13

Mediation Analysis with Ordinal and Nominal Outcomes

The best time to plan an experiment is after you've done it

- R.A. FISHER

INTRODUCTION

This chapter uses concepts from Chapters 5, 11 and 12 to address ordinal and nominal outcomes (and mediators) when evaluating interventions in an RET. As with previous chapters, you can use either limited information structural equation modeling (LISEM) and/or full information structural equation modeling (FISEM) for such models. In Chapter 12 on binary outcomes, I explicated LISEM using the modified linear probability model, probit-based modeling, and Bayesian modeling. In the interest of space, I am more selective here about the approaches I focus on, by and large using FISEM with a few exceptions. As with prior chapters, I address the three core questions of an RET, namely (1) is there an overall meaningful effect of the intervention on the outcome, (2) is there a meaningful effect of the intervention on the presumed mediators of intervention

effects on the outcome, and (3) do the mediators, in fact, meaningfully affect the outcome? I briefly introduced ordinal regression in Chapter 5. You may want to review that material before reading this chapter. The current chapter is long and not meant to be processed in a single sitting. Take your time reading and thinking through the different sections.

WORKED EXAMPLE FOR ORDINAL OUTCOMES

The main example for this chapter evaluates a program to increase the effectiveness of clinicians in treating depression and anxiety in young adults. An on-line program sought to train clinicians on three facets of therapeutic alliance. First, clinicians were taught how to better set common goals with their clients about treatment (a strategy known as goal alignment). Second, clinicians were taught how to better negotiate and reach agreement with clients about what needs to be done to accomplish those goals (a strategy known as task alignment). Third, clinicians were taught how to develop a positive bond with clients to promote client motivation and compliance (known as bonding). Each construct was measured near the end of treatment on multi-item scales completed by the patient about the clinician. Responses to items were made on 7 point agree-disagree scales: -3 = strongly disagree, -2 = moderately disagree, -1 = slightly disagree, 0 = neither agree nor disagree, 1 = slightly agree, 2 = moderately agree, 3 = strongly agree. Item scores were averaged across items to yield a total score for each person on each construct. The higher the score, the higher the goal alignment, task alignment, and bonding. Clinicians in the control condition engaged in treatment as usual (TAU) without the extra training.

Client symptom improvement was measured in multiple ways but the program staff and administrators were particularly interested in a commonly used four-point scale in clinic settings. The scale was completed by an independent clinician who rated the symptom improvement of the client, 1=no change or got worse since treatment initiation, 2 = minimally improved since treatment initiation, 3 = much improved since treatment initiation, and 4 = very much improved from treatment initiation. In the broader literature on multiple regression and SEM, simulation studies suggest that if a measure of a continuous construct makes 5 or more discriminations, then that often is sufficient precision to make reasonable inferences using methods that assume interval level data unless the metric is blatantly ordinal (see Chapter 3). In the current example, there are only 4 discriminations. As discussed in Chapter 3, it is possible to strategically select adverb qualifiers for scale points that promote interval level properties of a measure, but in this case, the adverbs appear ordinal relative to the underlying dimension of symptom improvement. I therefore approach the analysis assuming ordinal measurement with a small number of categories. When you design program evaluations, you typically are in

control of the measures you use. Why choose a measure that has suboptimal psychometric properties (only ordinality) when it often is straightforward to avoid doing so? Normally, I would discourage the use of a scale like the one used here, but let's suppose program administrators insisted on it given it is a standard scale in their clinics.

I include two covariates each measured at baseline, an index of clinician experience (CE1) and an index of overall clinician interpersonal skills (CS1), each ranging from 0 to 10 with higher scores indicating greater experience and greater interpersonal skills, respectively. In a typical program evaluation the list of covariates would be longer, but I use only these two to keep the example manageable. The total sample size was 600.

Figure 13.1 presents the RET logic model, omitting covariates to reduce clutter.

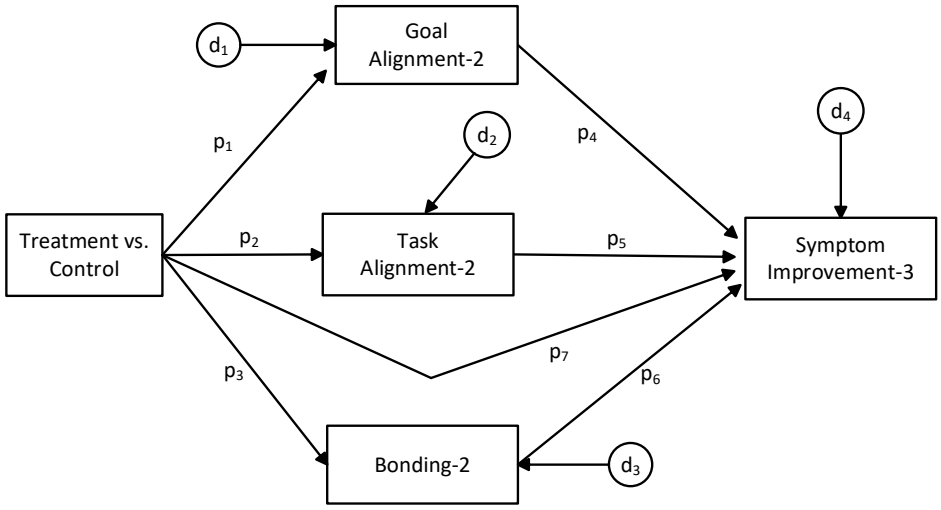

FIGURE 13.1. Clinician improvement example

Here are the model equations (note: T is the treatment condition, CE is clinical experience, CIS is clinician interpersonal skills; GA is goal alignment, TA is task alignment, BD is bonding, and IMP is symptom improvement; each followed by the number 1, 2 or 3 to indicate time of assessment):

$$GA2 = a_1 + p_1\ T + b_1\ CE1 + b_2\ CIS1 + d_1 \quad [13.1]$$

$$TA2 = a_2 + p_2\ T + b_3\ CE1 + b_4\ CIS1 + d_2 \quad [13.2]$$

$$BD2 = a_3 + p_3\ T + b_5\ CE1 + b_6\ CIS1 + d_3 \quad [13.3]$$

$$\text{Probit}(IMP3) = a_4 + p_4\,GA2 + p_5\,TA2 + p_6\,BD2 + p_7\,T + b_7\,CE1 + b_8\,CIS1 \qquad [13.4]$$

Equation 13.4 will make use of ordinal regression whereas the other equations are more traditional regression models with continuous outcomes.

Mplus uses by default what is known as a **cumulative link ordinal regression model**. When estimated using logit functions, it is called the **proportional odds model**. When analyzing ordinal models, there are two orientations I can take. One orientation is to model the continuous latent response, y*, thought to underlie the ordinal measure, as discussed in Chapter 5. In this case, the ordinal measure is viewed as a crude indicator of the underlying continuous dimension of improvement, with the latter being of primary interest. This approach emphasizes documenting program and mediator effects on y*. The second orientation is to model the four-category ordinal measure directly, not its underlying continuous dimension of symptom improvement. In this case, interest is with the four categories of the scale in their own right and statements about how the program affects the proportion of clients in the different scale categories. I refer to the former approach as the **latent response approach** and the latter as the **probability approach**. I consider both in this chapter. The former is much more straightforward in Mplus than the latter, but some would argue that it is less informative.

ORDINAL MODELING: OVERVIEW OF THE PROBABILITY APPROACH

As described in Chapter 5, the proportional odds model and the probit based analog of it rely on *a priori* defined "breakpoints" on the ordinal metric. When conducting cumulative link ordinal regression, we focus on k-1 pairs of categories of the outcome, where k is the number of categories on the response metric. In the current example, there are four categories, so I analyze 4-1 = 3 category pairs. The pairs in the proportional odds model are defined by combining categories on the outcome metric. Any given pair is defined as (a) a given target category and all categories above it versus (b) all categories below the target category. In our example, one "pair" is category 4 (very much improved) versus categories 1, 2, and 3 combined. Another pair is categories 3 and 4 (much improved or very much improved) combined versus categories 1 and 2 (no change or minimally improved) combined. The final pair is categories 2, 3, and 4 combined (either minimal, much or very much improved) versus category 1 (no change). There is a logic behind these pairings, the key to which is to think about them in terms of "break points," as I now explain.

One theoretically interesting break point is that for categories 2, 3, 4 (showing some degree of improvement) versus category 1 (not showing any improvement or getting

worse). I might find that individuals in the intervention condition are more likely to be in the "2, 3, and 4" combined category than individuals in the control group, or stated differently, that individuals in the treatment condition are less likely to be in the worst category (not showing any improvement or getting worse) than those in the control group. Such results would indicate a favorable program impact.

Suppose instead of the above break point, I examine a different one. I might define a substantively interesting break point as one that collapses categories 3 and 4 which represent non-trivial improvement (much improved or very much improved) versus the collapsed categories of 1 and 2, which represent trivial or no improvement. Suppose I find that individuals in the intervention condition are more likely to be in the combined "3 or 4" categories as compared to individuals in the control group or, stated differently, that individuals in the intervention condition are less likely to be in the bottom two categories that show no or trivial improvement than those in the control group. Such results also would reflect a favorable program effect.

Finally, suppose instead of the above break points, I examine yet a different one. I might define a break point as category 4 (very much improved) versus the combined categories "1, 2, and 3" (with each of the atter three categories representing not being "very much improved"). Suppose I again find that individuals in the intervention condition are more likely to be in the "much improved" category than individuals in the control group, or stated differently, that individuals in the intervention condition are less likely to be in the bottom three categories that show less than strong improvement than those in the control group. Such results also would reflect a favorable program effect.

Each of these "contrasts" are of substantive interest and it turns out that an ordinal regression analysis will provide perspectives on each of them by conducting separate logistic or probit regressions on each breakpoint. I can formalize the intervention versus control contrasts using different phrasings of them as follows:

Contrast 1: Test intervention versus control differences in the proportion of clients who score in the lowest outcome category, i.e., clients who showed no change or got worse since treatment initiation. This information is contained in the analysis of the first break point. The hypothesis is that this proportion will be lower in the intervention condition than in the control condition.

Contrast 2a: Test intervention versus control differences in the proportion of clients who scored in the combined lowest two outcome categories considered together, i.e., clients who showed no change or only minimal improvement. This information is contained in the analysis of the second break point. The hypothesis is that this proportion will be lower in the intervention condition than in the control condition.

Contrast 2b: Test intervention versus control differences in the proportion of clients who scored in the combined highest two outcome categories considered together, i.e., clients who were much improved or very much improved since treatment initiation. Note that this contrast is just the mirror image of Contrast 2a, phrased in the opposite way. It is redundant with Contrast 2a but in the program evaluation study for this example, program staff wanted the question answered when framed in this way as well. The hypothesis is that this proportion will be higher in the intervention condition than in the control condition.

Contrast 3: Test intervention versus control differences in the proportion of clients who were very much improved, the top category of the scale. This information is contained in the analysis of the third break point. The hypothesis is that this proportion will be higher in the intervention condition than in the control condition.

My goal is to test these contrasts and in doing so, gain perspectives on program effects. The probability approach as focused on breakpoints allows me to do so.

Cumulative link modeling can use different functions to conduct the above contrasts with the two most popular strategies being a logistic function and a probit function. The logit function is common and is associated with the term "proportional odds modeling" because of its reliance on odds. For reasons I discussed in Chapters 5 and 12 as well as later in this chapter, I prefer to use probit modeling.

The ordinal analysis of Equation 13.4 is implemented in Mplus using three subequations, each in the form of a binary probit model corresponding to a breakpoint analysis. Let $IMP3_{d1}$ be a dichotomous outcome where 0 = category 1 and 1 = the combined categories 2, 3 and 4, $IMP3_{d2}$ is 0 = categories 1 and 2 combined and 1 = categories 3 and 4 combined, and $IMP3_{d3}$ is 0 = categories 1, 2 and 3 combined and 1 = category 4. Note that the score of 1 in each of these outcome characterizations is the higher score(s) on the ordinal scale relative to those scored 0. The three probit-based subequations are thus:

$$\text{Probit}(IMP3_{d1}) = a_{4a} + p_{4a}\,GA2 + p_{5a}\,TA2 + p_{6a}\,BD2 + p_{7a}\,T + b_{7a}\,CE1 + b_{8a}\,CIS1 \qquad [13.5]$$

$$\text{Probit}(IMP3_{d2}) = a_{4b} + p_{4b}\,GA2 + p_{5b}\,TA2 + p_{6b}\,BD2 + p_{7b}\,T + b_{7b}\,CE1 + b_{8b}\,CIS1 \qquad [13.6]$$

$$\text{Probit}(IMP3_{d3}) = a_{4c} + p_{4c}\,GA2 + p_{5c}\,TA2 + p_{6c}\,BD2 + p_{7c}\,T + b_{ac}\,CE1 + b_{8c}\,CIS1 \qquad [13.7]$$

The mechanics of ordinal regression work with these subequations to make statements about the more general equation, namely Equation 13.4.

Cumulative link modeling makes a strong assumption about the focal breakpoint

equations in the analysis, namely it assumes that a path coefficient associated with a given predictor takes on the same value across the three subequations, e.g. $p_{4a} = p_{4b} = p_{4c}$; $p_{5a} = p_{5b} = p_{5c}$; $p_{6a} = p_{6b} = p_{6c}$ and so on. This is called the **parallel coefficient assumption**. The intercepts are allowed to differ across the equations, but not the path coefficients for any given predictor. As such, the ordinal analysis reports only a single path coefficient value for each predictor because once you know the value of the predictor coefficient for one subequation, you know its value across all of the subequations. These are the values that Mplus ultimately reports for Equation 13.4. Each of the separate intercepts also is reported but they are not assumed to be equal across the subequations. Thus, you will see a separate intercept for each subequation. Sometimes the parallel regression assumption is viable, sometimes not. I show you a way of testing it below. I make all of the above concrete shortly.

PRELIMINARY ANALYSES

In this section, I describe some of the preliminary analyses I routinely conduct when modeling ordinal outcomes whether I use the probability approach or the latent response approach. One of the first analyses I conduct is to evaluate the distribution of scores for the outcome and the predictors. I also evaluate the appropriateness of a probit function for the analysis and address issues of outliers and leverages. I then test the viability of the parallel coefficient assumption. My emphasis in the current section is not on repeating preliminary analyses I described in previous chapters (see the supplemental documents on my webpage for Chapters 11 and 12) but rather I highlight preliminary analyses that are particularly salient to ordinal modeling and that I have not yet covered. You, of course, would also perform the preliminary analyses described in Chapters 11 and 12, as appropriate.

The Outcome Distribution

It is important to examine the distribution of the ordinal outcome across the outcome categories with the idea of identifying categories that might have frequencies so small that the categories might be collapsed into an adjacent category prior to analysis. Simulation studies suggest that collapsing categories tends to have minimal impact on the estimation of the path/regression coefficients of ordinal models, instead primarily affecting model intercepts (see Peterson & Harrell 1988; Greenland 1994). Collapsing categories sometimes improves the viability of the parallel regression assumption, but it also can result in information loss relative to the contrasts of interest. Collapsing categories also can result in a loss of statistical power for tests of the path/regression

coefficients (Ananth & Kleinbaum, 1997; Manor et al., 2000; Murad, Fleishman, Sadetzki, Geyer & Freedman, 2003). In short, whether to collapse small-frequency categories is kind of a mixed bag.

Statisticians have raised concerns about the use of low base rate outcome categories in logit and probit modeling more generally, which are the primary engine of ordinal analyses via the subequations identified above. The concern is not so much with the small percentage of cases that occur in the smaller categories of the binary outcomes comprising the subequations but instead with the size of the absolute cell frequency (Allison, 2012). The presence of small cell frequencies can lead to bias in parameter estimates. A rough rule of thumb often offered for binary logit/probit regression is that there should be at least 10 individuals in the rarer category for each predictor (Vittinghoff & McCulloch, 2006). However, exceptions to this rule have been noted and much depends on other features of the data (Courvoisier, Combescure, Agoritsas, Gayet-Ageron & Perneger, 2011a, b; Steyerberg, Schemper, & Harrell, 2011). I discuss the use of logit/probit methods for small N in Chapter 28. Probably the best way to determine if your analysis will support a given sample size and cell frequency is to do so through a localized simulation, a strategy I discuss in Chapter 28.

For the IMP3 outcome, the percent and number of people in each category was 17% (n=102) for a score of 1, 31.2% (n=187), for a score of 2, 35.3% (n=213) for a score of 3, and 16.3% (n=98) for a score of 4. There does not seem to be a need to pursue collapsing.

Leverages

In Chapter 6, I discussed robust methods for identifying outliers (unusual scores on the outcome that might bias results) and leverages (unusual predictor profiles that might bias results). With sufficient frequencies at each level of the ordinal variable and the fact that large leverages typically drive many types of estimation problems in logit and probit modeling, I applied robust leverage analysis strategy to the predictors for Equation 13.4 (see the *Leverage analysis* program on my website and the video associated with the program for details). I identified 13 high leverage cases. For our numerical example, when I eliminated them from the overall model for the subsequent analyses I report in this chapter, none of my conclusions were affected. I therefore used all of the cases in the final analyses.

An effective analysis strategy that detects cases that are both outliers and that have notable leverages is one by Rousseuw and van Zomeren (1990; see Chapter 6). This method can be applied to Equations 13.1 to 13.3. I did so for the numerical example using the program called *Robust outlier analysis* on my website and did not find support for meaningful outlier-leverage problems in the data for these equations.

Appropriateness of Logit/Probit Modeling

Cumulative link ordinal regression relies on the analysis of multiple binary regression models. In the present case, there are three such binary regressions dictated by Equations 13.5 to 13.7. One approach to assessing the appropriateness of a given link function is to use the same methods I used in Chapter 12 (see the preliminary analysis supplement document for Chapter 12 on my webpage) but to apply them to the three binary outcomes defined by Equations 13.5 to 13.7. If the logit/probit function is affirmed as appropriate for each equation, one has increased confidence in the appropriateness of the function for the primary analyses. For our numerical example, I applied to each equation the polynomial regression approach, the partial residual plot approach, and the le Cessie–van Houwelingen–Copas–Hosmer approach described in the Chapter 12 supplement. I did not find evidence that the probit model would be inappropriate, so I move forward with it.

Parallel Coefficient Assumption

For logistic modeling using the proportional odds model, Mplus provides a statistic called the **Brant test**, which is an (imperfect) test of the parallel coefficient assumption. There is no corresponding test for probit-based ordinal modeling. An informal method for evaluating the assumption based on the logic of Asparouhov (2018) is to estimate in a single analysis the three binary probit subequations (Equations 13.5 to 13.7) without equality constraints on the coefficients and then to conduct contrasts of coefficient differences assumed to be zero in the larger ordinal regression. The relevant syntax for conducting these contrasts is shown in Table 13.1.

Table 13.1: Mplus Code for Parallel Regression Assumption

```
1.  TITLE: Parallel regression assumption test ;
2.  DATA: FILE IS c:\mplus\symptom.dat ;
3.  DEFINE:
4.    IF (IMP3 LE 1) THEN Y1 = 0 ELSE Y1 = 1 ;
5.    IF (IMP3 LE 2) THEN Y2 = 0 ELSE Y2 = 1 ;
6.    IF (IMP3 LE 3) THEN Y3 = 0 ELSE Y3 = 1 ;
7.  VARIABLE:
8.    NAMES ARE ID GA2 TA2 BD2 CE1 CIS1 T IMP3 ;
9.    USEVARIABLES ARE GA2 TA2 BD2 CE1 CIS1 T Y1 Y2 Y3 ;
10.   CATEGORICAL ARE y1 y2 y3 ;
11.   MISSING ARE ALL (-9999) ;
12. ANALYSIS:
13.   ESTIMATOR = ML ; LINK = PROBIT; BOOTSTRAP = 5000 ;
14. MODEL:
15.   Y1 on GA2 TA2 BD2 T CE1 CIS1  (p4a p5a p6a p7a b7a b8a)  ;
16.   Y2 on GA2 TA2 BD2 T CE1 CIS1  (p4b p5b p6b p7b b7b b8b)  ;
17.   Y3 on GA2 TA2 BD2 T CE1 CIS1  (p4c p5c p6c p7c b7c b8c)  ;
18. MODEL CONSTRAINT:
```

```
19.   NEW (p4a4b p4a4c p4b4c p5a5b p5a5c p5b5c p6a6b p6a6c p6b6c
20.        b7a7b b7a7c b7b7c b8a8b b8a8c b8b8c ) ;
21.   p4a4b = p4a-p4b ;
22.   p4a4c = p4a-p4c ;
23.   p4b4c = p4b-p4c ;
24.   p5a5b = p5a-p5b ;
25.   p5a5c = p5a-p5c ;
26.   p5b5c = p5b-p5c ;
27.   p6a6b = p6a-p6b ;
28.   p6a6c = p6a-p6c ;
29.   p6b6c = p6b-p6c ;
30.   b7a7b = b7a-b7b ;
31.   b7a7c = b7a-b7c ;
32.   b7b7c = b7b-b7c ;
33.   b8a8b = b8a-b8b ;
34.   b8a8c = b8a-b8c ;
35.   b8b8c = b8b-b8c ;
36. OUTPUT: SAMP RESIDUAL CINTERVAL(BOOTSTRAP) TECH4  ;
```

All of the code should be familiar from my discussion of Mplus programming in Chapters 11 and 12. Lines 4-6 define the binary outcome variables for Equations 13.5 to 13.7 using `DEFINE` subcommands. Y1 is $IMP3_{d1}$ in Equation 13.5, Y2 is $IMP3_{d2}$ in Equation 13.6, and Y3 is $IMP3_{d3}$ in Equation 13.7. Each of these created variables is listed last on the `USEVARIABLES` subcommand on Line 9, which is an Mplus requirement for newly created variables.[1] Line 14 requests a bootstrap analysis. Lines 18 through 35 request the pairwise tests of the coefficient differences for a given predictor across the three equations. There are 15 such tests, so one must take into account error rates for multiple contrasts.

