12

Mediation Analysis with Binary Outcomes

In theory, there is no difference between theory and practice. But, in practice, there is.

- ALBERT EINSTEIN

INTRODUCTION

This chapter builds on Chapter 11 by extending mediation analysis to binary outcomes and binary mediators. It develops key concepts for later chapters on ordinal outcomes, nominal outcomes, and count outcomes. I assume you are familiar with all previous chapters, but most importantly, Chapters 5, 10, and 11. I begin by providing core background material on binary regression followed by a description of the numerical example I use throughout the chapter. I then apply limited information estimation structural equation modeling (LISEM) to the example, first using the modified linear probability model and then using probit modeling. I then conduct analyses using full information structural equation modeling (FISEM), including probit based estimation, Bayesian modeling, and finally a FISEM modified linear probability model. Causal mediation frameworks that emphasize omnibus mediation tests are of lower priority in this book, but I consider them towards the end of the chapter and on my website. The Chapter is long and not meant to be processed in a single sitting.

As you will see, the different methods of analysis converge on the same conclusions for the example RET. This will not always be the case, but often it is. When analyzing your data, you may find one of the approaches or a variant of it is better suited to your purposes. For example, if you have a small sample size, you might want to use a limited information estimation framework that employs penalized likelihood (Firth) regression (a program available on my website) for the model equations with a binary outcome.[1] My goal is to broaden your statistical toolbox with a range of tools for analyzing binary outcomes so you can bring multiple perspectives to your data-based conclusions.

MEDIATION ANALYSIS WITH BINARY OUTCOMES

I first discuss some of the challenges of mediation analysis with binary outcomes. I revisit the concept of average marginal effects from Chapter 5 and then consider the role of covariates in logistic and probit regression. I discuss several underappreciated properties of probabilities in logit/probit models, including non-collapsibility. Be patient as I develop these background concepts. I will tie it all together in time.

[1] To watch a video that illustrates LISEM in a limited information context, click the link to the video associated with the analytic program HC2-BRL cluster regression on my webpage.

Average Marginal Effects for Binary Outcomes

Many analysts seek to document outcome probability differences between treatment and control groups and how outcome probabilities vary as a function of continuous mediators. There are multiple ways of doing so but one popular index (at least in some disciplines) is the average marginal effect (AME), which I introduced in Chapter 5 using a binary predictor as an example. Because I make use of AMEs in this chapter, I develop them in more detail here, considering first using AMEs for analyzing the relationship between a continuous mediator and a binary outcome. I then consider general properties of AMEs.

Figure 12.1 presents the probability of an outcome, Y, as a function of people's scores on a mediator, M, with M values ranging from 0 to 10. The figure shows both a probit and a logit function with positive associations between M and Y. Both relationships are non-linear and have different change dynamics when compared to linear functions. For linear functions, if an intervention increases the mediator by 1.0 unit, the proportion of people who perform the outcome should increase by the same amount no matter what people's scores are on the mediator. If the coefficient for the mediator in a linear model is 0.10, then this means that for every one unit increase in the mediator, the proportion of people performing the outcome should increase by 0.10. By contrast, for logit and probit functions, an increase of one unit on M has different effects on the probability of Y depending on where on the M dimension the increase occurs. For people with scores near 0, 1 or 2 in Figure 12.1, a one unit increase has little effect on the probability of Y. This also is true for people whose M scores are near 7, 8, or 9. The most dramatic changes occur for people whose M scores are near 5. This point is called the **inflection point** or where on the curve the maximum change occurs for a change in M.

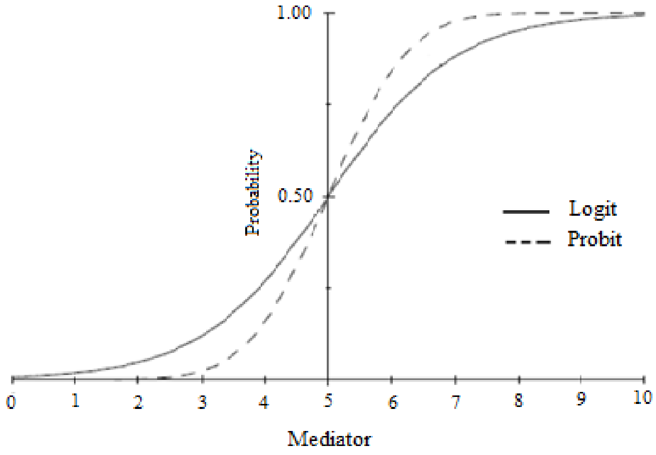

FIGURE 12.1. Logit and Probit Probability Functions in Binary Regression

Independent of whether the function is logit, probit, or linear in nature, program evaluators often want to know if a program changes a mediator, say, by one unit, how much will the proportion of people who perform the outcome increase or decrease *in the overall target population as a whole* irrespective of where on the M dimension they are. The AME helps to inform us of this property.

Here is an intuitive way of thinking about AMEs in the case of a continuous predictor and a binary outcome. Suppose the function that best represents the relationship between M and Y is a probit function. I can perform a probit analysis that regresses the outcome Y onto M, for which I might obtain the equation

$$\text{Probit}(Y_i) = -2.0 + .40\, M_i \qquad [12.1]$$

where the subscript *i* refers to individual *i*.[2] For each individual, I calculate a predicted probit score, $\hat{Y}_i$, by multiplying his or her M_i score by 0.40 and then adding to it -2.0, the intercept. I then convert this probit score to a predicted probability of having a 1 on Y for the individual by treating the predicted value of $\hat{Y}_i$ as a score in a cumulative standard normal distribution and transforming it accordingly, i.e.

$$p_i = \Phi(\hat{Y}_i) \qquad [12.2]$$

where p_i is the predicted probability for individual *i* given his or her value of M and Φ is the probit transformation function. This is the standard transformation for a probit to a probability. For example, if $\hat{Y}_i = 1.96$, p_i equals 0.975; if $\hat{Y}_i = 1.65$, then $p_i = 0.95$; if $\hat{Y}_i = 0$, then $p_i = 0.50$. I refer to this individual based probability as $p1_i$. Most binary regression software, including Mplus, allows users to convert probits to predicted probabilities.

To calculate the AME, after executing the above step for each individual, I add 1.0 to each person's M score and then use the original equation (Equation 12.1) to calculate a new predicted probit score for each individual by multiplying this incremented M_i score by 0.40 and adding -2.0 to it. I convert this new value to a probability using Equation 12.2, yielding $p2_i$ for each individual. The marginal effect for a given individual is defined as $p2_i - p1_i$, which is the change in the probability of performing the outcome behavior for the individual when his or her score shifts from M to M+1. The average of these differences across all individuals is the AME and it reflects the average change in the probability of Y given a one unit increase in M. Suppose I find the AME equals 0.098. Stated in percentage terms, if I increase M by one unit, the percentage of people engaging in or experiencing Y in the total sample should increase 9.8%, after taking into account the non-linear dynamics of probit.

[2] For a review of probit regression, see Chapter 5.

This description of AMEs does not map exactly onto the way that AMEs for continuous predictors are conceptualized and calculated in practice. Statisticians instead use derivatives and the concept of instantaneous rates of change (see Chapter 6 and Williams, 2020, and Wooldridge, 2010, for elaboration). There are several reasons why statisticians use the instantaneous change approach, one of which is that individuals with a score of 10 cannot have a score of 11 when we increase M by one unit because 10 is M's upper bound. Wooldridge (2010) describes a host of other reasons but I do not want to get sidetracked on them here. Sometimes the mapping of the instantaneous AME onto a unit change interpretation works fine but other times it falls apart, so some caution is required when interpreting an AME in terms of unit changes. However, the above represents an informal way of thinking about AMEs that makes intuitive sense to many researchers. I show in Appendix A how you can calculate AMEs by hand and in Mplus in ways that are reasonably faithful to the instantaneous change concept.

Inclusion of Covariates when Calculating AMEs

AMEs usually are calculated when there are multiple predictors in the equation, some of which might be other mediators and some of which might be covariates intended to adjust for confounding or to increase statistical power. For example, suppose I have three mediators and two covariates yielding the following equation:

$$\text{Probit(Y)} = a + b_1 \text{ M1} + b_2 \text{ M2} + b_3 \text{ M3} + b_4 \text{ C1} + b_5 \text{ C2} \qquad [12.3]$$

To calculate the AME for M1, I use the same method described above but I do not change any of the individuals' scores on M2, M3, C1, or C2 when calculating a given probit value for an individual. Rather, I calculate $\hat{Y}_i$ using the intercept and coefficients in Equation 12.3 first with M1 left untouched to obtain $p1_i$ and then with M1 incremented by 1 for everyone to obtain $p2_i$ but leaving all the other predictors at their original values in both cases.[3] In this sense, all of the other predictors take on the same values when I calculate $p1_i$ as they do when I calculate $p2_i$ so that the sole source of the difference between $p1_i$ and $p2_i$ is the increment in M1. However, keep in mind that the b_1 coefficient for M1 in Equation 12.3 has been calculated/defined so as to take into account the confounding influence of the other predictors per standard (probit) regression algorithms. Note also that estimates in probability differences over the distribution of observed covariate values is dependent on the covariate distributions of participants in the trial. For statistical details on the calculation of AMEs with covariates, see Wooldridge (2010). I return to the concept of AMEs shortly as they often come in handy when analyzing binary outcomes. Tuck this material away for now.

[3] Technically, I increment M1 not by one but a different constant – see Appendix A

Covariates in RETs with Binary Outcomes

As discussed in prior chapters, we often include covariates in RETs to (a) adjust for sampling imbalance, (b) increase statistical power and reduce margins of error, and/or (c) control for confounds that create biased estimates of causal coefficients. If the sample size is not small and randomization is effective, the use of covariates to adjust for sample imbalance is of lower priority when evaluating the effect of an intervention on an outcome. I now discuss the latter two uses of covariates in binary regression contexts.

Statistical Power and Precision

In social science research, the most common approach to increasing statistical power is to increase sample size. As discussed in prior chapters, an alternative strategy is to use covariates that reduce prediction error or "noise in the system" by increasing the amount of explained variance in the outcome. If such covariates are relatively uncorrelated with other predictors, including them often will increase power and reduce the magnitude of margins of errors. To illustrate this, I created simulated population data for a normally distributed continuous outcome, Y, that ranged from -5 to +5 (SD = 1) with a true mean difference between treatment and control groups of 0.50. I also created two covariates that were uncorrelated with each other and also uncorrelated with the treatment condition per the use of random assignment. Each covariate was correlated 0.35 with the outcome. I randomly selected 350 cases from the population, so some sampling error is present relative to these population values in what I am about to present. Table 12.1 presents three different analyses of the data. The first row regresses Y onto just the treatment dummy variable (0 = control, 1 = treatment), the second row adds the first covariate to the model and the third row has both covariates added. I used OLS linear regression to derive the coefficient values for the predictor(s). Note that the estimated effect of the treatment on the continuous Y is about the same in all three models, which should be the case given its zero or near zero correlation with the covariates. It is 0.50 give or take sampling error. The coefficient critical ratios for the intervention, T, increase as I move down the table rows, indicating greater statistical power as each covariate is added. This is because I am reducing the disturbance variance in the model. The 95% confidence intervals (margins of error), also decrease with each added covariate. These patterns all evolve from classic regression principles.

Table 12.1: Example of Increased Power and Reducing MOE by Adding Covariates

Model	Treatment Coefficient	Critical Ratio
T only	0.509 ±0.208	4.992
T + C1	0.494 ±0.185	5.123
T + C1 + C2	0.507 ±0.175	5.819

Let’s see what happens when I repeat this exercise but using logistic regression as applied to a binary outcome. I created population data that has a dichotomous Y with the population proportion of people who score 1 on Y in the intervention condition set to 0.60 and 0.35 for the controls, a difference of 0.25. The two covariates have the same properties as before and I randomly select a sample of 350 cases from the population. Table 12.2 presents the results for the same three models in Table 12.1.

Table 12.2: Logistic Regression Example with Covariates

Model	Treatment Coefficient	Critical Ratio	Odds Ratio	Ave Marginal Effect
T only	0.988 ±0.442	4.466	2.687	0.242 ±0.104
T+C1	1.066 ±0.462	4.604	2.903	0.242 ±0.100
T+C1+C2	1.271±0.510	4.985	3.565	0.246 ±0.093

Note that in contrast to Table 12.1, the logistic coefficients for the treatment and their associated odds ratios are not the same across the three models. Instead, the coefficients and odds ratios both get larger as the covariates are added. These changes in coefficient magnitude are *not* due to the control of confounds – none of the covariates are correlated with the other predictors so, by definition, they are not confounds. It turns out the addition of the covariates systematically affects the scaling factor in the latent variable representation of the logistic model per my discussion in Chapter 5 which has the consequence of inflating the estimated treatment effects. Specifically, the treatment effects are made more positive if the logistic coefficient is positive and they are made more negative if the coefficient is negative. Like Table 12.1, the critical ratios increase with covariate inclusion, suggesting there is greater statistical power as a result of including the covariates. However, the source of this increased power is primarily the concomitant inflated treatment effects that occur due to the change in the underlying scaling factor. Note also that the margins of errors tend to increase not decrease, again because of the changing metric of the underlying y*. These results illustrate the point of Norton et al. (2018) that I discussed in the Appendix of Chapter 5, namely that statements about the magnitude of odds ratios are "covariate bound" and require caution when generalizing to contexts that do not use covariates at all or that use different covariates.

It is because of this inflation property with the addition of uncorrelated or weakly correlated covariates that some researchers eschew the analysis of logistic/probit coefficients and odds ratios and prefer instead to focus on average marginal effects for binary regression. The last column of Table 12.2 presents the AMEs for the treatment condition in each logit model. The AMEs in this case are the estimated differences in the proportion of people who score Y = 1 in the treatment condition minus the corresponding proportion in the control condition. The AMEs are similar in all three models and map well onto the true population difference in proportions of 0.60 – 0.35 = 0.25, except for sampling error. The margins of error decrease with greater explained variance, much like traditional OLS regression. In short, the AMEs are better behaved than the logistic coefficients and odds ratios and easier to interpret, which is why many scientists prefer them (Mood, 2017).

In sum, when covariates are related to the outcome but relatively uncorrelated with other predictors, their inclusion in traditional regression usually increases statistical power, reduces the magnitude of margins of error, and does not alter estimates of the overall effect of the treatment. These properties do not hold for logit and probit regression. Usually, the inclusion of covariates will increase statistical power in logistic and probit regression but this is not always the case; the dynamics underlying the increase are more complex than for traditional OLS regression.

Parenthetically, the change in the scaling factor and the underlying y* for the latent propensity model associated with logit and probit regression also produces a phenomenon known as non-collapsibility. **Noncollapsibility** refers to the case where the measure of association (e.g., an odds ratio) conditioned on a covariate does not equal the overall association collapsed across the covariate, a property that can complicate interpretations. I explain this phenomenon in more detail later in this chapter using a concrete example.

Control of Confounds

As discussed in previous chapters, another reason to include covariates in regression models is to control for confounds that, left uncontrolled, can bias estimates of the magnitude of causal coefficients. In logistic, probit, and traditional regression, the nature of the relationship between the covariates and the other predictors determine whether coefficients for the other predictors increase or decrease when these covariates are statistically controlled. When predictive covariates are moderately to highly correlated with other predictors, then logistic coefficients, odds ratios, probit coefficients, and coefficients from the modified linear probability model (MLPM) all can change in either direction depending on the operative confounding dynamics. In this sense, covariates that represent true confounders of causal dynamics are important to include in all forms of analysis. Otherwise, our causal estimates are biased. Confound control is important for binary regression despite the challenges introduced by doing so as elaborated above.

Odds Ratios versus Probabilities

Yet another objection to odds ratios has been that they can be misleading as indices of effect size, independent of the above matters. I discussed this issue in Chapter 10 and encourage you to review that material. For example, an odds ratio of 2.0 could result from (a) a percentage for group 1 scoring a one on Y equaling 1% versus group 2 having a percentage of 0.5% (a percent difference of 0.5%), (b) a percentage for group 1 of 50% versus group 2 having a percentage of 33% (a percent difference of 27%), or (c) a percentage for group 1 of 80% versus group 2 having a percentage of 67% (a percent difference of 13%), making the ratio per se ambiguous. As you will see later in this chapter, my bias is to stay away from odds ratios and to work with probabilities.

Treatment Probability Differences and Covariates

There are additional properties of logit and probit regression for RET analyses worth noting. I illustrate these dynamics using logistic regression but they also apply to probit regression. Suppose I predict a binary outcome from a treatment dummy variable, T (0 = control, 1 = treatment), and a baseline continuous covariate, C, in which C is normally distributed with a mean of 0 and a standard deviation of 1. The covariate is uncorrelated with the treatment condition, as one would expect given random assignment. I created three scenarios for the population data to make my points. In one case, the proportion of cases (also called the **event rate**) for a score of 1 on Y is 0.10 in the control condition when the covariate equals its mean, namely zero; the second case is where this proportion is 0.30, and the third case is where the event rate equals 0.50. One possible function that might characterize the treatment minus control proportion difference is that it is the same no matter what the value of the covariate. For example, if I isolate people who have a score of -1 on C (one standard deviation below the mean of C) the proportion difference between the treatment and control might be 0.20. The same is true for people with a score of 0 on C as it is for people with a score of 1 on C. Such a pattern maps onto the function represented by a modified linear probability model and is a “main effect” model.

The dynamic is different for a logistic model where the logit function operates. Suppose the population logistic coefficients for T and the covariate, C, both equal 1.00. These coefficients translate into odds ratios for each variable of 2.72, i.e., the exponent of 1.00 is 2.72. It is interesting to explore how the risk *difference* for Y between the treatment versus control groups varies as a function of C for the logistic case in the above example. Table 12.3 shows the predicted probabilities from a logit model for the treatment and control conditions as values on C vary from -2 to +2 in increments of 1.0 as well as the relative risks (the probability for the intervention condition divided by the probability for the control condition) and odds ratios at each level of C. I present the results separately for each of the event rate scenarios.

Table 12.3: Treatment – Control Risk Difference

	Intervention Probability	Control Probability	Risk Difference	Relative Risk	Odds Ratio
Model 1 (ER = 0.10)					
C = -2	.039	.015	.024	2.651	2.718
C = -1	.100	.039	.061	2.546	2.718
C = 0	.232	.100	.132	2.320	2.718
C = 1	.451	.232	.219	1.944	2.718
C = 2	.691	.451	.240	1.532	2.718
Model 2 (ER = 0.30)					
C = -2	.136	.055	.081	2.484	2.718
C = -1	.300	.136	.164	2.203	2.718
C = 0	.538	.300	.238	1.794	2.718
C = 1	.760	.538	.222	1.412	2.718
C = 2	.896	.760	.136	1.179	2.718
Model 3 (ER = 0.50)					
C = -2	.269	.119	.150	2.256	2.718
C = -1	.500	.269	.231	1.859	2.718
C = 0	.731	.500	.231	1.462	2.718
C = 1	.881	.731	.150	1.205	2.718
C = 2	.953	.881	.072	1.081	2.718

Note that the population odds ratios for the effect of T on the outcome probability remain constant across the different values of C. This is a well-known and convenient statistical property of logistic regression that many researchers find attractive. However, if we examine the column of risk differences for T (i.e., the probability of the outcome for the intervention group minus the probability for the control group), we see that the values in it vary as the value of C varies. Stated another way, although the logistic model is framed in terms of "main effects" or "direct effects" of T and C on Y, the model actually implies an interaction between T and C as influencers of the *probability* of Y. If this was not the case, all of the risk differences would be the same in the risk difference column as we move across values of C, per the modified linear probability model. Furthermore, the nature of the "interaction" is complex depending on the event rate for Y, with none of interactions easily captured by classic product term representations of interactions. Are we truly able to specify *a priori* a viable substantive theory that predicts these risk difference patterns and how they shift as a function of event rates and values of C? Do we feel confident in the universality of such patterns to the point that we should assume by default they operate by always analyzing binary outcomes using logistic (or probit) regression? In my view, theory and data should dictate the function we choose to model risk difference patterns, not the convenience of a model that happens to produce equal odds ratios across C values when C is represented as a main effect.

