators of latent constructs. The analysis then focuses on linking the factors, components, or latent variable mediators to the outcomes in ways that the observed measures take on a more subsidiary role. We essentially "reduce" the data to a core set of factors, components or latent variables for purposes of RET analysis.

There are several limitations to such approaches. First, factor analysis and principal components analysis of data are not really forms of data reduction because one still must work with the covariance matrix for the full set of variables in order to apply a factor or component analysis to them in the context of an RET model. The curse of dimensionality remains unless one forms composites based on the factor or principal components analysis outside the focal modeling effort. Second, the strategy shifts the focus away from specific mediators to more abstract latent mediators that may be less substantively compelling and informative. Consider a study by Morrison et al. (1996) who explored the impact of 11 different behavioral beliefs that middle school children have about drinking alcohol on their intent to drink alcohol in the near future. Each belief was rated on a 5 point disagree to agree scale. The goal was to determine which particular behavioral beliefs were most influential in determining drinking intentions. The idea was that an intervention program would then target the four or five most influential beliefs. Some of the beliefs studied by Morrison et al. referred to advantages or positive consequences of drinking and others referred to disadvantages or negative consequences of drinking. Examples include (using the stem "drinking alcohol will…"): make me have fun; make me popular; make me feel more grown up; cause me to get a serious health problem like liver disease; get me in trouble; be hard to stop.

A factor analysis of the beliefs by Morrison et al. yielded a two factor solution; one factor had beliefs about the positive consequences of drinking alcohol loading on it and the other factor had beliefs about the negative consequences of drinking alcohol loading on it. The standardized loadings were equal to about 0.60 for items that "loaded" on a given factor, meaning the correlations between affectively common items were about $.60^2$ or about 0.35. Each separate behavioral belief conveys a specific consequence (e.g., would make me popular; would make me feel more grown up) that potentially can be addressed in an intervention. By contrast, the latent variables represent amorphous constructs of positive and negative affect that give little guidance for program content/design. Indeed, focusing on latent constructs alone implies that belief content per se does not matter; that it is only the affect of the belief that matters.[5]

The magnitude of the factor loadings also indicated that each belief has considerable unique variance relative to the latent construct on which it loads, indeed far more unique variance (about 65%) than common variance (about 35%). By focusing on the latent construct, one essentially ignores the unique explanatory variance of each

[5] I have found this conclusion not to be true in my own research; belief content matters).

individual belief, a strategy that might be ill-advised. Social scientists, in my opinion, are often too quick to factor analyze items (or form components of them) without appreciating how much they are giving up when they do so in the form of unique variance (see Goldberg, 1972, and Revelle et al., 2021, for similar viewpoints). To be sure, I am a fan of working with multi-indicator constructs, but I prefer scenarios where the indictors are interchangeable measures of the same construct, not predominately distinct constructs in their own right per the case of Morrison et al. (1996). Factor analysis, principal components analysis and latent variable formation have their place, but I urge you to think carefully before pulling that trigger.

Choosing Mediators Based on Links in a Mediational Chain

Barring the use of factor or principal components analysis, the question remains how one goes about reducing a large number of plausible mediators to a workable set for RET analysis while also maximizing the likelihood that the selected mediators are truly relevant to the outcome. This problem has been approached from two perspectives. One perspective seeks to reduce the number of plausible mediators by focusing just on the M→Y link of the mediational chain; a mediator is deemed relevant if exploratory analyses support the proposition that it is a meaningful determinant of the outcome. Faced with, say, 50 plausible mediators, I might discover based on preliminary analyses that only 8 of them are strongly associated with the outcome, so I decide to focus RET modeling on these 8 mediators. The second perspective focuses on the strength of *both* the T→M and M→Y links considered simultaneously to choose a subset of mediators to target in RET modeling; if data suggest that either one of the links is "broken," then the mediator is eliminated from consideration for RET modeling.

I generally have two goals when I conduct program evaluations. First, I want to identify what the program is "doing right" so that I can ensure those program efforts continue. Second, I want to determine ways I can improve the program to make it more effective. Central to both goals is identifying potentially changeable mediators that have meaningful M→Y links. Programs are "doing things right" if they address mediators that matter and represent changeable determinants of the outcome. Programs can be improved if they are failing to bring about change in a program-targeted mediator that is a key, changeable determinant of the outcome. Note that in both cases, having mediators with strong links to the outcome is essential. Once I bring such mediators into my formal RET model, I am then able to evaluate the T→M link for them, which is an integral part of all RETs.