Table 13.2 presents the coefficient values taken from the output for the three equations in the section called `MODEL RESULTS`. I expect the path coefficients to be reasonably comparable across the three columns within a given row of Table 13.2, recognizing there will be sampling error that causes some variation in them. With the possible exceptions of the T and CIS1 predictors, all of the coefficients seem reasonably homogenous.

Table 13.2: Probit Coefficients for Each Break Point

Predictor	$IMP3_{d1}$	$IMP3_{d2}$	$IMP3_{d3}$
GA2	1.02	0.80	0.72
TA2	0.71	0.77	0.66
BD2	0.86	0.71	0.64
T	0.53	-0.03	0.28

[1] This is not the case if an existing variable is transformed and the result stored in the same variable.

CE1	0.25	0.11	0.13
CIS1	0.58	0.13	0.27

Here is output for the pairwise significance tests:

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
New/Additional Parameters				
P4A4B	0.221	0.246	0.901	0.368
P4A4C	0.305	0.277	1.099	0.272
P4B4C	0.083	0.211	0.395	0.693
P5A5B	-0.064	0.243	-0.264	0.792
P5A5C	0.050	0.305	0.164	0.870
P5B5C	0.114	0.235	0.486	0.627
P6A6B	0.150	0.261	0.576	0.564
P6A6C	0.215	0.303	0.708	0.479
P6B6C	0.065	0.216	0.300	0.764
P7A7B	0.564	0.386	1.463	0.143
P7A7C	0.248	0.476	0.522	0.602
P7B7C	-0.316	0.341	-0.925	0.355
B7A7B	0.139	0.139	0.996	0.319
B7A7C	0.122	0.177	0.686	0.493
B7B7C	-0.017	0.129	-0.132	0.895
B8A8B	0.453	0.225	2.015	0.044
B8A8C	0.319	0.273	1.170	0.242
B8B8C	-0.134	0.212	-0.631	0.528

Only one of the contrasts was statistically significant. Given the number of contrasts performed, the result could be chance. In fact, I know it is given I created these hypothetical data using the Mplus simulation package, although I would not know this in practice. I address later how to deal with scenarios where the parallel regression assumption is untenable. For current purposes, I move forward with the analysis based on the overall trend in the tests. As well, I stress that this informal method and the significance tests themselves must be viewed with caution because conclusions assume adequate statistical power for the tests; low power would mean I might fail to reject the assumption of coefficient equivalence even when it is non-trivially violated. I also recommend you evaluate the contrasts using Mplus bootstrapping for sensitivity purposes. Indeed, this is true of all the non-Bayesian programs in this chapter.

As an aside, the coefficient for a predictor in the overall ordinal model typically will be a weighted average of the predictor coefficients for that predictor across the three different binary equations. We want to be careful of truly aberrant values within a row across columns of Table 13.2 because they can distort the overall coefficient estimate.

ORDINAL MODELING: APPLICATION OF THE PROBABILITY APPROACH

The probability approach characterizes intervention effects and mediator effects by focusing on breakpoint analysis using the cumulative link model for the ordinal outcome measure. For the numerical example, I outlined four contrasts within the three breakpoints (contrast 1, contrast 2a, contrast 2b, and contrast 3) that guide my analysis. I now walk you through analysis of the three questions of an RET for these contrasts, namely (1) does the intervention affect (and by how much) the outcome? (2) does the intervention affect and by how much each of the presumed mediators of program effects? and (3) do each of the mediators affect the outcome and by how much? Later, I approach the analysis using a different set of contrasts to show you the flexibility of ordinal modeling. I begin by evaluating the fit of the overall model in Figure 13.1.

Model Fit

The ML estimator with a probit link does not produce traditional global fit indices nor does it yield modification indices. To evaluate model fit, one can use the modification index strategy for LISEM discussed in Chapter 6 and illustrated in Chapter 12 but applied to this FISEM context. For example, to test if the omitted correlated disturbance between GA2 and TA2 has a statistically significant "modification index," I can add the correlation using the command `GA2 WITH TA2` and re-run the syntax to determine if the correlation is statistically significant. When I did so, the covariance for this parameter yielded a z value of 0.668 ($p < 0.51$) implying a modification index of $0.668^2 = 0.45$, which is trivial. When I used this strategy for other omitted parameters, I did not find any significant "modification indices."

Another strategy for evaluating specification error is to execute model syntax first using the WLSMV estimator instead of the maximum likelihood probit. WLSMV yields both global fit indices and localized stress indices for the model. I then use this information to explore ill fit in the maximum likelihood model because major points of stress from the WLSMV solution likely apply to the ML probit model as well. Technically, in FISEM, one should not shift estimators in this fashion. However, I have sometimes found the strategy informative on an informal, exploratory basis as I seek to identify specification error in my model.

Table 13.3 presents the Mplus syntax for the maximum likelihood probit ordinal regression as well as the modification to conduct an initial analysis based on WLSMV to help diagnose specification error.[2]

[2] I use ML instead of robust MLR, although the latter can be used as well. See my discussion in Chapter 12.

Table 13.3: Mplus Code for Ordinal Regression

```
1. TITLE: Ordinal regression with probit ;
2. DATA: FILE IS c:\mplus\symptom.dat ;
3. VARIABLE:
4.  NAMES ARE ID GA2 TA2 BD2 CE1 CIS1 T IMP3 ;
5.  USEVARIABLES ARE GA2 TA2 BD2 CE1 CIS1 T IMP3 ;
6.  CATEGORICAL ARE IMP3 ;
7.  MISSING ARE ALL (-9999) ;
8.  ANALYSIS:
9.  ESTIMATOR = ML ; LINK = PROBIT ;
10. !ESTIMATOR = WLSMV ;
11. MODEL:
12. GA2 on T CE1 CIS1 (p1 b1 b2)  ;
13. TA2 on T CE1 CIS1 (p2 b3 b4)  ;
14. BD2 on T CE1 CIS1 (p3 b5 b6)  ;
15. IMP3 on GA2 TA2 BD2 T CE1 CIS1 (p4 p5 p6 p7 b7 b8)  ;
16. [IMP3$1] (t1) ; [IMP3$2] (t2) ; [IMP3$3] (t3) ;
17. MODEL INDIRECT:
18. IMP3 IND T ;
19. OUTPUT: SAMP RESIDUAL STAND(STDY) CINTERVAL TECH4  ;
20. !OUTPUT: SAMP RESIDUAL STAND(STDY) MOD(ALL 4) CINTERVAL TECH4  ;
```

I comment out some of the lines using a ! and address them later. On Line 6, I declare IMP3 as an ordinal variable using the `CATEGORICAL` subcommand (Mplus internally discerns that IMP3 has more than two categories and treats it as ordinal). On Line 9, I use maximum likelihood as my estimator with a probit link (without the `LINK` command, Mplus uses the default logistic modeling for IMP3). Line 16 requests and labels the intercepts for each of the subequations identified earlier that Mplus works with internally. Mplus refers to these as thresholds, which equal the intercepts times minus 1. I describe later in the chapter why Mplus refers to them as thresholds instead of intercepts but for now, we just keep in mind that the intercepts are opposite signed thresholds. I label the threshold values from the three subequations on Line 16 using `t1`, `t2` and `t3`. The first threshold multiplied by -1 reflects the intercept for $IMP3_{d1}$ in Equation 13.5, the second threshold multiplied by -1 reflects the intercept for $IMP3_{d2}$ in Equation 13.6, and the third threshold multiplied by -1 reflects the intercept for $IMP3_{d3}$ in Equation 13.7. On the output Line 19, I omit modification indices and use standardization for endogenous variables, `STAND(STDY)`, per Chapter 12 and for reasons I discuss later.

To conduct the preliminary analysis using WLSMV, I remove the comment exclamation points on Lines 10 and 20 and then comment out the lines they replace by adding an exclamation point to Lines 9 and 19. The change in the OUTPUT line allows me to examine modification indices that are not available in maximum likelihood based probit analyses. (see Chapter 12 for a discussion of cautions to keep in mind when using

this approach). Here are the global fit statistics generated by the WLSMV analysis:

```
MODEL FIT INFORMATION

Chi-Square Test of Model Fit

          Value                              4.335
          Degrees of Freedom                     3
          P-Value                           0.2275

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                           0.027
          90 Percent C.I.                    0.000  0.079
          Probability RMSEA <= .05           0.704

CFI/TLI

          CFI                                0.999
          TLI                                0.996

SRMR (Standardized Root Mean Square Residual)

          Value                              0.008
```

All of the fit indices seem reasonable. There were no modification indices greater than 4 and the correspondence between the predicted and observed correlations was reasonable (not shown here).[3] Coupled with the other model fit diagnostics I already described, I am comfortable that the overall model is reasonably consistent with the data. I then restore the syntax in Table 13.3 to its original form for subsequent analyses, which I address later.

Question 1: The Effect of the Intervention on the Outcome

Meaningfulness Standards for the Total Program Effect

My first step is to set a meaningfulness standard for the total effect in order to determine if the program produces non-trivial effects (see Chapter 10). Setting such a standard is not straightforward when using ordinal regression because, per my discussion below, there are different ways of documenting the total effect. As a heuristic to help me think

[3] I examine the correlations in the Tech 4 output section called ESTIMATED CORRELATION MATRIX FOR THE LATENT VARIABLES and compare them with the observed correlations reported in the section ESTIMATED SAMPLE STATISTICS. The latter omits the ordinal outcome but the former includes it (but it is actually the y* version of it). Thus, one cannot compare predicted and observed correlations involving IMP3.

through total effect meaningfulness standards, I create a two way table with the different contrasts listed as rows and the treatment condition as columns, like this

Contrasts	Treatment Percent	Control Percent	Difference
C1: No change (1 vs. 2,3,4)	a	b	a-b
C2a: No change or minimal improve (1,2 vs. 2,3)	c	d	c-d
C2b: Much or very much improve (2,3 vs. 1,2)	e	f	e-f
C3: Very much improve (4 vs. 3,2,1)	g	h	g-h

My Mplus analyses will document the percentage of individuals who are in each primary contrast category and then perform the percentage difference contrasts indicated in the last column labeled "Difference." For example, for Contrast 1, I document the percentage of people in the intervention condition who show "no change or who got worse" (a score of 1) and the percentage of people in the control condition for whom this is the case. Obviously, I would hope that the percentage of people in the intervention condition who fall into this category is less than the percentage of people in the control condition who do so. How large a percentage difference for this contrast is meaningful and how large a difference is trivial? Answering such questions reveals my meaningfulness standard for this contrast. Ultimately, I apply the strategies and perspectives I discussed in Chapters 10 and 12 in consultation with program administrators and staff to evolve meaningfulness standards for each contrast in the table and to identify latitudes of meaningfulness, latitudes of no effect, and latitudes of ambiguity for those standards. In the interest of space, I do not elaborate this exercise here, but suppose I settle on an absolute percentage difference vis-à-vis the last column of the table greater than 5% for each contrast as being the cutoff for meaningfulness. I use this for my standards, accordingly.[4]

Although the four contrasts are linked to breakpoint analysis, I need to invoke some algebraic manipulations of model parameters in order to produce intuitive representations of them. I describe how to do so but later in the chapter I will consider other contrast approaches you can use that might better fit your needs.

Total Effect of the Program on the Outcome

Mplus reports the total effect of the intervention using the latent response framework, so it is not applicable to the probability approach. Program staff and administrators often

[4] In practice, it is not necessary for the standard to be the same for each contrast.

want characterizations of program effects using the specific categories of the outcome metric per se. This can be difficult to accomplish in FISEM with ordinal outcomes when the focus is on total effects. I find it easier to abandon the FISEM approach for this facet of RET analysis and to instead use LISEM, just as I did in Chapter 12 for probit modeling. In addition to circumventing certain statistical complications, doing so protects against specification error in the broader model that might bias total effect estimates. The primary disadvantage is that use of LISEM complicates decomposition of the total effect for analyzing the proportion of the total effect that different mediators account for. Such decompositions are, in my opinion, a low priority for purposes of program evaluation, so I am not inclined to let this get in the way.

I describe two methods for total effect analysis using LISEM based ordinal regression, one using profile analysis and another using average marginal effects. I introduced each of these methods in Chapter 12 for binary outcomes and extend them here to ordinal regression.

Probability/Proportion Differences: Profile Analysis. To perform the contrasts for total effects using LISEM, I step outside the full SEM model and zero in on a single equation that represents an ordinal regression of the outcome onto the dummy coded treatment condition variable (1 = intervention, 0 = control) and the operative covariates I want to control for:

$$\text{Probit(IMP3)} = a + p_1\,T + b_2\,CE1 + b_3\,CIS1 \qquad [13.8]$$

where *a* is an intercept and the coefficients are assumed to take on the same values across the subequations implied by the model. Table 13.4 presents the syntax for the analysis.

Table 13.4: Mplus Code for LISEM Total Effect Analysis

```
1.  TITLE: LISEM total effect analysis ;
2.  DATA: FILE IS c:\mplus\symptom.dat ;
3.  VARIABLE:
4.    NAMES ARE ID GA2 TA2 BD2 CE1 CIS1 T IMP3 ;
5.    USEVARIABLES ARE CE1 CIS1 T IMP3 ;
6.    CATEGORICAL ARE IMP3 ;
7.    MISSING ARE ALL (-9999) ;
8.  ANALYSIS:
9.    ESTIMATOR = ML ; LINK = PROBIT ;
10. MODEL:
11.   IMP3 on T CE1 CIS1 (p1 b1 b2 )  ;
12.   [IMP3$1] (t1) ;  [IMP3$2] (t2) ; [IMP3$3] (t3) ;
13. MODEL CONSTRAINT:
14.   NEW (c1probc c1probt c1diff
15.   c2aprobc c2aprobt c2adiff
```

```
16.   c2bprobc c2bprobt c2bdiff
17.   c3probc c3probt c3diff ) ;
18.   c1probc = 1-phi(-t1 + p1*0 + b1*3.036 + b2*7.044) ;
19.   c1probt = 1-phi(-t1 + p1*1 + b1*3.036 + b2*7.044) ;
20.   c1diff = c1probt - c1probc ;
21.   c2aprobc = 1-phi(-t2 + p1*0 + b1*3.036 + b2*7.044) ;
22.   c2aprobt = 1-phi(-t2 + p1*1 + b1*3.036 + b2*7.044) ;
23.   c2adiff = c2aprobt - c2aprobc ;
24.   c2bprobc = phi(-t2 + p1*0 + b1*3.036 + b2*7.044) ;
25.   c2bprobt = phi(-t2 + p1*1 + b1*3.036 + b2*7.044) ;
26.   c2bdiff = c2bprobt - c2bprobc ;
27.   c3probc = phi(-t3 + p1*0 + b1*3.036 + b2*7.044) ;
28.   c3probt = phi(-t3 + p1*1 + b1*3.036 + b2*7.044) ;
29.   c3diff = c3probt - c3probc ;
30. OUTPUT: SAMP RESIDUAL STANDARDIZED(STDY) CINTERVAL TECH4  ;
```

Most of the syntax should be familiar to you if you have read Chapters 11 and 12. Line 6 declares IMP3 as ordinal using the keyword `CATEGORICAL`. Line 12 specifies thresholds and labels for them. The workhorse portion of the program is the `MODEL CONSTRAINT` lines that implement the different contrasts. For each contrast, I need to assign specific values for the covariates at which to hold them constant and, in this instance, I do so using their sample mean values, i.e., their "typical" values. The means were 3.036 for `CE1` and 7.044 for `CIS1`. I consider the use of different covariate values shortly.

Lines 14 to 17 use the `NEW` subcommand to assign labels to each contrast or intermediate terms needed for the calculation of the contrast values. Recall that you can use any label you want, but it can be no longer than 8 characters. I begin all of the entries for the first contrast entries with the letters c1; the second contrast use entries that begin with c2a and c2b; and the last contrast uses entries that begin with c3.

Focusing on the first contrast, Line 18 calculates the control group's probability of being in category 1 of the outcome ("no change or got worse"). Line 19 calculates the corresponding probability for the intervention/treatment group. Line 20 calculates the difference between these two probabilities and is of primary interest. Considering each of these lines in more depth, the terms within the parentheses for Line 18 generate the predicted probit value for Equation 13.8 (which focuses on $IMP3_{d1}$) for the control group. The intercept is represented by `-t1` , where `t1` is the label in the main part of the program used for the first threshold. The coefficient p1 is the path coefficient associated with the treatment dummy variable in Equation 13.8. It is multiplied by 0 to reflect the control group. The coefficient b1 is the coefficient for the baseline covariate `CE1` and b2 is the baseline coefficient for `CIS1`. Each are multiplied by their respective mean values to hold them constant at these values. When the full expression is executed, the result is the predicted probit value for $IMP3_{d1}$ and when I apply the `PHI` function to it, Mplus

converts the predicted probit value to a probability. Note, however, that the probability in this case is for IMP_{d1} which has a score of 0 for category 1 and a score of 1 for categories 2, 3, and 4 combined. I want to isolate instead the probability for being in category 1 in order to execute Contrast 1 and I can accomplish this by subtracting the probability of the combined categories of 2, 3, and 4 from 1.0, which I have done in Line 18 (see the 1- entry at the beginning of the line). Line 19 follows the same process for the intervention group but now the coefficient p1 is multiplied by 1 to represent the intervention group. Line 20 differences the two category 1 probabilities.

Contrasts 2a and 2b follow this same format but I use `t2` in place of `t1` to reflect a focus on $IMP3_{d2}$. Contrast 3 does so as well but I use `t3` to reflect a focus on $IMP3_{d3}$. For contrasts 1 and 2a, the probabilities are subtracted from 1 and for contrasts 2b and 3, they are not. This orients the probabilities to the way the contrasts are phrased relative to Mplus' focus on $IMP3_{d1}$, $IMP3_{d2}$, and $IMP3_{d3}$.

Because the model is just-identified, issues of model fit are moot. The output for the contrasts in proportion form is in the output section `New/Additional Parameters:`

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
New/Additional Parameters				
C1PROBC	0.243	0.024	10.304	0.000
C1PROBT	0.018	0.004	4.317	0.000
C1DIFF	-0.225	0.022	-10.021	0.000
C2APROBC	0.739	0.026	28.866	0.000
C2APROBT	0.226	0.023	9.688	0.000
C2ADIFF	-0.514	0.031	-16.807	0.000
C2BPROBC	0.261	0.026	10.181	0.000
C2BPROBT	0.774	0.023	33.254	0.000
C2BDIFF	0.514	0.031	16.807	0.000
C3PROBC	0.017	0.004	3.930	0.000
C3PROBT	0.237	0.025	9.329	0.000
C3DIFF	0.220	0.024	9.275	0.000

I summarize the results in Table 13.5 using percents instead of proportions, i.e. the proportions are multiplied by 100.

Table 13.5: LISEM Total Effects based on Profile Analysis

Contrast Categories	Treatment Percent	Control Percent	Difference
C1: No change	1.8 ±0.8	24.3 ±4.8	-22.5 ±4.4*
C2a: No change or minimal improve	22.6 ±4.3	73.9 ±5.2	-51.4 ±6.4*
C2b: Much or very much improved	77.4 ±4.6	26.1 ±5.2	51.4 ±6.4*
C3: Very much improved	23.7 ±5.0	1.7 ±0.8	22.0 ±4.8*

Table notes: * $p < 0.05$; C2a and C2b are statistically redundant

The percent of control individuals in the lowest category of the IMP3 scale (no change) was 24.3% ±4.8 as compared with 1.8% ±0.8 of individuals in the intervention group, a difference that was statistically significant (critical ratio (CR) = 10.02, $p < 0.05$). The percent of control individuals in the lowest two categories of the IMP3 scale (no change and minimal change combined) was 73.9% ±5.2 as compared with 22.6% ±4.6 in the treatment group, a difference that was again statistically significant (CR = 16.81, $p < 0.05$). The percent of control individuals in the highest two categories of the IMP3 scale (much or very much improved) was 26.1% ±5.2 as compared with 77.4% ±4.6 in the treatment group, a difference that was statistically significant (CR = 16.81, $p < 0.05$). Finally, the percent of control individuals in the highest category of the IMP3 scale (very much improved) was 1.7% ±0.8 as compared with 23.7% ±5.0 in the treatment group, a difference that was statistically significant (CR = 9.27, $p < 0.05$). In practice, researchers often highlight treatment versus control percent differences in the lowest and highest outcome categories, namely contrasts C1 and C3 (Agresti & Tarantola, 2018). However, given the particular descriptors of the IMP metric, I felt program staff and administrators might be interested in the C2 contrasts as well. Also, I calculated the MOEs in Table 13.5 using the "double the standard error" heuristic. You can use bootstrapping to check for confidence interval asymmetry and adjust them accordingly. All of the effects described in Table 13.5 exceeded their meaningfulness standards.

As noted in Chapter 12, the percentage differences in Table 13.5 can change depending on the values at which the covariates are held constant. In the above analysis, I set the covariates equal to their mean values, but I can explore variants of the syntax in which I hold the covariates constant at different values, such as low or high scores on them. Of interest is how much the magnitude of the total effect is predicted to change as I move from one predictor profile to the next. In Chapter 12, I illustrated these types of

analyses for probit regression with a binary outcome. The ideas directly extend to the case of ordinal regression. Table 13.6 show the results when I held constant CE1 and CIS1 at low scores, namely their 25th quantiles (2.38 and 6.67) whereby the clinicians being trained are relatively inexperienced and have lower interpersonal skills. The same statistical significance trends are apparent as those for the mean centered analyses, but the percentage values and differences change somewhat.

I repeated the analyses a third time but now using covariate values that mapped onto the 75th quantiles of the covariates (3.72 and 7.37). In this case, I evaluate treatment versus control differences for clinicians who are experienced and interpersonally skilled. The same trends of statistical significance were apparent (see Table 13.7).