It is, in part, because of such risk difference dynamics that Pischke (2012) refers to logit and probit models as "red herrings" and argues instead for using linear functions as an initial referent when modeling data and then accommodating deviations from linearity as theory and data dictate. The spirit of Pischke's argument is to (a) focus on probability/risk differences because they are intuitive and straightforward, (b) adopt an initial theoretical frame of constant risk difference across values of predictors/covariates, and then (c) let substantive considerations and data rather than the convenience of a statistical theory drive modifications to these initial orientations. I sympathize with Pischke's orientations.[4] To be sure, I see nothing wrong with applying logit or probit models if they adequately approximate the true operating dynamics at the level of probabilities and proportions. It is just that, in my opinion, there are substantive implications of using such models that may be theoretically questionable in certain contexts. The same argument, of course, can be made for the MLPM but to me, the MLPM often makes more sense as a starting point.

[4] For an example of how a logistic regression with two "main effect" continuous predictors implies an interaction between the predictors for outcome probabilities, watch the video on surface plots on my webpage. The effect is striking.

Non-Collapsibility Revisited

I now return to the concept of non-collapsibility to further justify my focus in this chapter on probabilities and probability differences. The Food and Drug Administration (FDA) recently published a document on the use of covariates in randomized trials that emphasizes the potential problem of non-collapsibility with covariate control (FDA, 2023). I adapt an example from their report to make my points. Suppose I conduct a two arm RCT (intervention versus control) to convince people to vote in favor of a referendum on mandatory vaccinations for school attendance. The outcome is scored 1 = success, person voted for the referendum and 0 = failure, person did not vote for the referendum. Half the population for this study is male and half is female. Here is a table that summarizes the results of an RCT for the intervention as a function of biological sex as well as collapsing across biological sex (the row labeled "Combined"):

Biological Sex	Intervention Success Rate	Control Success Rate	Percent Difference	Odds Ratio
Female	80.0%	33.3%	46.7%	8.0
Male	25.0%	4.0%	21.0%	8.0
Combined	52.5%	18.7%	33.8%	4.8

There are notable features of these results. First, the odds ratio for males and for females are identical, 8.0. Given this result, one would think that if I calculated the odds ratio for the combined data, the odds ratio also would equal 8.0 given that half the population is male and half is female. But note that this is not the case; the combined group odds ratio is [0.525/(1-0.525)]/[0.187/(1-0.187)] = 4.8. This is why the odds ratios are said to be non-collapsible. Note, by contrast, the average of the two percent differences (46.7% +21.0%)/2 = 33.8%) does, in fact, equal the average percent difference for the two groups combined. The results for the percent differences are collapsible, which is a desirable statistical property and makes causal statements more straightforward. If I conducted a logistic regression for the above data and included gender as a covariate, the odds ratio for the intervention effect would be 8.0. This value is indeed a valid estimate of the odds ratio for the two subgoups. However, if I am interested in the odds ratio for the population as a whole collapsing across biological sex and I conduct a logistic regression omitting gender as a covariate, the odds ratio for the treatment effect would be 4.8, despite the fact that biological sex does not act as a confound. If one study includes biological sex as a covariate when evaluating the treatment and another study does not, the results of the odds ratios in the studies cannot necessarily be meaningfully compared because they are focused on different estimands. This makes it difficult to build cumulative bodies of knowledge across studies or across multiple equations because conclusions are covariate dependent even when no confounding is present.

Second, note that the percent difference reflecting intervention effectiveness is much larger for females than it is for males. For females, the intervention minus control percent is 46.7%. For males, the differences is 21.0%. This suggests the presence of an interaction or moderator effect. Despite this, the odds ratios for males and females are equal, both being 8.0. The moderating effect of biological sex on treatment effectiveness is masked when intervention effectiveness is framed using odds ratios as opposed to percent differences. If I use relative risks, the interaction is characterized *opposite* to what the probabilities suggest; for males the relative risk is .25/.04 = 6.25 but for females it is 0.80/0.33 = 2.40, making it appear the intervention is more effective for males than females. Such phenomena as well as those discussed earlier lead me to prefer probabilities and probability differences to odds, odds ratios and relative risks when characterizing effects in RETs with binary outcomes. Others feel differently.

BROADER PERSPECTIVES ON MODELING BINARY OUTCOMES

As discussed in Chapter 5, for research with binary outcomes, some investigators treat the outcome as a crude indicator of an underlying continuous variable, y*, that represents the propensity to experience or engage in the outcome. Theoretical interest is with making causal statements about this underlying continuous variable, but one is stuck with a crude binary indicator of it. In such cases, one typically wants to know the overall effect of the program on y*, the effects of mediators on y*, and the effects of the program on those mediators. This is the heart of the latent propensity model underlying logistic and probit regression. A major challenge in such cases is dealing with the arbitrariness of the metric of y* and the unverifiable assumptions one must make about the parameters of the propensity model to make inferences about it (see Chapter 5 and Kuha & Mills, 2018). To me, there is a simple solution to these challenges. If you are truly interested in y* dynamics, then instead of using a crude binary indicator of y*, use a more sophisticated strategy and obtain a measure of y* that is a continuous or many-valued indicator of the underlying latent propensity; then analyze the data using the methods outlined in Chapter 11 for continuous outcomes. By doing so, most of the challenges of conducting mediational analyses with binary outcomes vanish because the binary outcomes vanish.

Having said that, some interventionists are not interested in a latent propensity to experience or engage in an outcome but are interested in the dichotomous outcome in its own right. I might want to know, for example, whether a program increases the number of people obtaining a flu shot, not their propensity to obtain a flu shot. I might want to know whether a program increases whether people vote for a referendum, not their propensity to do so. In such cases, we seek to understand the effect of the program on outcome probabilities and how program mediators affect those probabilities. One should not, the argument goes, be sidetracked by ambiguous constructs like "a propensity to perform an outcome, y*." Rather, we want to model the outcome probabilities directly.

Yet another possibility is that an evaluator is interested in understanding *both* an underlying propensity as well as how that propensity translates into a dichotomous response. If this is the case, I again recommend obtaining a direct measure of y* as well as a measure of the relevant binary outcome. The traditional latent propensity framework assumes y* completely mediates the effects of its causes on the dichotomous outcome by means of a threshold function. Is this empirically supported by the data given that we would now have a means of evaluating the proposition with both reasonable measures of y* and the dichotomy? Are there group differences in the operative threshold value? What is the nature of the residual distribution for equations predicting y* and are there group differences in these distributions? These questions also can be tested empirically with both types of measures in hand.

The bottom line is that you need to decide what your focus is and orient your psychometric and analytic strategies accordingly.

WORKED EXAMPLE WITH A BINARY OUTCOME

I illustrate approaches to analyzing binary outcomes in this chapter using an example with continuous mediators. The data are available on the resources tab of my website. The example focuses on parental communication with young adolescent, middle school children about reasons not to have sex at this time in their lives. About 60% of parents of middle school youth in the United States have never talked with their child about sex. Research suggests that parents often are reluctant to do so because they do not feel they have enough knowledge about sex and birth control to adequately discuss the topic. Parents also tend to feel that such discussions will be embarrassing for both them and their child. My example focuses on a program aimed at parents to encourage them to discuss issues surrounding not engaging in sex at this time in their lives by addressing three factors, (1) educating parents about the advantages of engaging in such conversations, (2) providing parents with the knowledge they feel they need to have effective conversations, and (3) teaching parents strategies to reduce embarrassment. The target mediators were measured on multi-item inventories in which each item was rated on 7 point disagree-agree scales: -3 = strongly disagree, -2 = moderately disagree, -1 = slightly disagree, 0 = neither agree nor disagree, 1 = slightly agree, 2 = moderately agree, 3 = strongly agree. Scores were averaged across items; higher scores indicated (1) higher levels of perceived advantages of engaging in the conversations, (2) higher levels of perceived knowledge, and (3) beliefs that conversing about sex would be embarrassing.

The outcome measure was whether the parent engaged in a meaningful conversation about sex and pregnancy with his or her child in the ensuing 9 months after program participation. This was assessed by self-reports from the adolescent child of the parent at a follow-up interview. The outcome was scored 0 = parent did not engage in a conversation versus 1 = parent engaged in a conversation. Each of the mediators was measured at baseline and again at program completion. The control group received exposure to materials on an unrelated topic. The covariates measured at baseline were the biological sex of the adolescent (0 = male, 1 = female), and the overall quality of parent-adolescent communication in general. The latter used a multi-item scale with each item measured on a -3 to +3 disagree-agree metric, averaged across items. Higher scores indicate higher quality communication. In a real evaluation, there would be a longer list of covariates, but I use only two to keep the example manageable. The N was 1,500.

The RET model, absent covariates, is in Figure 12.2. The binary outcome has a disturbance term in the diagram, but often it is omitted. One justification for its inclusion is that it signifies there are determinants of the outcome other than the mediators in the model and the intervention. Another justification is that it can be thought of as the disturbance term in the latent propensity version of binary regression, per Chapter 5. Reasons for omitting it is that the statistical theory for logit and probit regression that is tied to the generalized linear model does not focus on individual scores on the outcome, only on the group-derived conditional probabilities of Y across predictor profiles. As such, there are no individual-level disturbances. It also turns out that the variance of the Y scores at a given predictor profile is completely determined by the mean of Y; mathematically, the variance of dichotomous 0-1 Y scores equals the Y mean times one minus that mean. Unlike traditional regression where the mean and variances are assumed independent, the mean and variance in a binary regression model are dependent, so the conditional variance reflecting disturbances is redundant. You will see cases in the literature where the disturbance term is included and cases where it is not.

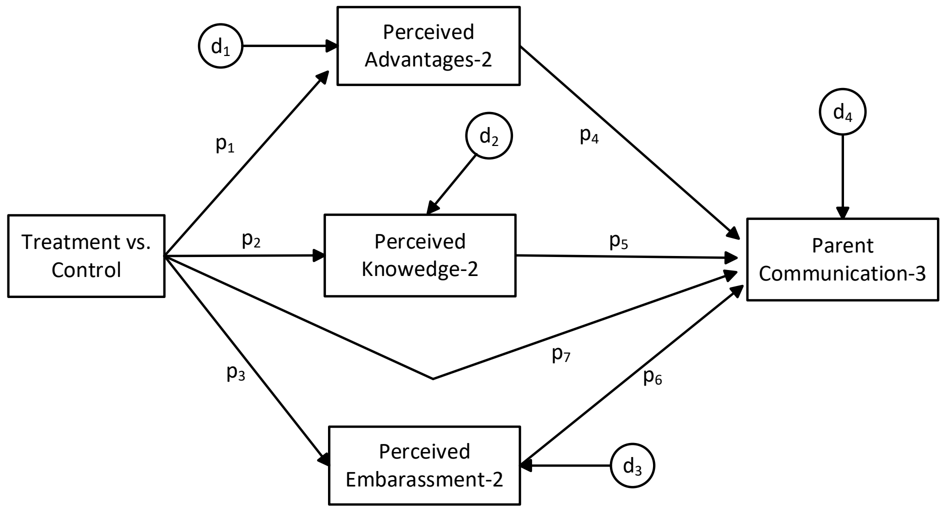

FIGURE 12.2. Parent communication example

The equations for the model, including the covariates, are (note: TREAT = the treatment condition, BS is biological sex, CQ is the general quality of communication between parent and child at baseline; PA is perceived advantages; PK is perceived knowledge, PE is perceived embarrassment, and COM is parent communication; each followed by the number 1, 2 or 3 to indicate time of assessment):

$$PA2 = a_1 + p_1\ TREAT + b_1\ BS1 + b_2\ CQ1 + b_3\ PA1 + d_1 \quad [12.4]$$

$$PK2 = a_2 + p_2\ TREAT + b_4\ BS1 + b_5\ CQ1 + b_6\ PK1 + d_2 \quad [12.5]$$

$$PE2 = a_3 + p_3\ TREAT + b_7\ BS1 + b_8\ CQ1 + b_9\ PE1 + d_3 \quad [12.6]$$

$$COM3 = a_4 + p_4\ PA2 + p_5\ PK2 + p_6\ PE2 + p_7\ TREAT + b_{10}\ BS1 + b_{11}\ CQ1 \quad [12.7]$$

The left hand term for Equation 12.7 takes on different forms depending on the link function implied by the chosen method of analysis, such as probit versus logistic regression, as will become apparent later. For now, I express the equation generically.

PRELIMINARY ANALYSES

I provide a document on my webpage that presents the preliminary analyses I typically pursue when my RET has a binary outcome. In that document, I show how I evaluate the potential applicability of a logit, probit and/or MLPM approach. In the current example. I found each approach to be viable. I also evaluated response distributions, conducted leverage analyses, and evaluated heteroscedasticity. Heteroscedasticity refers to variance homogeneity for the disturbance term of the latent response underlying probit and logit regression. See the preliminary analyses document on my webpage for details. There is considerable useful information in this document, so check it out.

LISEM ANALYSIS: THE MODIFIED LINEAR PROBABILITY MODEL

In this section, I consider a straightforward approach to analyzing the numerical example, namely the use of the MLPM in a LISEM context. I recognize there are methodologists who object to this analytic strategy but as I have argued in prior chapters, each approach has strengths and weaknesses. Recall from Chapter 5 that the MLPM involves predicting a binary outcome from a set of predictors using traditional OLS or maximum likelihood analysis much like you would do in a traditional regression analysis or SEM. There are no logit or probit functions involved. The binary outcomes are analyzed directly but a robust Huber-White estimator is used to accommodate non-normality and variance heterogeneity that results from the binary outcome. The modified linear probability model is not the same as the linear probability model that is so often criticized in the statistical literature. Keep in mind also that the MLPM is viable only if outcome probabilities reasonably approximate a linear relationship to the continuous predictors in the model. If the functions are non-linear, then adaptations need to be made.

Model Fit

As a first step, I want to evaluate the overall fit of the model depicted in the influence diagram in Figure 12.2. However, in LISEM the traditional global fit statistics do not apply. Instead, I need to use the strategy outlined in Chapters 8 and Chapter 11 for LISEM based on independence tests. I used the program *graph theory* (DAGitty) on my website to identify the implied independencies by the model in Figure 12.2. For example, one model-based independence model assumption is that the association between perceived advantages at posttest and perceived embarrassment at posttest should be zero if I hold constant (a) perceived advantages at baseline, (b) the treatment condition, (c) biological sex, and (d) communication quality at baseline. I calculated this partial correlation and found it to equal 0.014 with a p value of 0.59, which is consistent with model predictions. The graph theory program identified 24 different independence conditions after taking into account all covariates and variables. I evaluated each independence condition to gain perspectives on model fit, i.e., whether the independence assumption made by the model was born out in the data. Given the number of contrasts involved, one likely will want to take into account chance associations by using one of the strategies discussed in Chapter 6. Across the contrasts, the data were model supportive. My preference when evaluating model fit is to work at the level of such localized tests rather than the omnibus C statistic described in Chapter 8. You can download the DAGitty code I used to identify independence conditions for this example from my website on the *Resources* tab. Given minimal meaningful violations of the independence conditions implied by the model, I moved forward with it accordingly.

Total Effect of the Intervention on the Outcome

The first question I ask is if the program affects the outcome and by how much. I must first, however, set a meaningfulness standard for evaluating the total effect and then I need to estimate the total effect relative to that standard.

Meaningfulness Standard for the Program Total Effect

In Chapter 10, I outlined a strategy for setting meaningfulness standards when working with a client to evaluate a program. Here, I illustrate a different approach that uses instead the logic a researcher might use to develop a meaningfulness standard and who is developing a program for widespread population use rather than for a specific client in a more constrained population setting. My goal is to show you different ways of thinking about meaningfulness.

Suppose I envision my intervention as being adopted by middle schools throughout the United States. Middle school youth typically are between the ages of 12 to 14, with studies indicating that students who engage in sex at such young ages typically are at a higher risk of experiencing an unintended pregnancy in the future during high school. Indeed, parenthood is the leading reason that adolescent girls drop out of middle or high school.

There are approximately 12 million youth in middle schools nationwide. If my program can reach a parent for, say, half of these students, it will impact about 6 million parents. If I increase by just 5% the number of these 6 million parents who talk meaningfully with their children about reasons not to engage in sex at this time in the adolescents' lives, then this reflects an increase of about 300,000 parents potentially using more effective communication strategies. Over the course of, say, three or so "generations" of students passing through middle school every few years, this translates into about a million parents. Thus, increasing by a minimum of 5% the number of parents who talk effectively with their middle school child about sex seems a reasonable goal. Of course, I need to keep in mind that just because parents talk with their adolescent child about not having sex at such a young age does not mean the child won't have sex anyway; I need to weigh the empirical evidence that such communication does indeed make a difference and is worth the resource investment.

Having laid out the above logic as I zero in on a meaningfulness standard of 5% or so, I personally would not and should not be content to "armchair" this matter on my own. Rather, I should seek the opinions of policymakers, school principals, school staff, experts in adolescent sexual behavior, experts in family communication, parents, and adolescents to explore the issue with them, say, in the context of a well-designed qualitative research program. I also should seek input from diverse constituencies, including different types of middle schools. Based on this foundational work, latitudes of meaningfulness, effect ambiguity, and no effect can emerge to guide my research (see Chapters 2 to 10 for a discussion of such latitudes). For the present example, I will use a value of 5% as the upper bound cutoff for the latitude of no effect, i.e., population increases less than 5% will be deemed trivial. I will use as the latitude of effect ambiguity (a gray area in terms of effect size meaningfulness; some people see it as meaning while others do not) change values between 5% to 8%; and for the latitude of meaningful effects, I will use change values of 8% or higher.

Estimation of the Total Effect

The total effect of the program on the outcome can be estimated by regressing the binary outcome onto the dummy coded treatment variable in conjunction with the baseline covariates, biological sex (BS1) and communication quality (CQ1). When a binary outcome is scored 0 and 1, the mean of the outcome equals the proportion of people who score 1 on the outcome, in this case the proportion of parents who talk meaningfully with their child about not having sex. The use of the MLPM in this regression context is analogous to using an analysis of covariance (ANCOVA) with parental communication (COM3) as the outcome, the treatment condition (TREAT) as the factor, and BS1 and CQ1 as covariates. Traditional ANCOVA uses the general linear model for its calculations and essentially mean centers all covariates when calculating the adjusted outcome means. By mean centering covariates, the covariate adjusted outcome means are for people with "typical" scores on the covariates, i.e., people who are "average" on them relative to the sample data. As discussed in Chapter 6, mean centering a covariate makes a score of 0 on the transformed covariate correspond to the mean score on the original covariate metric. The equation for the model using centered covariates is:

$$\text{COM3} = a + p_1\,\text{TREAT} + b_2\,\text{CQ1}_C + b_3\,\text{BS1}_C$$

where the subscript C indicates mean centered scores. The intercept is the mean outcome when all predictors equal zero. In this case, a score of 0 on TREAT is the control group and scores of 0 on the covariates are the covariate means. Thus, the intercept is the predicted outcome mean for the control group for people who are “typical” on the covariates. Although mean centering biological sex (BS1) may seem unusual, I discuss in Chapter 6 the rationale for doing so. The covariate adjusted mean for the intervention group is the intercept plus p_1 because p_1 reflects the added effect of the intervention on the outcome relative to the control. I can execute the MLPM using Mplus code and it appears in Table 12.4.[5] As in other chapters, the actual Mplus code does not include line numbers; I use them here for referencing.