I tend to give less emphasis to screening mediators using omnibus indirect effects (T→M→Y) because I then run the risk of inadvertently excluding mediators with strong

M→Y links that the program has not been effective in changing. To be sure, if my primary goal is to gain an understanding of the mechanisms that account for the effect of a distal variable on an outcome, then it does indeed make sense to screen mediators on the omnibus T→M→Y links. But if my goal is to both understand the mechanisms and to maximize insights for program improvement, then individual link analysis is informative. Again, in the ideal world I would much rather perform detailed RET modeling on *all* plausible mediators, but the curse of dimensionality can dictate otherwise. To gain control over the curse, I am sometimes forced to reduce the set of mediators that my RET modeling can focus on.

Choosing Mediators Based on the M→Y Link

For mediator reduction based on M→Y links, the subset methods I discussed earlier (forward regression, backward regression, stepwise regression, lasso regression, all-possible regressions) may seem like viable approaches for doing so but they typically are not viable without large N, i.e., they suffer from the curse of dimensionality. This is because they require analysis of the full covariance matrix among all mediators and their covariates, which is implausible in high dimensional scenarios. Predictor screening methods are popular in the data mining and machine learning literatures but most of these methods prioritize prediction over explanation. As such, their algorithms may not be helpful for cases where causation is a priority. I prefer instead methods that allow me to exercise theoretical/substantive judgment during the screening process rather than relying strictly on prediction per se, so I concentrate my discussion on such methods here.

Some data mining and machine learning methods that focus on variable selection draw upon techniques used in the social sciences but under the guise of different labels. For example, one method, called **sure independence screening** (SIS; Fan & Lv, 2008) selects predictors for consideration if the predictor is larger than a pre-determined squared correlation cut-off value, such as $r^2 = 0.05 = 5\%$ explained variance, or the top 10% of predictors with the highest squared r. After this first stage of variable screening or even during the screening stage itself, causal theory is invoked to reduce the number of target mediators to include only those that are likely causally relevant. As well, organizational schemes can be imposed on the data to assist the screening process. For example, for the positive youth development RET described above, I might apply the SIS method separately to variables within each of the 15 different categories and then select a subset of variables from each category that are the strongest predictors of Y. If a category fails to have any mediators nominated by the SIS criteria, then all variables within that category are omitted.

The SIS method can be suboptimal because it ignores correlations among the various mediators and because of the problem of confounds biasing correlation magnitudes. One adaptation of the method is to first reduce the number of variables using SIS and then use dominance analysis as a second stage of variable reduction applied to those mediators that survive the first iteration of screening. Covariates to control for confounds can be included in the dominance analysis as illustrated in the video for dominance analysis on my webpage. The best predictors that survive both the first and second screening steps are then used in the final model. The dominance analysis is spared the curse of dimensionality because of the first step of variable reduction.

An example of an interesting variable screening/selection method that adapts the above ideas but without the SIS screen has been proposed by Cai, Tsay and Chen (2009). The approach uses linear models in an all-possible regressions framework. Suppose I have, say, 56 plausible mediators and I want to choose the linear model that best predicts the outcome. There are 2^{56}-1 possible linear models to evaluate in an all possible regressions context, which is unworkable. Cai et al. propose to first divide the predictor pool into non-overlapping smaller groups, say, eight groups of 7 predictors each. Assignment of predictors to a given group can be either random or theory based. For each set of 7 predictors, there are 2^{7}-1 or 127 different linear equations one can analyze in an all-possible regressions framework, which is workable. For a given predictor set, you apply a penalty-based regression method to all possible regressions within a set to isolate the "best" (most predictive) model within each set. For example, you might use OLS regression in which fit is evaluated using a BIC criterion.

Next, the "winning" equation from each of the 8 sets of predictors are identified. You can use the program for all possible regressions on my website to accomplish such identification. The predictors in each "winning" equation are then set aside into their own category. This smaller predictor pool is then re-grouped into subgroups, just like in the first step. For example, the new pool might have 32 predictors, which are then divided into 4 sets of 8 predictors each. The best fitting model within each of these reformulated groups is identified using the all possible regressions strategy for each of the four sets and the predictors from each of the "winning" equations across the 4 sets are set aside into their own predictor pool. During this step, it often is possible to bring theory and substantive considerations to bear when choosing which variables to merge. The process then repeats itself iteratively until no new variables are added to the grand winning predictor pool. At the final step, the grand winning predictor pool is analyzed to using the all possible regressions strategy to identify the best fitting model within it, again with an appropriate penalty function for model complexity. For a concrete example of the approach, see the document on all possible regressions on my webpage on the Resources tab under Chapter 17. Cai at al. (2009) describe variations of the strategy and show it outperforms a number of more computationally demanding strategies widely used in data mining contexts. There exist other strategies for variable screening in data mining (e.g., the one covariate at a time method; decision trees), but I do not review them here; see Cai et al., (2009), Chudek et al., (2018). and Brick, Koffer, Gerstorf, and Ram (2018). A popular method in the machine learning literature for assessing relative variable importance is SHAP (Shapley Additive exPlanation) analysis, but the method has been found to have non-trivial limitations (Huang & Marques-Silva, 2024).