Table 13.6: LISEM Total Effect Analysis at 25th Quantiles for Covariates

Contrast Categories	Treatment Percent	Control Percent	Difference
C1: No change	6.8 ±2.4	46.1 ±6.8	-39.3 ±6.2*
C2a: No change or minimal improve	43.8 ±7.0	89.2 ±3.6	-45.4 ±6.2*
C2b: Much or very much improved	56.2 ±7.0	10.8 ±3.6	45.4 ±6.2*
C3: Very much improved	9.5 ±3.4	0.3 ±0.2	9.1 ±3.2*

Table 13.7: LISEM Total Effect Analysis at 75th Quantiles for Covariates

Contrast Categories	Treatment Percent	Control Percent	Difference
C1: No change or got worse	0.4 ±0.2	10.2 ±3.2	-9.8 ±3.0*
C2a: No change or minimal improve	9.2 ±3.0	52.6 ±6.4	-43.4 ±6.0*
C2b: Much or very much improved	90.8 ±3.0	47.4 ±6.4	43.4 ±6.0*
C3: Very much improved	44.5 ±6.6	6.3 ±2.4	38.2 ±6.0*

Table notes: * $p < 0.05$; C2a and C2b are statistically redundant

Average Marginal Effects. Another approach to documenting treatment-control differences for the total effect is to calculate the average marginal effect for each of the

contrasts. I discussed average marginal effects in Chapters 6 and 12. Here is the output for the probit equation from the analysis in Table 13.3 that I need to make use of when I program Mplus to calculate the AMEs:

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

 IMP3       ON
    T                 1.394      0.098     14.292      0.000
    CE1               0.498      0.053      9.358      0.000
    CIS1              0.725      0.094      7.731      0.000

Thresholds
    IMP3$1            5.923      0.650      9.119      0.000
    IMP3$2            7.261      0.666     10.901      0.000
    IMP3$3            8.729      0.682     12.800      0.000
```

These results yield the following three probit subequations:

$$\text{Probit}(IMP3_{d1}) = -5.923 + 1.394\ T + .498\ CE1 + .725\ CIS1$$

$$\text{Probit}(IMP3_{d2}) = -7.261 + 1.394\ T + .498\ CE1 + .725\ CIS1$$

$$\text{Probit}(IMP3_{d3}) = -8.729 + 1.394\ T + .498\ CE1 + .725\ CIS1$$

I use these equations to calculate AMEs for the contrasts using the syntax in Table 13.8.

Table 13.8: Syntax for Calculating AMEs

```
1.  TITLE: LISEM total effect analysis ;
2.  DATA: FILE IS c:\mplus\symptom.dat ;
3.  DEFINE:
4.  !Contrast 1
5.    PROBIT1C = -5.923 + 1.394*0 + .498*CE1 + .725*CIS1 ;
6.    PROB1C = 1-PHI(PROBIT1C) ; !C1 probability for control group
7.    PROBIT1T = -5.923 + 1.394*1 + .498*CE1 + .725*CIS1 ;
8.    PROB1T = 1-PHI(PROBIT1T) ; !C1 probability for treat group
9.    IME1 = PROB1T-PROB1C ;
10. !Contrast 2A
11.   PROBIT2AC = -7.261 + 1.394*0 + .498*CE1 + .725*CIS1 ;
12.   PROB2AC = 1-PHI(PROBIT2AC) ; !C2A probability for control group
13.   PROBIT2AT = -7.261 + 1.394*1 + .498*CE1 + .725*CIS1 ;
14.   PROB2AT = 1-PHI(PROBIT2AT) ; !C2A probability for treat group
15.   IME2A = PROB2AT-PROB2AC ;
16. !Contrast 2B
17.   PROBIT2BC = -7.261 + 1.394*0 + .498*CE1 + .725*CIS1 ;
18.   PROB2BC = PHI(PROBIT2BC) ; !C2B probability for control group
19.   PROBIT2BT = -7.621 + 1.394*1 + .498*CE1 + .725*CIS1 ;
```

```
20.    PROB2BT = PHI(PROBIT2BT) ; !Probability for treat group
21.    IME2B = PROB2BT-PROB2BC ;
22. !Contrast 3
23.    PROBIT3C = -8.729 + 1.394*0 + .498*CE1 + .725*CIS1 ;
24.    PROB3C = PHI(PROBIT3C) ; !C3 probability for control group
25.    PROBIT3T = -8.729 + 1.394*1 + .498*CE1 + .725*CIS1 ;
26.    PROB3T = PHI(PROBIT3T) ; !Probability for treat group
27.    IME3 = PROB3T-PROB3C ;
28. VARIABLE:
29.    NAMES ARE ID GA2 TA2 BD2 CE1 CIS1 T IMP3 ;
30.    USEVARIABLES ARE IME1 IME2A IME2B IME3 PROB1C PROB1T
31.    PROB2AC PROB2AT PROB2BC PROB2BT PROB3C PROB3T;
32.    MISSING ARE ALL (-9999) ;
33.    ANALYSIS:
34. ESTIMATOR = ML ;  TYPE = BASIC ;
35. OUTPUT: !use defaults on output
```

I described the logic of this syntax in the Appendix of Chapter 12. For contrast 1 in Line 5, I set the value for T to 0 for everyone (see the term 1.394*0) and in Line 7, I set it to 1 for everyone (see the term 1.394*1). I calculate the predicted probit value and then converted this to a probability in Lines 6 and 8. I subtract some of the probabilities from 1 (see Lines 6, 8, 12 and 14) per my previous discussion about isolating the contrast categories that map onto how I phrased the contrasts. I next calculate the difference between the two probabilities to obtain the individual marginal effect. I repeat the process for each contrast and then have the program average values to yield the AMEs (Line 34).

Mplus produces warning messages in this analysis but they can be ignored. I obtain the means of `IME1`, `IME2`, `IME3`, and `IME4` from the output called `RESULTS FOR BASIC ANALYSIS`. The averages that Mplus calculates include the proportions that are differenced as part of the AME. The results from the output are shown in Table 13.9 in percentage form. A disadvantage of the Mplus syntax is that it does not yield confidence intervals or significance tests for the AMEs. I provide a program on my webpage called *AMEs: Ordinal-multinomial* that does so in a LISEM context. However, it uses a different parametrization of the contrasts, which I describe in the Appendix.

In sum, the program leads to significant reductions of patients in the lowest category of symptom change (no change) as well as reductions of patients in the lowest two categories (no change or minimal change). The program increases the proportion of patients in the highest two categories (much improved or very much improved) as well as the highest category of symptom change (very much improved). The magnitude of the effects are meaningful. Keep in mind that the AMEs are conceptually distinct from the conditional total effects approach that uses panel analysis, although their results converged on the same conclusions in this case.

Table 13.9: AMEs for Total Effect Analysis

Contrast Categories	Treatment Percent	Control Percent	Difference
C1: No change	4.8	28.7	-23.9
C2a: No change or minimal improve	37.5	69.8	-32.2
C2b: Much or very much improve	62.5	30.2	32.2
C3: Very much improve	28.2	4.6	23.5

Parenthetically, there is a third, model-free way of evaluating the total effect of the RET based on contingency tables in an LISEM context. I provide a document called *Supplementary Total Effect Analysis for Ordinal Modeling* that describes this method on the resources tab of my webpage in Chapter 13. Both of the approaches I outline here for documenting the total effect of the intervention on an ordinal outcome (LISEM modeling and AMEs) have strengths and weaknesses. I like to examine each of them to gain multiple perspectives on the data.

Question 2: Effect of the Intervention on the Mediators

Meaningfulness Standard for Intervention Effect on Mediators

Each of the mediators were measured on multi-item scales in which patients rated each scale item on 7 point agree-disagree metrics (-3 = strongly disagree, -2 = moderately disagree, -1 = slightly disagree, 0 = neither agree nor disagree, 1 = slightly agree, 2 = moderately agree, 3 = strongly agree). The item responses were then averaged across items for the scale. Suppose that after carefully reviewing item content on each scale with program staff and clinicians as well as making use of other concepts discussed in Chapter 10, I determine that a shift of half a scale unit (0.50) on the overall score for each of the mediators, considered separately, would be meaningful. This value, then, becomes my meaningfulness standard for the mediators. Note that in this case I am using the same standard for each mediator but this need not be the case.

Analysis of Intervention Effects on Mediators

There are three relevant equations for assessing intervention effects on the mediators, one equation per mediator (see Equations 13.1 to 13.3). Here is the relevant output for the GA2 mediator from the overall analysis that used the syntax in Table 13.3.

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
GA2	ON				
T		0.989	0.037	26.851	0.000
CE1		0.179	0.020	8.867	0.000
CIS1		0.261	0.039	6.623	0.000

All variables but T are covariates. The coefficient of interest is that for T, which reflects the covariate adjusted mean difference between the treatment and control conditions. The difference on the -3 to +3 metric of goal alignment was 0.99 ±0.07 (CR = 26.85, $p < 0.05$). The lower limit of the 95% confidence interval for the intervention versus control mean difference was 0.92, which is larger than the meaningfulness standard of 0.50. This suggests the effect of the intervention on GA2 was meaningful. The comparable output for TA2 and BD2 yielded covariate adjusted mean differences between the intervention and control conditions of 0.93 ±0.07 (CR = 25.80, $p < 0.05$) and -0.015 ±0.07 (CR = 0.42, $p < 0.68$), respectively. Evaluation of the lower limits of the relevant confidence intervals revealed that the mean difference for TA2 was meaningful but the mean difference for BD2 was not.[5]

In sum, the analyses suggest the program was effective in changing goal alignment and task alliance. However, the program did not meaningfully affect positive bonding. The program designers need to revisit their intervention and figure out better ways of teaching clinicians how to bond with their clients.

Question 3: Effect of the Mediators on the Outcome

Meaningfulness Standards for Mediator Effects on Outcome

To set meaningfulness standards for the effects of the mediators on the outcome, we need to think about the following questions:

Contrast M1: What is a meaningful shift in the proportion of clients in the lowest outcome category, (i.e., clients who show no change in symptoms) given a *k* unit increase in a mediator?

Contrast M2a: What is a meaningful shift in the proportion of clients in the lowest two categories (i.e., clients who show no or only minimal improvement) given a *k* unit

[5] The separate covariate adjusted posttest means for the control and intervention conditions can be estimated by re-running the Table 13.3 syntax but adding mean centering with DEFINE commands and then examining the intercepts for each mediator to isolate the control group means. Then do the same but re-run the program reverse scoring the treatment condition dummy variable.

increase in a mediator?

Contrast M2b: What is a meaningful shift in the proportion of clients in the highest two outcome categories (i.e., clients who were much improved or very much improved) given a *k* unit increase in a given mediator. (As before, this is redundant with contrast M2a, just framed differently)

Contrast M3: What is a meaningful shift in the proportion of clients who were very much improved, the top category of the scale, given a *k* unit increase in a mediator.

In traditional regression analyses, *k* is usually set to 1.0, e.g., for an OLS regression coefficient, we usually interpret its magnitude as how much the mean of Y is predicted to change given a one unit change in the predictor.

Answering the above questions by directly referencing the probit coefficient for a mediator predicting the outcome is challenging because the coefficients are non-intuitive and because of possible non-linear relationships between mediators and outcome probabilities. The strategy I use is similar to the one I described in Chapter 12 for mediator-outcome effects for the analysis of probit modeling of binary outcomes.

Consider contrast 1 for the goal alignment mediator, GA2. I first consult the standard I set earlier for this contrast for the total effect. That standard was a change of 5%. I do not expect one mediator to account for all of the total effect; I instead might expect each mediator to carry an equal share of the load in producing the total effect. Given three mediators, I divide the 5% total effect standard by 3, which yields 0.017 or 1.7%. Next, I need to factor into my standard what I think is a reasonable amount of change in the GA2 mediator that I can expect of the intervention. In the previous section, I found that the intervention changed GA2 by about 1.00 units (to be exact, the effect of the intervention on GA2 was estimated to be 0.99 ±0.07). Using 1.00 as an index of the amount of change in GA2 I can reasonably expect, in order to produce a 0.017 reduction in the proportion of people who show no improvement (category 1), the probit coefficient for GA2 needs to translate into a proportional change of 0.017/1.00 = 0.017, or 1.7% in the desired direction. I use this as my meaningfulness standard for this contrast for the GA2 mediator. For the mediator TA2, the amount of change the intervention brought about in it was also close to 1.00 (it was 0.93 ±0.07), so I decide to use the same meaningfulness standard for it. For BD2, I encounter a complication. The complication is that the effect of the intervention on BD2 was virtually nil. Obviously, the program designers need to alter what they are doing to bring about change in BD2. If they do so, what magnitude of change can I reasonably expect the intervention to have on BD2? Suppose after discussions with relevant staff, we decide that a reasonable guess is about

1.0 units on the -3 to +3 metric of BD2. This yields a meaningfulness standard of 0.017/1.00 = 0.017. For elaboration on this logic, see Chapter 10. Essentially, I set the standard based on the overall proportional change I desire for the contrast, the amount of change in the mediator that I can reasonably expect from the intervention, the number of mediators, and the proportionate share of the contrast effect that I want the mediator to account for relative to the other mediators. I decide in this study to apply the same standards to the other contrasts as well, but with sign adjustments to reflect the desired change direction.

I describe two strategies for evaluating mediator effects on the outcome, profile analysis and average marginal effects.

Profile Analysis

For profile analysis of mediator effects on the outcome, I use Equation 13.4 which I repeat here for convenience and with a probit link indicated:

$$\text{Probit}(IMP3) = a_4 + p_4\,GA2 + p_5\,TA2 + p_6\,BD2 + p_7\,T + b_7\,CE1 + b_8\,CIS1$$

I re-run the syntax in Table 13.3 for this equation but I add a `MODEL CONSTRAINT` command and associated subcommands to it. The logic of these commands is to estimate the proportion or percentage of clients who show symptom improvement for strategically defined predictor profiles. By comparing estimates of different pairs of profiles, I can make inferences about the effect of each mediator on the outcome. For a given pair of profiles targeting a mediator for a given contrast (say contrast 1), the first profile is defined by typical scores of the control group on each variable in the equation. The second profile is the same but it increases the value of the target mediator, say GA2, by one unit. By formally comparing the change in proportions for these two profiles, I gain a sense of the effect of a one unit change in GA2 on outcome probabilities/proportions for contrast 1 using control group proportions as a base.

For each mediator, I calculated the "typical" or mean control group value at the posttest and they generally were near 0, which is the "neither agree nor disagree" point on their respective metrics. I therefore used the values of 0 to define initial "typical" scores on the mediators for the first profile. I use the equivalent of mean centering for the baseline covariates by multiplying their coefficients by their respective means calculated across the total sample because these measures were taken before randomization and likely reflect the means from the general population from which study participants were sampled. I set T = 0 to reflect the control group in the first profile, again to capture the "natural" state of the study population. In order to keep all values for the second profile except that for the target mediator (in this case, GA2, which I increase by 1) equivalent, I

set T = 0 for the second profile as well, i.e. the profile values for the first profile are GA2= 0, TA2 = 0, BD2 = 0, T = 0, CE1 = 3.036, CIS1 = 7.044 and for the second profile they are GA2= 1, TA2 = 0, BD2 = 0, T = 0, CE1 = 3.036, CIS1 = 7.044. Note that the only difference in the two profiles is that I increased GA2 one unit in the second profile.

Table 13.10 presents the syntax for the GA2 mediator for each of the four contrasts.

Table 13.10: Code for Effect of Mediator on Outcome Profile Analysis

```
1.  TITLE: Ordinal regression with probit profile analysis 1 ;
2.  DATA: FILE IS c:\mplus\symptom.dat ;
3.  VARIABLE:
4.  NAMES ARE ID GA2 TA2 BD2 CE1 CIS1 T IMP3 ;
5.    USEVARIABLES ARE GA2 TA2 BD2 CE1 CIS1 T IMP3 ;
6.    CATEGORICAL ARE IMP3 ;
7.    MISSING ARE ALL (-9999) ;
8.  ANALYSIS:
9.  ESTIMATOR = ML ; LINK = PROBIT ;
10. MODEL:
11.   GA2 on T CE1 CIS1 (p1 b1 b2)  ;
12.   TA2 on T CE1 CIS1 (p2 b3 b4)  ;
13.   BD2 on T CE1 CIS1 (p3 b5 b6)  ;
14.   IMP3 on GA2 TA2 BD2 T CE1 CIS1 (p4 p5 p6 p7 b7 b8)  ;
15.   [IMP3$1] (t1) ; [IMP3$2] (t2) ; [IMP3$3] (t3) ;
16. MODEL INDIRECT:
17. IMP3 IND T ;
18. MODEL CONSTRAINT:
19. NEW (c1pm0 c1pm1 c1diff c2apm0 c2apm1 c2adiff
20.      c2bpm0 c2bpm1 c2bdiff c3pm0 c3pm1 c3diff ) ;
21. !CONTRAST 1
22.  c1pm0 = 1-phi(-t1+p4*0+p5*0+p6*0+p7*0+b7*3.036+b8*7.044) ;
23.  c1pm1 = 1-phi(-t1+p4*1+p5*0+p6*0+p7*0+b7*3.036+b8*7.044) ;
24.  c1diff = c1pm1-c1pm0 ;   !prob difference
25. !CONTRAST 2A
26.  c2apm0 = 1-phi(-t2+p4*0+p5*0+p6*0+p7*0+b7*3.036+b8*7.044) ;
27.  c2apm1 = 1-phi(-t2+p4*1+p5*0+p6*0+p7*0+b7*3.036+b8*7.044) ;
28.  c2adiff = c2apm1-c2apm0 ; !prob difference
29. !CONTRAST 2B
30.  c2bpm0 = phi(-t2+p4*0+p5*0+p6*0+p7*0+b7*3.036+b8*7.044) ;
31.  c2bpm1 = phi(-t2+p4*1+p5*0+p6*0+p7*0+b7*3.036+b8*7.044) ;
32.  c2bdiff = c2bpm1-c2bpm0 ; !prob difference
33. !CONTRAST 3
34.  c3pm0 = phi(-t3+p4*0+p5*0+p6*0+p7*0+b7*3.036+b8*7.044);
35.  c3pm1 = phi(-t3+p4*1+p5*0+p6*0+p7*0+b7*3.036+b8*7.044);
36.  c3diff = c3pm1-c3pm0 ;    !prob difference
37. OUTPUT: SAMP RESIDUAL STAND(STDY) CINTERVAL TECH4  ;
```

When TA2 is the target (not shown in the syntax), I manipulate the value of TA2 for p5 across the two profiles for a contrast and when BD2 is the target (also not shown in the syntax), I manipulate the value of BD2 for p6. Here are the GA2 results:

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

New/Additional Parameters
    C1PM0             0.230      0.026      8.967      0.000
    C1PM1             0.062      0.017      3.554      0.000
    C1DIFF           -0.168      0.020     -8.476      0.000
    C2APM0            0.783      0.024     32.408      0.000
    C2APM1            0.492      0.052      9.515      0.000
    C2ADIFF          -0.291      0.044     -6.625      0.000
    C2BPM0            0.217      0.024      9.004      0.000
    C2BPM1            0.508      0.052      9.824      0.000
    C2BDIFF           0.291      0.044      6.625      0.000
    C3PM0             0.007      0.002      2.915      0.004
    C3PM1             0.047      0.014      3.298      0.001
    C3DIFF            0.040      0.013      3.145      0.002
```

Table 13.11 summarizes the above results but expressed as percentages, that is, I multiply each of the proportions by 100. I also provide results for when I ran comparable syntax for the other two mediators.

Table 13.11 Effects of Mediators on Outcome for Profile Analysis

	Contrast	Percent at Score of 1	Percent at Score of 0	Difference
GA2	C1: No change	6.2 ±3.4	23.0 ±5.2	-16.8 ±4.0*
	C2a: No change or minimal improve	49.2 ±10.4	78.3 ±4.8	-29.1 ±8.8*
	C2b: Much or very much improve	50.8 ±10.4	21.7 ±4.8	29.1 ±8.8*
	C3: Very much improve	4.7 ±2.8	0.7 ±0.4	4.0 ±2.6*
TA2	C1: No change	7.6 ±4.0	23.0 ±5.2	-15.4 ±4.0*
	C2a: No change or minimal improve	53.5 ±10.2	78.3 ±4.8	-24.7 ±8.6*
	C2b: Much or very much improve	46.5 ±10.2	21.7 ±4.8	24.7 ±8.6*
	C3: Very much improve	3.7 ±2.4	0.7 ±0.4	3.1 ±2.3*
	C1: No change	7.7 ±4.2	23.0 ±5.2	-15.3 ±4.2*

BD2	C2a: No change or minimal improve	53.8 ±10.4	78.3 ±5.8	-24.4 ±8.8*
	C2b: Much or very much improve	46.2 ±10.4	21.7 ±4.8	24.4 ±8.8*
	C3: Very much improve	3.7 ±2.4	0.7 ±0.4	3.0 ±2.0*

Table notes: * $p < 0.05$; C2a and C2b are statistically redundant

A one unit increase in goal alignment leads to statistically significant reductions of clients in the lowest category of "no change in symptoms" as well as in the lowest two categories that imply no or minimal treatment response. A one unit increase in goal alignment also leads to significant increases of clients in the highest category of the symptom measure (very much improved) as well as in the highest two categories that both imply treatment responsive symptom change. The bottom two sections of Table 13.12 show the results when I did comparable profile analyses for TA2 and BD2. All of the mediators targeted by the program were meaningfully relevant to the outcome, although the meaningfulness standard for C3 was not unambiguously met if one considers the lower limit of the 95% confidence interval of the percent differences. Using the joint significance test, GA2 and TA2 are both declared as non-zero mediators of the effects of the treatment on the outcome, but this is not the case for BD2. The analysis of the direct effect of the treatment independent of the mediators (not shown here) did not yield results supportive of the effect, which is consistent with the overall ordinal analysis in which its coefficient was statistically non-significant.