Table 12.4: Syntax for MLPM for Total Effect with LISEM

```
1. TITLE: MLPM Analysis of communication  ;
2. DATA: FILE IS c:\mplus\communication.dat ;
3. DEFINE:
4. CENTER BS1 CQ1 (GRANDMEAN) ;
5. VARIABLE:
6. NAMES ARE ID COM3 PA2 PK2 PE2 CQ1 PA1 PK1 PE1 TREAT BS1 ;
7. USEVARIABLES ARE COM3 TREAT BS1 CQ1 ;
8. MISSING ARE ALL (-9999) ;
9. ANALYSIS:
10. ESTIMATOR = MLR ;
11. MODEL:
12. COM3 ON TREAT BS1 CQ1 (p1 b1 b2) ;
13. [COM3] (pcontrol) ;
14. OUTPUT: SAMP STANDARDIZED(STDYX) MOD(ALL 4) RESIDUAL
15. CINTERVAL TECH4 ;
```

There are several noteworthy features of the code. Line 3 illustrates a new command called `DEFINE`, which tells Mplus I want to transform one or more of the input variables. Line 4 is a subcommand of `DEFINE`. It tells Mplus I want to mean center the variables named after the keyword `CENTER` and then specifies the type of centering to execute. I therefore use this command to mean center the covariates, which, as noted, makes the intercept in the equation equal to the covariate adjusted outcome proportion for the control group.

Line 13 asks Mplus to calculate the intercept for the equation that predicts `COM3`, which Mplus does by default anyway; however, I specify it in the code so I can assign a label to it for future use. I also label the coefficients on Line 12, using the letter b to signify a nuisance variable coefficient and a p for a coefficient that is meaningful in the model.

[5] Per Chapter 5, there are implementations of Huber-White estimation that sometimes work better for smaller sample sizes, most notably the HC3 version (Long & Ervin, 2000). Mplus does not provide HC3 as an option. If your sample size is <250, you might be better served using the HC3 implementation on my webpage.

The estimated, single equation model is just-identified, so issues of model fit are moot. Here is the relevant output:[6]

```
MODEL RESULTS

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

COM3     ON
    TREAT              0.190      0.025      7.561      0.000
    BS1                0.112      0.025      4.468      0.000
    CQ1                0.081      0.032      2.522      0.012
```

The intercept from the output was 0.374 ±0.034, which expressed as a percentage is 37.4% ±3.4%. This is the estimated covariate adjusted control group percent for COM3 given that I mean centered the covariates. I can calculate the corresponding percent for the intervention group from the information on the Mplus output, but in order to obtain standard errors for the intervention proportion/percent, it is easier to re-score the TREAT variable so that the intervention group becomes the reference group, re-run the analysis, and then the new intercept is the covariate adjusted COM3 mean/proportion for the intervention group. I reverse code TREAT by adding the following command after line 4:

```
TREAT = ABS(TREAT-1) ;
```

The intercept in the output of the revised code was 0.564 ±0.036, which expressed as a percentage is 56.4% ±3.6%. The path coefficient for TREAT on the output is 0.19 ±0.050 ($z = 7.56$, $p < 0.05$), which expressed as a percentage is 19.0 ±5.0%. The z test is a test of the covariate adjusted proportion/percent difference for the groups defined by TREAT, i.e., it is the estimated total effect of the program. Thus, the program increased effective communication from 37.4% to 56.4%. Technically, the value of 0.19 is the group difference when the covariates equal their mean. It turns out that for the MLPM, this difference will be the same no matter what values the covariates are held constant at. As you will see, this is not the case for logit/probit modeling.

I can convert the treatment and control group probabilities to a relative risk and a number needed to treat (NNT) index using the formulae from Chapter 10. The relative risk divides the estimated probability for COM3 for the treatment group (0.564) by the corresponding probability for the control group (0.374), yielding a value of 1.51. It is

[6] A warning message appears on the output about a non-positive definite first order product matrix. This warning can be ignored for all chapter examples; see the document in Chapter 11 on the Resources tab of my webpage.

about one and half times more likely that parents exposed to the intervention will talk with their child about sex and pregnancy than parents in the control condition. The NNT is 1.0 divided by the treatment probability minus the control probability. It was 1/(.564-.374) = 5.27. In general, the program needs to "treat" just over 5 parents to have one additional parent engage in communication about sex and pregnancy as compared with doing nothing. Thinking on a grander scale, for every 1,000 parents who participate in the program, about 1,000/5.27 = 190 more of them will engage in communication with their child as compared to doing nothing, i.e., as compared to treatment as usual (TAU).

The standard for a meaningful effect was a proportion difference greater than or equal to 0.08. The confidence interval for the total effect from the Mplus output was 0.14 to 0.24. Because the lower limit of this interval exceeds the effect size standard, I conclude the program effect was indeed meaningful. Confidence intervals for a proportion are often asymmetric. One can apply bootstrapping to obtain asymmetric intervals. I calculated a percentile bootstrap confidence interval and found comparable results for the confidence interval to those reported here. I recommend that you make a habit for all programs discussed in this chapter and most other chapters of conducting sensitivity analyses by pursuing bootstrap analyses of them in addition to the more traditional methods of estimation.

In sum, based on LISEM using the MLPM, the program increased the proportion of parents who talk with their children about sex and pregnancy by about 19%, from 37% to 56%, a difference I determined to be meaningful. Unfortunately, about 44% of parents who engage in the program do not talk effectively with their children about sex.[7] In general, outreach needs to engage about five parents to have one of them engage in communication about sex as compared to just doing nothing. The program seems to "work" but there is room for improvement.

Effect of the Intervention on the Targeted Mediators

The next set of questions focus on whether and to what degree the program affects the targeted mediators.

Meaningfulness Standards for Intervention Effects on the Mediators

To define a standard for meaningfulness for the effects of the program on the presumed mediators, I used the strategy discussed in Chapters 10 and 11 as implemented in the program called *effect size standards* on my website. I describe the steps for the current numerical example in Appendix B. A meaningful effect for perceived advantages is an

[7] This is 100% - 56% = 44%. In ways, 44% is a type of "exceptions to the rule" index because it reflects people who participated in the program but who did not communicate with their child.

absolute mean difference of 0.22 or more; for perceived knowledge it is 0.24 or more; and for perceived embarrassment it is 0.19.

Estimation of Program Effects on the Mediators

There are three relevant equations for this facet of program evaluation, one for each mediator. Because the mediators are continuous, I use robust maximum likelihood to evaluate program effects on them. The analysis of each equation takes the form of an ANCOVA with a continuous outcome, which I implement using linear regression in Mplus. I focus first on Equation 12.4 for the perceived advantages mediator, which I repeat here for convenience:

$$PA2 = a_1 + p_1\ TREAT + b_1\ BS1 + b_2\ CQ1 + b_3\ PA1 + d_1$$

Table 12.5 presents the syntax for the analysis.

Table 12.5: Syntax for Intervention Effect on Mediator for MLPM

```
1.  TITLE: MLPM analysis of program effects on mediators  ;
2.  DATA: FILE IS c:\mplus\communication.dat ;
3.  DEFINE:
4.  CENTER CQ1 PA1 BS1 (GRANDMEAN) ;
5.  VARIABLE:
6.  NAMES ARE ID COM3 PA2 PK2 PE2 CQ1 PA1 PK1 PE1 TREAT BS1 ;
7.  USEVARIABLES ARE PA2 CQ1 PA1 TREAT BS1 ;
8.  MISSING ARE ALL (-9999) ;
9.  ANALYSIS:
10. ESTIMATOR = MLR ;
11. MODEL:
12. PA2 ON TREAT BS1 CQ1 PA1 (p1 b1-b3)  ;
13. OUTPUT: SAMP STANDARDIZED(STDYX) RESIDUAL
14. CINTERVAL TECH4 ;
```

Lines 3 and 4 mean center the covariates so the intercept reflects the covariate adjusted mediator mean for the control group when the covariates equal their “typical” values.

The single equation model is just identified so issues of model fit are moot. Here is the relevant output for our purposes:

```
MODEL RESULTS

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

 PA2      ON
    TREAT              0.823      0.020     41.818      0.000
    BS1                0.097      0.020      4.912      0.000
    CQ1                0.092      0.025      3.678      0.000
    PA1                0.355      0.025     14.178      0.000

 Intercepts
    PA2                0.063      0.014      4.547      0.000
```

The intercept is the PA2 covariate adjusted control group mean and it equals 0.06 ±0.03. This is indexed on the -3 to +3 disagree to agree scale with anchors -3 = strongly disagree, -2 = moderately disagree, -1 = slightly disagree, 0 = neither agree nor disagree, 1 = slightly agree, 2 = moderately agree, 3 = strongly agree. I re-run the analysis using a reverse coded TREAT via the strategy described earlier to obtain the PA2 covariate adjusted mean and margin of error for the treatment group. It was 0.89 ±0.03. The covariate adjusted mean difference between the treatment and control groups is the coefficient associated with TREAT on the output. It was 0.82 ±0.04 ($z = 41.82$, $p < 0.05$), which equals 0.89 – 0.06.[8] The 95% confidence limit for the covariate adjusted mean difference was 0.78 to 0.86. (output not shown here but it is in the output section labeled CONFIDENCE INTERVALS OF MODEL RESULTS). The lower limit of the confidence interval exceeds the standard for meaningfulness (≥0.22), so I conclude the program had a meaningful effect on PA2.

Using comparable code for PK2, the covariate adjusted control group mean was 0.02 ±0.03. and for the intervention group it was 0.86 ±0.03. The mean difference as reflected by the coefficient associated with TREAT was 0.84 ±0.04 ($z = 41.66$, $p < 0.05$). The 95% confidence limit for the covariate adjusted mean difference was 0.80 to 0.88. The lower limit of this confidence interval exceeds the effect size standard for meaningfulness for PK2 (≥0.24), so I conclude the program had a meaningful effect on it.

For PE2, the covariate adjusted control group mean was -0.062 ±0.03 and for the intervention group it was -0.055 ±0.03. The mean difference as reflected by the coefficient associated with TREAT was 0.007 ±0.04 ($z = 0.33$, $p < 0.74$). The 95% confidence limits for the covariate adjusted mean difference were -0.03 to 0.046. The confidence interval is fully contained within the latitude of no effect, so I conclude the program effect on the embarrassment mediator was functionally zero. This program facet

[8] Minor disparities across analyses are due to rounding error.

needs to be reworked.

On the resources tab of my web page for Chapter 10 I described how to calculate standardized effect sizes for the effects of a program on a continuous mediator. You can consult that document and apply the same procedures here. As examples, the probability of exceptions to the rule for `PA2` was 0.06, for `PK2` it was 0.06, and for `PE2` it was 0.49.

I replicated the analyses for program effects on `PA2`, `PK2`, and `PE2` using trimmed mean and MM regression for sensitivity purposes; my conclusions were the same.

In sum, the program had meaningful effects on perceived advantages and perceived knowledge but not perceived embarrassment. For perceived advantages and perceived knowledge, the control group means on the -3 to +3 disagree to agree scales tended to be near the midpoint (neither agree nor disagree). The program shifted the means to values closer to a scale value of 1.0 (slightly agree), more specifically values near 0.80. Keep in mind that a person's score on these scales is the average of multiple items, so the shifts in total score means represent movement across the constellation of scale items. In this sense, a change of 0.80 units is no small feat. By the same token, there is room to make the agreement associated with perceived advantages and perceived knowledge stronger.

Effects of the Mediators on the Outcome

My next task is to evaluate estimates of the extent to which the targeted mediators meaningfully impact the outcome and to determine whether and by how much program effects exist on the outcome independent of the mediators. I first need to set a meaningfulness standard for the causal effect of each mediator on the outcome. The analysis that addresses mediator effects on outcomes focuses on Equation 12.7, which I repeat here for convenience:

$$\text{COM3} = a_4 + p_4\,\text{PA2} + p_5\,\text{PK2} + p_6\,\text{PE2} + p_7\,\text{T} + b_{10}\,\text{BS1} + b_{11}\,\text{CQ1}$$

Consider the mediator PA2, the perceived advantages of talking with one's child about delaying sex. For the MLPM, the coefficient p_4 reflects the change in the proportion of parents who talk with their children about sex and pregnancy given a one unit increase in PA2 on its -3 to +3 metric holding constant the other predictors in the equation. Suppose p_4 equals 0.05. This means that if I increase PA2 by 1.0, the proportion of parents who talk with their child should increase by 0.05 or by about 5%. I argued earlier that the meaningfulness standard for the total program effect was a proportion increase of 0.08. I do not expect PA2 to carry the entire load in producing this change, but I want it to carry its share relative to the other mediators. With two other mediators, yielding a total of three mediators, I might reason that PA2 should produce at least 1/3 of the desired total intervention effect, namely 0.08/3 = 0.027. Next, I need to factor in what I think is a reasonable amount of change in PA2 that I can expect of the intervention. In the previous section, I found that the intervention changed PA2 by about 0.85 units (to be exact, the effect of the intervention on PA2 was estimated to be 0.82 ±0.04). Using 0.85 as an index of the amount of change in PA2 I can reasonably expect, then to produce a 0.027 increase in communication, PA2 needs to have a p_4 path coefficient of at least 0.027/.85 = 0.032, or, rounded to two decimals, 0.03. I use this as my meaningfulness standard for the population value of p_4. I outlined the above logic and variants of it in Chapter 10, so consult that material for elaboration.

When I applied similar logic to the PK2 mediator, whose metric also is -3 to +3, I came up with the same meaningfulness standard for the population value of p_5, namely 0.03. When I applied the approach to the PE2 mediator, whose metric also is -3 to +3, I encountered a complication, namely the effect of the intervention on PE2 was virtually nil. Obviously, the program designers need to alter what they are doing to bring about change in PE2. If they do so, what magnitude of change can I reasonably expect the intervention to have on PE2? Suppose after discussions with relevant staff, I decide that a reasonable guess is about 0.85 units on the -3 to +3 metric of PE2. This also yields a meaningfulness standard of 0.03 for PE2, but it is negatively signed because PE2 is negatively associated with the outcome. The population p_6 coefficient needs to be -0.03 or less.

Meaningfulness standards for mediators do not have to be equal across the mediators, but that turned out to be the case here. The size of the standard will, in part, be a function of the metric of the mediator. If a mediator is scored from 0 to 100, it obviously will have a different standard than a mediator that is scored from -3 to +3. Also, when setting the standard on the basis of the amount of expected change in the mediator that is reasonable, I must keep in mind that this also is not set in stone. There is subjectivity involved but at least we can identify the areas of ambiguity that can then be discussed and debated as we search for a standard or set of standards to apply. Again, see Chapter 10 for details.

The Mplus syntax that allows me to evaluate Equation 12.7 uses the programming principles already described and is presented in Table 12.6.

Table 12.6: Syntax for MLPM Mediator Effect Analysis with LISEM

```
TITLE: MLPM analysis of mediator effects on outcome  ;
DATA: FILE IS c:\mplus\communication.dat ;
VARIABLE:
NAMES ARE ID COM3 PA2 PK2 PE2 CQ1 PA1 PK1 PE1 TREAT BS1 ;
USEVARIABLES ARE PA2 PK2 PE2 CQ1 COM3 TREAT BS1 ;
```

```
MISSING ARE ALL (-9999) ;
ANALYSIS:
ESTIMATOR = MLR ;
MODEL:
COM3 ON PA2 PK2 PE2 TREAT BS1 CQ1 (p4-p7 b10 b11) ;
OUTPUT: SAMP STANDARDIZED(STDYX) MOD(ALL 4) RESIDUAL
CINTERVAL TECH4 ;
```

I do not use mean centering of the covariates in the above syntax because it is unnecessary for purposes of answering the questions being addressed.

The model for this single equation analysis is just-identified, so model fit is moot. Here is the output for the coefficients for Equation 12.7:

```
MODEL RESULTS
                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 COM3      ON
    PA2                0.117      0.030      3.874      0.000
    PK2                0.110      0.029      3.769      0.000
    PE2               -0.141      0.030     -4.672      0.000
    TREAT              0.003      0.043      0.065      0.949
    BS1                0.090      0.025      3.574      0.000
    CQ1                0.056      0.032      1.736      0.083
```

For perceived advantages of talking with one's child about sex and birth control (PA2), for every one unit that scores on its -3 to +3 metric increase, the proportion of parents subsequently engaging in conversations is predicted to increase by 0.12 ±0.06 holding constant the other predictors. The coefficient was statistically significant ($z = 3.87$, $p < 0.05$). Roughly the same effect size was observed for perceived knowledge about sex, as well as for perceived embarrassment, but the latter relationship was inverse. All three targeted mediators yielded meaningful effects vis-à-vis the effect size standards I set, namely coefficients of 0.03 after taking into account sampling error. The 95% confidence interval for PA2 was 0.06 to 0.18, for PK2 it was 0.05 to 0.17, and for PE2 it was -0.20 to -0.08 (not shown here).

The treatment itself had a statistically non-significant and seemingly trivial impact on communication over and above the three targeted mediators (coefficient = 0.003 ±0.08, $z = 0.07$, $p < 0.95$). However, the confidence interval for the effect was -0.08 to +0.08, which overlaps the latitude of effect ambiguity, so some caution must be maintained before dismissing it.

Parenthetically, Mplus reports a squared R for the equation on its output (in the STANDARDIZED STDYX section), which is often called a pseudo-R squared when outcomes are binary. For the MLPM, it is referred to as **Efron's R squared** (Efron, 1978). I am not a fan of pseudo-R squares. In the case of the MLPM, individual variability about a conditional mean/proportion is determined by the mean/proportion itself. Given this, the magnitude of the squared R decreases as conditional proportions approach 0.50 in a study because this is where variability about the mean/proportion is largest. In ways, the overall event rate that your outcome happens to have impacts the squared R you will observe. Corrections for this phenomenon have been suggested for pseudo-R squares (see Long & Freese, 2006), but for the MLPM, the squared R as an index of explained variance is of limited use.

Based on the above analyses using the MLPM, I conclude that each of the mediators targeted by the program are relevant to parental communication. The treatment does not appear to have much effect on parental communication over and above the mediators, but some caution is required in dismissing this possibility entirely.

Overall Conclusions for LISEM Analyses Based on the MLPM

In sum, the RET found that the program meaningfully increased parent-adolescent communication about sex and birth from a level of about 38% in the control group to 57% in the program group, an increase of 19%. Each of the mediators targeted by the program were non-trivially related to the outcome, but the program meaningfully affected only two of them, parental perceived advantages of talking with their child about sex and birth control, and parental perceived knowledge about sex and birth control. The RET affirmed the importance of addressing parents' anticipated embarrassment, but the program failed to reduce feelings of such potential embarrassment in a meaningful way. A new approach is needed for this program segment. The program did not seem to affect parental communication through mechanisms other than the targeted mediator.

Using the logic of the joint significance test, I would conclude from the analyses that perceived advantages and perceived knowledge mediate some of the effect of the program on the outcome, but that perceived embarrassment does not. However, such omnibus statements are not as helpful as the more specific link-by-link analyses. For discussion of how to calculate omnibus mediation tests using the MLPM in an LISEM context, see the document on omnibus tests on the resources tab of my webpage.

LISEM ANALYSIS: THE PROBIT MODEL

In this section, I describe how to apply a probit model to the communication example but in a LISEM context. The reason someone might pursue such a strategy instead of a FISEM probit based analysis is if the sample size or data context cannot support a complex multi-equation model using simultaneous estimation. LISEM has the ability to apply statistical tools that are not available in FISEM, such as Firth regression, and offers the option of mixing different estimation strategies for different parts of the model. The current numerical example can support FISEM, but I apply LISEM to it here to show the basic logic of the LISEM probit approach.