Choosing Mediators Based on the Omnibus T→M→Y Link

Although I am not a fan of approaches that rely on the omnibus T→M→Y links to screen possible mediators, such a strategy has been suggested by van Kesteren and Oberski (2019). In this approach, one calculates the omnibus indirect effect for each plausible mediator one mediator at a time in its own separate model. By analyzing the data one mediator at a time, you avoid the problem of high dimensionality that would result from including a large number of mediators in same model. Note that each omnibus effect has the same metric, namely, the mean Y difference between the treatment and control groups). We therefore can order the different omnibus mediational effects in terms of their magnitude across analyses. For the 56 mediator scenario described earlier, I would conduct 56 analyses, one for each mediator. Figure 17.4 shows the influence diagram I would use for a given analysis. The model includes measured confounds to control for bias. The omnibus mediational effect for a given mediator is reflected by the product of paths *a* and *b*.[6] Path *c* reflects the impact of the treatment on Y of all omitted mediators on the outcome other than the target mediator being evaluated. It generally is not of interest in the selection process but it is important to include to control for correlations between the target mediator and the omitted mediators.

[6] If the outcome is binary, the omnibus indirect effect for a given mediator can be indexed using proportion or probability differences between the treatment and control conditions through the mediator using the causal mediation framework or a linear probability model per Chapter 12.

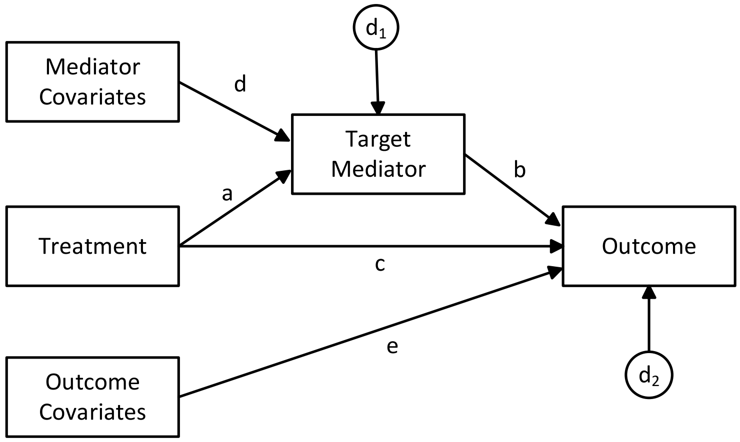

FIGURE 17.4. One mediator at a time approach

Table 17.5 presents the Mplus code I would use for the above mediator screening strategy but with some programming shortcuts to make the code more efficient when conducting the multiple separate runs. The example has 56 plausible mediators (med1 to med56), one outcome (behave1), one covariate for the outcome (bbehave1), and one (baseline) covariate per mediator (covm1-covm56). I would run the code 56 times, with the only changes being in Line 4, which I explain shortly.

Table 17.5: Mplus Code for One Mediator at a Time Analysis

```
1. TITLE: ANALYSIS OF ONE MEDIATOR AT A TIME ;
2. DATA: FILE IS c:\mplus\ret\chap17Ma.txt ;
3. DEFINE:
4. y=behave1 ; m=med1; t=treat ; cm1=covm1 ; cy1=bbehave1 ;
5. VARIABLE:
6. NAMES ARE
7. behave1 behave2 med1-med56 covm1-covm56 bbehave1 bbehave2 treat ;
8. USEVARIABLES ARE
9. y m cm1 cy1 t ;
10. MISSING ARE ALL(-9999) ;
11. ANALYSIS:
12. ESTIMATOR=MLR ;
13. MODEL:
14. y on m t cy1 ;
15. m on t cm1 ;
16. MODEL INDIRECT:
17. y IND t ;
18. OUTPUT:
19. SAMP STDYX MOD(All 4) ; !RESIDUAL CINTERVAL TECH4 ;
```

On Line 4, I use the `DEFINE` statement to isolate the outcome, mediator, and treatment variable I want to focus the analysis on. This is the only line I need to edit for the successive runs of the program as I work through the 56 plausible mediators and their covariates. On Line 7, I use Mplus hyphens for naming the 56 mediators and their covariates. The Mplus output will print out an estimate of the omnibus indirect effect (path *a* times path *b*) for the target mediator and its estimated standard error and p value in the output section called `TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS`. These effects and their margins of error are then summarized across mediators in a table with 56 rows, one per mediator. The mediators with the largest indirect effects become candidates for ultimate selection into my final model.