To explore the generalizability of the GA2 effects across different profile contexts, I repeated the above analysis but now holding the two non-manipulated mediators in the profiles, TA2 and BD2, constant at values of their 75th quantile. I also used the 75th quantiles for the two baseline covariates. I held T constant at a value of 1. Note that all of these changes in the non-target predictors and covariates are slanted towards increasing overall symptom improvement; so the one unit manipulation of GA2 is now occurring in a context that is with patients experiencing higher levels of symptom improvement due to more favorable profile values on the TA2, BD2, CE1, and CIS1 variables. Table 13.12 presents the relevant syntax for the GA2 mediator analysis.

Table 13.12: Code for Effects of Mediator on Outcome Profile Analysis 2

```
1.  TITLE: Ordinal regression with probit profile analysis 2 ;
2.  DATA: FILE IS c:\mplus\symptom.dat ;
3.  VARIABLE:
4.  NAMES ARE ID GA2 TA2 BD2 CE1 CIS1 T IMP3 ;
5.    USEVARIABLES ARE GA2 TA2 BD2 CE1 CIS1 T IMP3 ;
6.    CATEGORICAL ARE IMP3 ;
```

```
7.    MISSING ARE ALL (-9999) ;
8.  ANALYSIS:
9.  ESTIMATOR = ML ; LINK = PROBIT ;
10. MODEL:
11.   GA2 on T CE1 CIS1 (p1 b1 b2)  ;
12.   TA2 on T CE1 CIS1 (p2 b3 b4)  ;
13.   BD2 on T CE1 CIS1 (p3 b5 b6)  ;
14.   IMP3 on GA2 TA2 BD2 T CE1 CIS1 (p4 p5 p6 p7 b7 b8)  ;
15.   [IMP3$1] (t1) ; [IMP3$2] (t2) ; [IMP3$3] (t3) ;
16. MODEL INDIRECT:
17. IMP3 IND T ;
18. MODEL CONSTRAINT:
19. NEW (c1pm0 c1pm1 c1diff c2apm0 c2apm1 c2adiff
20.      c2bpm0 c2bpm1 c2bdiff c3pm0 c3pm1 c3diff ) ;
21. !CONTRAST 1
22.  c1pm0 = 1-phi(-t1+p4*0+p5*1+p6*0.37+p7*1+b7*3.1775+b8*7.3700) ;
23.  c1pm1 = 1-phi(-t1+p4*1+p5*1+p6*0.37+p7*1+b7*3.1775+b8*7.3700) ;
24.  c1diff = c1pm1-c1pm0 ;   !prob difference
25. !CONTRAST 2A
26.  c2apm0 = 1-phi(-t2+p4*0+p5*1+p6*0.37+p7*1+b7*3.1775+b8*7.3700) ;
27.  c2apm1 = 1-phi(-t2+p4*1+p5*1+p6*0.37+p7*1+b7*3.1775+b8*7.3700) ;
28.  c2adiff = c2apm1-c2apm0 ; !prob difference
29. !CONTRAST 2B
30.  c2bpm0 = phi(-t2+p4*0+p5*1+p6*0.37+p7*1+b7*3.1775+b8*7.3700) ;
31.  c2bpm1 = phi(-t2+p4*1+p5*1+p6*0.37+p7*1+b7*3.1775+b8*7.3700) ;
32.  c2bdiff = c2bpm1-c2bpm0 ; !prob difference
33. !CONTRAST 3
34.  c3pm0 = phi(-t3+p4*0+p5*1+p6*0.37+p7*1+b7*3.1775+b8*7.3700) ;
35.  c3pm1 = phi(-t3+p4*1+p5*1+p6*0.37+p7*1+b7*3.1775+b8*7.3700) ;
36.  c3diff = c3pm1-c3pm0 ;    !prob difference
37. OUTPUT: SAMP RESIDUAL STAND(STDY) CINTERVAL TECH4  ;
```

The upper section of Table 13.13 summarizes results for the GA2 mediator for the new profile analysis and the second section presents the original profile analysis I conducted prior to this.

Table 13.13 Effects of Mediators on Outcome for Second Profile Analysis

	Contrast Category	Percent at Score of 1	Percent at Score of 0	Difference
Profile 2 GA2	C1: No change	0.3 ±0.1	2.4 ±1.8	-2.1 ±1.6*
	C2a: No change or minimal improve	10.5 ±3.6	32.5 ±10.4	-22.0 ±8.6*
	C2b: Much or very much improve	89.5 ±3.6	67.5 ±10.4	22.0 ±8.6*
	C3: Very much improve	33.0 ±6.6	10.8 ±5.6	22.3 ±4.8*

Profile 1 GA2	C1: No change	6.2 ±3.4	23.0 ±5.2	-16.8 ±4.0*
	C2a: No change or minimal improve	49.2 ±10.4	78.3 ±4.8	-29.1 ±8.8*
	C2b: Much or very much improve	50.8 ±10.4	21.7 ±4.8	29.1 ±8.8*
	C3: Very much improve	4.7 ±2.8	0.7 ±0.4	4.0 ±2.6*

Table notes: * $p < 0.05$; C2a and C2b are statistically redundant

There are several striking differences in the two analyses. Note, for example, the mediator effect in the "no change" category (contrast 1) is much larger in the original profile analysis than this profile analysis. This is because of the non-linear nature of the probit analysis with respect to probabilities and the fact that in the second profile analysis, few individuals occur in the "no change" category because of the favorability of the contextual variables for symptom improvement in general. The same dynamic is at work for the last contrast but with the first profile analysis showing a weaker effect than the second profile analysis.

Two points are worth reiterating here. First, conditional probability based mediation effects with logit or probit regression can be context dependent and sometimes they can be complicated in form. In Chapter 12 on binary outcomes, I noted that Pischke (2012) argues that we often do not know if the probit or logit model is the "right model" and argues that one can get into just as much if not more trouble when we choose the wrong non-linear function than when we wrongly apply a linear function. The possibility of specification error should always lead you to approach your conclusions with humility. Second, as emphasized by Harrell (2021) and many other methodologists, if you use logit or probit modeling, then cases can arise where it is misleading to characterize a mediator effect (or a total effect) using a single number, at least when estimates are based on conditional effects. I show below how average marginal effects can circumvent this issue.

Average Marginal Effects

A second way of documenting mediator effects on the outcome is to use average marginal effects. To implement this method, I again use results for Equation 13.4 from the original ordinal regression model and the syntax in Table 13.3. Here is the relevant output from that analysis that I will make use of in the AME analyses:

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

IMP3       ON
   GA2                0.801      0.107      7.485      0.000
```

```
    TA2                0.692      0.119      5.828      0.000
    BD2                0.684      0.114      6.017      0.000
    T                  0.174      0.174      0.997      0.319
    CE1                0.147      0.067      2.198      0.028
    CIS1               0.295      0.106      2.775      0.006

 Thresholds
    IMP3$1             1.786      0.795      2.245      0.025
    IMP3$2             3.305      0.805      4.105      0.000
    IMP3$3             5.000      0.811      6.169      0.000
```

These results yield the following three probit subequations:

$$IMP3_{d1} = -1.786 + .801\ GA2 + .692\ TA2 + .684\ BD2 + .174\ T + .147\ CE1 + .295\ CIS1$$

$$IMP3_{d2} = -3.305 + .801\ GA2 + .692\ TA2 + .684\ BD2 + .174\ T + .147\ CE1 + .295\ CIS1$$

$$IMP3_{d3} = -5.000 + .801\ GA2 + .692\ TA2 + .684\ BD2 + .174\ T + .147\ CE1 + .295\ CIS1$$

I use these three subequations equations to calculate the AMEs for the mediator contrasts using the method of Cameron and Trivedi. Table 13.14 presents the syntax for the GA2 mediator. In this code, I take advantage of the fact that Mplus executes the `DEFINE` statements sequentially to re-use the `PROBIT` name in succeeding statements. I use the basic logic from the Appendix of Chapter 12 on binary outcomes. Note that I again strategically use the 1 – `PHI` function to frame each contrast in the way it is worded.

Table 13.14 Syntax for AME for Effects of Mediators on Outcomes

```
1.  TITLE: AME for effects of GA2 mediator on outcome ;
2.  DATA: FILE IS c:\mplus\symptom.dat ;
3.  DEFINE:
4.   DELTA = SQRT(0.538)/1000 ; !divide SD of GA2 by 1000
5.   !Calculate base probabilities for the contrasts
6.   !Do not differentiate contrasts 2a and 2b yet
7.   PROBIT = -1.786+.801*GA2+.692*TA2+.684*BD2+.174*T+.147*CE1+.295*CIS1;
8.   PROB1C1 = 1-PHI(PROBIT) ;    !base prob for contrast 1
9.   PROBIT = -3.305+.801*GA2+.692*TA2+.684*BD2+.174*T+.147*CE1+.295*CIS1;
10.  PROB1C2 = 1-PHI(PROBIT) ;    !base prob for contrast 2A and 2B
11.  PROBIT = -5.000+.801*GA2+.692*TA2+.684*BD2+.174*T+.147*CE1+.295*CIS1;
12.  PROB1C3 = PHI(PROBIT) ;    !base prob for contrast 3
13.  !Increment for Contrasts
14.  GA2=GA2+DELTA ;
15.  !Incremented probabilities for contrasts
16.  PROBIT = -1.786+.801*GA2+.692*TA2+.684*BD2+.174*T+.147*CE1+.295*CIS1;
17.  PROB2C1 = 1-PHI(PROBIT) ;  !incremented prob for contrast 1
18.  PROBIT = -3.305+.801*GA2+.692*TA2+.684*BD2+.174*T+.147*CE1+.295*CIS1;
```

```
19.  PROB2C2 = 1-PHI(PROBIT) ;  !incremented prob for contrast 2A and 2B
20.  PROBIT = -5.000+.801*GA2+.692*TA2+.684*BD2+.174*T+.147*CE1+.295*CIS1;
21.  PROB2C3 = PHI(PROBIT) ;    !incremented prob for contrast 3
22.  !Calculate individual marginal effects
23.  IMEC1 =(PROB2C1-PROB1C1)/DELTA ;    !individual me contrast 1
24.  IMEC2A =(PROB2C2-PROB1C2)/DELTA ;   !individual me contrast 2A
25.  !below I set contrast 2B to be opposite signed me for contrast 2A
26.  IMEC2B = -(PROB2C2-PROB1C2)/DELTA ; !individual me for contrast 2B
27.  IMEC3 = (PROB2C3-PROB1C3)/DELTA ;   !individual me contrast 3
28. VARIABLE:
29.  NAMES ARE ID GA2 TA2 BD2 CE1 CIS1 T IMP3 ;
30.  USEVARIABLES ARE IMEC1 IMEC2A IMEC2B IMEC3 ;
31.  MISSING ARE ALL (-9999) ;
32. ANALYSIS:
33. ESTIMATOR=ML ;  TYPE=BASIC ;
34. OUTPUT: !use defaults on output
```

For GA2, the average marginal effects on the output for the four contrasts were -0.12, -0.19, 0.19, and 0.12, respectively. Roughly speaking, for every unit that GA2 increases, (a) the proportion of clients in the lowest outcome category (no change in symptoms) is predicted to decrease by 0.12, (b) the proportion of clients in the lowest two outcome categories (i.e., clients who showed no change or who showed only minimal improvement) is predicted to decrease by 0.19, (c) the proportion of clients in the highest two outcome categories (clients who were much improved or very much improved) is predicted to increase by 0.19, and (d) the proportion of clients who were very much improved, the top category of the scale, is predicted to increase by 0.12. When I re-ran the program to focus on TA2, the corresponding AMEs were -0.11, -0.17, 0.17 and 0.10, respectively. For BD2, they were -0.11, -0.16, 0.16, and 0.10.

I adapted the syntax in Table 13.14 to calculate the AMEs for the independent effects of the treatment condition on the contrasts over and above the three mediators. I deleted lines 4 and 14, removed the `/DELTA` terms from lines 25 to 28, changed `.174*T` to `.174*0` in lines 7, 9 and 11, and changed `.174*T` to `.174*1` in lines 16, 18 and 20. Consistent with the overall ordinal analysis that revealed a statistically non-significant independent effect for the treatment effect over and above the mediators, the AMEs were trivial in magnitude. (-0.03, -0.04, 0.04 and 0.03, respectively).

The method of calculating AMEs in Mplus does not yield significance tests nor confidence intervals which is a drawback. The program on my website called *AMEs: Ordinal-Multinomial* calculates marginal effects for ordinal regression with standard errors and confidence intervals but it applies to LISEM contexts and uses an alternative parameterization for ordinal regression that I provide in the Appendix. Without confidence intervals, I cannot formally determine if the lower bound of the marginal

effect confidence interval exceeds the meaningfulness standards I set for each mediator. The point estimates of the AMEs did, in fact, exceed the meaningfulness standards.

Concluding Comments on Estimating the Effects of Mediators on the Outcome

In sum, both methods for documenting the effects of mediators on the outcome have value. AMEs are useful because they do not rely on conditional effects in the way that panel analysis does. They yield a single number that captures the effect of a unit increase in the predictor on the proportion of individuals who embrace the outcome after collapsing across but taking into account operative non-linearities and covariates. Panel analysis explores how mediator effects differ as a function of predictor contexts.

Concluding Comments on the Probability Approach

The probability approach to ordinal regression provides a wealth of information about RET dynamics. One complaint about the approach is that it provides too much information. A goal of ordinal regression, the argument goes, is to simplify analyses by specifying a single set of coefficients that apply to each of the underlying subequations. By turning to detailed evaluations of specific category proportions and category proportion differences in the form of contrasts, the analysis becomes too complicated and difficult to summarize. It is for this reason that some methodologists prefer the latent response approach to ordinal regression that I describe in the next section.

I have eschewed logistic ordinal regression in favor of probit ordinal regression because I think the probit approach has statistical advantages over the logistic approach (Norton, 2012; Muthén, Muthén & Asparouhov, 2016). For probit analysis, the latent response variable formulation assumes a disturbance term that is normally distributed with a variance of 1.0. This gives it a statistical edge for a variety of statistical theories as compared to logistic regression that assumes a disturbance term with a standard logistic distribution and a variance of 3.29. For example, these characteristics make probit modeling more amenable to Bayesian estimation than logit modeling.

The approach I have outlined is themed around contrasts that correspond to breakpoints on the ordinal metric. There are other parameterizations you can use, such as contrasts focused on the proportion of individuals in each separate response category as a function of the treatment condition and mediator status. The Appendix describes how to execute such a parameterization and represents a different approach you might consider.

ORDINAL MODELING: THE LATENT RESPONSE APPROACH

An alternative approach to the analysis of an ordinal outcome is to conceptualize the

ordinal measure as a crude indicator of an underlying continuous latent variable and to focus modeling on that continuous latent variable. When the outcome variable has a single indicator that has ordinal properties, this strategy uses the latent response framework outlined in Chapter 5 for ordinal regression. The approach does not rely on "breakpoint" analysis but instead emphasizes how a person's standing on the underlying continuous latent response variable translates into the endorsement of a given category on the observed ordinal metric. In our numerical example, assume there is a continuous latent variable, IMP3*, that is the true symptom outcome of interest. The four-point response scale, IMP3, reflects this continuous construct, but the measure is a crude representation of it. In principle, one can specify a set of rules by which people's location on IMP3* translates into responses on the four point rating scale. I might formulate a rule that if a person's score on IMP3* is below a certain threshold value, that I call τ_1, then the response made on the rating scale by the person will fall into category 1, "no change or got worse." If a person's score on IMP3* is above that threshold but below a second threshold, τ_2, then the response made on the rating scale will fall into category 2, "minimal improvement." If a person's score on IMP3* is above threshold τ_2, but below a third threshold, τ_3, then the response made on the rating scale will fall into category 3, "much improved." Finally, if the person's score on IMP3* is above threshold τ_3, the response on the rating scale will fall into category 4, "very much improved." If the outcome measure has *k* levels, there are *k*-1 thresholds. Figure 13.2 presents the dynamic graphically, showing the location of 4 different people on the underlying IMP3* and the response they would make (using a dashed arrow) on the rating scale.

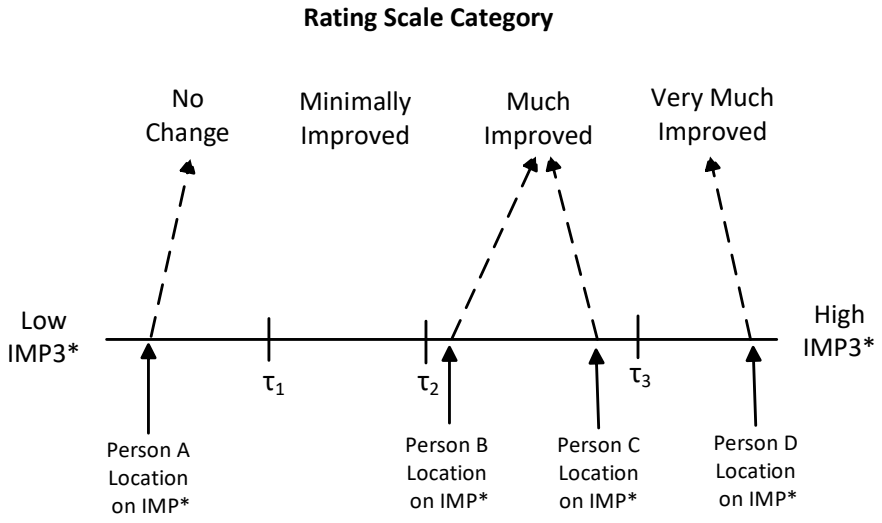

FIGURE 13.2. Illustration of Thresholds

The latent response approach estimates the threshold values that apply based on the observed data. However to do so, the model must make the assumption that the underlying continuous latent variable is normally distributed, an assumption that may not be viable.[6] The framework refocuses Equation 13.4 in the model to be:

$$IMP3^* = p_4\,GA2 + p_5\,TA2 + p_6\,BD2 + p_7\,T + b_7\,CE1 + b_8\,CIS1 + d \qquad [13.11]$$

where IMP3* is the latent continuous variable underlying the observed measure IMP, and *d* is a disturbance term that is assumed to have a mean of 0, a variance of 1 and to be normally distributed. Note there is no intercept term because it is assumed to be zero, an assumption needed for the model to be statistically identified when there is only a single indicator of IMP3*. Equation 13.11 is a traditional linear expression for a continuous outcome and the task at hand is to estimate the coefficients in the equation.

A major difficulty with the latent response formulation is that the metric of the latent response variable, IMP3*, is arbitrary while the variance of *d* is fixed at 1.0. In Chapter 5, I discussed some of the challenges that such conditions pose for effective data analysis. I do not repeat that discussion here but it is worth noting again that because the metric of IMP3* is arbitrary, it makes interpretation of the path coefficients for it difficult. A path coefficient estimates how much a one unit increase in a predictor changes the mean of Y holding constant the other predictors in the equation. But if the metric of Y is arbitrary, we cannot make heads or tails of the coefficients other than their sign and statistical significance. It is analogous to predicting income from, say, the number of years of education to try to determine the "worth" of a year of education, but without knowing if income is measured in units of dollars, in units of thousands of dollars, in pesos, or whatever. We might determine that education is indeed related to income in a positive way, but determining the worth of a year of education eludes us without knowing the units in which income is measured.

One work-around for this problem is to use standardized metrics in which you standardize your key endogenous variables. I will adopt this approach and illustrate yet additional challenges with using it when I consider each RET question. Table 13.15 presents the Mplus syntax I use for the analysis. All of the syntax should be familiar to you. Note, however, my use of endogenous standardization (`STAND(STDY)`) on Line 18. I also use bootstrapping for reasons I describe shortly.

[6] This is the assumption with probit modeling. With logistic modeling, the latent variable is assumed to take on a standard logistic distribution.

Table 13.15 Latent Response Analysis of Ordinal Outcome

```
1.  TITLE: Latent response analysis ;
2.  DATA: FILE IS c:\mplus\symptom.dat ;
3.  VARIABLE:
4.  NAMES ARE ID GA2 TA2 BD2 CE1 CIS1 T IMP3 ;
5.    USEVARIABLES ARE GA2 TA2 BD2 CE1 CIS1 T IMP3 ;
6.    CATEGORICAL ARE IMP3 ;
7.    MISSING ARE ALL (-9999) ;
8.  ANALYSIS:
9.  ESTIMATOR = ML ; LINK = PROBIT ; BOOTSTRAP = 2000 ;
10. MODEL:
11.   GA2 on T CE1 CIS1 (p1 b1 b2)  ;
12.   TA2 on T CE1 CIS1 (p2 b3 b4)  ;
13.   BD2 on T CE1 CIS1 (p3 b5 b6)  ;
14.   IMP3 on GA2 TA2 BD2 T CE1 CIS1 (p4 p5 p6 p7 b7 b8)  ;
15.   [IMP3$1] (t1) ; [IMP3$2] (t2) ; [IMP3$3] (t3) ;
16. MODEL INDIRECT:
17.   IMP3 IND T ;
18. OUTPUT: SAMP RESIDUAL STAND(STDY) CINTERVAL(BOOTSTRAP) TECH4  ;
```

I do not review output for model fit because I already did so in the context of the probability approach to ordinal modeling. The model fit diagnostics are identical in that analysis to this one. Given that I have a good fitting model, I turn to addressing the three key questions of an RET. Before doing so, however, I address the specification of meaningfulness standards. Because my instantiation of the latent response approach relies on partially standardized and fully standardized solutions, I need to express my meaningfulness standards in standardized terms. Some of the parameters of interest in my analysis compare partially standardized means on IMP3* or the mediators for the treatment versus control groups. The question becomes what is a meaningful standardized mean difference? As discussed in Chapter 10, a common standard for Cohen's *d* is that a *d* of 0.20 is a "small" effect, a *d* of 0.50 is a "medium" effect, and a *d* of 0.80 or greater is a "large" effect but I also stressed how these standards can be arbitrary and misleading. Many researchers seek to have at least a "medium" effect size on mediators and outcomes, which translates into a meaningfulness standard of 0.50 for a standardized mean difference. You, of course, might adjust this upward or downward depending on the particulars of your evaluation context, taking into account such factors as the number of people affected by the target outcome, the impact on the quality of their lives, the severity and reversibility (or positiveness and sustainability) of the outcome, the vulnerability of the affected population, and the costs and organizational readiness to bring about change, among others. I will use a meaningfulness standard of 0.50 for our numerical example to illustrate the latent response approach, but do so being fully cognizant of its limitations.

. The latent response approach also works with fully standardized coefficients when mapping the effects of the mediators on the outcome using Mplus. As such, one also needs meaningfulness standards for them. Acock (2014) suggests that an absolute standardized regression/path coefficient less than 0.20 is weak, one between 0.20 and 0.50 is moderate, and one greater than 0.50 is strong.[7] However, these standards also are somewhat arbitrary and can be shifted upward or downward depending on your evaluation context. I will use a standard of 0.20 for the current example.

I often find it difficult to develop meaningfulness standards using standardized metrics when working with project administrators and staff. Most administrators and staff simply are not comfortable with standardized metrics. This is one reason I often use the probability approach in place of or in conjunction with the latent response approach.

Question 1: The Effects of the Intervention on the Outcome

Mplus reports a total effect for program impact based on the analysis of IMP3*. As discussed in Chapter 12, the most informative version of this test standardizes IMP3* but not the treatment predictor T, which results by using the `STAND(STDY)` option on the output line of the syntax. IMP3* then has a variance of 1.0 and can be thought of as being standardized. Here is the result from the Mplus output for the syntax in Table 13.15:

```
STANDARDIZED TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS FOR
LATENT RESPONSE VARIABLES

STDY Standardization

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

Effects from T to IMP3

  Total                0.956      0.053     18.206      0.000
```

The mean difference for the standardized IMP3* as a function of the treatment and control conditions was 0.96 ±0.10, CR = 18.21, $p < 0.05$). The magnitude of the total effect is not assumed to change with changing values of the covariates when working with latent propensities (which was not true of the probability approach). Here are the bootstrapped confidence intervals for the effect (which I edited to fit better on the current page formats):

[7] These guidelines and the reliance on standardized regression coefficients are not applicable if suppression dynamics are evident for a given coefficient.