Mplus offers four estimation strategies for probit modeling (a) WLSMV, (b) robust maximum likelihood, (c) maximum likelihood, and (d) Bayesian. WLSMV is a form of weighted least squares that uses a diagonal weight matrix but with standard errors and mean- and variance-adjusted chi-square test statistics based on a full weight matrix. See the Mplus technical appendix on the Mplus website for details (accessible through my website on the syntax tab). The option MLR, coupled with a link subcommand to invoke probit modeling, uses robust maximum likelihood with sandwich estimation. The option ML, also coupled with a link subcommand to invoke probit modeling, uses traditional maximum likelihood. I use ML here but discuss choosing among the four methods when I consider FISEM based probit analysis. For cautions about using sandwich estimators (MLR) for logit/probit models, see Greene (2012, p. 692-693). Although MLR may not be helpful for the probit portion of your analyses per the discussion by Greene, it still might be worth using to accommodate non-normality and heteroscedasticity in the non-probit portions of your model. You make this choice based on data, as I discuss later.

Model Fit

I evaluate overall model fit using the same strategy for independence testing as described for the MLPM. The results for the MLPM should be comparable but not identical to those for probit-based modeling. For example, one of the independence assumptions of the communication model is that COM3 and perceived knowledge at baseline (PK1) should be independent holding constant PA1, PE2, PK2, TREAT, CQ1 and BS1. This means that if I conduct a probit regression that regresses COM3 onto PK1, PA1, PE2, PK2, TREAT, CQ1 and BS1, the coefficient for PK1 should be statistically non-significant. This was indeed the case (the Wald coefficient for PK1 was 2.28, $p < 0.14$). Across all of the independence contrasts, the data for the communication example were model supportive.

Total Effect of the Intervention on the Outcome

I describe two approaches to evaluate the total effect of the intervention on the outcome in an LISEM context, (1) traditional probit regression coupled with profile analysis and (2) the analysis of average marginal effects.

Probit Regression with Profile Analysis

To evaluate the overall effect of the program on parental communication, I use the same effect size standard as for the MLPM, i.e., a proportion increase ≥ 0.08. Harrell (2021) questions the practice of summarizing a total effect with a single number in logit and probit modeling because such models are inherently non-linear. As I discussed above, for probabilities the treatment predictor often interacts with covariates despite the fact that a "main effect" model is fit at the level of logits or probits, per Table 12.3. I use a different programming strategy than before but without mean centering. I discuss why later.

I begin by fitting the equation

$$\text{Probit(COM3)} = a + p_1\, T + b_2\, CQ1 + b_3\, BS1 \qquad [12.8]$$

The relevant code is in Table 12.7.

Table 12.7: Syntax for Probit Profile Analysis for Total Effect with LISEM

```
1. TITLE: Probit-based total effect analysis of communication  ;
2. DATA: FILE IS c:\mplus\communication.dat ;
3. VARIABLE:
4.    NAMES ARE ID COM3 PA2 PK2 PE2 CQ1 PA1 PK1 PE1 TREAT BS1 ;
5.    USEVARIABLES ARE COM3 TREAT BS1 CQ1 ;
6.    MISSING ARE ALL (-9999) ;
7.    CATEGORICAL ARE COM3 ;
8. ANALYSIS:
9.    ESTIMATOR = ML ; LINK=PROBIT ;
10. MODEL:
11.   COM3 ON TREAT BS1 CQ1 (p1 b1 b2) ;
12.   [COM3$1] (thresh) ;
13. MODEL CONSTRAINT:
14.   NEW(CPROBIT TPROBIT CPROB TPROB DIFF) ;
15.   CPROBIT = -thresh + p1*0 + b1*0.525 + b2*0.012  ;
16.   TPROBIT = -thresh + p1*1 + b1*0.525 + b2*0.012  ;
17.   CPROB = PHI(CPROBIT);
18.   TPROB = PHI(TPROBIT) ;
19.   DIFF = TPROB-CPROB ;
20. OUTPUT: SAMP STANDARDIZED(STDYX) RESIDUAL CINTERVAL TECH4 ;
```

The first 6 lines should be familiar. Line 8 declares the variable `COM3` as ordinal/categorical by using the `CATEGORICAL` subcommand. Mplus automatically detects if the listed variable is binary and, if it is, knows to invoke specialized binary regression methods. Only endogenous variables can be specified on the `CATEGORICAL` subcommand. You do not specify exogenous variables as binary, such as a dummy predictor.

For binary outcomes and probit regression, Mplus estimates thresholds rather than intercepts (the intercept is the threshold times -1) and specifies thresholds using different syntax than for intercepts. I discussed the concept of thresholds in Chapter 5 and do not repeat that discussion here. Consult that material, as needed. On Line 12, the command to estimate the threshold for the binary outcome is `[COM3$1] (thresh)`. The $ sign indicates a threshold is to be estimated followed by a number that indicates the number of the threshold to be estimated (in some models, there is more than one threshold for a given variable; for binary regression models, it will always be 1). I label the parameter `thresh`.

Key to the total effect analysis are the subcommands for the `MODEL CONSTRAINT` command on Line 13. These lines estimate, based on the probit equation, the predicted proportion of parents in the control group who communicated with their child about sex and pregnancy (Lines 15 and 17) and the predicted proportion of parents in the intervention group who communicated with their child (Lines 16 and 18), both holding constant the two covariates. Line 19 subtracts the former from the latter to document the total effect. Line 14 indicates the names to give each new parameter. There are 5 new parameters named `CPROBIT`, `TPROBIT`, `CPROB`, `TPROB`, and `DIFF`. I can use any names but they cannot exceed 8 characters or violate Mplus naming conventions. The estimation of these parameters will *not* affect model fit. Line 15 calculates a predicted probit score for the control group (hence the label `CPROBIT`). The first term on Line 15 is the label I gave to the threshold value and by taking the negative value of it, it becomes the intercept. The label p1 refers to the label I gave to the probit coefficient for the `TREAT` predictor and by multiplying it by 0, I invoke the control group (because 0 = the control group and 1 = the treatment group on the dummy variable). The label b2 refers to the coefficient for baseline communication quality and I multiply it by its mean score, 0.012. This means that I hold communication quality constant at its "typical" score in the sample, i.e., its sample mean. The label b1 refers to the coefficient for biological sex. Per my discussion in Chapter 6, I multiply it by its mean which is analogous to mean centering it. Although this latter operation for a binary predictor is a reasonable strategy for classic linear regression (and the MLPM), it is controversial for logit and probit modeling (see Hanmer & Ozan, 2013; Muller & MacLehose, 2014) because of the non-linear nature of these models. I address this controversy below when I consider profile analysis. For now, we think of the resulting value of `CPROBIT` as the covariate adjusted probit score for individuals in the control group who have "typical" values on CQ1 and where we take into account the typicalness of the categories representing biological sex. Similarly, we think of the resulting value of `TPROBIT` as the covariate adjusted probit score for individuals in the intervention group who have "typical" values on CQ1 and where we take into account the typicalness of the categories representing biological sex.

Line 16 is identical to Line 15 except I multiply the coefficient for TREAT (p1) by 1 instead of 0 to signify the treatment group. The resulting value of TPROBIT is the covariate adjusted probit score for individuals in the treatment group using “typical” values on the covariates. Lines 17 and 18 invoke the PHI function in Mplus that converts each of the probit scores to probabilities/proportions; Line 20 calculates the difference between the two probabilities/proportions to yield the estimated total effect. Because the model is just-identified, I omit requests on the output line for fit statistics. Also, I cannot ask for modification indices when using the MODEL CONSTRAINT line; Mplus will give me an error message. It is irrelevant here because the focused model is just-identified.

Here is the output from the section called New/Additional Parameters:

```
                                                   Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value
New/Additional  Parameters
    CPROBIT           -0.325      0.047     -6.971      0.000
    TPROBIT            0.163      0.046      3.512      0.000
    CPROB              0.373      0.018     21.114      0.000
    TPROB              0.565      0.018     30.861      0.000
    DIFF               0.192      0.025      7.561      0.000
```

The predicted proportion of parents who communicated with their child was 0.37 ±0.04 in the control group and 0.56 ±0.04 in the treatment group, a difference of 0.19 ±0.05 ($z = 7.56$, $p < 0.05$). The estimated effect of the treatment on the outcome is close to the result obtained using the MLPM. The confidence interval for the total effect was 0.14 to 0.24. The lower limit of the confidence interval exceeded the effect size standard for meaningfulness of a 0.08 proportion increase, so I conclude the effect is meaningful. These results replicated when I used bootstrapping instead of traditional ML estimation.

Following Harrell’s recommendations, I next use variants of the MODEL CONSTRAINT code to examine treatment versus control proportion differences for different combinations of covariate values, i.e., using profile analysis. Here is the code where I held biological sex constant at “males” (BS1 = 0) and baseline communication quality at a value of -0.30, which is near its 20th quantile thus representing a “low” score on prior communication quality:

```
MODEL CONSTRAINT:
  NEW(CPROBIT TPROBIT CPROB TPROB DIFF) ;
  CPROBIT = -thresh + p1*0 + b1*0 + b2*(-0.30)  ;
  TPROBIT = -thresh + p1*1 + b1*0 + b2*(-0.30)  ;
  CPROB = PHI(CPROBIT);
  TPROB = PHI(TPROBIT) ;
  DIFF = TPROB-CPROB ;
```

The predicted proportion of parents who communicated with their child was 0.29 ±0.04 in the control group and 0.48 ±0.05 in the treatment group, a difference of 0.18 ±0.05 ($z = 7.52$, $p < 0.05$). Although the separate proportions shift downward in the two groups, as would be expected given the impact of the covariates on communication, the proportion difference between the treatment and control groups changes only slightly. Table 12.8 shows results for different combinations of biological sex crossed with communication quality at values near its 20th, 50th, and 80th quantiles (values of -0.30, 0.0 and 0.30). You can see that for these different profiles, the treatment minus control proportion differences are close to one another. This will not always be the case but it is convenient for characterizing total effects; it seems fair to say that the total effect of the program is to increase communication by about 19% ±5%.

Table 12.8: Treatment-Control Proportion Differences as a Function of Covariates

	Males			Females		
CQ1	Treat	Cntrl	Difference	Treat	Cntrl	Difference
-0.30	0.48	0.29	0.18 ±0.05*	0.59	0.40	0.19 ±0.05*
0.00	0.50	0.32	0.19 ±0.05*	0.62	0.43	0.19 ±0.05*
0.30	0.53	0.34	0.19 ±0.05*	0.64	0.45	0.19 ±0.05*

Parenthetically, the values for CQ1 that I used are covariate values that interest me in their own right and that map onto low, moderate, and high scores for the CQ1 distribution. I am *not* making statements about the total effect of the treatment at the values of CQ1 at their 20th, 50th and 80th *population* quantiles per se. I am evaluating the generalizability of the total treatment effect across these three particular CQ1 values (-.30, 0 and .30) and biological sex. The same is true when I mean centered the two covariates in the original analysis, i.e. I evaluate the total effect at the values of the covariate *sample* means.

The advantage of including covariates in the analysis of the total effect is that doing so usually increases statistical power for the test of the program effect. As well, if you truly believe a probit function applies (or that a logit function applies if you are using logistic regression), then despite the fact that you are evaluating a main effect probit model at the level of probits/logits, it is possible that at the level of probabilities, the total effect varies as a function of the covariates in an interactive/moderator sense. By including the covariates and conducting profile analyses on the probabilities, you gain insights into this possibility.

I mean centered the two covariates in the original probit analysis but I also indicated that this practice is controversial when applying logit and probit models. I digress here to address this controversy. One objection to mean centering a dummy covariate is that with logit or probit transformations to probabilities, the covariate adjusted intercept in Equation 12.8, can (but does not necessarily) distort the true overall event rate for the control group. I did not find this to be true in the numerical example, but it can happen (see Hanmer & Ozan, 2013; Muller & MacLehose, 2014).[9] Such distortions tend to be less problematic when the covariate is weakly or only moderately associated with the outcome (e.g., a correlation between the covariate and the outcome less than 0.30). Nevertheless, mean centering the binary covariate, the argument goes, can lead to mischaracterizations of intercept based event rates in logit and probit models. A second argument against the use of mean centering for binary covariates is that the magnitude of treatment effects when framed in probability units in logit/probit regression are conditioned on the values the covariates take on. When analyzing a profile or comparing two profiles, a profile with a mean centered binary covariate does not exist in practice; the covariate score is either one or zero, so it does not make sense to evaluate a profile that is somewhere in between these values. Note that this argument also can apply to the mean of a continuous covariate, i.e., there might not be anyone in the sample or population who has a score exactly equal to the mean thereby potentially rendering the profile meaningless. Some purists insist that the profiles studied should occur in the data and with a reasonable frequency. The counterargument to this position is that it sometimes is reasonable to use values for non-existent profiles if doing so provides substantive insights into the dynamics of interest.

My own view is that profile analyses are important to pursue when focusing on probabilities in the context of logit or probit regression. You will need to decide the values for the different predictor profiles you want to evaluate based on substantive considerations. If you use mean centering for binary covariates in the original model analysis, you should be cautious about interpreting the model intercept. It may or may not be meaningful. Carefully constructed profile analyses probably are the best way to make statements about the total effect, as recommended by many methodologists (e.g., Hanmer & Ozan, 2013; Muller & MacLehose, 2014; Harrell, 2021), if you truly believe a logit or probit model applies.

Average Marginal Effects

Another way of expressing the program total effect is to use average marginal effects (AMEs). I used the R program for average marginal effects on my website in conjunction

[9] In the current example, when I used mean centering for biological sex, the estimated probability of a 1 on the outcome for the control condition was 0.37, which was indeed equal to the event rate for the control group.

with Equation 12.8 and found the following AMEs:

```
Average marginal effects

          AME     SE       z      p  lower  upper
    bs1 0.1120 0.0251 4.4618 0.0000 0.0628 0.1612
     cq 0.0812 0.0324 2.5042 0.0123 0.0176 0.1447
  treat 0.1897 0.0251 7.5595 0.0000 0.1405 0.2389
```

The average marginal effect for the treatment dummy variable was 0.19 ±0.05 ($z = 7.56$, $p < 0.05$), which comports with the results for the MLPM. The lower limit of the 95% confidence interval for the AME (0.14 to 0.24) exceeds the meaningfulness standard for the total effect (≥ 0.08). Note that I do not need to calculate separate AMEs for different predictor profiles because the AME is defined in a way that it takes into account covariate values when estimating the total effect (see my earlier discussion of covariates in the calculation of AMEs). AMEs represent what we call **marginal effects** that collapse across covariates (albeit in a unique way) when characterizing the total effect of the intervention; profile analyses, by contrast, represent what are called **conditional effects** that hold the covariates constant at specific values for each profile and then compares different pairs of profiles. AMEs tell us how much change to expect in the proportion of parents who communicate with their child as a function of the intervention taking into account any non-linearities in the component probability curves and the distributions of the covariate values that likely exist in the population. By contrast, profile analysis tells us how two or more profiles of people defined by specific values on the predictors differ in the proportion of parents who communicate with their children. Each approach is informative, but in different ways, which is why I like to use them both.

The average marginal effect program on my website does not provide the marginal means for the treatment and control groups that are differenced to obtain the AME. These can be obtained using the R program on my website called *Profile analysis*. Here is the program output for first the intervention condition and then the control condition:

```
Average adjusted prediction values for profile
 at(treat) Prediction      SE  lower  upper
         1      0.564 0.01806 0.5286 0.5993

Average adjusted prediction values for profile
 at(treat) Prediction      SE  lower  upper
         0     0.3743 0.01742 0.3401 0.4084
```

The estimated marginal proportion for the intervention group is 0.56 ±0.04 and for the control group it is 0.37±0.03. Note that these values are close to the estimates for the MLPM. This generally is true of the MLPM when applied to binary outcomes.

There is a third method you can use to estimate the total effect of the intervention that uses the causal mediation framework as implemented in Mplus. I describe this method later in the context of FISEM. I generally prefer the above two methods to it.

Effect of the Intervention on the Targeted Mediators

The analysis of the effect of the program on the targeted mediators to address the question of whether and by how much the program impacts the mediators uses the identical methods as the MLPM with an MLR estimator because the endogenous outcome variables (the mediators) are the same in both the MLPM and the probit analyses. This represents a case where I shift estimators to meet analytic requirements vis-à-vis LISEM. I do not repeat the analyses here given their redundancy with the ones I reported earlier in the context of the MLPM.

Effects of the Mediators on the Outcome

My final task is to evaluate if the targeted mediators are related to the outcome and whether there are program effects on the outcome over and above the targeted mediators. I do so using probit analysis applied to Equation 12.7 which I repeat here for convenience but with the probit link indicated:

$$\text{Probit(COM3)} = a_4 + p_4\ \text{PA2} + p_5\ \text{PK2} + p_6\ \text{PE2} + p_7\ \text{TREAT} + b_{10}\ \text{BS1} + b_{11}\ \text{CQ1}$$

Specifying a meaningfulness standard for each mediator using the path coefficients p_4, p_5 and p_6 in the above equation is challenging because probit coefficients are not intuitive and because of the potential non-linear relationship between mediators and outcome probabilities. I approach the task by invoking the same general meaningfulness standards I used for the MLPM for mediator effects on the outcome. Recall that the meaningfulness standard for PA2 was a probability or proportion increase of 0.03 or greater in the outcome for each unit increase in PA2, for PK2 it was the same, and for PE2 it was a probability or proportion decrease of -0.03 or less in the outcome for each unit increase in PE2. Rather than focus on the probit coefficients per se to evaluate the effects of the mediators on the outcome, I focus on probabilities but in the context of probit modeling.

Table 12.9 presents the Mplus syntax for the probit analysis of the above equation.