This strategy does not work if there are causal relationships among the mediators because it would then ignore indirect effects of a mediator through another mediator. If you believe such effects exist and if the sample size permits, you can formally bring the additional mediators into the target mediator model so that the effects of the target mediator on the outcome through the other mediators are taken into account vis-à-vis SEM modeling. The approach also assumes the absence of interaction effects among the mediators but these also can be incorporated into the model if need be via SEM. Finally, the approach assumes the absence of correlated disturbances, but these also can be added if necessary if one uses SEM.

Another way of thinking about this approach is that you are dividing a model with the 56 mediators into smaller subsets of more manageable SEM models (usually containing one mediator) and then based on analyses of these submodels, you make decisions about what mediators to include in the final model.

Serang et al., (2017) suggest a similar approach to the above but based on regularized estimation of SEM models that shrink small path coefficients to zero, much like lasso regression. Regularization occurs on the T→M paths independent of the M→Y paths and vice versa. van Kesteren and Oberski (2019) are critical of this approach on the grounds it fails to address both T→M and M→Y simultaneously, which is necessary if the focus is on omnibus indirect effects (see also Jacobucci, Brandmaier & Kievit, 2018). van Kesteren and Oberski (2019) introduce a modification called the **coordinate-wise mediation filter** that more directly works with omnibus indirect effects. An initial R version of their approach is available on the github website, but it is not yet widely available. It still needs improvement relative to matters of covariate inclusion, convergence criteria, computer speed, cutoff selection criteria, and error rates. Also, evidence for the method's superiority to the one-mediator-at-a-time approach is not strong.

Inferential Tests after Exploratory Mediator Analysis

When using exploratory methods to identify mediators, one must be careful with significance tests surrounding those mediators when used in one's final RET model. Many methodologists argue that p values in such cases are problematic. According to this argument, in traditional model tests that do not use preliminary exploratory tests, the statistical theory underlying the generation of a p value, confidence intervals, and standard errors applies to the case where a model/null hypothesis is a priori specified and then fit to the data, much like we do when we conduct a multiple regression analysis and decide a priori what predictors to include in the regression equation. The statistical theory works with a sampling distribution that assumes random sampling from the population. The theory was not derived under the presumption that an imperfect "screening" step is first performed, a step that brings with it its own assumptions and weaknesses. By including such a "screening" step, one essentially alters the sampling distribution of the model parameters in unknown ways, which, in turn, can then undermine the p values and confidence intervals.

Critics of preliminary screening steps also worry about overfitting one's data in which we interpret a post hoc chance effect or an effect that is sample specific as meaningful and generalizable to the population when, in fact, it mainly reflects random noise in the analytic process. To combat this latter problem, methodologists often suggest the use of cross validation strategies. One variant of cross validation is to randomly divide one's sample data in half, then conduct the exploratory screening on each half independent of the other half. Only mediators that are identified as being relevant in both samples are screened into the model for use in the final RET analyses. A disadvantage of this method is that the exploratory analyses use sample sizes that are 50% smaller than the full sample and, hence, are subject to more sampling error than if the full sample is used for the exploratory analyses. If your sample size is small to begin with, this can undermine the exploratory mediator analyses because results are more likely to be unstable across the different randomly defined samples (also known as **folds** in the data mining literature).

In machine learning and data mining, you will see frequent reference to variants of cross validation and the splitting of data into **training data**, in which the machine learns the optimal prediction strategy, and **test data**, in which one tests how accurate the machine's predictions are when applied to new data. The term **feature selection** is often used to refer to predictor selection during the training phase. Refinement of the prediction equation during the training phase is often called **tuning**. My discussion here is focused on cross validation as used during the training phase for purposes of feature inclusion or elimination. I mention this terminology here because it is easy to get confused over the different jargon used and foci in the machine learning and data mining literatures.

CONCLUDING COMMENTS

In this chapter, I addressed two topics (1) given a set of mediators, how does one order their relative importance, and (2) if one has a large number of plausible mediators, how does one apply data reduction strategies to reduce them to a workable set for more detailed modeling. The concept of "relative importance" is somewhat amorphous and there are many ways of defining it. I personally believe the best way to develop standards for asserting relative importance is to adapt the methods described in Chapter 10 for determining meaningfulness standards, but the field has done so too infrequently. Rather, the tendency is to focus on context free statistical criteria grounded in linear modeling.

In my opinion, you need to approach this enterprise with caution and realize that the concept of predictor importance in RETs goes well beyond simple statistics (King, 1986). None of the methods I have discussed in this chapter are problem free. As you zero in on the use of particular method, it is important you keep in mind its strengths and weaknesses and draw conclusions accordingly.

My emphasis in this chapter has been on the analysis of continuous outcomes. For a discussion of extensions to binary, ordinal, count and nominal outcomes, see the Resources tab of my webpage under the current chapter.