```
STDY Standardization

                    Lower 2.5%    Estimate    Upper 2.5%

Effects from T to IMP3

  Total                 0.851       0.956       1.057
```

For a confidence interval of a standardized coefficient, the interval often is asymmetric, hence my use of bootstrapping when using the latent response approach. In this case, the interval is relatively symmetric. The lower limit of the confidence interval for the standardized IMP3* difference between the treatment and control conditions was 0.85, so the total effect would be declared meaningful given a meaningfulness standard of 0.50.

Question 2: Effect of the Intervention on the Mediators

The analysis of the effect of the treatment condition on the mediators can be pursued using either unstandardized or standardized metrics. For the former, the same format and statistics as that for the probability approach is used. Some methodologists prefer to use the standardized endogenous variable approach to keep matters of standardization consistent across the three RET questions and that is what I do here. The results for the analysis of the effects of the intervention on the mediators in this case are

```
STDY Standardization

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

 GA2        ON
    T                1.349      0.039     34.602      0.000
    CE1              0.244      0.027      9.044      0.000
    CIS1             0.356      0.052      6.807      0.000

 TA2        ON
    T                1.284      0.041     30.963      0.000
    CE1              0.324      0.023     13.924      0.000
    CIS1             0.302      0.045      6.702      0.000

 BD2        ON
    T               -0.029      0.069     -0.415      0.678
    CE1              0.357      0.033     10.664      0.000
    CIS1             0.504      0.068      7.453      0.000
```

These results, like the total effect, are interpreted as partially standardized covariate adjusted mean differences between the intervention and control conditions, analogous to a Cohen's *d* (Karlson, 2015). For example, the effect of the intervention on GA2 was to raise it, relative to the control group, by 1.35 standard deviations of GA2 (CR = 34.60, $p < 0.05$). Here are the confidence intervals for the three mediators:

```
STDY Standardization

                  Lower 2.5%    Estimate    Upper 2.5%

 GA2       ON
    T                1.265       1.349        1.423
    CE1              0.191       0.244        0.296
    CIS1             0.257       0.356        0.457

 TA2       ON
    T                1.202       1.284        1.362
    CE1              0.280       0.324        0.371
    CIS1             0.212       0.302        0.391

 BD2       ON
    T               -0.163      -0.029        0.106
    CE1              0.292       0.357        0.422
    CIS1             0.367       0.504        0.635
```

The lower limits of the 95% confidence intervals for GA2 and TA2 both exceed the meaningfulness standards, leading me to conclude the intervention meaningfully affected these two mediators. This was not the case for BD2. Because the confidence intervals for GA2 and TA2 do not contain the value of zero, they can be deemed to be statistically significant ($p < 0.05$). If you prefer to work with the unstandardized metrics of the mediators, then you would examine the output in the `MODEL RESULTS` section and interpret the statistics in the traditional ways I have laid out in prior chapters.

Question 3: Effects of the Mediators on the Outcome

To document the effects of the mediators on IMP3*, I again examine the path coefficients reported in the section `STDY Standardization`. It turns out the mediators are endogenous variables in the broader model so they too are standardized in this output section (T, the treatment condition, is not standardized because it is exogenous). This means that the reported coefficients for the mediators are fully standardized per traditional regression analyses and are interpreted much like we would any standardized coefficient in a regression context. Here are the coefficients:

```
                                                    Two-Tailed
```

		Estimate	S.E.	Est./S.E.	P-Value
IMP3	ON				
GA2		0.351	0.044	8.011	0.000
TA2		0.300	0.050	6.014	0.000
BD2		0.211	0.034	6.220	0.000
T		0.104	0.101	1.028	0.304
CE1		0.088	0.040	2.181	0.029
CIS1		0.176	0.063	2.795	0.005

For GA2, for every one standard deviation that GA2 increases, IMP3* is predicted to increase by 0.35 standard deviations of IMP3* (CR = 8.01, margin of error (MOE) = ±0.09, $p < 0.05$) holding the other predictors in the equation constant. For TA2, for every one standard deviation that TA2 increases, IMP3* is predicted to increase by 0.30 standard deviations of IMP3* (CR = 6.01, MOE = ±0.10, $p < 0.05$) holding the other predictors constant. For BD2, for every one standard deviation that BD2 increases, IMP3* is predicted to increase by 0.21 standard deviations of IMP3* (CR = 6.22, MOE = ±0.07, $p < 0.05$) holding the other predictors constant. The direct effect of the treatment condition on IMP3* over and above the mediators was not statistically significant (standardized coefficient = 0.10, CR = 1.03, MOE = ±0.20, $p = 0.30$). Because the underlying latent variable IMP3* is assumed to be a linear function of the predictors, the magnitude of the mediator effects do not change with changing values of the covariates (which was not the case for the probability approach).

Here are the confidence intervals for the standardized coefficients that can be used to help make meaningfulness judgments for the effects of mediators on IMP3* (these should be checked against bootstrapped results):

STDY Standardization

		Lower 2.5%	Estimate	Upper 2.5%
IMP3	ON			
GA2		0.263	0.351	0.436
TA2		0.201	0.300	0.396
BD2		0.145	0.211	0.278
T		-0.092	0.104	0.293

The lower limits of the 95% confidence interval exceed the meaningfulness standard of 0.20 for both GA2 and TA2, so these mediators meaningfully affect the outcome. The confidence interval for BD2 (0.14 to 0.28) overlaps the meaningfulness standard so I cannot conclude with confidence that BD2 meaningfully affects symptom improvement. To be sure, the sample estimate of 0.21 for this coefficient is suggestive of a meaningful

effect, but the presence of sampling error makes it such that I cannot be strongly (95%) confident that the effect of BD2 on IMP3* is meaningful.

Omnibus Mediation Analysis

Using the joint significance test in the latent response framework, I would conclude that GA2 and TA2 are both non-zero mediators of the effects of the treatment on the outcome, but that this is not the case for BD2 because the treatment failed to statistically significantly impact it. In Mplus, traditional omnibus mediation tests are generated by the `COM3 IND T` command on Line 17 of Table 13.15. The section of the output where the mediation tests are reported is titled `TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS`. Mplus reports results using IMP3*, the latent propensity and that is what is of primary interest to us. Given this, I want to focus on the results reported in the `STDY Standardization` section. Here is the output:

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

Effects from T to IMP3

  Total               0.956      0.053     18.206      0.000
  Total indirect      0.853      0.091      9.366      0.000

  Specific indirect 1
    IMP3
    GA2
    T                 0.473      0.062      7.601      0.000

  Specific indirect 2
    IMP3
    TA2
    T                 0.386      0.067      5.797      0.000

  Specific indirect 3
    IMP3
    BD2
    T                -0.006      0.015     -0.414      0.679

  Direct
    IMP3
    T                 0.104      0.101      1.028      0.304
```

As examples, the estimated standardized mean difference between the treatment and control conditions through the causal chain T→GA2→IMP3* is 0.47 ±0.13 (CR = 7.60,

$p < 0.05$). The corresponding estimate for the chain T→TA2→IMP3* is 0.39 ±0.13 (CR = 5.80, $p < 0.05$) and for the chain T→BD2→IMP3* it is -0.01 ±0.03 (CR = 0.41, $p < 0.68$). You can use these estimates to perform decomposition analyses of the total effects to determine the proportion of the total effect of T on the standardized IMP3* that each mediator is responsible for. I do not pursue these analyses because I personally do not find such indices compelling and, as noted in Chapter 10, they tend not be stable.

Parenthetically, it is possible to apply causal mediation analysis to the mediators, one mediator at a time, as I illustrated in Chapter 12. However, there is little consensus on how best to parameterize such effects (but see VanderWeele, Zhang, & Limb, 2016), so given current knowledge, I think it best to focus instead on IMP3* and the traditional omnibus mediation analysis that goes with it if you are interested in such tests.

Concluding Comments on the Latent Response Approach

The latent response approach leads to the same overall conclusions as the probability approach for our numerical example but it is much more succinct and simpler to apply. This is both a strength and a weakness. To me, the biggest weakness of the approach is the arbitrary metric of the latent propensity, y*, that forces us to standardize it and then rely on standardized metrics that can be difficult to assign meaningfulness standards to. One way of addressing this limitation is to include benchmarks in one's evaluation study that represent meaningful outcomes that can then be linked to the standardized metric of y* (Kazdin, 2003). For example, in the numerical example, I might include measures of variables that reflect functioning in everyday life that are related to anxiety and depression symptom improvement and which then can be used to help calibrate y* against them.

ORDINAL MEDIATORS AND LATENT VARIABLES

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

By definition, ordinal mediators are endogenous because there is a causal path emanating from the treatment (dummy) variable to the mediator and/or from another mediator to the ordinal mediator. At the same time, the ordinal mediator is a cause of the outcome and/or possibly another mediator. Figure 13.3 shows a classic influence diagram of this scenario.

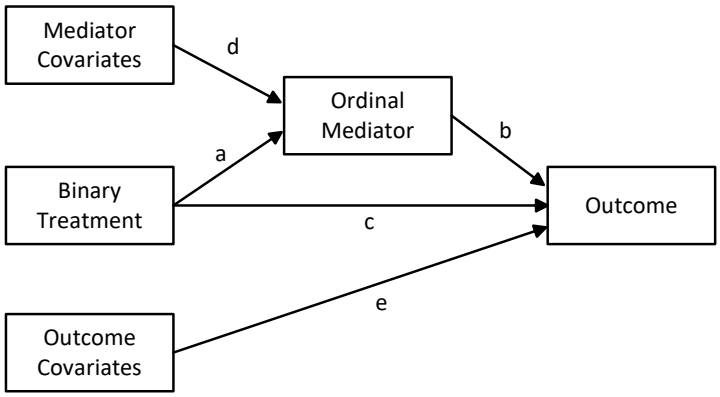

FIGURE 13.3. Classic single mediator model

The dual role of the ordinal mediator as both a predictor and a dependent variable creates challenges for FISEM analyses. For LISEM, which splits up the model and works with one equation at a time or model segments, matters are more straightforward but with some sacrifices. An issue in both frameworks is how to treat the ordinal mediator when it takes on the role of a predictor in the mediational chain, i.e., when you estimate path *b* in Figure 13.3. Different approaches have been suggested. The first and perhaps most controversial approach is to treat the ordinal scaled variable as an interval scaled variable per traditional regression analysis, i.e., to ignore the non-intervalness of the measure: You regress the outcome onto the mediator per traditional regression modeling and interpret the coefficient as you would any regression coefficient. Critics argue that this strategy is hypocritical because if you judged the measure was sufficiently non-interval to use ordinal regression when the mediator is a dependent variable (per path *a*), how can you turn around and say it is sufficiently interval when it is a predictor (path *b*)? Maybe you can make such a case but doing so can be tricky.

A second strategy for representing an ordinal predictor in a linear equation is to code it using dummy variables. A common dummy variable approach uses what is known as **staircase coding**. Suppose I have an ordinal mediator with k = 5 categories that are assigned the numbers 1, 2, 3, 4 or 5 and I am predicting a continuous outcome, Y. With staircase coding, I create k - 1 = 4 dummy variables that are defined such that when Y is regressed onto them, the coefficients for the dummy variables reflect mean outcome differences between successive categories on the ordinal predictor. In traditional dummy coding, coefficients reflect the mean outcome difference for the group scored 1 on the dummy variable versus the reference group. Staircase coding is different. Consider the data in Table 13.16 which illustrates the spirit of staircase coding:

Table 13.16 Example of Staircase Coding

			Dummy Variables				
ID	Outcome (Y)	Ordinal Predictor	D1	D2	D3	D4	Mean Y
1	1.00	1	0	0	0	0	
2	2.00	1	0	0	0	0	2
3	3.00	1	0	0	0	0	
4	4.00	2	1	0	0	0	
5	5.00	2	1	0	0	0	5
6	6.00	2	1	0	0	0	
7	8.00	3	1	1	0	0	
8	9.00	3	1	1	0	0	9
9	10.00	3	1	1	0	0	
10	13.00	4	1	1	1	0	
11	14.00	4	1	1	1	0	14
12	15.00	4	1	1	1	0	
13	20.00	5	1	1	1	1	
14	21.00	5	1	1	1	1	21
15	22.0	5	1	1	1	1	

Individuals with a score of 1 on the ordinal predictor are assigned zeros for all four dummy variables. Individuals with a score of 2 are assigned a 1 on the first dummy variable and 0s on the remaining dummy variables. Individuals with a score of 3 are assigned a 1 on the first two dummy variables and 0s on the remaining dummy variables. Individuals with a score of 4 are assigned a 1 on the first three dummy variables and then a 0 on the remaining dummy variables. Finally, individuals with the highest score are assigned all 1s on the dummy variables. When Y is regressed onto the four dummy variables, the intercept equals the mean Y for individuals with a score of one on the ordinal predictor; the coefficient for D1, the first dummy variable, equals the predicted Y mean difference for individuals who score 2 on the ordinal predictor minus those who score 1; the coefficient for D2 equals the predicted Y mean difference for individuals who score 3 on the ordinal predictor minus those who score 2; the coefficient for D3 equals the Y mean difference for individuals who score 4 on the ordinal predictor minus those who score 3; and the coefficient for D4 equals the Y mean difference for individuals who score 5 on the ordinal predictor minus those who score 4 on the predictor.

Here is the regression output for the above data where I re-labeled the dummy variables with subscripts to reflect the mean difference their coefficients reflect:

Predictor	Coefficient	Standard Error	t Value	p Value	Mean Difference
Intercept	2.00	.58	3.46	.006	-
D_{2-1}	3.00	.82	3.67	.004	5.0 - 2.0 = 3.0
D_{3-2}	4.00	.82	4.90	.001	9.0 - 5.0 = 4.0
D_{4-3}	5.00	.82	6.12	.000	14.0 - 9.0 = 5.0
D_{5-4}	7.00	.82	8.57	.000	21.0 - 14.0 = 7.0

The coefficients make evident how the mean of Y changes with each shift in the ordinal predictor from one category to the next highest category. Note also that if desired, I can include covariates in the regression model to obtain covariate adjusted means. To interpret the effects of the ordinal predictor on the outcome, one interprets each of the coefficients for the dummy variables for the predictor. For example, the coefficient for D_{2-1} was 3.00. When the ordinal predictor shifts from category 1 to category 2, the mean of Y increases by 3.0 units ($t = 3.46$, MOE = ±1.64, $p < 0.05$). The coefficient for D_{3-2} was 4.00. When the ordinal predictor shifts from category 2 to category 3, the mean of Y increases by 4.0 units ($t = 3.67$, MOE = ±1.64, $p < 0.05$). And so on.

A third strategy for using ordinal predictors is to score the ordinal metric to be an interval level metric using either the midpoint of its categories or an algorithm that makes logical sense relative to an interval level metric. Suppose single young adults are asked how often they used a condom across their instances of sexual intercourse during the past 3 months with the following 7-point response alternatives:

____ never, hardly at all (1 to 19% of the time)
____ a small part of the time (20% to 39%)
____ about half the time (40% to 59%)
____ most of the time (60% to 79%)
____ almost all of the time (80% to 99%)
____ always (100% of the time).

The midpoint of the respective categories are 10%, 30%, 50%, 70%, 90% and 100%. The researcher would assign the midpoint score for the category that respondents mark and then analyze the data as if the responses are interval level. Granted, there is a certain crudeness to the measure because of the coarseness of the categories, but the coarseness may not matter that much for the particular questions addressed in the study

(or maybe it will be).

As another example, suppose for a frequency judgment of how often a young adult has smoked marijuana in the last year, the following response metric is used

_____ daily
_____ a few times a week
_____ once a week
_____ a few times a month
_____ once a year
_____ never

The response category a respondent marks can be roughly translated by the researcher into the following numbers: daily = 365; a few times a week = 3 times 52 weeks or 156; once a week = 1 times 52 weeks or 52; a few times a month = 3 times 12 months or 36; once a year = 1; and never = 0, and then analyzed as an approximately interval level measure.

For LISEM, the focus is on documenting and evaluating the magnitude of the path coefficient for each separate link in the mediational chain of Figure 13.3 using any one of many available statistical tools and then using the joint significance test (JST) to evaluate the null hypothesis of no mediation across the full mediational chain (see Chapter 9). Ordinal mediators pose no special problems in such cases. I now describe LISEM-based methods you might use for a range of scenarios with ordinal mediators.

Scenario 1: If the mediator is ordinal 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 13.3. Score the ordinal predictor using any of the aforementioned strategies, such as staircase coding, as appropriate. In a second analysis, regress M onto T using ordinal regression to isolate the probit coefficient *a*. Apply either the probability approach or the latent propensity approach to the ordinal regression analysis. Include measured covariates in all analyses as dictated by theory. Evaluate the magnitude of each link in the mediational chain and then apply the joint significance test.

Scenario 2: If the mediator is ordinal and the outcome is continuous, use robust OLS or robust maximum likelihood to regress Y onto M and T to isolate path coefficients *b* and *c* in Figure 13.3. Score the ordinal predictor using any of the aforementioned strategies, such as staircase coding, as appropriate. In a second analysis, regress M onto T using ordinal regression to isolate the probit coefficient *a*. Apply either the probability approach or the latent propensity approach to the ordinal regression analysis. Include measured covariates in all analyses, as dictated by theory. Evaluate the magnitude of each

link in the mediational chain and then apply the joint significance test.

Scenario 3: If the mediator is ordinal and the outcome is ordinal, use ordinal regression to regress Y onto M and T to isolate path coefficients *b* and *c* in Figure 13.3. Score the ordinal predictor using any of the aforementioned strategies, such as staircase coding, as appropriate. In a second analysis, regress M onto T using ordinal regression to isolate the probit coefficient *a.* Apply either the probability approach or the latent propensity approach to the ordinal regression analysis. Include measured covariates in all analyses, as dictated by theory. Evaluate the magnitude of each link in the mediational chain and then apply the joint significance test.

Scenario 4: If the mediator is ordinal and the outcome is nominal, use multinomial regression to regress Y onto M and T to isolate path coefficients *b* and *c* in Figure 13.3. Score the ordinal predictor using any of the aforementioned strategies, such as staircase coding, as appropriate. In a second analysis, regress M onto T using ordinal regression to isolate the probit coefficient *a.* Apply either the probability approach or the latent propensity approach to the ordinal regression analysis. Include measured covariates in all analyses, as dictated by theory. Evaluate the magnitude of each link in the mediational chain and then apply the joint significance test.

Scenario 5: If the mediator is ordinal 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* in Figure 13.3. Score the ordinal predictor using any of the aforementioned strategies, such as staircase coding, as appropriate. In a second analysis, regress M onto T using ordinal regression to isolate the probit coefficient *a.* Apply either the probability approach or the latent propensity approach to the ordinal regression analysis. Include measured covariates in all analyses, as dictated by theory. Evaluate the magnitude of each link in the mediational chain and then apply the joint significance test.

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 options 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 small sample appropriate statistical methods outside of Mplus (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.

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 or zero. 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.

With FISEM, the analytic flexibility is more constrained. Consider, for example, the case of an ordinal 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) 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 ordinal mediator needs to function as an exogenous variable in the M→Y portion of the model but as a logit or probit ordinal outcome variable in the T→M portion.