Table 12.9: Syntax for Probit-Based Mediator Effects on Outcome

```
1. TITLE: Probit-based profile analysis  ;
2. DATA: FILE IS c:\mplus\communication.dat ;
3. VARIABLE:
4. NAMES ARE ID COM3 PA2 PK2 PE2 CQ1 PA1 PK1 PE1 TREAT BS1 ;
5. USEVARIABLES ARE COM3 PA2 PK2 PE2 CQ1 TREAT BS1 ;
6. CATEGORICAL ARE COM3 ;
7. MISSING ARE ALL (-9999) ;
8. ANALYSIS:
9. ESTIMATOR = ML ; LINK=PROBIT
10. MODEL:
11. COM3 ON PA2 PK2 PE2 TREAT BS1 CQ1 (p4 p5 p6 p7 b10 b11) ;
12. [COM3$1] (thresh) ;
13. OUTPUT: SAMP STANDARDIZED(STDYX) CINTERVAL TECH4 ;
```

All of the syntax should be familiar. On Line 11, I specify the equation to be analyzed. I do not request modification indices on the output line because the model is just-identified. Here are the results from the `MODEL RESULTS` section:

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

 COM3        ON
    PA2                0.313      0.083       3.785      0.000
    PK2                0.298      0.081       3.696      0.000
    PE2               -0.375      0.082      -4.593      0.000
    TREAT             -0.001      0.113      -0.012      0.990
    BS1                0.242      0.067       3.600      0.000
    CQ1                0.150      0.087       1.710      0.087
```

The coefficients for the three mediators are statistically significant. The coefficient for the treatment variable, `TREAT`, was statistically non-significant for its impact on communication over and above the three targeted mediators (coefficient = -0.001 ±0.23, z = 0.01, p < 0.99). As noted, evaluating effect sizes using the probit coefficients per se is daunting, so instead I turn to profile analyses and average marginal effects.

Profile Analysis

I described the profile analysis approach earlier for total effects. Here, I strategically define profiles based on Equation 12.7 using the `MODEL CONSTRAINT` command. I present in Table 12.10 eight profiles that I explored for the PA2 mediator, which I then repeated for the other mediators (ignore the last two columns of Table 12.10 for now, which contain results that I discuss later). Each sequential pair of lines in the table represents a contrast between two profiles. For a given contrast, I increase PA2 by one unit, from 0 (first line) to 1 (second line) and determine how this affects the probability of a 1 on the outcome. I chose a value of 0 for PA2 in the first line because the mean and median PA1 scores at baseline were close to 0. A value of 0 thus represents a "natural" level of perceived advantages that parents typically bring to the study prior to their participation in the intervention. It also maps onto the mean value of PA2 for the control condition. The second line increases the PA2 value by 1 unit for purposes of contrasting it with the prior line so that I can evaluate how much the probability of communication shifts for a one unit increase in PA2. Each line pair holds constant the other variables in the equation to a specific value to reflect the effect of a unit change in PA2 in different contexts. The first two contrasts (profiles 1 versus 2 and profiles 3 versus 4) are for males and the second two contrasts (profiles 5 versus 6 and profiles 7 versus 8) are for females. Two of the contrasts examine the impact of a one unit increase in PA2 on communication for individuals where the treatment condition dummy variable is set to 0 (the control condition; see profiles 1 versus 2 and profiles 5 versus 6) while the other two contrasts examine the same effect but for individuals where the treatment condition dummy variable is set to 1 (the intervention condition). In some contrasts, I set the other two mediators, PK2 and PE2, to values that were close to the mean of the baseline PK1 and PE1 measures, respectively (both values equaled 0 on their -3 to +3 metric). In other contrasts, I set the values for PK2 and PE2 to be near their post-intervention mean scores. I held communication quality constant at a value near the midpoint of its distribution (which is a score of 0), but in practice I might also explore other values for it. There are no set rules about which profiles to explore. You should choose ones that are substantively interesting and that you think might be revealing.

The key motivation for the contrasts is to determine the extent to which the impact of a unit change in PA2 on communication varies as a function of the different contexts defined by the other study variables. In both logit and probit models, the respective effects across the contrasts could be close to one another or they could vary considerably. Some methodologists suggest conducting profile contrasts for many contexts whereas others prefer fewer contrasts that are strategically chosen. I used the latter approach.

The second to the last column of Table 12.10 shows the estimated proportion of parents who communicated with their child about sex and pregnancy for the profile in question and the last column shows the difference in proportions for the two profiles comprising a contrast pair. Table 12.11 presents the Mplus syntax I used to generate these results. Note that the effects of PA2 on the outcome vary little across the contrast pairs.

Table 12.10: LISEM Probit-Based Profile Analysis

Profile	PA2	PK2	PE2	TREAT	BS1	CQ1	Proportion who Communicate	Proportion Difference
1	0	0	0	0	0	0	0.304	-
2	1	0	0	0	0	0	0.421	0.117*
3	0	.85	0	1	0	0	0.397	-
4	1	.85	0	1	0	0	0.521	0.124*
5	0	0	0	0	1	0	0.394	-
6	1	0	0	0	1	0	0.517	0.124*
7	0	.85	0	1	1	0	0.493	-
8	1	.85	0	1	1	0	0.616	0.123*

* p < 0.05

Table 12.11: Syntax for Probit Model Profile Analysis with LISEM

```
TITLE: Probit-based profile analysis  ;
DATA: FILE IS c:\mplus\communication.dat ;
VARIABLE:
  NAMES ARE ID COM3 PA2 PK2 PE2 CQ1 PA1 PK1 PE1 TREAT BS1 ;
  USEVARIABLES ARE COM3 PA2 PK2 PE2 CQ1 TREAT BS1 ;
  CATEGORICAL ARE COM3 ;
  MISSING ARE ALL (-9999) ;
 ANALYSIS:
   ESTIMATOR = ML ; LINK=PROBIT
 MODEL:
   COM3 ON PA2 PK2 PE2 TREAT BS1 CQ1 (p4 p5 p6 p7 b10 b11) ;
   [COM3$1] (thresh) ;
   [BS1] (m1) ; [CQ1] (m2) ;
 MODEL CONSTRAINT:
   NEW(PROBIT1 PROBIT2 PROBIT3 PROBIT4 PROBIT5 PROBIT6 PROBIT7 PROBIT8
   PROB1 PROB2 PROB3 PROB4 PROB5 PROB6 PROB7 PROB8 DIFF1 DIFF2 DIFF3
   DIFF4);
   PROBIT1 = -thresh+p4*0+p5*0+p6*0+p7*0+b10*0+b11*0;
   PROBIT2 = -thresh+p4*1+p5*0+p6*0+p7*0+b10*0+b11*0;
   PROBIT3 = -thresh+p4*0+p5*0.85+p6*0+p7*1+b10*0+b11*0;
   PROBIT4 = -thresh+p4*1+p5*0.85+p6*0+p7*1+b10*0+b11*0;
   PROBIT5 = -thresh+p4*0+p5*0+p6*0+p7*0+b10*1+b11*0;
   PROBIT6 = -thresh+p4*1+p5*0+p6*0+p7*0+b10*1+b11*0;
   PROBIT7 = -thresh+p4*0+p5*0.85+p6*0+p7*1+b10*1+b11*0;
```

```
25.   PROBIT8 = -thresh+p4*1+p5*0.85+p6*0+p7*1+b10*1+b11*0;
26.   PROB1 = PHI(PROBIT1);
27.   PROB2 = PHI(PROBIT2) ;
28.   PROB3 = PHI(PROBIT3);
29.   PROB4 = PHI(PROBIT4) ;
30.   PROB5 = PHI(PROBIT5);
31.   PROB6 = PHI(PROBIT6) ;
32.   PROB7 = PHI(PROBIT7);
33.   PROB8 = PHI(PROBIT8) ;
34.   DIFF1 = PROB2-PROB1 ;
35.   DIFF2 = PROB4-PROB3 ;
36.   DIFF3 = PROB6-PROB5 ;
37.   DIFF4 = PROB8-PROB7 ;
38. OUTPUT: SAMP STANDARDIZED(STDYX) CINTERVAL TECH4 ;
```

Most of the syntax in Table 12.11 should be familiar. On Line 11, I specify the equation to be analyzed. Lines 14 to 37 have the same “profile analysis” format as Table 12.9 where I applied profile analysis for the total effect. Here is the output:

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

New/Additional  Parameters
    PROBIT1           -0.512      0.060     -8.597      0.000
    PROBIT2           -0.199      0.100     -1.989      0.047
    PROBIT3           -0.260      0.090     -2.897      0.004
    PROBIT4            0.053      0.060      0.878      0.380
    PROBIT5           -0.270      0.058     -4.668      0.000
    PROBIT6            0.043      0.092      0.467      0.641
    PROBIT7           -0.018      0.095     -0.188      0.851
    PROBIT8            0.295      0.057      5.137      0.000
    PROB1              0.304      0.021     14.587      0.000
    PROB2              0.421      0.039     10.741      0.000
    PROB3              0.397      0.035     11.467      0.000
    PROB4              0.521      0.024     21.764      0.000
    PROB5              0.394      0.022     17.691      0.000
    PROB6              0.517      0.037     14.063      0.000
    PROB7              0.493      0.038     12.964      0.000
    PROB8              0.616      0.022     28.074      0.000
    DIFF1              0.117      0.032      3.611      0.000
    DIFF2              0.124      0.032      3.856      0.000
    DIFF3              0.124      0.033      3.764      0.000
    DIFF4              0.123      0.033      3.746      0.000
```

The predicted probit values are in the first 8 rows, followed by the conversion of those values to probabilities/proportions in the second 8 rows. The last four rows subtract the first probability of the contrast pair from the second probability of the contrast pair. Each of the probability differences are statistically significant and near the value of 0.12. The results also were comparable when I bootstrapped the data. Confidence intervals are presented on the Mplus output but I do not show them here in the interest of space. They can be used to add margins of error (MOEs) to the results in Table 12.10. A sense of the MOEs is obtained by doubling the standard errors (the `S.E.` column) of each statistic.

I would then repeat the profile analyses focusing on PK2 and PE2.

Average Marginal Effects

An alternative to profile analysis for analyzing mediator effects on outcomes is to use average marginal effects (AMEs). This approach is much more succinct than the profile analysis approach. I used the average marginal effects program on my website to calculate the AMEs for each mediator in Equation 12.7. Here is the program output:

```
Average marginal effects

           AME     SE        z       p   lower   upper
   bs1  0.0899 0.0247  3.6464 0.0003  0.0416  0.1382
   cq1  0.0555 0.0323  1.7148 0.0864 -0.0079  0.1188
   pa2  0.1161 0.0302  3.8395 0.0001  0.0568  0.1753
   pe2 -0.1392 0.0297 -4.6882 0.0000 -0.1973 -0.0810
   pk2  0.1106 0.0294  3.7598 0.0002  0.0529  0.1682
 treat -0.0005 0.0418 -0.0120 0.9904 -0.0825  0.0815
```

The AME for perceived advantages was 0.12 ±0.06 ($z = 3.84$, $p < 0.05$); for perceived knowledge it was 0.11 ±0.06 ($z = 3.76$, $p < 0.05$); and for perceived embarrassment, it was -0.14 ±0.06 ($z = 4.69$, $p < 0.05$). Each value is interpreted, roughly, as the estimated change in the proportion of parents who engage in communication with their child about sex and pregnancy given a unit increase in the mediator in question. They represent marginal effects. Note that the AME values are close to the coefficient values of the MLPM, which is often the case. The 95% confidence intervals were 0.06 to 0.18 for perceived advantages, 0.05 to 0.17 for perceived knowledge, and -0.20 to -0.08 for perceived embarrassment. All of the estimated effects were meaningful after taking into account sampling error and using the same meaningfulness standards I outlined for the MLPM.

The AME for the effect of the intervention on parental communication independent of the target mediators was -0.005 ±0.08 ($z = 0.01$, $p < 0.99$), with a 95% confidence interval of -0.08 to 0.08. It is not significantly predictive of parental communication over and above `PA2`, `PK2` and `PE2`.

Overall Conclusions for Probit-Based LISEM

The overall conclusions for the probit-based LISEM are fundamentally the same as those for the MLPM so I do not repeat them here. Using the logic of the joint significance test, I would again conclude from the analyses that perceived advantages and perceived knowledge both meaningfully mediate some of the effect of the program on the outcome, but that perceived embarrassment does not. Note that for the latter, my more detailed analyses showed that lack of mediation for perceived embarrassment was not due to perceived embarrassment being irrelevant to parental communication. Rather, it was because the intervention failed to meaningfully change perceived embarrassment. This dynamic would be missed if one focuses only on the omnibus mediational parameter. For a discussion of how to calculate omnibus mediation tests using probit-based modeling in an LISEM context, see the document on omnibus tests on the resources page.

A Note on Supplemental Profile Analyses

Before leaving this section, I want to mention one side note on the profile analyses described above with probit based modeling of RETs. It sometimes will be useful to conduct supplemental profile analyses in your data beyond the ones I have highlighted to further understand the implications of different patterns of total effects or of the effects of mediators on the outcome. Consider the probit equation that I analyzed earlier using the syntax in Table 12.9:

$$\text{Probit(COM3)} = a_4 + p_4\ \text{PA2} + p_5\ \text{PK2} + p_6\ \text{PE2} + p_7\ \text{TREAT} + b_{10}\ \text{BS1} + b_{11}\ \text{CQ1}$$

I or my clients might want to know what the predicted proportion of parents who talk with their child about sex and pregnancy would be if the intervention achieved scores of 1 in the desired direction on their -3 to +3 metrics on all three mediators simultaneously, i.e., a 1 on PA2, a 1 on PK2 and a -1 on PE2. Or, what the proportion would be if the intervention had more modest effects of 0.33 in the desired direction on all three mediators when considered multivariately? Or, what the proportion would be if the intervention focuses on just two mediators and the program brings perceptions on those mediators to values of 1.0 in the desired direction, how do different combinations of two mediators impact the proportion of parents who discuss sex and pregnancy with their child (e.g., PA2 and PK2 versus PA2 and PE2 versus PK2 and PE2). Or, I might want to know if the intervention targets individuals in the control condition who are particularly low on all three mediators (scores of 1 in the undesired direction) and we manage to bring their scores to near 0 for all three mediators vis-à-vis the intervention, what will the effects be on the proportion of those parents who discuss sex and pregnancy with their child? Here is the syntax that you can use to explore these types of questions using profile analysis for females setting communication quality to a score of 0, which is near its mean value:

```
TITLE: Probit-based profile analysis  ;
DATA: FILE IS c:\mplus\communication.dat ;
VARIABLE:
NAMES ARE ID COM3 PA2 PK2 PE2 CQ1 PA1 PK1 PE1 TREAT BS1 ;
  USEVARIABLES ARE COM3 PA2 PK2 PE2 CQ1 TREAT BS1 ;
  CATEGORICAL ARE COM3 ;
  MISSING ARE ALL (-9999) ;
ANALYSIS:
  ESTIMATOR = ML ; LINK=PROBIT
MODEL:
  COM3 ON PA2 PK2 PE2 TREAT BS1 CQ1 (p4 p5 p6 p7 b10 b11) ;
  [COM3$1] (thresh) ;
MODEL CONSTRAINT:
   NEW (PROFILE1 PROFILE2 PROFILE3 PROFILE4 PROFILE5
   PROFILE6 PROFILE7 PROFILE8 ) ;
   PROFILE1 = PHI(-thresh+p4*0+p5*0+p6*0+p7*0+b10*1+b11*0) ;
   PROFILE2 = PHI(-thresh+p4*1+p5*1+p6*(-1)+p7*1+b10*1+b11*0) ;
   PROFILE3 = PHI(-thresh+p4*.33+p5*.33+p6*(-.33)+p7*1+b10*1+b11*0) ;
   PROFILE4 = PHI(-thresh+p4*1+p5*1+p6*0+p7*1+b10*1+b11*0) ;
   PROFILE5 = PHI(-thresh+p4*1+p5*0+p6*(-1)+p7*1+b10*1+b11*0) ;
   PROFILE6 = PHI(-thresh+p4*0+p5*1+p6*(-1)+p7*1+b10*1+b11*0) ;
   PROFILE7 = PHI(-thresh+p4*(-1)+p5*(-1)+p6*(1)+p7*0+b10*1+b11*0) ;
   PROFILE8 = PHI(-thresh+p4*0+p5*0+p6*0+p7*1+b10*1+b11*0) ;
 OUTPUT: SAMP STANDARDIZED(STDYX) CINTERVAL TECH4 ;
```

All of this syntax should be familiar to you. To conduct the analysis for males, I repeat the above syntax but set the score for `b10` to 0 for the profiles. Here are the results I found for males and females considered separately:

	Female %	Male %
Profile 1: PA2=0, PK2=0, PE2=0, TREAT=0	39.4 ±4.4	30.4 ±4.2
Profile 2: PA2=1, PK2=1, PE2=-1, TREAT=1	76.3 ±6.0	68.2 ±7.2
Profile 3: PA2=.33, PK2=.33, PE2==.33, TREAT=1	52.2 ±6.8	42.5 ±6.8
Profile 4: PA2=1, PK2=1, PE2=0, TREAT=1	63.3 ±4.4	53.9 ±4.8
Profile 5: PA2=1, PK2=0, PE2=-1, TREAT=1	66.2 ±8.4	56.9 ±9.6
Profile 6: PA2=0, PK2=1, PE2=-1, TREAT=1	65.6 ±8.8	56.3 ±9.4
Profile 7: PA2=-1, PK2=-1, PE2=1, TREAT=0	10.4 ±6.0	6.7 ±4.2

Profile 8: PA2=0, PK2=0, PE2=0, TREAT=1	39.3 ±8.8	30.4 ±7.8

Although I did not do so, one can apply significance tests that compare any given pair of profiles, as I did earlier for my other profile analyses. I leave it to you to impose substantive interpretations onto the above results but having the ability to pursue such comparative profile analyses is a strength. In essence, as you think about different counterfactuals in the context of program evaluation, profile analysis can be useful for exploring them.

FISEM ANALYSIS: THE PROBIT MODEL

In this section, I show you how to use FISEM with probit modeling to analyze an RET. An important decision in such analyses is whether to use traditional maximum likelihood estimation with a probit link or to use WLSMV estimation (I show Bayesian estimation later in the chapter). The advantage of WLSMV is that it yields a fuller set of global fit indices as well as modification indices. The more traditional maximum likelihood approach does not. WLSMV, however, can be problematic when estimating exogenous means and correlations between the exogenous variables in your model. It also handles missing data less elegantly than the ML strategy; it uses pairwise deletion of missing data whereas the ML strategy uses FIML. WLSMV generally yields larger standard errors than ML and, hence, it tends to have less statistical power. It also is not as robust to normality violations when continuous variables are in the model. Technically, WLSMV is not even a pure form of FISEM, but as implemented in Mplus it does capture the spirit of FISEM as defined in this book. For more about WLSMV, see Muthén et al. (1997). Given the above, I generally prefer to use the ML estimator to the WLSMV estimator.

Having said that, when I conduct preliminary analyses for an FISEM probit analysis, I often use the WLSMV estimator as a way of exploring possible specification error in my model to identify major points of stress I should be concerned about. These stress points are likely to apply to my maximum likelihood based probit model as well, so as a preliminary analysis, I find WLSMV based runs to be helpful. Technically, in FISEM, one should not shift estimators in this way willy nilly. However, sometimes I find the strategy of conducting preliminary analyses using WLSMV as a form of checks on specification error to be helpful.[10] When I used WLSMV for our numerical example, all of the global fit indices pointed to good model fit. The syntax for the WLSMV model is in Table 12.12. Most of it follows from already developed programming principles.

[10] Because WLSMV handles missing data differently than FISEM maximum likelihood with a probit link, if you have considerable missing data, the utility of WLSMV as a preliminary check on specification error can degenerate.

Table 12.12: Mplus Syntax for FISEM Probit-Based Model Using WLSMV

```
1.  TITLE: FISEM Probit analysis of RET using WLSMV  ;
2.  DATA: FILE IS c:\mplus\communication.dat ;
3.  VARIABLE:
4.    NAMES ARE ID COM3 PA2 PK2 PE2 CQ1
5.    PA1 PK1 PE1 TREAT BS1 ;
6.  USEVARIABLES ARE COM3 PA2 PK2 PE2 CQ1
7.    PA1 PK1 PE1 TREAT BS1 ;
8.  CATEGORICAL ARE COM3 ;
9.  MISSING ARE ALL (-9999) ;
10. ANALYSIS:
11.   ESTIMATOR = WLSMV ;
12. MODEL:
13.   PA2 ON BS1 CQ1 TREAT PA1  ;
14.   PK2 ON BS1 CQ1 TREAT PK1  ;
15.   PE2 ON BS1 CQ1 TREAT PE1  ;
16.   COM3 ON PA2 PK2 PE2 TREAT BS1 CQ1 ;
17. OUTPUT: SAMP STANDARDIZED(STDYX) MOD(ALL 4) RESIDUAL CINTERVAL TECH4;
```

Line 11 specifies the WLSMV estimator. I do not need to specify the probit link because WLSMV uses it by default.

Table 12.13 presents the Mplus syntax for the main analysis that uses maximum likelihood.

Table 12.13: Mplus Syntax for FISEM Probit-Based Model Using ML

```
1.  TITLE: FISEM Probit analysis of RET  ;
2.  DATA: FILE IS c:\mplus\communication.dat ;
3.  VARIABLE:
4.    NAMES ARE ID COM3 PA2 PK2 PE2 CQ1
5.    PA1 PK1 PE1 TREAT BS1 ;
6.  USEVARIABLES ARE COM3 PA2 PK2 PE2 CQ1
7.    PA1 PK1 PE1 TREAT BS1 ;
8.  CATEGORICAL ARE COM3 ;
9.  MISSING ARE ALL (-9999) ;
10. ANALYSIS:
11.   ESTIMATOR=ML ; LINK=PROBIT ;
12. MODEL:
13.   PA2 ON BS1 CQ1 TREAT PA1 (b1 b2 p1 b3) ;
14.   PK2 ON BS1 CQ1 TREAT PK1 (b4 b5 p2 b6) ;
15.   PE2 ON BS1 CQ1 TREAT PE1  (b7 b8 p3 b9) ;
16.   COM3 ON PA2 PK2 PE2 TREAT BS1 CQ1 (p4 p5 p6 p7 b10 b11) ;
17.   [COM3$1] (thresh) ;
18. MODEL INDIRECT:
19.   COM3 IND TREAT ;
20. OUTPUT: SAMP STANDARDIZED(STDY) RESIDUAL CINTERVAL TECH4;
```

All of the syntax should be self-explanatory. Line 20 asks for standardized output but note instead of `STDYX` I use `STDY`. I explain why shortly.