For FISEM with an ordinal mediator, I recommend you use Bayes SEM. Bayes modeling offers flexibility for how you treat the mediator, providing you with two options. By specifying the option `MEDIATOR=LATENT` on the `ANALYSIS` command, Mplus will invoke the latent response formulation of the mediator, 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 is the linear regression of Y onto the continuous mediator-based m* for the ordinal mediator. 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` command and the `IND` subcommand in this scenario. 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 an ordinal mediator you can use the

statement `MEDIATOR=OBSERVED` in place of `MEDIATOR=LATENT`. In this case, Mplus will apply the traditional probit based ordinal regression for the T→M link but then it treats the ordinal mediator as if it is interval scaled using the numbers assigned by the researcher to the metric for the M→Y link rather than regressing Y onto m*. 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 FISEM statistical theory does not support it. The number of possible permutations between the ordinal 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).

In practice, I sometimes find I can accomplish my analytic goals with an ordinal mediator strictly within an FISEM framework. Other times I need to move to an LISEM framework. Still other times I use a blend of the two approaches.

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 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. However, sometimes this is not the case.

Sometimes multiple ordinal variables serve as indicators of the same latent construct. When a measurement model is structured as such, the underlying latent variable is treated as a continuous variable. For a good discussion of ordinal multiple indicators of latent variables, see Brown (2015). I provide an example of such a case in the document on *Using Ordinal Multiple Indicators* on my webpage.

CONCLUDING COMMENTS ON THE ANALYSIS OF ORDINAL OUTCOMES

If a measure of a variable is blatantly ordinal, we may need to invoke specialized analytical methods for modeling purposes. Do not do so lightly. Sometimes researchers get into more trouble using ordinal regression than if they just analyzed the data using traditional regression that assumes roughly interval level measurement. As long as the metric of the outcome is not "too ordinal," then the data often can be meaningfully analyzed using traditional methods (see Chapter 3). I urge you to think long and hard about the magnitude of ordinality of your measures before rushing into ordinal regression

modeling, which can be complex, sample size demanding, and subject to its own set of unrealistic underlying assumptions (e.g., parallel coefficients, normality of latent variables) even though intervalness of the outcome metric is not one of them. Ordinal regression often has lower statistical power than traditional OLS regression and often requires larger N to produce stable results (Maxwell, 2000; Whitehead, 1993; Campbell, Julious & Altman, 1995; Taylor, West & Aiken, 2006).

When the Parallel Coefficient Assumption Fails

Suppose the assumption of coefficient equality across subequations is found to be untenable in your ordinal modeling. What remedial actions can you take? One approach is to pursue instead what is known as a **partial model** that relaxes the equality constraints for one or more of the predictors. A partial model forces equality constraints for user-defined subsets of predictors across the subequations, but allows coefficients of other predictors of your choice to vary. This sacrifices some of the parsimony of traditional ordinal regression. For the proportional odds model, the approach is called **partial proportional odds modeling**. There is a substantial literature on these unconstrained counterparts of more traditional constrained models. Interested readers are referred to Fullerton (2009). Mplus does not offer partial proportional odds modeling but they can be pursued in SEM using LISEM. See the program on my website called *ordinal regression* for illustrations of this approach in the video associated with that program.

Another strategy for dealing with violations of the parallel coefficient assumption is to abandon ordinal regression altogether and to use multinomial based SEM that treats outcome categories as if they are unordered and that therefore does not make the parallel coefficient assumption. I illustrate this strategy and provide a detailed example below.

Other Forms of Ordinal Regression

There are many forms of ordinal regression other than the proportional odds model or the probit-based version of it that I have considered in this chapter. One reasonably popular alternative is known as the **adjacent category model.** For this model, the "pairs" comprising the binary subequations or ordinal regression are defined differently than in the proportional odds model. The pairs are based on the adjacent categories of the outcome, i.e., category 2 versus category 1; category 3 versus category 2; and category 4 versus category 3. The subequations for our numerical example focus on comparing (a) the likelihood of individuals being in category 2 versus the likelihood of them being in category 1, (b) the likelihood of individuals being in category 3 versus the likelihood of them being in category 2, and (c) the likelihood of individuals being in category 4 versus

the likelihood of them being in category 3. This translates into a set of three subequations that typically are analyzed using logistic-based ordinal modeling instead of probit modeling.

Suppose I predict a four category outcome from a treatment condition dummy variable, T, and two continuous covariates, CE1 and CIS1, and I let π_j represent the probability of being in category *j* of the four categories, the adjacent category logistic equations that invoke the parallel coefficient assumption are:

Category 2 vs. Category 1: $\ln(\pi_2 / \pi_1) = a_1 + p_1 T + b_1 CE1 + b_2 CIS1$ [13.12]

Category 3 vs. Category 2: $\ln(\pi_3 / \pi_2) = a_2 + p_1 T + b_1 CE1 + b_2 CIS1$ [13.13]

Category 4 vs. Category 3: $\ln(\pi_4 / \pi_3) = a_3 + p_1 T + b_1 CE1 + b_2 CIS1$ [13.14]

An important feature of adjacent category modeling is that it focuses on **local odds** which is distinct from the proportional odds model or its probit equivalent. Consider Equation 13.12. If I focus on just individuals in categories 1 and 2, then the probability of being in category 1 (π_1) must, by definition, equal 1 minus the probability of being in category 2 because if you are in category 1, you are not in category 2 and vice versa. Stated another way, $\pi_1 = 1 - \pi_2$ and if I substitute $1 - \pi_2$ for π_1 on the left hand side of Equation 13.12, I obtain $\pi_2 / (1 - \pi_2)$, which is the classic definition of an odds. If I take the natural log of this odds, I have a logistic model. Because the definitions of π in the adjacent category model are localized to just individuals in the two target categories, the adjacent category model is said to focus on local odds. This is not the case for the proportional odds model.

In equation 13.12, p_1 equals the log odds difference of being in category 2 as opposed to category 1 for the treatment group minus the control group, holding constant CE1 and CIS1; the coefficient b_1 is the number of units that the log odds of being in category 2 (as opposed to category 1) changes given a one unit increase in CE1, holding constant T and CIS1; and so on. As noted, like the proportional odds model, the value of the coefficients for a given predictor are assumed to be constant across subequations, hence I use the same coefficient designation in all three equations, namely p_1 appears in all three equations as does b_1 and b_2.

Mplus does not offer SEM modeling with the adjacent category model. However, you can apply it with LISEM using software that offers single equation adjacent category modeling. For an example, see the video for the *ordinal regression* program on my website.

Another popular type of ordinal model is called a **continuation ratio model**. Like the adjacent category model, continuation ratio models work with local odds. In one

instantiation, known as a **forward stopping model**, a category pair is defined by the probability associated with category *j*, π_j, versus the combined π for all categories greater than *j*. This yields the following pairings for a four category outcome, expressed as ratios:

1 versus 2, 3, and 4: ln ($\pi 1$ / ($\pi 2 + \pi 3 + \pi 4$))

2 versus 3 and 4: ln ($\pi 2$ / ($\pi 3 + \pi 4$))

3 versus 4: ln ($\pi 3$ / ($\pi 4$))

If the categories of the outcome represent a forward progression on some dimension of interest, then this model essentially examines the odds of stopping or "stagnating" in that forward sequence at a certain "stage" or category level. For example, suppose the outcome reflects different levels of mastery of mathematics, with higher categories/numbers indicating greater levels of mastery. I might want to model the odds that people will stagnate at category 1 (versus move forward), the odds they will stagnate at category 2 (versus move forward from there), and the odds they will stagnate at category 3 versus move forward from there. The above subequations accomplish this. As with most ordinal regression, the coefficients again are assumed to satisfy the parallel coefficient assumption. Mplus does not offer this ordinal model but it also can be applied to RETs using LISEM with the *ordinal regression* program on my website.

In sum, there are a range of ordinal regression models available. See Agresti (2010), Fullerton and Xu (2020), Garson (2014), and Yee (2010) for introductions.

NOMINAL/ORDINAL OUTCOMES: THE MULTINOMIAL MODEL

In this section, I illustrate the analysis of an RET when the outcome is nominal or treated as such. For example, a researcher might study the effects of an educational program about treatment options for cancer patients on the type of therapy patients choose from among four options. In this case, the outcome of chosen therapy type is nominal with four categories or levels. In this section, I will reanalyze the data from the previous example on symptom improvement but ignoring the ordinal properties of the outcome and treating each category as being of interest in its own right. This approach is often taken when no viable ordinal model can be fit to the ordinal outcome, usually because of violations of the assumptions of parallel coefficients. The method uses multinomial logistic regression as its analytic cornerstone and applies to any nominal outcome or an ordinal outcome treated as nominal. Mplus does not offer probit modeling in this case but it can be found in R.

Like ordinal regression, multinomial regression uses a series of subequations.

However, unlike ordinal regression, it does not assume coefficients for a given predictor are equal across the subequations; there is no assumption of parallel coefficients. For an outcome variable with *k* levels, there are *k*-1 subequations. In our numerical example, the outcome variable has four categories so there are three sub-equations. The first category is "no improvement," the second category is "minimal improvement," the third category is "much improved" and the fourth category is "very much improved." The subequations focus on local odds and, by default in Mplus, target the following pairs:

$$\text{Category 1 vs. Category 4: } \ln(\text{Odds}_{1 \text{ vs. } 4}) = a_1 + p_1\,T + b_1\,CE1 + b_2\,CIS1 \qquad [13.15]$$

$$\text{Category 2 vs. Category 4: } \ln(\text{Odds}_{2 \text{ vs. } 4}) = a_2 + p_2\,T + b_3\,CE1 + b_4\,CIS1 \qquad [13.16]$$

$$\text{Category 3 vs. Category 4: } \ln(\text{Odds}_{3 \text{ vs. } 4}) = a_3 + p_3\,T + b_5\,CE1 + b_6\,CIS1 \qquad [13.17]$$

I will use the symbol system for defining the dichotomous outcomes in each of the above equations such that $d_{1,4}$ is the outcome of being in category 1 versus category 4, $d_{2,4}$ is the outcome of being in category 2 versus category 4, and $d_{3,4}$ is the outcome of being in category 3 versus category 4. These binary outcomes are modeled using log odds, hence the use of logistic modeling. For the analysis of an RET, we are not so much interested in these equations per se but rather in the probability implications derived from the equation coefficients, which I show below.

In the interest of space, I do not review preliminary analyses for multinomial modeling. They parallel those of ordinal regression but without the parallel coefficient preliminary analyses. For details, see the Resources tab on my webpage for the current chapter.

The core syntax for the analysis with a nominal outcome is shown in Table 13.17. To this syntax, I will be adding `MODEL CONSTRAINT` commands for purposes of contrast analysis which I explain later.

Table 13.17 Multinomial Analysis

```
1.  TITLE: Multinomial logistic analysis ;
2.  DATA: FILE IS c:\mplus\symptom.dat ;
3.  VARIABLE:
4.    NAMES ARE ID GA2 TA2 BD2 CE1 CIS1 T IMP3 ;
5.    USEVARIABLES ARE GA2 TA2 BD2 CE1 CIS1 T IMP3 ;
6.    NOMINAL ARE IMP3 ;
7.    MISSING ARE ALL (-9999) ;
8.  ANALYSIS:
9.    ESTIMATOR = ML ;
10. MODEL:
11.   GA2 on T CE1 CIS1 (p1 b1 b2)  ;
12.   TA2 on T CE1 CIS1 (p2 b3 b4)  ;
```

```
13.   BD2 on T CE1 CIS1 (p3 b5 b6)  ;
14.   IMP3#1 on GA2 TA2 BD2 T CE1 CIS1 (p4a p5a p6a p7a b7a b8a)  ;
15.   IMP3#2 on GA2 TA2 BD2 T CE1 CIS1 (p4b p5b p6b p7b b7b b8b)  ;
16.   IMP3#3 on GA2 TA2 BD2 T CE1 CIS1 (p4c p5c p6c p7c b7c b8c)  ;
17.   [IMP3#1](a1) ; [IMP3#2] (a2) ;  [IMP3#3] (a3) ;
18. OUTPUT: SAMP RESIDUAL STANDARDIZED(STDY) CINTERVAL TECH4  ;
```

Most syntax should be familiar. Line 6 declares the outcome variable as nominal. Line 9 indicates the maximum likelihood estimator which, by default, invokes logistic regression. Lines 14 to 16 specify the three subequations of interest using the notation # followed by a number to reference each equation. Mplus numbers the equations from 1 to 3 and uses the default binary outcomes noted above. Line 17 specifies the intercepts for the three subequations and assigns them each a label. Line 18 omits the modification indices because they are not allowed for this type of model.

In ordinal regression that uses the probability approach, I defined a set of four contrasts based on metric break-points to evaluate intervention effects. In the Appendix to this chapter, I presented an alternative contrast framework that compared the intervention and control groups on each response category of the outcome per se. With multinomial modeling, the tradition is to use the latter approach, namely focus on each response category, so that is what I do here.

To be explicit about the contrasts in models with nominal outcomes, consider Table 13.18. The cells in the first column are conditional probabilities that are the proportion of people in the treatment group who are in each outcome response category. The cells in the second column are the corresponding probabilities for the control group. The contrasts shown in the last column are the difference between the cell proportions in a given row. They are the respective proportion differences between people in the treatment group minus the proportion of people in the control group for the response category in question. My working hypotheses for our numerical example are that the intervention proportions will be greater than the control proportions for categories 3 and 4 (“much improved” and “very much improved”) but the reverse will be true for categories 1 and 2 (“no change” and “minimal improvement”).

Table 13.18: Mplus Code for LISEM Total Effect Analysis

	Treatment	Control	Contrast
No Change (C1)	P(C1|T=1)	P(C1|T=0)	P(C1|T=1) - P(C1|T=0)
Minimal improvement (C2)	P(C2|T=1)	P(C2|T=0)	P(C2|T=1) - P(C2|T=0)
Much improved (C3)	P(C3|T=1)	P(C3|T=0)	P(C3|T=1) - P(C3|T=0)
Very much improved (C4)	P(C4|T=1)	P(C4|T=0)	P(C4|T=1) - P(C4|T=0)

Model Fit

The maximum likelihood estimator applied to logistic regression does not produce traditional global fit indices, nor does it yield modification indices or formal tests of the disparities between predicted and observed covariances/correlations. To evaluate model fit, one can use the analog modification index strategy for LISEM discussed in Chapter 8. I did not find any significant "modification indices" in this regard.

The output yields a predicted correlation matrix in the Tech 4 section called `ESTIMATED CORRELATION MATRIX FOR THE LATENT VARIABLES`. This can be inspected and compared with the observed correlations on a cell-by-cell basis to identify large disparities. As with ordinal regression, the correlations with the various IMP3 variables in this matrix should be ignored as they represent y* constructs for the subequations. In general, the analyses supported the model.

Question 1: The Effects of the Intervention on the Outcome

Mplus does not report total effects for nominal outcomes. I therefore use a LISEM approach to document them, just as I did with ordinal regression. I can employ profile analysis or average marginal effects to gain perspectives on the total effect of a program on a nominal outcome. I consider each approach, in turn. Both approaches consider the four contrasts outlined In Table 13.18. For each contrast, I use the same meaningfulness standard as in the prior ordinal regression analysis, namely for a contrast effect to be meaningful, there must be a difference between the intervention and control groups of at least 5% of clients who marked the metric category.

Profile Analysis

The core guiding equation for profile analyses of the total effect of the intervention on the nominal outcome predicts the nominal outcome from the treatment effect dummy

variable, T, and the two covariates CE1 and CIS1. Table 13.19 presents the syntax I used for the LISEM analysis of the total effect of the intervention on the outcome.

Table 13.19: Mplus Code for LISEM Total Effect Analysis for Nominal Outcome

```
TITLE: LISEM total effect analysis for nominal outcome ;
DATA: FILE IS c:\mplus\symptom.dat ;
VARIABLE:
  NAMES ARE ID GA2 TA2 BD2 CE1 CIS1 T IMP3 ;
  USEVARIABLES ARE  CE1 CIS1 T IMP3 ;
  NOMINAL ARE IMP3 ;
  MISSING ARE ALL (-9999) ;
ANALYSIS:
  ESTIMATOR=ML ;
MODEL:
  IMP3#1 on T CE1 CIS1 (p1a b7a b8a)  ;
  IMP3#2 on T CE1 CIS1 (p1b b7b b8b)  ;
  IMP3#3 on T CE1 CIS1 (p1c b7c b8c)  ;
  [IMP3#1](a1) ; [IMP3#2] (a2) ; [IMP3#3] (a3) ;
MODEL CONSTRAINT:
  NEW (PRED1C PRED2C PRED3C PROB1C PROB2C PROB3C PROB4C SUMC
  PRED1T PRED2T PRED3T PROB1T PROB2T PROB3T PROB4T SUMT
  DIFF1 DIFF2 DIFF3 DIFF4);
!Generate predicted odds for controls as intermediate terms
  PRED1C = exp(a1+p1a*0+b7a*3.036+b8a*7.044) ;
  PRED2C = exp(a2+p1b*0+b7b*3.036+b8b*7.044) ;
  PRED3C = exp(a3+p1c*0+b7c*3.036+b8c*7.044) ;
  SUMC = PRED1C+PRED2C+PRED3C+1;
!Generate predicted control probabilities for the four categories
  PROB1C = PRED1C/SUMC ;
  PROB2C = PRED2C/SUMC ;
  PROB3C = PRED3C/SUMC ;
  PROB4C = 1/SUMC ;
!Generate predicted odds for treatment as intermediate terms
  PRED1T = exp(a1+p1a*1+b7a*3.036+b8a*7.044) ;
  PRED2T = exp(a2+p1b*1+b7b*3.036+b8b*7.044) ;
  PRED3T = exp(a3+p1c*1+b7c*3.036+b8c*7.044) ;
  SUMT = PRED1T+PRED2T+PRED3T+1;
!Generate predicted treatment probabilities for the four categories
  PROB1T = PRED1T/SUMT ;
  PROB2T = PRED2T/SUMT ;
  PROB3T = PRED3T/SUMT ;
  PROB4T = 1/SUMT ;
!Calculate differences in probabilities
  DIFF1 = PROB1T-PROB1C ;
  DIFF2 = PROB2T-PROB2C ;
  DIFF3 = PROB3T-PROB3C ;
  DIFF4 = PROB4T-PROB4C ;
```

```
44. OUTPUT: SAMP RESIDUAL STANDARDIZED(STDY) CINTERVAL TECH4  ;
```

Lines 3 and 4 mean center the covariates and Line 9 declares IMP3 as nominal. I explain the `MODEL CONSTRAINT` commands shortly. All other commands should be self-explanatory. Note that I provide labels for the various intercepts, path coefficients and covariate coefficients, which I make use of in the `MODEL CONSTRAINT` commands.

I use the `MODEL CONSTRAINT` commands to define the four contrasts in Table 13.18. I use an inelegant programming strategy here but it makes explicit the computational mechanics. Consider first Lines 20 to 23. These lines calculate the predicted odds for the control group for each of the three subequations. In Lines 20, 21 and 22, I insert labels and values for the predictors and covariates into each subequation that map onto the definition of the respective equation. Specifically, I multiply the covariate coefficients by their respective mean values (for CE1 the mean is 3.036 and for CIS1 the mean is 7.044) which has the same effect as mean centering them. AS per ordinal modeling, these reflect "typical scores" on the covariates. I multiply the path coefficients for the treatment variable in the three subequations (p1a, p1b, and p1c) by 0 to represent the control group. The predicted log odds result from the expressions in the parentheses and then I take the exponent of them using the `exp` function in Mplus to convert them to odds. In Line 23, I calculate an intermediate term for later use, namely the sum of the predicted odds across the three subequations plus 1. In Lines 25 to 27, I divide each predicted odds by the intermediate term to obtain the control group conditional probabilities for the first k-1 categories. I then define the conditional probability for the reference group (category k) as 1.0 divided by this intermediate value (for the mathematical rationale of these manipulations, see Muthén et al., 2016). I then repeat the entire process for the treatment condition but I now multiply the respective path coefficients for T by 1.0 to reflect my focus on the treatment condition. In lines 40 to 43 I calculate the probability differences for the four target contrasts and these are what I am primarily interested in.

The output for the contrasts is in the section called `New/Additional Parameters`:

```
MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value
New/Additional Parameters
         PRED1C       12.579      4.900      2.567      0.010
         PRED2C       26.117      9.629      2.712      0.007
         PRED3C       15.374      5.623      2.734      0.006
         PROB1C        0.228      0.030      7.675      0.000
         PROB2C        0.474      0.033     14.335      0.000
         PROB3C        0.279      0.029      9.490      0.000
```

PROB4C	**0.018**	**0.006**	**2.829**	**0.005**
SUMC	55.070	19.464	2.829	0.005
PRED1T	0.031	0.015	2.074	0.038
PRED2T	1.008	0.191	5.273	0.000
PRED3T	2.137	0.347	6.155	0.000
PROB1T	**0.007**	**0.003**	**2.179**	**0.029**
PROB2T	**0.241**	**0.027**	**8.860**	**0.000**
PROB3T	**0.512**	**0.032**	**16.068**	**0.000**
PROB4T	**0.239**	**0.028**	**8.465**	**0.000**
SUMT	4.176	0.493	8.465	0.000
DIFF1	**-0.221**	**0.029**	**-7.581**	**0.000**
DIFF2	**-0.233**	**0.044**	**-5.330**	**0.000**
DIFF3	**0.232**	**0.044**	**5.302**	**0.000**
DIFF4	**0.221**	**0.028**	**7.962**	**0.000**

I highlight in red the output lines that are of most interest and I summarize them in Table 13.20 but with the proportions/probabilities multiplied by 100 to put them in percentage format. The percent of control individuals in the lowest category of the IMP3 scale (no change) was 22.8% ±6.0 as compared with 0.7% ±0.6 in the treatment group, a difference that was statistically significant (CR = 7.58, $p < 0.05$). The percent of control individuals in second category of the IMP3 scale (minimal change) was 47.4% ±6.6 as compared with 24.1% ±5.4 in the treatment group, a difference that was statistically significant (CR = 5.33, $p < 0.05$). The percent of control individuals in the third category of the IMP3 scale (much improved) was 27.9% ±5.8 as compared with 51.2% ±6.4 in the treatment group, a difference that was statistically significant (CR = 5.30, $p < 0.05$). Finally, the percent of control individuals in the highest category of the IMP3 scale (very much improved) was 1.8% ±1.2 as compared with 23.9% ±5.6 in the treatment group, a difference that also was statistically significant (CR = 7.96, $p < 0.05$). I calculated the MOEs here using the "double the standard error" heuristic. You can use bootstrapping to check for confidence interval asymmetry and obtain more precise estimates of them.