Model Fit

I first examine global fit indices on the output for the WLSMV analysis to determine if the data are in accord with the overall model in Figure 2.2. The chi square test of perfect population model fit was statistically non-significant ($\chi^2(12) = 13.22$, $p < 0.35$), which is consistent with good model fit. The RMSEA was 0.008; the upper limit of the 90% confidence interval for it was 0.028; the CFI was 1.00; the standardized RMR was 0.013. All suggest good fit.

Next, I examine localized fit indices reported on the output for the WLSMV analysis. Mplus does not provide formal tests of the difference between predicted and observed covariances for probit modeling but it does provide a matrix of disparities between predicted and observed correlations in the section called `Residuals for Correlations for the Joint Model` on the output. Inspection of this matrix on a cell-by-cell basis did not reveal notable disparities (e.g., a deviation between predicted and observed correlations of 0.08 or greater). For the modification indices, there was only one that was above 4. It was to include perceived advantages at baseline (`PA1`) in the equation predicting perceived embarrassment at posttreatment (`PE2`) from `PE1`, `BS1`, `CQ1`, and `TREAT`. If I take into account the number of modification indices tested in total (24), this result could readily be chance induced. As well, it does not make theoretical sense, nor does including it affect any of my major conclusions. Finally, I know for a fact that the result is a Type I error given the way I generated the simulated data, although I would not have such knowledge in practice.

In the ML based analyses, I can pursue tests of model fit using the strategy I elaborated for the LISEM probit analyses that focuses on predicted independence conditions. I do not pursue these analyses here because you already know how to do so based on the material I presented earlier in this chapter. The tests were generally supportive of reasonable model fit.

Total Effect of the Intervention on the Outcome

The first question I address is whether and to what extent the intervention affected the outcome. For a binary outcome, we typically want to characterize for our clients the total effect of the intervention in the form of a proportion difference. Calculating the proportion difference is not straightforward in either FISEM probit modeling or FISEM logit modeling. In this section, I first show you how Mplus approaches the calculation of total program effects using FISEM but invoking the latent propensity framework to probit analysis discussed in Chapter 5 (which you may want to review at this point in the current chapter). I then discuss strategies for characterizing the more practical proportion difference between the treatment and control groups.

Total Effect using the Latent Propensity Framework

The total effect of the program on the outcome is generated by Mplus using the `COM3 IND T` command on Line 22 of Table 12.13. The section of the output where the total effect is reported is `TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS.` In this section, Mplus reports the estimated total effect of the program on the y* latent propensity. Unfortunately, it is difficult to interpret the reported result because the metric of y* is arbitrary and difficult to interpret (see Chapter 5). One solution to this problem suggested by Karlson (2015) is to standardize y* so that across all individuals, y* has a mean of 0 and a SD = 1.0. By using the option `STDY` instead of `STDYX` on the output line of the Mplus syntax, Mplus standardizes y* when calculating the total effect but not `TREAT`, the treatment dummy variable. The result is in the subsection called `STDY Standardization`. Here is the output:

```
STDY Standardization
                                                         Two-Tailed
                    Estimate       S.E.     Est./S.E.      P-Value

  Total                0.471      0.060       7.911       0.000
  Total indirect       0.472      0.086       5.490       0.000
```

The treatment-control group mean difference on the underlying propensity to talk with one’s child about sex was, in standard deviation units, 0.47 ±0.12, which is statistically significant (z = 7.91, $p < 0.05$). Karlson (2015) notes that the value of 0.47 is analogous to a Cohen’s *d* but note that it uses a different referent standard deviation than Cohen’s *d* per my discussion in Chapter 10; so, the indices are analogous but not equivalent.

A disadvantage of the Karlson approach with RETs is that the standard deviation of y* may not be all that meaningful, hence it may be not be a reasonable referent for standardizing the y* mean difference. The standard deviation of y* reflects variability in y* from a hypothetical population of individuals in which approximately half of the population has been exposed to an artificial intervention to increase or decrease y* and half not (the control group). This is not a very realistic population, but perhaps it applies to a context of interest to you. Also, most clients for evaluations have a difficult time understanding the Karlson metric (or Cohen’s *d* for that matter) so for them it is not particularly intuitive. I prefer characterizing total effects as proportion differences.

Total Effect as a Proportion Difference

As noted, obtaining an estimate of the total program effect in the form of a proportion difference in probit based FISEM is not straightforward, especially if there are complex relationships between the mediators. I personally think it is easiest to approach the question of total effects using the limited information probit approach described earlier vis-à-vis the equation

$$\text{probit(COM3)} = a + p_1\,T + b_2\,CQ1 + b_3\,BS1$$

Shifting to a LISEM equation violates the principle of model coherence among path coefficients in FISEM but it usually gets the job done for practical purposes. I do not restate here how to estimate the total effect with LISEM probit applied to the above equation; consult the text from the earlier sections in this chapter that did so. One can use either a profile analysis approach or an approach based on AMEs, or both.

Parenthetically, a third LISEM approach that you should be aware of for estimating the total effect uses the causal mediation framework as implemented in Mplus. The strategy I describe here only works for models in which there are no causal relationships among the mediators, which is the case for the current example. It also presumes that all major confounds are controlled, either by design or through statistical controls vis-à-vis the inclusion of covariates. To implement the approach, I select one of the mediators of interest, say PA2 (the posttreatment perceived advantages mediator). I then form a single mediator causal model per Figure 12.4. Note that the model should include measured confounds to control for confound bias. The omnibus mediational effect in this model for PA2 is captured, in theory, by the product of paths *a* and *b*. Path *c* reflects the impact of the treatment on Y of all omitted mediators on the outcome other than the particular target mediator being evaluated. This includes the other measured mediators in the full model because these mediators are explicitly omitted from the current analysis even though we have measures of them. It turns out that Mplus can estimate the total intervention effect from this model using the information from the various paths and does so in proportion form by invoking the causal mediation framework described in Chapter 8 (for details, see Muthén, Muthén & Asparouhov, 2016).

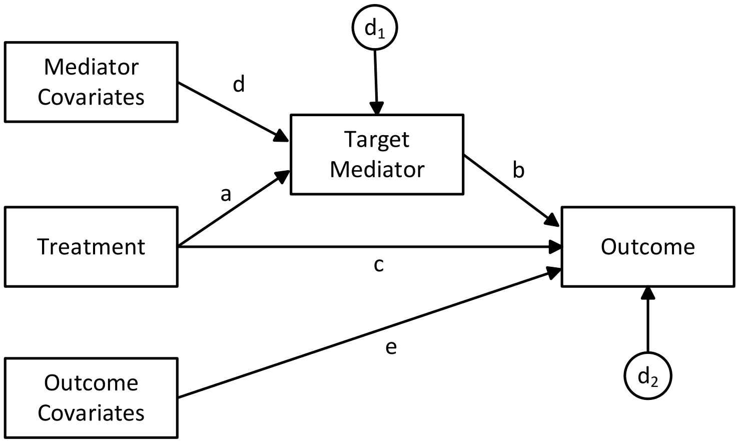

FIGURE 12.4. Single mediator causal mediation model

Table 12.14 presents the Mplus code I use to perform the analysis.

Table 12.14: Mplus Syntax for Causal Mediation Total Effect

```
1.  TITLE: Causal mediation analysis for probit model  ;
2.  DATA: FILE IS c:\mplus\communication.dat ;
3.  DEFINE:
4.  CENTER CQ1 BS1 PA1 (GRANDMEAN) ;
5.  VARIABLE:
6.    NAMES ARE ID COM3 PA2 PK2 PE2 CQ1
7.    PA1 PK1 PE1 TREAT BS1 ;
8.  USEVARIABLES ARE COM3 PA2 CQ1 PA1 TREAT BS1 ;
9.  CATEGORICAL ARE COM3 ;
10.  MISSING ARE ALL (-9999) ;
11. ANALYSIS:
12.   ESTIMATOR=ML ; LINK=PROBIT ;
13. MODEL:
14.   PA2 ON BS1 CQ1 TREAT PA1  ;
15.   COM3 ON PA2 TREAT BS1 CQ1 ;
16. MODEL INDIRECT:
17.   COM3 IND PA2 TREAT ;
18. OUTPUT: SAMP STANDARDIZED(STDYX) RESIDUAL CINTERVAL TECH4 ;
```

On Lines 3 and 4, I use the DEFINE statement to mean center all the covariates in the analysis. This is important because Mplus assumes meaningful zeros for covariates when applying the causal mediation framework and we want to set the zero points to variable means. The remaining syntax should be self-explanatory except for Line 17. This line specifies the customary indirect effect (COM3 IND TREAT) but note that I have added the PA2 target mediator between the outcome (COM3) and the distal predictor (TREAT). This tells Mplus to use the causal mediation approach. Here is the key output:

Table 12.15: Causal mediation output

```
Effects from TREAT to COM3

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

  Tot natural IE       0.101      0.026      3.848      0.000
  Pure natural DE      0.092      0.036      2.516      0.012
  Total effect         0.193      0.025      7.593      0.000
```

The relevant portion of the output for us is the Total effect row (I explain the other rows later when I consider omnibus indirect effects). The entry in that row in the Estimate column is the estimated proportion difference in the outcome for the intervention group minus the control group, 0.193 (critical ratio = 7.59, $p < 0.05$). The output also provides confidence intervals (not shown here) so you can calculate a margin of error for the total effect. Note that the result is quite close to the results for the MLPM LISEM analysis as well as the probit-based LISEM analyses from earlier.

I used the PA2 mediator for the above analysis. I can repeat the analysis using the other mediators (PK2 or PE2) in place of PA2 and each should yield the same total effect or a value very close to it. That was the case in the present example.

In sum, you can use any of or all three of the above methods for estimating the total effect of the intervention on the outcome, as appropriate. Based on the analyses I conducted, the overall effect of the intervention was to increase the proportion of parents communicating with their children about sex and birth control by about 0.19 ±0.05, a statistically significant difference. The 95% confidence interval was 0.14 to 0.24. The lower limit of this interval exceeds the meaningfulness standard of ≥0.08, so I conclude the program had a meaningful effect. If I rely on the analysis of y*, I also conclude for a statistically significant total effect but I need to then specify a meaningful effect size standard for the y* intervention versus control difference in y* units, which I personally find difficult to do.

Effect of the Intervention on the Target Mediators

The next question asks what the effect of the intervention is on each mediator. I decide to use the same meaningfulness standards as those for the LISEM using MLPM because they make logical sense. There are three relevant equations I need to consult from executing the syntax in Table 12.13, one equation per mediator (see Equations 12.4 to 12.7). Here is the relevant output for PA2 from the analysis:

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

    BS1               0.097      0.020      4.929      0.000
    CQ1               0.092      0.026      3.602      0.000
    TREAT             0.823      0.020     41.761      0.000
    PA1               0.355      0.026     13.896      0.000
```

The coefficient of interest is that for TREAT, which reflects the covariate adjusted mean difference between the treatment and control conditions. The difference (on the -3 to +3 metric of perceived advantages) was 0.82 ±0.04 (z = 41.76, p < 0.05). The 95% confidence interval for the mean difference was 0.78 to 0.86. The lower limit of this interval exceeds the meaningfulness standard for PA2 (≥0.22), so I conclude the program had a meaningful effect on PA2.

The output for PK2 and PE2 yielded covariate adjusted mean differences between the treatment and control conditions of 0.84 ±0.04 (z = 41.36, p < 0.05) and 0.007 ±0.04 (z = 0.37, p < 0.72), respectively. The 95% confidence interval for the PK2 mean difference was 0.80 to 0.88. The lower limit of this interval exceeds the meaningfulness standard for PK2 (≥0.22), so I conclude the program had a meaningful effect on PK2. The 95% confidence interval for the PE2 mean difference was -0.032 to 0.046. The interval is contained within the latitude of no effect, so I conclude the effect of the program on PE2 was functionally nil.

In sum, the FISEM analysis suggests the program was effective in bringing about meaningful change in parental perceived advantages of engaging in conversations with their child and judgments of their perceived knowledge for doing so. However, the program did not meaningfully affect anticipated embarrassment of such conversations.

In the above analyses, I did not estimate the separate covariate adjusted mediator means for each of the two treatment conditions. I can do so with a simple modification of the syntax in Table 12.13 and then re-running the syntax. The change adds the DEFINE: command after Line 2 followed by the command that mean centers the covariates, namely CENTER CQ1 BS1 PA1 PK1 PE1 (GRANDMEAN);. On the output, the intercept for PA2, which equaled 0.06 ±0.03, is the covariate adjusted PA2 mean for the control group. If I reverse score TREAT and re-run the analysis, the intercept is the covariate adjusted PA2 mean for the intervention group, which was 0.88 ±0.03. The difference between the two values is 0.82, which is the coefficient for TREAT. I can do the same for the other mediators, namely PK2 and PE2, as desired.

Effects of the Mediators on The Outcome

To evaluate the effects of the target mediators on the outcome, I focus on Equation 12.7, which I repeat here for convenience

$$\text{Probit(COM3)} = a_4 + p_4\,\text{PA2} + p_5\,\text{PK2} + p_6\,\text{PE2} + p_7\,\text{T} + b_{10}\,\text{BS1} + b_{11}\,\text{CQ1}$$

As with probit-based LISEM, I can test the statistical significance of the coefficients from the probit-based coefficients that result from executing the syntax in Table 12.15. Here are the results from the `MODEL RESULTS` section:

```
                                                     Two-Tailed
                    Estimate       S.E.   Est./S.E.    P-Value
 COM3      ON
    PA2                0.313      0.083      3.785      0.000
    PK2                0.298      0.081      3.696      0.000
    PE2               -0.375      0.082     -4.592      0.000
    TREAT             -0.001      0.113     -0.012      0.991
    BS1                0.242      0.067      3.600      0.000
    CQ1                0.150      0.087      1.710      0.087
```

All three of the mediators had a statistically significant effect on the outcome but the treatment variable did not do so over and above the mediators. However, as before, the probit coefficient values are difficult to interpret. My preference is to use the meaningfulness standards I specified for the MLPM. I can do if I invoke either average marginal effects (AMEs) or profile analysis when evaluating Equation 12.7.

Mplus does not offer an option for calculating AMEs but you can stay within the FISEM framework and estimate AMEs using the Cameron and Trivedi (2010) approach explicated in Appendix A. The example in the Appendix presents the Mplus syntax for `PA2` in the current example. The AME for perceived advantages was 0.11, for perceived knowledge it was 0.11 and for perceived embarrassment it was -0.13. The AME for the effect of the treatment condition independent of the other predictors was effectively zero. These results are similar to those for the probit-based LISEM analysis.

For the profile analysis, I use the syntax in Table 12.13 but I add `MODEL CONSTRAINT` commands in the spirit of the profile analyses I described for probit-based LISEM. As with the LISEM analysis, I did not find that the proportion differences changed much as a function of varying covariate contexts.

In sum, the conclusions about the three mediators are fundamentally the same for the FISEM probit-based model as they were for the LISEM analyses. Each of the targeted mediators seem relevant.

Overall Conclusions for Probit-Based FISEM

The overall conclusions for the probit-based FISEM are fundamentally the same as those for the LISEM analyses so I do not repeat them here. Using the logic of the joint significance test, I would again conclude from the analyses that perceived advantages and perceived knowledge both mediate some of the effect of the program on the outcome, but that perceived embarrassment does not. For discussion of how to calculate omnibus mediation tests and effect sizes using probit-based modeling in an FISEM context, see the document on omnibus tests on the resources tab of my webpage. Also, keep in mind that, as illustrated with LISEM probit analysis, it is possible to conduct supplemental profile analyses using probit-based FISEM to explore various counterfactuals that may be of interest for the particular program evaluation you are pursuing.

FISEM ANALYSIS: BAYESIAN MODELING

In this section, I illustrate Bayes modeling for the communication example. Bayes analysis with a binary outcome in Mplus relies on probit modeling by default. I used the syntax in Table 12.13 with the following modifications, I (a) changed Line 11 to `ESTIMATOR=BAYES; BITERATIONS=100000(50000); BCONVERGENCE=.01 ;`, (b) removed `SAMP` and `MOD (ALL 4)` from the output line and changed `CINTERVAL` to `CINTERVAL(HPD)` and added `TECH8` and (c) added `PLOT: TYPE=PLOT2;` to the last line of the program. To save space, I do not describe in detail the strategies for answering the three fundamental questions of an RET because the strategies are nearly identical to what I presented for probit-based FISEM. I only highlight selected results. I assume you have read the material on Bayesian SEM in Chapter 8.

Model Fit

The model yielded a PSR = 1.00 for the parameter with the largest PSR at the final iteration of the Baeysian algorithm, suggesting model convergence. There were no problematic Kolmogorov-Smirnov values reported by Mplus. Inspection of the Mplus generated plots, per Chapter 8, also suggested good convergence. The posterior predictive p-value was 0.48, which is consistent with a good model fit. The 95% confidence interval for the difference between the observed and the replicated chi-square values was -22.29 to 22.90, which also is consistent with good model fit. Bayesian models with categorical outcomes do not yield CFI or RMSEA statistics, so these indices could not be evaluated. Nor are modification indices provided. In Mplus, Bayes output shows the predicted correlations between variables (in the section called `ESTIMATED CORRELATION MATRIX FOR THE LATENT VARIABLES`). When I inspected these correlations and compared them to the observed correlations on a cell-by-cell basis, I did not find any notable disparities. Overall, the model seems to be consistent with the data. Given the dearth of fit indices, I often conduct a preliminary model fit analysis using the WLSMV estimator rather than the Bayesian estimator and resolve blatant ill fit using the full armament of global and local fit indices that come with probit based WLSMV. Once resolved, I then conduct my Bayesian analysis and examine the fit indices it provides. This approach is not ideal but I find it to be a helpful check on specification error, much like I did earlier with maximum likelihood based probit analysis. In the present case, there was no evidence for specification error.

Total Effect of Intervention on the Outcome

In the FISEM probit analysis, I used the `STANDARDIZED(STDY)` option on the output line to evaluate the total program effect on the standardized y* means for the treatment versus control condition. This approach also is used for Bayes estimation. From the same section of the output as the FISEM probit model, the estimated standardized mean difference on y* for the intervention minus control conditions was 0.47 (95% credible interval = 0.35 to 0.58, $p < 0.05$). This difference is interpreted just as it is for the FISEM probit-based analysis, namely it is the standardized mean difference between the treatment and control conditions, analogous to a Cohen's *d*.

For the maximum likelihood based FISEM probit approach, I recommended shifting to a limited information estimation strategy to express overall proportion differences between the treatment and control groups for purposes of documenting the total effect of the intervention. I make the same recommendation for the Bayesian analysis. You use the same syntax as in Table 12.7 (with the accompanying profile analyses) and Table 12.14 but change the estimator from maximum likelihood to Bayes and modify the `OUTPUT` line, exactly as I did above. Here are the results for the profile based analyses from the modified syntax from Table 12.7:

```
                               Posterior  One-Tailed         95% C.I.
                    Estimate       S.D.      P-Value   Lower 2.5%  Upper 2.5%  Sig

New/Additional  Parameters
    CPROBIT           -0.325       0.046      0.000      -0.416      -0.235      *
    TPROBIT            0.163       0.046      0.000       0.072       0.255      *
    CPROB              0.373       0.018      0.000       0.339       0.407      *
    TPROB              0.565       0.018      0.000       0.529       0.600      *
    DIFF               0.192       0.025      0.000       0.142       0.242      *
```

The estimated proportion of parents in the treatment condition who talked with their children about sex when the covariates are at their "typical" values was 0.57 (95% credible interval = 0.53 to 0.60). The corresponding proportion for the control condition was 0.37 (95% credible interval = 0.34 to 0.41). The difference was 0.20 (95% credible interval = 0.15 to 0.25, p < 0.05). As with the probit-based FISEM, I would evaluate how robust the total effect estimate across different values of the covariates using profile analysis. I also need to be cautious about intercept interpretation when centering binary covariates, per my earlier discussion.

Here are the results for the analysis based on the causal mediation framework per the modified syntax from Table 12.14 (with appropriate changes for Bayes estimation) and where the relevant information is in the final row for the `Total effect`:

```
                              Posterior  One-Tailed         95% C.I.
                   Estimate       S.D.      P-Value   Lower 2.5%  Upper 2.5%   Sig

Effects from TREAT to COM3

    Tot natural IE      0.101       0.026       0.000       0.049       0.152       *
    Pure natural DE     0.092       0.036       0.005       0.021       0.164       *
    Total effect        0.193       0.025       0.000       0.143       0.243       *
```

The entry in that row in the `Estimate` column is the estimated proportion difference in the outcome for the intervention group minus the control group, 0.193 (95% credible interval = 0.14 to 0.24, p < 0.05). Mplus does not provide the component proportions used to create the proportion difference.

All of these results are close to the results for the MLPM LISEM analysis, the probit-based LISEM analysis and the ML FISEM probit analysis.

Effects of the Intervention on the Target Mediators

There are three equations for evaluating program effects on the mediators, one equation per mediator. Here is the relevant (edited) output for `PA2` for the syntax that mean centered the baseline covariates:

```
                            Posterior          95% C.I.
               Estimate       S.D.      Lower 2.5%   Upper 2.5%    Sig
PA2 ON
    BS1           0.097        0.020        0.058        0.136       *
    CQ1           0.092        0.026        0.042        0.142       *
    TREAT         0.822        0.020        0.783        0.861       *
    PA1           0.355        0.026        0.305        0.405       *
```

All variables but `TREAT` are covariates. The coefficient for `TREAT`, which reflects the covariate adjusted mean difference between the treatment and control conditions was

0.82 (95% credible interval = 0.78 to 0.86, p < 0.05). Comparable analyses for PK2 and PE2 yielded covariate adjusted mean differences between the treatment and control conditions of 0.84 (95% credible interval = 0.80 to 0.88) and 0.00 (95% credible interval = -0.03 to 0.05). Again, I apply the same strategies for interpreting the effects as for probit FISEM using ML.

Effects of the Mediators on the Outcome

The analysis to determine the effects of the mediators on the outcome uses the same approach as FISEM probit modeling. Here are the Equation 12.7 probit coefficients from executing the Bayes modified syntax in Table 12.13.:

```
                           Posterior           95% C.I.
               Estimate      S.D.      Lower 2.5%   Upper 2.5%   Sig

COM3       ON
   PA2           0.313       0.083      0.000       0.147       0.471       *
    PK2          0.298       0.081      0.000       0.144       0.462       *
    PE2         -0.375       0.081      0.000      -0.534      -0.214       *
    T            0.000       0.114      0.498      -0.220       0.226
    BS1          0.243       0.067      0.000       0.111       0.372       *
    CQ1          0.149       0.088      0.045      -0.023       0.321
```

All three of the mediators yielded statistically significant coefficients, but the coefficient for the independent effect of TREAT over and above the mediators was not statistically significant. These coefficients, as before, are difficult to interpret. For the probit-based FISEM, I described two methods for obtaining more intuitive indices, one using average marginal effects as outlined in Appendix A and the other using a profile analysis. Both of these methods can be applied in the same way for the Bayesian analysis. I leave these analyses as an exercise for you. However, they yielded results very similar to the traditional probit-based FISEM.

Overall Conclusions for Bayesian FISEM

Bayesian modeling yielded much the same results as the probit-based FISEM as well as the LISEM approaches despite employing a different statistical philosophy. I used uninformative priors in my analyses, but it is possible to use informative priors, as appropriate. Although I do not discuss it, one also can use a LISEM version of Bayesian SEM by analyzing separately the model equations (Equations 12.3 to 12.7), like I did in Chapter 8. The omnibus mediation tests in Bayesian modeling are approached exactly as they are for the FISEM probit-based analyses. See the document on omnibus indirect effects for the current chapter on the resources tab of my web page for details.

FISEM ANALYSIS: THE MODIFIED LINEAR PROBABILITY MODEL

In this section, l illustrate FISEM analysis using the MLPM. Because of the controversial nature of the linear probability model, you will not find many applications of the MLPM in an FISEM context. Lindon and Karlson (2012) applied such an approach in a small scale simulation of a binary treatment variable, a binary mediator, and a binary outcome. They found that the linear probability model (using OLS, not the MLPM) showed non-trivial bias when estimating the omnibus indirect effect when the outcome variable had a base rate of 95% in the 0 category of the outcome. However, this also was true of the other methods in their simulation that included methods proposed by Imai et al. (2010a, b, c), propensity methods by VanderWeele (2009), and y* standardized methods. Also, the focus of the simulation was not on Type I or Type II errors nor confidence interval coverage but instead on the ratio of the indirect effect to the total effect, an index that has been shown to be problematic (see Chapter 11). Barrett and colleagues (Barret, 2019; Barrett, Lockhart & Cruz, 2022) have applied binary marginal effects analysis to mediation modeling. To the extent the MLPM yields estimates that closely approximate marginal effects for binary outcomes (Angrist & Pischke, 2009), this supports the use of the MLPM in such SEM contexts. It, of course, is not appropriate to use the MLPM when the true functions linking variables are non-linear, just as it is not appropriate to use a probit or logit model when the true functions are not probit or logit. Even when linear functions are present, we simply do not know much about how the MLPM performs in FISEM contexts. As you will see, for the current example, it produces much the same results as the other forms of analysis.

Table 12.15 presents the relevant Mplus syntax for the MLPM.

Table 12.15: Mplus Syntax for the Modified Linear Probability Model

```
TITLE: Using the MLPM  ;
DATA: FILE IS c:\mplus\communication.dat ;
DEFINE:
CENTER PA1 PK1 PE1 CQ1 BS1 (GRANDMEAN) ;
VARIABLE:
NAMES ARE ID COM3 PA2 PK2 PE2 CQ1 PA1 PK1 PE1 TREAT BS1 ;
USEVARIABLES ARE COM3 PA2 PK2 PE2 CQ1 PA1 PK1 PE1 TREAT BS1 ;
MISSING ARE ALL (-9999) ;
ANALYSIS:
ESTIMATOR = MLR ;
MODEL:
PA2 ON BS1 CQ1 TREAT PA1 (b1 b2 p1 b3) ;
PK2 ON BS1 CQ1 TREAT PK1 (b4 b5 p2 b6) ;
PE2 ON BS1 CQ1 TREAT PE1 (b7 b8 p3 b9) ;
```

```
COM3 ON PA2 PK2 PE2 TREAT BS1 CQ1 (p4-p7 b10 b11) ;
[COM3] (a1) ;
MODEL INDIRECT:
COM3 IND TREAT ;
OUTPUT: SAMP STANDARDIZED(STDYX) MOD(ALL 4) RESIDUAL
CINTERVAL TECH4 ;
```

The syntax is similar to the probit-based syntax in Table 12.14 but the `CATEGORICAL` command is left out and the intercept of `COM3` is estimated rather than a threshold.[11]

Model Fit

For overall model fit, the chi square test of perfect model fit in the population was statistically non-significant ($\chi^2(12) = 12.25$, $p < 0.43$), which is consistent with good model fit. The RMSEA was 0.004, the upper limit of the 90% confidence interval for the RMSEA was 0.027; the CFI was 1.00, and the standardized RMR was 0.013. All indices point towards satisfactory fit.

For localized tests of the difference between each predicted and observed variance/covariance, of the 55 z tests, only two were statistically significant. Given the number of tests, these effects could be chance induced (in fact, I know they are because I generated the simulated data). For the modification indices, four of the 45 indices yielded values above 4.0. They were theoretically vacuous and also reflect chance. The points of stress raise warnings but theory leads me to conclude they are not consequential.

Total Effect of the Intervention on the Outcome

The total effect of the program is taken from the output section `TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS` that results from the `COM3 IND TREAT` command. It is the treatment outcome mean (which is a proportion, given 0, 1 scoring) minus the control group mean and equaled 0.19 ±0.05 ($z = 7.62$, $p < 0.05$).

Effects of the Intervention on the Target Mediators

The effects of the program on the mediators are reported directly on the output for the three equations where `PA2`, `PK2`, and `PE2` are endogenous. Here are the results for PA2 taken from the `MODEL RESULTS` section:

[11] A warning message appears on the output about a non-positive definite first order product matrix. This warning can be ignored for all chapter examples; see the document in Chapter 11 on the Resources tab of my webpage.

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

 PA2       ON
    BS1                0.097      0.020      4.912      0.000
    CQ1                0.092      0.025      3.679      0.000
    TREAT              0.823      0.020     41.818      0.000
    PA1                0.355      0.025     14.180      0.000
```

The covariate adjusted mean difference between the treatment and control conditions was 0.82 ±0.04 (z= 41.82, p < 0.05). Using the intercepts as indicators of the covariate adjusted means and using reverse scoring of `TREAT`, I find that the covariate adjusted mean for the control group was 0.06 ±0.03 and for the treatment group it was 0.88 ±0.03. For `PK2`, the covariate adjusted mean for the control group was 0.02 ±0.03, for the treatment group it was 0.86 ±0.03, and the mean difference was 0.84 ±0.04, z= 41.66, p < 0.05. For `PE2`, the covariate adjusted mean for the control group was -0.06 ±0.03, for the treatment group it was -0.06 ±0.03, and the mean difference was 0.00 ±0.04, z= 0.33, p < 0.74. The lower limits of the 95% confidence intervals for `PA2` and `PK2` both exceeded the meaningful standards, but this was not the case for `PE2`. These results are similar to the prior analyses using probit based FISEM and LISEM.

Effects of the Mediators on the Outcome

The estimated effects of the mediator on the outcome come directly from the output:

```
                                                          Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value
COM3      ON
    PA2                0.117      0.030      3.877      0.000
    PK2                0.110      0.029      3.772      0.000
    PE2               -0.141      0.030     -4.671      0.000
    TREAT              0.003      0.043      0.060      0.952
    BS1                0.090      0.025      3.574      0.000
    CQ1                0.056      0.032      1.736      0.083
```

For `PA2`, the coefficient was 0.12 ±0.06 (z=3.88, p < 0.05). For every one unit that `PA2` increases, the proportion of parents who communicate with their child increases by 0.12, holding the other predictors constant. For `PK2`, the coefficient was 0.11 ±0.06, z = 3.77, p < 0.05. For `PE2`, the coefficient was -0.14 ±0.05, z = 4.67, p < 0.05. Bootstrapping can be used to evaluate asymmetry of the margins of error. The independent effect of the treatment on the outcome was trivial (coefficient = 0.003 ±0.09, z = 0.06, p < 0.96).

In the FISEM probit model, I calculated AMEs for each mediator. In the MLPM, the coefficients themselves equal the AMEs, so there are no additional computations to be done. The values were close to the AMEs for the FISEM probit analysis. As well, there is no need to conduct profile analyses because the effect of a given mediator on outcome probabilities is assumed to be the same no matter what the values of the other mediators or covariates are. The analyses indicate that all three mediators are relevant to parental communication, consistent with the other analyses. All of them exceeded the meaningfulness standards taking into account sampling error via the confidence intervals.

Overall Conclusions for the MLPM using FISEM

Although the MLPM is controversial and viewed by some with skepticism, it brings with it the flexibility of FISEM analyses described in Chapter 11 for continuous outcomes. This includes the ability to accommodate causal relationships among mediators, the ability to accommodate error structures and straightforward analyses of the three program questions. As I have stressed, to use the MLPM one must be convinced that linear approximations between variables exist and that offending predicted probabilities are dealt with appropriately so they do not compromise the analysis. If conditions are conducive, the approach offers flexibility and circumvents some of the non-trivial challenges of the more complex non-linear models of logit and probit.

SUPPLEMENTAL ANALYSES

In Chapter 11, I conducted supplemental analyses following the primary RET analysis that included sensitivity checks, checks for biasing effects of measurement error, consideration of competing models, testing moderator-based specification error, and replicating results with bootstrapping. I do not illustrate these analyses in the current chapter in the interest of space, but you will want to routinely pursue them no matter the metric of your mediators and outcomes. You can use Chapter 11 as a model. You also should not hesitate to use the tools in the LISEM toolbox to help gain perspectives on RET dynamics, even if you are using an FISEM approach. For example, in Chapter 8 I discussed the use of quantile regression as a way to evaluate quantile treatment effects (QTEs), an approach that is not available in FISEM. I can still make use of this tool for the current example by examining QTEs for the effects of the program on each of the mediators. Figure 12.5 presents a quantile plot using the program from my website as applied to the perceived advantages mediator. The graph shows the effect of the program on perceived advantages is uniform across the perceived advantages continuum.

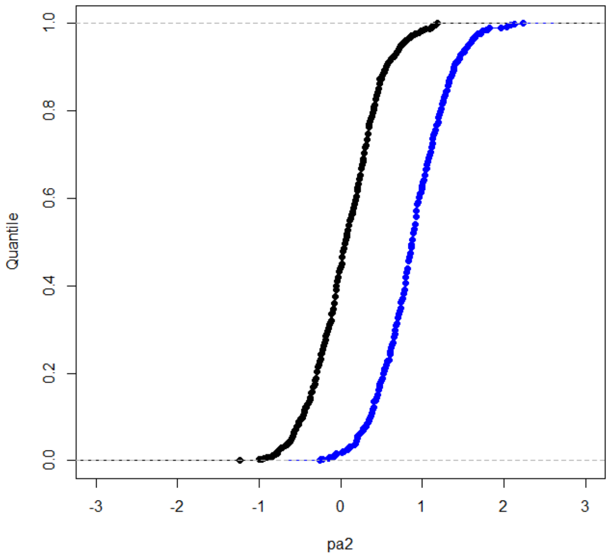

FIGURE 12.5. Quantile treatment effect plot for perceived advantages

As another example, for sensitivity reasons and to provide reassurances about outliers and leverages, I conducted an outlier-leverage resistant MM regression on Equations 12.4 to 12.7. I used the program for MM regression on my website for each of the following Equations[12]:

$$PA2 = a_1 + p_1\ TREAT + b_1\ BS1 + b_2\ CQ1 + b_3\ PA1 + d_1 \qquad [12.4]$$

$$PK2 = a_2 + p_2\ TREAT + b_4\ BS1 + b_5\ CQ1 + b_6\ PK1 + d_2 \qquad [12.5]$$

$$PE2 = a_3 + p_3\ TREAT + b_7\ BS1 + b_8\ CQ1 + b_9\ PE1 + d_3 \qquad [12.6]$$

The outlier adjusted mean differences between the treatment and control groups were 0.82 ($t(1495) = 40.91$, $p < 0.05$) for PA2, 0.82 ($t(1495) = 40.91$, $p < 0.05$) for PK2, and 0.0002 ($t(1495) = 0.01$, ns) for PE2, all of which comport well with the primary analyses.

THE CASE OF BINARY MEDIATORS AND LATENT VARIABLES

You may encounter situations where one or more of your mediators is binary. My focus to this point has been exclusively on binary outcomes. In this section, I consider strategies for analyzing models with binary mediators.

[12] The technique cannot be used with dichotomous outcomes, so I did not use it for Equation 12.7.

By definition, binary mediators are endogenous because there is a causal path emanating from the treatment (dummy) variable to the mediator or from another mediator to the binary mediator. At the same time, the binary mediator is a cause of the outcome and/or another mediator. Figure 12.6 shows a classic influence diagram of this scenario.

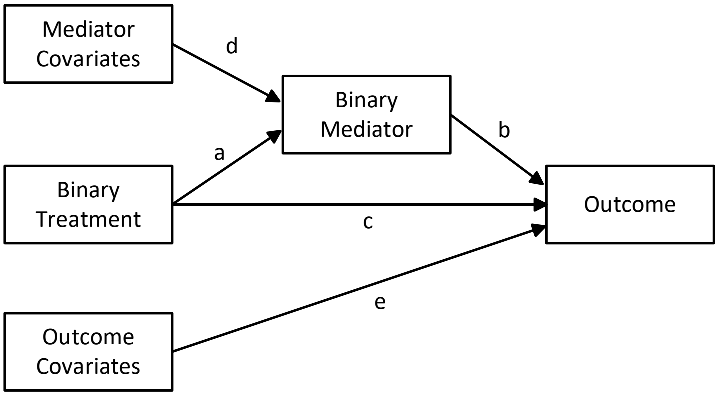

FIGURE 12.6. Classic single mediator model

The dual role of the binary mediator as both a predictor variable and a dependent variable creates challenges for FISEM approaches that prioritize the analysis of omnibus mediation effects. For LISEM, the focus is on documenting and evaluating the magnitude of the path coefficient for each separate link in the mediational chain in Figure 12.6 using any one of many available statistical tools in our statistical toolbox and then using the joint significance test (JST) to evaluate the null hypothesis of no mediation across the full mediational chain (see Chapter 9). Binary mediators pose no special issues in such cases. I now describe LISEM-based methods you might use for a range of scenarios with binary mediators.

Scenario 1: If the mediator is binary and the outcome is binary, use logit or probit regression (or the MLPM) to regress Y onto M and T to isolate path coefficients *b* and *c* in Figure 12.6 and do the same when regressing M onto T to isolate path coefficient *a*. Include measured covariates in all analyses, as appropriate. Evaluate the magnitude of each path coefficient and then apply the JST to coefficients *a* and *b*.

Scenario 2: If the mediator is binary and the outcome is continuous, use traditional linear regression with a robust estimator to regress Y onto M and T (both dummy coded) to isolate path coefficients *b* and *c*. Apply logit or probit regression (or the MLPM) when regressing M onto T to isolate path coefficient *a*. Include measured covariates in all analyses, as appropriate. Evaluate the magnitude of each path coefficient and then apply the JST to coefficients *a* and *b*.

Scenario 3: If the mediator is binary and the outcome is ordinal, use ordinal regression to regress Y onto M and T to isolate path coefficients *b* and *c* and use logistic or probit regression (or the MLPM) when regressing M onto T to isolate path coefficient *a*. Include measured covariates in all analyses, as appropriate. Evaluate the magnitude of each path coefficient and then apply the JST to coefficients *a* and *b*.

Scenario 4: If the mediator is binary and the outcome is nominal, use multinomial regression to regress Y onto M and T to isolate path coefficients *b* and *c* and apply logit or probit regression (or the MLPM) when regressing M onto T to isolate path coefficient *a*. Include measured covariates in all analyses, as appropriate. Evaluate the magnitude of each path coefficient and then apply the JST to coefficients *a* and *b*.

Scenario 5: If the mediator is binary and the outcome is time until an event occurs, use survival analysis to regress Y onto M and T to isolate path coefficients *b* and *c* and apply logit or probit regression (or the MLPM) when regressing M onto T to isolate path coefficient *a*. Include measured covariates in all analyses, as appropriate. Evaluate the magnitude of each path coefficient and then apply the JST to coefficients *a* and *b*.

Each of these analyses can be conducted within Mplus in the spirit of LISEM so you can take advantage of the modern missing data algorithms, robust estimation, and bootstrapping offered by Mplus. If you have a latent variable with multiple indicators for your outcome, then you can bring the latent variable and its indicators into the analysis vis-à-vis standard Mplus programing (as illustrated in Chapter 11). If your sample size is too small to accommodate asymptotic theory, then you can use a small sample appropriate statistical method outside of Mplus, such as Firth regression as provided on my website (see Chapter 28). If you want to adjust for measurement error in Y but you do not have multiple indicators, you can consider using the single indicator strategies for error correction outlined in the document on my web page for Chapter 3. For calculating AMEs, complications are introduced with the presence of latent variables but there are workarounds (see Appendix A).