Table 13.20 Contrast Results for Total Effect using Covariate Mean Scores

IMP3 Category	Treatment Percent	Control Percent	Difference
No change	0.7 ±0.6	22.8 ±6.0	-22.1 ±5.8*
Minimal improvement	24.1 ±5.4	47.4 ±6.6	-23.3 ±8.8*
Much improved	51.2 ±6.4	27.9 ±5.8	23.2 ±8.8*
Very much improved	23.9 ±5.6	1.8 ±1.2	22.1 ±5.6*

As noted in Chapter 12, the percentage differences in Table 13.21 can change depending on the values at which the covariates are held constant. In the above analysis, I set the covariates equal to their mean values. I might also examine total effects when both covariates are at their 25th quantiles and also at their 75th quantiles. To do so, change the `MODEL CONSTRAINT` lines to the covariate values on which you want to focus. For the 25th quantile, I set the clinical experience covariate value to 2.38 and the clinician interpersonal skills value to 6.67 for each subequation. After executing the syntax, I changed the CE1 covariate value to 3.72 and the CIS1 value to 7.37 to represent the 75th quantiles. Note that when the covariates are at their low values (the 25th quantiles), this tends to push scores downward toward the no improvement category because the profile reflects therapists with initial limited experience and possibly substandard clinical skills at baseline. When the covariates are at their high values (the 75th quantiles), this tends to push scores upward toward the very much improved category because the profile reflects therapists with considerable experience and initially solid clinical skills. Tables 13.21a and 13.21b show the results of the two analyses.

Table 13.21a Contrast Results for the 25th Quantile Total Effect Analysis

IMP3 Category	Treatment Percent	Control Percent	Difference
No change	3.7 ±2.8	52.9 ±8.6	-49.1 ±9.4*
Minimal improvement	38.3 ±7.6	34.7 ±7.4	3.6 ±9.8
Much improved	47.9 ±7.6	12.1 ±4.4	35.8 ±7.8*
Very much improved	10.1 ±4.4	0.4 ±0.4	9.7 ±4.2*

Table 13.21b Contrast Results for the 75th Quantile Total Effect Analysis

IMP3 Category	Treatment Percent	Control Percent	Difference
No change	0.1 ±0.1	7.0 ±3.6	-6.9 ±3.6*
Minimal improvement	12.3 ±4.4	43.8 ±7.4	-31.5 ±7.6*
Much improved	43.5 ±7.2	43.2 ±7.4	0.3 ±9.6
Very much improved	44.1 ±7.8	6.1 ±3.6	38.0 ±8.8*

The overall trend towards the program making a difference is apparent, but there are indeed differences in category percentage differences across the tables.

Average Marginal Effects

A second approach to documenting treatment-control differences is to calculate the average marginal effect for each of the four contrasts. I use the same programming logic as that developed for analysis of total effect AMEs for ordinal regression. To write the syntax, I need to first run the program in Table 13.18 to obtain the logistic based subequations to work with. Here are the relevant results from the output:

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

IMP3#1     ON
   T                -6.000      0.584    -10.281      0.000
   CE1              -2.003      0.238     -8.426      0.000
   CIS1             -3.113      0.426     -7.303      0.000


IMP3#2     ON
   T                -3.254      0.395     -8.229      0.000
   CE1              -1.147      0.176     -6.524      0.000
   CIS1             -1.535      0.318     -4.827      0.000

IMP3#3     ON
   T                -1.973      0.369     -5.344      0.000
   CE1              -0.630      0.153     -4.127      0.000
   CIS1             -1.036      0.280     -3.697      0.000

Intercepts
   IMP3#1           30.542      3.137      9.735      0.000
   IMP3#2           17.555      2.418      7.259      0.000
   IMP3#3           11.941      2.135      5.594      0.000
```

The three subequations appear as `IMP3#1`, `IMP3#2`, and `IMP3#3`. The entries are log-odds. From the output, the three equations can be summarized as

$$\ln(\text{Odds}_{1 \text{ vs. } 4}) = 30.542 + -6.000\ T + -2.003\ CE1 + -3.113\ CIS1 \quad [13.18]$$

$$\ln(\text{Odds}_{2 \text{ vs. } 4}) = 17.555 + -3.254\ T + -1.147\ CE1 + -1.535\ CIS1 \quad [13.19]$$

$$\ln(\text{Odds}_{3 \text{ vs. } 4}) = 11.941 + -1.973\ T + -0.630\ CE1 + -1.036\ CIS1 \quad [13.20]$$

Table 13.22 shows the syntax for the average marginal effects for these three equations.

Table 13.22: Mplus Syntax for Average Marginal Effects in Multinomial Model

```
1.  TITLE: AMEs for Total Effect ;
2.  DATA: FILE IS c:\mplus\symptom.dat ;
3.  DEFINE:
4.  !Generate odds for controls as intermediate terms
5.    ODDS1C = exp(30.542+(-6.000*0)+(-2.003*CE1)+(-3.113*CIS1)) ;
6.    ODDS2C = exp(17.555+(-3.254*0)+(-1.147*CE1)+(-1.535*CIS1)) ;
7.    ODDS3C = exp(11.941+(-1.973*0)+(-0.630*CE1)+(-1.036*CIS1)) ;
8.    SUMC = ODDS1C+ODDS2C+ODDS3C+1;
9.  !Generate control probabilities for the four categories
10.   PROB1C = ODDS1C/SUMC ;
11.   PROB2C = ODDS2C/SUMC ;
12.   PROB3C = ODDS3C/SUMC ;
13.   PROB4C = 1/SUMC ;
14. !Generate odds for treatment as intermediate terms
15.   ODDS1T = exp(30.542+(-6.000*1)+(-2.003*CE1)+(-3.113*CIS1)) ;
16.   ODDS2T = exp(17.555+(-3.254*1)+(-1.147*CE1)+(-1.535*CIS1)) ;
17.   ODDS3T = exp(11.941+(-1.973*1)+(-0.630*CE1)+(-1.036*CIS1)) ;
18.   SUMT = ODDS1T+ODDS2T+ODDS3T+1;
19. !Generate treatment probabilities for the four categories
20.   PROB1T = ODDS1T/SUMT ;
21.   PROB2T = ODDS2T/SUMT ;
22.   PROB3T = ODDS3T/SUMT ;
23.   PROB4T = 1/SUMT ;
24. !Calculate différences in probabilities
25.   IME1 = PROB1T-PROB1C ;
26.   IME2 = PROB2T-PROB2C ;
27.   IME3 = PROB3T-PROB3C ;
28.   IME4 = PROB4T-PROB4C ;
29. VARIABLE:
30.   NAMES ARE ID GA2 TA2 BD2 CE1 CIS1 T IMP3 ;
31.   USEVARIABLES ARE IME1 IME2 IME3 IME4 PROB1C PROB1T
32.     PROB2C PROB2T PROB3C PROB3T PROB4C PROB4T;
33.   MISSING ARE ALL (-9999) ;
34. ANALYSIS:
35.   ESTIMATOR = ML ;  TYPE = BASIC ;
36.   OUTPUT: !use defaults on output
```

I have already described the essential features of AME calculation in Chapter 12 and in the current chapter when I introduced Table 13.8. In Table 13.23, Lines 3-7 use the `DEFINE` command to specify the three subequations for calculating the predicted odds for the control condition because I assign everyone a score of T = 0. Lines 15 through 17 do the same for the treatment condition because I assign everyone a score of T= 1. Lines 10 to 13 convert these predicted odds to probabilities for the control condition and Lines 20 to 23 do so for the treatment condition. Lines 25-28 define the differences between the

treatment versus control probabilities to create individual marginal effects, called IMEs. The remainder of the code asks Mplus to calculate the mean of the IMEs to yield the average marginal effects.

Here is the core output from the analysis, which shows the average marginal effects for each response category (the mean of the IMEs), the proportion of controls who are in each response category (PROB1C through PROB4C), and the proportion of those exposed to the treatment in each response category (PROB1T through PROB4T):

```
ESTIMATED SAMPLE STATISTICS

      Means
         IME1          IME2          IME3          IME4          PROB1C
         _______       _______       _______       _______       _______
          -0.263        -0.137         0.163         0.237         0.297

      Means
         PROB1T        PROB2C        PROB2T        PROB3C        PROB3T
         _______       _______       _______       _______       _______
           0.035         0.383         0.246         0.276         0.438

       Means
         PROB4C        PROB4T
         _______       _______
           0.044         0.281
```

Table 13.23 organizes the above information into a table in the form of percentages by multiplying the entries by 100.

Table 13.23: AMEs as Percentages for Total Effect of Intervention on the Outcome

Response Categories	Treatment Percent	Control Percent	Difference (AME)
No change	3.5	29.7	-26.3
Minimal improvement	24.6	38.3	-13.7
Much improved	43.8	27.6	16.3
Very much improve	28.1	4.4	23.7

I am unable to calculate standard errors for the statistics within Mplus. Because the analysis is LISEM-based and the contrasts in my program *AMEs: Ordinal-multinomial* on my website map onto the current parameterization, I can use the program to obtain the AMEs but with p values, and confidence intervals. Here is the (edited) output:

```
Average marginal effects

 Group  Term Contrast Estimate Std. Error      z Pr(>|z|)   2.5 %  97.5 %
     1 TREAT    1 - 0  -0.2621     0.0242 -10.826  < 0.001 -0.3095 -0.21462
     2 TREAT    1 - 0  -0.1370     0.0364  -3.764  < 0.001 -0.2083 -0.06565
     3 TREAT    1 - 0   0.1625     0.0373   4.358  < 0.001  0.0894  0.23553
     4 TREAT    1 - 0   0.2366     0.0257   9.190  < 0.001  0.1861  0.28701
```

I highlight in red the AME estimates, p values, and 95% confidence intervals. The `Group` column lists the metric categories of IMP3. The upper limits of the confidence intervals for the first two contrasts are less than their meaningfulness standards and the lower limits of the confidence intervals for the last two contrasts are greater than their meaningfulness standards, so I conclude the effects of the intervention on each response category is meaningful.

In sum, it is apparent across the two sets of analyses that the program leads to non-trivial symptom improvement but the way this is reflected in the data depends on the vantage point one takes, either as a profile analysis or as average marginal effects or both. It is up to you as the program evaluator to decide how best to frame the data to program staff and administrators for purposes of discussion.

Question 2: Effect of the Intervention on the Mediators

The second question asks what the effect of the intervention is on the three mediators. I will use the same meaningfulness standards as those from the ordinal regression model, which was a mean difference of half a scale unit (0.50) for each mediator (recall that each mediator was measured on a metric of -3 to +3).

There are three relevant equations for assessing intervention effects on the mediators, one equation per mediator (see Equations 13.1 to 13.3). Here is the relevant output after I run the multinomial model syntax in Table 13.17:

```
MODEL RESULTS

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

 GA2        ON
    T                0.989      0.037     26.753      0.000
    CE1              0.179      0.019      9.235      0.000
    CIS1             0.261      0.037      7.111      0.000
 TA2        ON
    T                0.934      0.037     25.582      0.000
    CE1              0.236      0.019     12.309      0.000
    CIS1             0.220      0.036      6.061      0.000
```

```
BD2        ON
   T                -0.015      0.036     -0.416      0.677
   CE1               0.185      0.019      9.867      0.000
   CIS1              0.261      0.035      7.363      0.000
```

All variables listed under a mediator but T are covariates. The coefficient of interest is that for T, which reflects the covariate adjusted mean difference between the treatment and control conditions. The difference for goal alignment was 0.99 ±0.07 ($z = 26.85$, $p < 0.05$). The lower limit of the 95% confidence interval for the intervention versus control mean difference was 0.92, which is larger than the meaningfulness standard of 0.50. This suggests the effect of the intervention on GA2 was meaningful. The results for TA2 and BD2 yielded covariate adjusted mean differences between the intervention and control conditions of 0.93 ±0.07 ($z = 25.80$, $p < 0.05$) and -0.015 ±0.07 ($z = 0.42$, $p < 0.68$), respectively. Evaluation of the lower limits of the confidence intervals revealed that the mean difference for TA2 was meaningful but the mean difference for BD2 was not.

In sum, the analyses suggest the program was effective in changing goal alignment and task alliance. However, the program did not meaningfully affect positive bonding. The program designers need to revisit their intervention and figure out better ways of teaching clinicians to bond with their clients.

Question 3: Effects of the Mediators on the Outcome

The third question addressed in RETs is whether and to what extent the mediators affect the outcome. To answer this question, I will use the same meaningfulness standards that I used in my ordinal regression analyses, namely a one unit change in the mediator must be associated with a 0.017 proportion change in the outcome, i.e., a change that translates into a 1.7% change in the percentage of clients showing improvement as reflected by category analysis. I describe two strategies for evaluating mediator effects on the outcome, profile analysis and average marginal effects.

Profile Analysis

For profile analysis, I focus on the subequations for the three binary IMP3 variables as predicted from the three mediators and the treatment condition for the multinomial analysis in Table 13.17. The subequations are:

$$\ln(\text{Odds}_{1 \text{ vs. } 4}) = a_1 + p_{4a}\,\text{GA2} + p_{5a}\,\text{TA2} + p_{6a}\,\text{BD2} + p_{7a}\,\text{T} + b_7\,\text{CE1} + b_{8a}\,\text{CIS1} \quad [13.21]$$

$$\ln(\text{Odds}_{2 \text{ vs. } 4}) = a_2 + p_{4b}\,\text{GA2} + p_{5b}\,\text{TA2} + p_{6b}\,\text{BD2} + p_{7b}\,\text{T} + b_{7b}\,\text{CE1} + b_{8b}\,\text{CIS1} \quad [13.22]$$

$$\ln(\text{Odds}_{3 \text{ vs. } 4}) = a_3 + p_{4c}\,\text{GA2} + p_{5c}\,\text{TA2} + p_{6c}\,\text{BD2} + p_{7c}\,\text{T} + b_{7c}\,\text{CE1} + b_{8c}\,\text{CIS1} \quad [13.23]$$

Again, as a reminder, there are no equality constraints imposed on the coefficients. To apply the profile analysis approach, I use the core syntax in Table 13.17 but I add a MODEL CONSTRAINT command and associated subcommands to do the profile analysis. I first need to make decisions about the profiles to explore. The basic logic is to estimate the proportion or percentage of clients who show symptom improvement for strategically defined predictor profiles. By then comparing the estimates of different pairs of profiles, I make inferences about the effect of each mediator on the outcome.

For a given pair of profiles targeting a mediator for a given contrast (say contrast 1), I decide to define the first profile by typical scores of the control group on each variable in the equation. The second profile is the same but it increases the value of the target mediator, say GA2, by one unit. By formally comparing the change in proportions for these two profiles, I gain a sense of the effect of a one unit change in GA2 on outcome probabilities/proportions for contrast 1 using control group proportions as a base. This is essentially the same strategy I articulated for ordinal regression.

For each mediator, I calculated the "typical" or mean control group value at the posttest and they generally were near 0, which is the "neither agree nor disagree" point on their respective metrics. I therefore used the values of 0 to define initial "typical" scores on the mediators for the first profile. I used the equivalent of mean centering for the baseline covariates by multiplying their coefficients by their respective means calculated across the total sample because these measures were taken before randomization and likely reflect the means from the general population from which study participants were sampled. I set T = 0 to reflect the control group in the first profile, again to capture the "natural" state of the study population. In order to keep all values for the second profile except that for the target mediator (in this case, GA2, which I increase by 1) equivalent, I set T = 0 for the second profile as well, i.e. the profile values for the first profile are GA2= 0, TA2 = 0, BD2 = 0, T = 0, CE1 = 3.036, CIS1 = 7.044 and for the second profile they are GA2= 1, TA2 = 0, BD2 = 0, T = 0, CE1 = 3.036, CIS1 = 7.044.

Table 13.24 presents the syntax for the GA2 mediator for each of the four contrasts that constitute the core of my analysis, i.e., the proportion of people in each response category of the outcome.

Table 13.24: Code for Effect of Mediator on Outcome Profile Analysis

```
1.  TITLE: Multinomial logistic analysis ;
2.  DATA: FILE IS c:\mplus\symptom.dat ;
3.  VARIABLE:
4.    NAMES ARE ID GA2 TA2 BD2 CE1 CIS1 T IMP3 ;
5.    USEVARIABLES ARE GA2 TA2 BD2 CE1 CIS1 T IMP3 ;
6.    NOMINAL ARE IMP3 ;
7.    MISSING ARE ALL (-9999) ;
```

```
8.  ANALYSIS:
9.    ESTIMATOR = ML ;
10. MODEL:
11.   GA2 on T CE1 CIS1 (p1 b1 b2)  ;
12.   TA2 on T CE1 CIS1 (p2 b3 b4)  ;
13.   BD2 on T CE1 CIS1 (p3 b5 b6)  ;
14.   IMP3#1 on GA2 TA2 BD2 T CE1 CIS1 (p4a p5a p6a p7a b7a b8a)  ;
15.   IMP3#2 on GA2 TA2 BD2 T CE1 CIS1 (p4b p5b p6b p7b b7b b8b)  ;
16.   IMP3#3 on GA2 TA2 BD2 T CE1 CIS1 (p4c p5c p6c p7c b7c b8c)  ;
17.   [IMP3#1](a1) ; [IMP3#2] (a2) ;  [IMP3#3] (a3) ;
18. MODEL CONSTRAINT:
19.   NEW (PRED1P1 PRED2P1 PRED3P1 PROB1P1 PROB2P1 PROB3P1 PROB4P1 SUMP1
20.   PRED1P2 PRED2P2 PRED3P2 PROB1P2 PROB2P2 PROB3P2 PROB4P2 SUMP2
21.   DIFF1 DIFF2 DIFF3 DIFF4);
22. !Generate odds for profile 1 as intermediate terms
23.   PRED1P1 = exp(a1+p4a*0+p5a*0+p6a*0+p7a*0+b7a*3.036+b8a*7.044) ;
24.   PRED2P1 = exp(a2+p4b*0+p5b*0+p6b*0+p7b*0+b7b*3.036+b8b*7.044) ;
25.   PRED3P1 = exp(a3+p4c*0+p5c*0+p6c*0+p7c*0+b7c*3.036+b8c*7.044) ;
26.   SUMP1 = PRED1P1+PRED2P1+PRED3P1+1;
27. !Generate probabilities for profile 1 for the four categories
28.   PROB1P1 = PRED1P1/SUMP1 ;
29.   PROB2P1 = PRED2P1/SUMP1 ;
30.   PROB3P1 = PRED3P1/SUMP1 ;
31.   PROB4P1 = 1/SUMP1 ;
32. !Generate odds for profile 2 as intermediate terms
33.   PRED1P2 = exp(a1+p4a*1+p5a*0+p6a*0+p7a*0+b7a*3.036+b8a*7.044) ;
34.   PRED2P2 = exp(a2+p4b*1+p5b*0+p6b*0+p7b*0+b7b*3.036+b8b*7.044) ;
35.   PRED3P2 = exp(a3+p4c*1+p5c*0+p6c*0+p7c*0+b7c*3.036+b8c*7.044) ;
36.   SUMP2 = PRED1P2+PRED2P2+PRED3P2+1;
37. !Generate probabilities for profile 2 for the four categories
38.   PROB1P2 = PRED1P2/SUMP2 ;
39.   PROB2P2 = PRED2P2/SUMP2 ;
40.   PROB3P2 = PRED3P2/SUMP2 ;
41.   PROB4P2 = 1/SUMP2 ;
42. !Calculate differences in probabilities
43.   DIFF1 = PROB1P2-PROB1P1 ;
44.   DIFF2 = PROB2P2-PROB2P1 ;
45.   DIFF3 = PROB3P2-PROB3P1 ;
46.   DIFF4 = PROB4P2-PROB4P1 ;
47. OUTPUT: SAMP RESIDUAL STANDARDIZED(STDY) CINTERVAL TECH4  ;
```

When TA2 is the target (not shown in the syntax), I manipulate the value of TA2 for p5 across the two profiles for a contrast and when BD2 is the target (also not shown in the syntax), I manipulate the value of BD2 for p6. Here are the GA2 results:

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
New/Additional Parameters				
PRED1P1	24.423	11.152	2.190	0.029
PRED2P1	63.441	27.310	2.323	0.020
PRED3P1	28.261	11.813	2.392	0.017
PROB1P1	**0.209**	**0.031**	**6.620**	**0.000**
PROB2P1	**0.542**	**0.036**	**15.116**	**0.000**
PROB3P1	**0.241**	**0.030**	**7.995**	**0.000**
PROB4P1	**0.009**	**0.004**	**2.400**	**0.016**
SUMP1	117.125	48.805	2.400	0.016
PRED1P2	0.773	0.507	1.525	0.127
PRED2P2	8.698	4.195	2.074	0.038
PRED3P2	10.623	4.720	2.251	0.024
PROB1P2	**0.037**	**0.017**	**2.123**	**0.034**
PROB2P2	**0.412**	**0.065**	**6.349**	**0.000**
PROB3P2	**0.504**	**0.066**	**7.686**	**0.000**
PROB4P2	**0.047**	**0.020**	**2.385**	**0.017**
SUMP2	21.094	8.846	2.385	0.017
DIFF1	**-0.172**	**0.025**	**-6.845**	**0.000**
DIFF2	**-0.129**	**0.061**	**-2.117**	**0.034**
DIFF3	**0.262**	**0.060**	**4.367**	**0.000**
DIFF4	**0.039**	**0.018**	**2.214**	**0.027**

I highlight in red the output lines that are of most interest and summarize them for GA3 in the top portion of Table 13.25 but with the proportions multiplied by 100 to put them in percentage format.

Table 13.25 Effects of Mediators on Outcome for Profile Analysis

	Contrast	Percent at Score of 1	Percent at Score of 0	Difference
GA2	C1: No change	3.7 ±3.4	20.9 ±6.2	-17.2 ±5.0*
	C2: Minimal improvement	41.2 ±13.0	54.2 ±7.2	-12.9 ±12.2*
	C3: Much improved	50.4 ±13.2	24.1 ±6.0	26.2 ±12.0*
	C4: Very much improved	4.7 ±4.0	0.9 ±0.8	3.9 ±3.6*
TA2	C1: No change	5.8 ±4.8	20.9 ±6.2	-15.1 ±5.4*
	C2: Minimal improvement	39.2 ±12.8	54.2 ±7.2	-15.0 ±12.0*
	C3: Much improved	50.6 ±13.2	24.1 ±6.0	26.5 ±12.2*
	C4: Very much improved	4.4 ±4.0	0.9 ±0.8	3.6 ±3.5*

BD2	C1: No change	4.9 ±4.4	20.9 ±5.2	-15.9 ±5.2*
	C2: Minimal improvement	42.7 ±13.2	54.2 ±7.2	-11.4 ±12.2*
	C3: Much improved	48.1 ±13.4	24.1 ±6.0	24.0 ±12.2*
	C4: Very much improved	4.2 ±3.6	0.9 ±0.8	3.3 ±3.2*

A one unit increase in goal alignment leads to statistically significant reductions of clients in the lowest category of "no change in symptoms" as well as in the category for minimal treatment response. A one unit increase in goal alignment also leads to significant increases of clients in the highest category of the symptom measure (very much improved) as well as in the next highest category of much improved.

The bottom two sections of Table 13.25 show the results when I did comparable profile analyses for TA2 and BD2. All of the mediators targeted by the program were meaningfully relevant to the outcome. Using the joint significance test, GA2 and TA2 are both declared as non-zero mediators of the effects of the treatment on the outcome, but this is not the case for BD2. The analysis of the direct effect of the treatment independent of the mediators (not shown here) did not yield results supportive of the effect, which is consistent with the overall ordinal analysis in which its coefficient was statistically non-significant.

I will next want to explore the generalizability of the GA2, TA2 and BD2 mediator effects across other profile contexts, just as I did for ordinal regression. In the interest of space, I do not do so here, but the process is the same as for ordinal regression. For example, recall that in the ordinal regression analysis, I repeated the GA2 analyses but I held TA2 and BD2 constant at values of their 75th quantile. I also used the 75th quantiles for the two baseline covariates. I held T constant at a value of 1.

Average Marginal Effects

A second way of documenting mediator effects on the outcome is to use average marginal effects. To implement this method, I again use results from the original multinomial regression model and the syntax in Table 13.18. Here is the relevant output from that analysis:

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

  IMP3#1     ON
    GA2             -3.453      0.527     -6.551      0.000
    TA2             -2.926      0.509     -5.751      0.000
    BD2             -3.031      0.524     -5.785      0.000
    T               -1.222      0.819     -1.492      0.136
    CE1             -0.669      0.287     -2.336      0.020
    CIS1            -1.479      0.481     -3.074      0.002

 IMP3#2     ON
    GA2             -1.987      0.380     -5.224      0.000
    TA2             -1.969      0.397     -4.959      0.000
    BD2             -1.828      0.389     -4.694      0.000
    T               -0.297      0.596     -0.498      0.619
    CE1             -0.274      0.214     -1.285      0.199
    CIS1            -0.470      0.365     -1.286      0.198

 IMP3#3     ON
    GA2             -0.978      0.324     -3.022      0.003
    TA2             -0.903      0.339     -2.667      0.008
    BD2             -0.901      0.334     -2.700      0.007
    T               -0.533      0.542     -0.983      0.326
    CE1             -0.169      0.185     -0.913      0.361
    CIS1            -0.457      0.315     -1.452      0.146

 Intercepts
    IMP3#1          15.643      3.566      4.387      0.000
    IMP3#2           8.291      2.752      3.013      0.003
    IMP3#3           7.071      2.379      2.972      0.003
```