One criticism of the LISEM approach is that it does not yield a quantifiable, intuitive index of the magnitude of the omnibus mediation effect linking T to Y through a given mediator, M. I agree with this criticism but I find it to be minor if my focus is on program evaluation, namely if I want to figure out how to strengthen an intervention or why an intervention is not working well. The more micro-level link-by-link analyses provide the specific information I need to make suggestions for program improvement, which is less true of the omnibus indirect tests. Also, if I know a given link in the mediational chain is "broken" (i.e. functionally zero or trivial in magnitude), I know for a fact that the omnibus mediational index must also be weak unless the other links in the chain are so strong that they overwhelm a weak, non-zero link. Another way of saying this is that in most cases, once I have a good sense of the strength and meaningfulness of the individual links in a mediational chain, I also have a good qualitative sense of the strength of the overall omnibus effect. For elaboration, see Chapter 17.[13]

With FISEM, the analytic flexibility is more constrained. Consider, for example, the case of a binary mediator and a continuous outcome. To calculate an omnibus mediational effect linking T to Y through M using coefficient multiplication, we must contend with the fact that the coefficient for path *a* is on a different scale (it reflects the effects of a unit change in the predictor using probits or log odds) than the coefficient for path *b* (which reflects the effects of a unit change in the predictor using means). Mixing coefficient scaling can complicate the interpretation of the coefficient product and the statistical theory for estimating standard errors and significance tests for it. In addition, the binary mediator needs to function as an exogenous dummy variable in the M→Y portion of the model but as a logit or probit based binary outcome variable in the T→M portion of the model.

For FISEM and a binary mediator, Bayes modeling offers flexibility for how you treat the binary mediator, providing you with two options. By specifying the option `MEDIATOR=LATENT` on the `ANALYSIS` command in a Bayes analysis, Mplus invokes the y* formulation of the binary mediator, i.e.., it focuses on the latent response underlying the observed m, namely m*. The coefficient for T→M is the (covariate adjusted) m* mean difference between the intervention and control groups for the continuous m* and the coefficient for M→Y where Y is continuous is the linear regression of Y onto m* for the binary mediator.[14] The omnibus indirect effect is then the product of these two coefficients expressed in units of change in the mean continuous outcome Y. Mplus performs a complete mediation and total effect analysis via the `MODEL INDIRECT`

[13] One must be careful not to be deceived by the strength of individual effects relative to overall effects by always keeping in mind the nature of multiplicative functions at play. I am reminded of the classic example where a complex machine has 50 working parts that are each essential to the machine's operation. If each part has a probability of successfully functioning of 0.99 and failures for one part are independent of failures for the other parts, then the probability of successful function is 0.99^50 = 0.60. Mediation analyses usually involve a small number of links in a given mediational chain so they are not subject to extremes per the machine example, but the dynamics should be kept in mind.

[14] It turns out that for the T→M link, the probit coefficients under the y* conceptualization will be identical to the coefficients in the more traditional probability formulation of probit regression. In the former, the tradition is to interpret partially standardized coefficients in which y* is standardized to 1.0 whereas in the latter, one engages in probability-based interpretations as outlined in this chapter.

command and the IND subcommand in this scenario. If you have a single mediator, you also can request the causal mediation analysis for the total effect. The MEDIATOR=LATENT option is the default for Bayes analysis in Mplus so if you specify nothing, this is what will be invoked. Remember that when working with m* it usually is best to focus on endogenous standardization (STAND(STDY) in Mplus).

Alternatively, for the Bayes model with a binary mediator you can use the statement MEDIATOR=OBSERVED in place of MEDIATOR=LATENT. In this case, Mplus will apply the traditional probit regression for the T→M link but then treats the binary mediator as a 0-1 dummy variable (given it is scored 0-1) for the M→Y link rather than regressing Y onto m*. If Y is continuous, it is analogous to treating M as a two valued dummy variable when predicting Y in a traditional linear regression model. This can produce different results than using m* as the predictor. In some cases, Mplus may not conduct the omnibus mediation test nor the total effect analysis, instead printing the message MODEL INDIRECT IS NOT AVAILABLE FOR SOME VARIABLES. This is because the statistical theory does not support it. The number of possible permutations between the binary mediator and the type of outcome analyzed (e.g., continuous, binary, ordinal, count) and the ways of handling them in FISEM can be rather involved. For discussions of this topic, see Muthén, Muthén and Asparouhov (2016).

It also is possible to use a causal mediation framework based on potential outcomes within Mplus for a binary mediator but this approach is limited to the case of a single mediator, which is suboptimal for RET analysis (see Muthén, 2011; Rijnhart et al., 2023; Schuster et al., 2023; Stanghellinia & Kateri, 2023).

In practice, I sometimes find I can accomplish my analytic goals strictly within an FISEM framework. Other times I need to move to an LISEM framework. Still other times I use a blending of the two approaches. An example of such blending was my treatment in this chapter of the estimation of the total program effect in terms of intervention versus control group proportion differences for probit-based RET models. Mplus offers great deal of flexibility in this regard.

A final point I want to discuss here concerns the fact that the numerical example I used throughout this chapter did not have latent variables with multiple indicators. All of the variables of substantive interest were captured by single indicators with straightforward calculations of sample means, sample standard deviations, and sample covariances and correlations. With the introduction of multiple indicator latent variables into a model, the estimation of path coefficients remains analytically straightforward but the incorporation of latent means and latent intercepts can be complex. Often the mean structure of data is unimportant and we can effectively analyze data for purposes of testing a model by assigning arbitrary values to the means and intercepts. For example, in many FISEM applications, latent variable means and intercepts are fixed to values of zero without any disruption of our ability to test the hypothesized model. When working with latent variables, clients sometimes have difficulty understanding such arbitrariness and I need to make choices when presenting results that reduce confusion. The numerical example in Chapter 11 used latent variables and illustrates instances where this was the case, such as when I characterized the total effect of the intervention in Figure 11.4. In future chapters, I revisit the issue of mean structures for certain model forms where the mean structures are not arbitrary.

CONCLUDING COMMENTS

In this chapter, I analyzed the numerical example from both LISEM and FISEM perspectives and from multiple vantage points. The results across the different forms of analysis were similar and the same conclusions were made in all cases. This will not always occur in practice. However, the exercise illustrates that different analytic approaches each with different strengths and weaknesses often can be used in similar contexts. As with Chapter 11, I do not argue that you should apply all of the methods of analysis I covered to your RET data. My goal was to provide you with a toolbox of analytic strategies so you can choose the ones that are appropriate for answering your empirical questions and to encourage you to come at your data from multiple angles.

I did not discuss the causal mediation framework in much detail because its strength is in analyzing omnibus mediation effects and decomposing those omnibus effects into omnibus indirect and direct effects. I present material on these approaches on the resources tab for the current chapter on my webpage. Such omnibus tests, in my opinion, just do not have much information gain beyond the analyses I recommend when evaluating programs. As I have already said, this is not to say that such tests do not have a place in mediation modeling more generally nor that they never will be relevant to program evaluations. It is just that for the vast majority of program evaluations, I think they can complicate matters unnecessarily.

I discussed in depth probit modeling and modified linear probability modeling, shying away from logistic models and their reliance on odds ratios. Of course, there are other forms of binary regression one can use that assume different functions from these models in terms of the relating continuous predictors to outcome probabilities, including log-log models, complementary log-log models, and log binomial models, to name but a few. Of these, the log binomial model is often used because it is amenable to modeling relative risks, which is popular in some disciplines. A complication of log binomial regression for binary outcomes is that the p values for a predictor can change when the outcome variable is reverse coded, from 0-1 to 1-0, an undesirable property, especially if assignment to the zero and one categories is arbitrary. Algorithms for the log binomial model also often fail to converge with continuous predictors. It sometimes is recommended to use an alternative algorithm, called Zou's (2004) modified Poisson method with robust Huber-White estimation. However, I find that even that approach has its share of shortcomings.

Pischke (2012), a proponent of the MLPM, argues that the reason we have so many different model functions "is just a statement to the fact that we don't know what the 'right' model is. Hence, there is a lot to be said for sticking to a linear regression function as compared to a fairly arbitrary choice of a non-linear one." He argues that one can get into just as much if not more trouble when we choose the wrong non-linear function than when we apply a wrong linear function, the former of which often occurs when people blindly adopt a heuristic like "if the outcome is binary, use logistic regression." Obviously, the best approach is to choose the right function when modeling data. My own orientation is that linear models often are a reasonable starting point, but I also do not hesitate to deviate from or augment them if theory and data so dictate. For a discussion of the many different binary regression model forms, see Wooldridge (2010).

APPENDIX A: CALCULATION OF AVERAGE MARGINAL EFFECTS

In this Appendix, I show how to calculate average marginal effects (AMEs) within Mplus. The method does not yield significance tests or confidence intervals for the AMEs, which is a disadvantage.

Cameron and Trivedi (2010) show the AME for a continuous M can be estimated manually as follows:

1. Calculate $p1_i$ as described in the main text.

2. Increment the value of M for each individual by a very small amount. Cameron and Trivedi recommend increasing it by the standard deviation of M divided by 1,000. I call this increment delta. So, to M, add delta.

3. Calculate $p2_i$ using this incremented value of M.

4. Define each individual's marginal effect, IME, as $(p2_i – p1_i)$ / delta.

5. Calculate the average of the IMEs. This value is the AME.

The same approach is used for logistic equations, but the predicted values based on the derived logistic equation, $\hat{Y}_i$, are translated into probabilities using the following formula:

$$p_i = \exp(\hat{Y}_i) / (1 + \exp(\hat{Y}_i)) \qquad [A.1]$$

where exp is the exponent function. For example, if the predicted logit for an individual is $\hat{Y}_i = 1.11$, the individual's predicted probability is 3.034/(1 + 3.034) = 0.752.

If M is binary, then the AME for Y relative to M is calculated using the first, intuitive method I described but where scores on M for everyone are set to 0 when calculating $\hat{Y}_i$ for p1 and scores for everyone are set to 1 when calculating $\hat{Y}_i$ for p2. Other than that, the calculations are the same.

Using the chapter numerical example, I first run the syntax in Table 12.13 but I eliminate lines 3-5 so there is no mean centering. Here is relevant output from `MODEL RESULTS` for the `COM3` equation that I will make use of:

```
                                                     Two-Tailed
                    Estimate       S.E.    Est./S.E.    P-Value
 COM3      ON
    PA2                0.313      0.083      3.785      0.000
    PK2                0.298      0.081      3.696      0.000
    PE2               -0.375      0.082     -4.592      0.000
    TREAT             -0.001      0.113     -0.012      0.991
    BS1                0.242      0.067      3.600      0.000
    CQ1                0.150      0.087      1.710      0.087

Thresholds
    COM3$1             0.512      0.060      8.597      0.000
```

The probit linear equation (flipping the sign of the threshold) is

Probit(COM3) = -.512 + .313 PA2 + .298 PK2 + -.375 PE2 + -.001 T + .242 BS1 +
.150 CQ1

From the descriptive statistics section of the output (titled `UNIVARIATE HIGHER-ORDER MOMENT DESCRIPTIVE STATISTICS`) I note that the variance of `PA2` is 0.335, for `PK2` it is 0.347, and for `PE2` it is 0.171. Here is the syntax I use to calculate the AME for `PA2`:

Table A.1: Mplus Syntax for AME in a Probit-Based Model

```
TITLE: AME analysis for PA2  ;
DATA: FILE IS c:\mplus\communication.dat ;
DEFINE:
DELTA=SQRT(0.335)/1000 ; !divide SD of PA2 by 1000
PROBIT1=-.512+0.313*PA2+0.298*PK2-0.375*PE2-0.001*TREAT+0.242*BS1+
0.150*CQ1;
PROB1=PHI(PROBIT1) ;
PA2=PA2+DELTA ; !increase PA2 by a small amount
PROBIT2=-.512+0.313*PA2+0.298*PK2-0.375*PE2-0.001*TREAT+0.242*BS1+
0.150*CQ1;
PROB2=PHI(PROBIT2) ;
IME=(PROB2-PROB1)/DELTA ;
VARIABLE:
NAMES ARE ID COM3 PA2 PK2 PE2 CQ1 PA1 PK1 PE1 TREAT BS1 ;
USEVARIABLES ARE IME ;
MISSING ARE ALL (-9999) ;
ANALYSIS:
ESTIMATOR = ML ; TYPE=BASIC ;
OUTPUT: !use defaults on output
```

Most of the Cameron and Trivedi method is implemented in the `DEFINE` command. `DEFINE` commands are executed sequentially by Mplus and I take advantage of this property for my calculations. In Line 4, I define delta as the standard deviation of PA2 divided by 1000. In Lines 5 and 6, I calculate a probit value for each individual by applying the probit equation to their raw scores. In Line 7, I convert this probit value to a probability using the PHI function. In Line 8, I increment everyone's PA2 score by delta. I calculate the probit for it (Lines 9 and 10) and convert it to a probability (Line 11). Line 12 calculates the individual marginal effect as the difference between the two probabilities divided by delta. All that is left is to average these scores, which is what Lines 13-19 do. Line 18 adds TYPE=BASIC, which informs Mplus I want it to calculate the means of all the variables on the USEVARIABLES line (Line 13). The mean of IME reported on the output was 0.116, which is the AME. It is possible to generate standard errors through bootstrapping but for AMEs in Mplus, it is complicated and time consuming. After making appropriate changes in the syntax, I found the AME for PK2 was 0.11 and for PE2 it was -0.13. These values are close to what I observed for the LISEM analyses.

The AME for the direct effect of TREAT on the outcome is calculated using the syntax in Table A.2. Notice on Line 4 that I substitute the value of 0 for TREAT, which has the effect of setting everyone's score to 0. On line 6, I substitute the value of 1 for TREAT, which has the effect of setting everyone's score to 1. The remaining syntax follows from Table A.1.

Table A.2: Mplus Syntax for AME for the Treatment Condition

```
1. TITLE: AME analysis for TREAT  ;
2. DATA: FILE IS c:\mplus\communication.dat ;
3. DEFINE:
4. PROBIT1 = -.512+0.313*PA2+0.298*PK2-0.375*PE2-0.001*0+0.242*BS1+0.150*CQ1;
5. PROB1 = PHI(PROBIT1) ;
6. PROBIT2 = -.512+0.313*PA2+0.298*PK2-0.375*PE2-0.001*1+0.242*BS1+0.150*CQ1;
7. PROB2 = PHI(PROBIT2) ;
8. IME = (PROB2-PROB1) ;
9. VARIABLE:
10. NAMES ARE ID COM3 PA2 PK2 PE2 CQ1 PA1 PK1 PE1 TREAT BS1 ;
11. USEVARIABLES ARE PROB1 PROB2 IME ;
12. MISSING ARE ALL (-9999) ;
13. ANALYSIS:
14. ESTIMATOR = ML ; TYPE = BASIC ;
15. OUTPUT: !use defaults on output
```

The AME for TREAT was effectively <0.001 which is consistent with prior analyses that showed TREAT has a trivial impact on COM3 over and above the mediators and covariates. The means for PROB1 and PROB2 are the component statistics that feed into the IME difference, i.e., IME = PROB2-PROB1.

Average Marginal Effects and Latent Variables

Estimating marginal effects in models with mediators or covariates that are latent variables is difficult in Mplus. The easiest approach translates the multiple indicators for a latent variable into a single indicator using one of the methods described below and then to use either the program on my website or the Mplus syntax I provided above to calculate the relevant AMEs. This is suboptimal because it sacrifices corrections for measurement error by using multiple indicators, but if your indicators have reasonably high reliabilities, the strategy at least gets you a ballpark estimate of the AME.

One strategy for generating a single indicator from multiple indicators of a latent variable is to form a composite of the indicators using the methods discussed in Chapter 3, such as by averaging the indicators. You then can obtain AMEs, confidence intervals, and margins of error using the average marginal effects program on my website or using the Mplus syntax approach presented above. If the indicators have different metrics, consider using percent of maximum possible (POMP) scoring discussed in Chapter 3 or some form of standardization. Another possibility is to choose what you think is the best indicator of the latent variable indicators and conduct the analysis using only that indicator (Hayduk & Littvay, 2012). Yet another possibility is to generate factor scores for the latent variables and then use the factor scores in place of the latent variables in the single indicator approach. You can save the original input data from any Mplus program with latent variables as well as the factor scores for the latent variables in the model by adding the syntax

```
SAVEDATA: FILE IS dataplusfs.dat; SAVE=FSCORES;
```

after the OUTPUT line of your program. FILE IS tells Mplus the file name for the saved data (you can use any name and tag you want; the file will be saved in the same folder as the syntax file you run). SAVE=FSSCORES tells Mplus to save the original data plus the factor scores for the latent variables. Be sure to look in the section SAVEDATA INFORMATION in the output of the initial run with the latent variables to see the order in which the variables are saved in the new file. In addition to the factor scores, Mplus saves standard errors for them, which you can ignore. With factor scores, the standard errors, significance tests, and confidence intervals for the AMEs generated by the average marginal effect program on my webpage become questionable because they do not take into account sources of sampling error in the factor scores; these statistics must be used with recognition of this limitation, if at all. For saving the Bayesian equivalent of factor scores in Bayes analyses, see the Mplus users guide.

The Mplus syntax I provide for calculating AMEs in a FISEM framework with mediators or covariates as latent variables will not work.

APPENDIX B: SETTING MEANINGFULNESS STANDARDS FOR THE NUMERICAL EXAMPLE

In this Appendix, I describe the approach for defining meaningfulness standards for (a) the effect of the intervention on each mediator, and (b) the effect of the mediators on the outcome.

Meaningfulness Standard for Effect of Intervention on Mediators

For each mediator, I take into account the overall meaningfulness standard for the total effect, namely a proportion increase of 0.08, and the fraction of that effect that I feel each mediator should account for. I decide to require each mediator to share equal responsibility with the two other mediators to produce the minimal meaningful change in parental communication, which yields a desired overall minimal meaningful effect per mediator on the outcome of 0.08/3 = 0.027. For each mediator, I also note the value of the path coefficient from the mediator to the outcome, M→O, that resulted from regressing the outcome onto the mediators in a MLPM analyses. This latter information helps me contextualize the needed intervention effect on the mediator relative to the strength of the effect of the mediator on the outcome, per Chapters 10 and 11. Table B.1 presents the numbers used in the calculations, with the derived standard for meaningfulness shown in the last column.

Table B.1: Meaningfulness Standard for Intervention Effect on Mediator

Link of Interest	COM3 Standard	Fraction	M→O	Derived Standard
TREAT→PA2	0.08	.33	0.12	≥0.22
TREAT→PK2	0.08	.33	0.11	≥0.24
TREAT→PE2	0.08	.33	-0.14	≤-0.19

Meaningfulness Standard for Effect of Mediators on Outcome

I again used the strategy discussed in Chapters 10 and 11 using the program on my website *effect size standards*. The minimal meaningful change in parental communication (a proportion increase of at least 0.08) serves as a referent and the fraction of that effect I want each mediator to account for is a third of it, given three explicit mediators. Taking into account the program's likely ability to bring about change in each mediator based on the collected data, I calculate the effect size standards based on calculations in Table B.2 (see the derived standards in the last column).

Table B.2: Meaningfulness Standard for Mediator Effect on Outcome

Link of Interest	COM3 Standard	Fraction	T→M	Derived Standard
PA2→COM3	0.08	.33	0.82	≥0.03
PK2→COM3	0.08	.33	0.84	≥0.03
PE2→COM3	0.08	.33	-0.83[a]	≤-0.03

[a] I substituted a value of plausible change for actual change for the perceived embarrassment mediator. I used the opposite signed average of the other two mediators, which seems reasonable given the metrics and context.

I recognize that not everyone will agree with the approaches I take in this Appendix; if not, use whatever approach you are comfortable with taking into account my discussion of effects sizes in Chapter 10.