These results yield the following three logistic subequations:

$$\ln(\text{Odds}_{1 \text{ vs. } 4}) = 15.643 + (-3.453)\,\text{GA2} + (-2.926)\,\text{TA2} + (-3.031)\,\text{BD2} + (-1.222)\,\text{T} + (-.669)\,\text{CE1} + (-1.479)\,\text{CIS1}$$

$$\ln(\text{Odds}_{2 \text{ vs. } 4}) = 8.291 + (-1.987)\,\text{GA2} + (-1.969)\,\text{TA2} + (-1.828)\,\text{BD2} + (-.297)\,\text{T} + (-.274)\,\text{CE1} + (-.470)\,\text{CIS1}$$

$$\ln(\text{Odds}_{3 \text{ vs. } 4}) = 7.071 + (-.978)\,\text{GA2} + (-.903)\,\text{TA2} + (-.901)\,\text{BD2} + (-.533)\,\text{T} + (-.169)\,\text{CE1} + (-.457)\,\text{CIS1}$$

I use these equations to calculate the AMEs for the mediator contrasts using the method of Cameron and Trivedi. Table 13.26 presents the syntax for the GA2 mediator which uses the same logic I described for calculating AMEs for ordinal regression.

Table 13.26 Syntax for AME for Effects of Mediators on Outcomes

```
TITLE: AME for effects of GA2 mediator on outcome ;
DATA: FILE IS c:\mplus\symptom.dat ;
DEFINE:
DELTA = SQRT(0.538)/1000 ; !divide SD of GA2 by 1000
!Generate base odds as intermediate terms
  ODDS1B = exp(15.643+(-3.453)*GA2+(-2.926)*TA2+(-3.031)*BD2+(-1.222)*T+
           (-.669)*CE1+(-1.479)*CIS1) ;
  ODDS2B = exp(8.291+(-1.987)*GA2+(-1.969)*TA2+(-1.828)*BD2+(-.297)*T+
           (-.274)*CE1+(-.470)*CIS1) ;
  ODDS3B = exp(7.071+(-.978)*GA2+(-.903)*TA2+(-.901)*BD2+(-.533)*T+
           (-.169)*CE1+(-.457)*CIS1) ;
  SUMB = ODDS1B+ODDS2B+ODDS3B+1;
!Generate base probabilities for the four categories
  PROB1B = ODDS1B/SUMB ;
  PROB2B = ODDS2B/SUMB ;
  PROB3B = ODDS3B/SUMB ;
  PROB4B = 1/SUMB ;
!Increment mediator for contrasts
  GA2=GA2+DELTA ;
 !Generate incremented odds as intermediate terms
  ODDS1I = exp(15.643+(-3.453)*GA2+(-2.926)*TA2+(-3.031)*BD2+(-1.222)*T+
           (-.669)*CE1+(-1.479)*CIS1) ;
  ODDS2I = exp(8.291+(-1.987)*GA2+(-1.969)*TA2+(-1.828)*BD2+(-.297)*T+
           (-.274)*CE1+(-.470)*CIS1) ;
  ODDS3I = exp(7.071+(-.978)*GA2+(-.903)*TA2+(-.901)*BD2+(-.533)*T+
           (-.169)*CE1+(-.457)*CIS1) ;
  SUMI = ODDS1I+ODDS2I+ODDS3I+1;
!Generate incremental probabilities for the four categories
  PROB1I = ODDS1I/SUMI ;
  PROB2I = ODDS2I/SUMI ;
  PROB3I = ODDS3I/SUMI ;
  PROB4I = 1/SUMI ;
!Calculate differences in probabilities, incremented minus base
  IME1 = (PROB1I-PROB1B)/DELTA ;
  IME2 = (PROB2I-PROB2B)/DELTA ;
  IME3 = (PROB3I-PROB3B)/DELTA ;
  IME4 = (PROB4I-PROB4B)/DELTA ;
VARIABLE:
  NAMES ARE ID GA2 TA2 BD2 CE1 CIS1 T IMP3 ;
  USEVARIABLES ARE IME1 IME2 IME3 IME4 ;
  MISSING ARE ALL (-9999) ;
ANALYSIS:
  ESTIMATOR = ML ;  TYPE = BASIC ;
  OUTPUT: !use defaults on output
```

For GA2, the average marginal effects on the output for the four contrasts were -0.13, -0.07, 0.09, and 0.11, respectively. For every unit that GA2 increases, (a) the

proportion of clients in the lowest outcome category (no change in symptoms) is predicted to decrease by 0.13, (b) the proportion of clients in the second lowest category (minimal improvement) is predicted to decrease by 0.07, (c) the proportion of clients in the third highest outcome category (much improved) is predicted to increase by 0.09, and (d) the proportion of clients who were very much improved, the top category of the scale, is predicted to increase by 0.11. When I re-ran the program to focus on TA2, the corresponding AMEs were -0.09, -0.10, 0.09 and 0.10, respectively. For BD2, they were -0.10, -0.07, 0.08, and 0.10.

When I adapted the syntax in Table 13.26 to calculate the AMEs for the independent effects of the treatment condition on the contrasts over and above the three mediators. I deleted lines 4 and 18, removed the `/DELTA` terms from lines 33 to 36, changed `T` to `0` in lines 6, 8 and 10, and changed `T` to `1` in lines 20, 22 and 24. Consistent with the overall multinomial analysis that revealed a statistically non-significant independent effect for the treatment effect over and above the mediators, the AMEs were trivial in magnitude.

The method of calculating AMEs in Mplus does not yield significance tests nor confidence intervals for them, which is a drawback. The program AMEs: Ordinal-multinomial on my website calculates AMEs for multinomial regression but it does so in a LISEM context. To gain some sense of significance tests and confidence intervals, I applied my program to the multinomial equation predicting IMP3 from GA2, TA2, BD2, the treatment condition, CE1 and CIS1. Here are the (edited) results, which, it turns out yield AME estimates that are quite close to the AMEs based on FISEM that I calculated in Mplus:

```
Average marginal effects

Group  Term Contrast Estimate Std. Error      z Pr(>|z|)     2.5 %   97.5 %
    1 GA2       dY/dX -0.12753     0.0262 -4.866  < 0.001 -0.178890 -0.07616
    2 GA2       dY/dX -0.07054     0.0384 -1.837  0.06622 -0.145806  0.00472
    3 GA2       dY/dX  0.08940     0.0392  2.281  0.02256  0.012575  0.16622
    4 GA2       dY/dX  0.10867     0.0263  4.132  < 0.001  0.057122  0.16021
    1 TA2       dY/dX -0.09165     0.0243 -3.768  < 0.001 -0.139322 -0.04398
    2 TA2       dY/dX -0.10306     0.0389 -2.650  0.00805 -0.179282 -0.02684
    3 TA2       dY/dX  0.09195     0.0406  2.265  0.02354  0.012367  0.17153
    4 TA2       dY/dX  0.10276     0.0279  3.683  < 0.001  0.048081  0.15744
    1 BD2       dY/dX -0.10670     0.0261 -4.085  < 0.001 -0.157898 -0.05550
    2 BD2       dY/dX -0.07249     0.0393 -1.845  0.06507 -0.149508  0.00453
    3 BD2       dY/dX  0.07929     0.0403  1.968  0.04903  0.000337  0.15825
    4 BD2       dY/dX  0.09990     0.0275  3.628  < 0.001  0.045935  0.15386
    1 TREAT     1 - 0 -0.06085     0.0405 -1.502  0.13302 -0.140240  0.01854
    2 TREAT     1 - 0  0.06823     0.0613  1.113  0.26554 -0.051878  0.18834
```

```
3 TREAT    1 - 0 -0.04937     0.0636 -0.776  0.43763 -0.174044  0.07530
4 TREAT    1 - 0  0.04200     0.0443  0.948  0.34309 -0.044822  0.12881
```

I again highlight key results in red. The column called `Contrast` describes the mathematical formulation used to calculate the marginal effect with dY/dX indicating instantaneous change for a continuous variable and 0-1 indicating discrete change for a dummy variable. In general and with a few exceptions, the results affirm the effects of the mediators on the outcome for each response category.

Concluding Comments on the Effects of Mediators on the Outcome

In sum, both profile analysis and average marginal effects are useful for documenting the effects of the mediators on the outcome. Each method addresses the matter from a different vantage point. AMEs are useful because they do not rely on conditional effects in the way that panel analysis does. Panel analysis explores how mediator effects on outcomes differ as a function of predictor contexts, i.e., predictor profiles.

Parenthetically, given the above results and using the joint significance test, both goal alignment and task alliance would be declared as providing non-zero mediation of the effect of the intervention on symptom improvement, but this would not be the case for bonding because the program had a trivial effect on bonding, which breaks the mediational chain.

Nominal Mediators and Latent Variables

You may encounter situations where one or more of your mediators is nominal with three or more categories. As with ordinal variables, nominal mediators are endogenous because there is a causal path emanating from the treatment (dummy) variable to the mediator or from another mediator to the ordinal mediator. At the same time, the nominal mediator is a cause of the outcome and/or another mediator. The dual role of the nominal mediator as both a predictor variable and a dependent variable creates challenges for FISEM approaches but the situation is more straightforward for LISEM. For LISEM, the focus is on documenting and evaluating the magnitude of the path coefficient(s) for each separate link in the mediational chain using any one of many available statistical tools and then using the joint significance test (JST) to evaluate the null hypothesis of no mediation across the full mediational chain (see Chapter 9). Nominal mediators pose no special challenges in such cases. I now describe LISEM-based methods you might use for a range of scenarios with nominal mediators.

Scenario 1: If the mediator is nominal and the outcome is nominal, use multinomial logistic regression as outlined in this chapter to regress Y onto the dummy variables for

M and the dummy variable for T to isolate the effect M→Y and the direct effect of T on Y holding M constant. In a second analysis, regress M onto T using multinomial logistic regression to isolate the effect T→M. Document all effects from the multinomial logistic regressions using the relevant *a priori* defined probability differences. Include measured covariates in all analyses as dictated by theory. Evaluate the statistical significance and magnitude of each link in the respective mediational chains and then apply the joint significance test, as appropriate.

Scenario 2: If the mediator is nominal and the outcome is continuous, use robust OLS or robust maximum likelihood to regress Y onto the dummy variables for M and T to isolate the effect M→Y and the direct effect of T on Y holding M constant. In a second analysis, regress M onto T using multinomial logistic regression to isolate the effect T→M. Document all effects from the multinomial logistic regressions using the relevant *a priori* defined probability differences. Include measured covariates in all analyses as dictated by theory. Evaluate the statistical significance and magnitude of each link in the respective mediational chains and then apply the joint significance test, as appropriate.

Scenario 3: If the mediator is nominal and the outcome is ordinal, use ordinal regression to regress Y onto the dummy variables for M and T to isolate the effect M→Y and the direct effect of T on Y holding M constant. In a second analysis, regress M onto T using multinomial logistic regression to isolate the effect T→M. Document all effects from the multinomial logistic regressions using the relevant *a priori* defined probability differences. Include measured covariates in all analyses as dictated by theory. Evaluate the statistical significance and magnitude of each link in the respective mediational chains and then apply the joint significance test, as appropriate.

Scenario 4: If the mediator is nominal and the outcome is binary, use one of the binary regression models from Chapter 12 to regress Y onto the dummy variables for M and T to isolate the effect M→Y and the direct effect of T on Y holding M constant. In a second analysis, regress M onto T using multinomial logistic regression to isolate the effect T→M. Document all effects from the multinomial logistic regressions using the relevant *a priori* defined probability differences. Include measured covariates in all analyses as dictated by theory. Evaluate the statistical significance and magnitude of each link in the respective mediational chains and then apply the joint significance test, as appropriate.

Scenario 5: If the mediator is nominal and the outcome is time until an event occurs, use survival analysis to regress Y onto the dummy variables for M and T to isolate the effect M→Y and the direct effect of T on Y holding M constant. In a second analysis, regress M onto T using multinomial logistic regression to isolate the effect T→M. Document all

effects from the multinomial logistic regressions using the relevant *a priori* defined probability differences. Include measured covariates in all analyses as dictated by theory. Evaluate the statistical significance and magnitude of each link in the respective mediational chains and then apply the joint significance test, as appropriate.

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 (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.

With FISEM, the analytic flexibility is more constrained. Muthén (2011) discusses in detail the case of nominal mediators in such cases. The joint analysis of a nominal variable as an outcome in one equation and as a predictor in another equation is handled in Mplus by using mixture analysis with a nominal latent class variable that is defined by the observed nominal M. Latent class membership is known and treated using the `KNOWNCLASS` feature of Mplus. For a continuous outcome, Y, the mean Y changes across the classes. An interaction between T and M allows the direct influence of X on Y vary over the latent classes. Maximum-likelihood estimation is used for the omnibus mediation effect using `MODEL CONSTRAINT` commands. For details, see Muthén (2011).

Concluding Comments on Multinomial Modeling

Nominal outcomes are not as common as binary, ordinal, and continuous outcomes in the social and health sciences. They can be readily handled in RETs using Mplus either with FISEM or LISEM. Another use of multinomial modeling is when an ordinal regression model fails to fit the data or its underlying assumptions are violated. In such cases, one can treat the response categories as nominal and still gain considerable insights into RET dynamics because the multinomial model is more flexible; it does not make the parallel coefficient assumption that is made in ordinal regression. Indeed, when I analyzed the ordinal outcome example in the present chapter in this way, I was able to extract a wealth of information relative to the effects of an intervention on an outcome.

CONCLUDING COMMENTS

Ordinal modeling is typically invoked when a measure has blatantly ordinal measurement properties. In my experience, many researchers are too quick to treat a measure as ordinal without appreciating the core psychometrics of interval and ordinal measurement. It is a mistake to think of measures as being *either* interval *or* ordinal. As I discussed in Chapter 3, intervalness is a matter of degree; sometimes a measure is ordinal in character but it reasonably approximates intervalness. Other times, the measure deviates from intervalness so much that it interferes with conclusion validity when analyzed as if it is interval. As someone who has specialized in attitude measurement, I find it particularly galling when researchers refer to a Likert scale as any single item rating scale with adverb qualifiers attached to it, insisting that the measure must be ordinal. First, Likert scaling is an elegant multi-item psychometric approach to attitude scale construction. It involves item generation, item screening by invoking theoretical tracelines, and specific scoring algorithms for estimating a person's true underlying attitude. Rensis Likert would turn in his grave if he knew that virtually every adjective-based single item rating scale is called a "Likert scale." Many single item rating scales approximate intervalness enough that they can be effectively analyzed as if they are interval. In fact, for some questions, the ordinality can be weak with only trivial ramifications for the inferences we make.

When we shift to ordinal modeling, we escape assumptions of strict intervalness of our outcome measure but this comes at a cost of making other assumptions that might be problematic. For example, traditional ordinal modeling makes assumptions about parallel coefficients across equations and it assumes a specific distribution shape for the continuous variable thought to underlie the ordinal metric. Sparse cell frequencies in the ordinal measure can introduce problems as can assumptions about thresholds. Usually, you as a researcher are in control of the measures you use when you design an RET for purposes of program evaluation. My advice is to get your psychometric house in order during the planning stages of an evaluation and stay away from blatantly ordinal measures. Sometimes this may not be possible, but it is an ideal you should strive for. To do good program evaluation, you need to be a good psychometrician or have a good one on your research team.

In this chapter, I outlined two approaches to ordinal regression, both relying on a probit version of a proportional odds model. The approaches are cumulative link in character and emphasize break points on the ordinal metric. One approach (the probability approach) provides extensive information about group differences on the categories defined by the break points. The other approach (the latent response approach) focuses on the continuous latent construct thought to underlie the ordinal metric and

explores predictors of that continuous variable. The latter approach has the advantage of being simple and straightforward, but it relies on standardized metrics, which clients sometimes have difficulty understanding. The former approach works with intuitive proportions and percentages, but it often defies simple summarization. You should use the approach that you feel is best for the particular evaluation context you are in. In my work, I usually use both approaches because I like to come at issues from multiple perspectives.

If the assumptions of the proportional odds approach are non-trivially violated, then one option is to use a LISEM based adjacent category model. My webpage has the software you need to do so. Alternatively, you can shift to multinomial modeling that does not make some of the strong assumptions of ordinal regression and execute the analysis in Mplus in the way illustrated in the current chapter. Indeed, for a probability approach with an ordinal outcome, there is something to be said for just using multinomial modeling so that one avoids the parallel coefficient assumption altogether.

When writing this chapter, I debated whether to orient my presentation of the probability approach around break-point analysis and contrasts surrounding those break points or to orient the chapter around the alternative parameterization that I outline in the Appendix. Ultimately, I decided to do the former because of its coherence with the underlying statistical theory but the fact is that I usually, in practice, apply the approach in the Appendix because I find it simpler and easier to interpret. In the final analysis, when faced with an outcome with an ordinal metric, I can use any of the following approaches: (1) probit based breakpoint ordinal modeling with the probability approach parameterized using breakpoint contrasts, (2) probit based breakpoint ordinal modeling with the probability approach parameterized using separate outcome response categories (per the Appendix), (3) probit based breakpoint ordinal modeling with the latent response approach, or (4) multinomial modeling. Furthermore, I can apply these methods in either a FISEM or a LISEM context, and for probit models, using either a frequentist or Bayesian framework. I ultimately draw on one or more of these various tools from my statistical toolbox depending on the substantive context in which I am doing my research.

APPENDIX: ALTERNATIVE PARAMETERIZATIONS OF ORDINAL REGRESSION

In this appendix, I show you how to apply the probit version of the proportional odds model to contrasts that focus on each response category of the ordinal outcome rather than the contrasts centered on break points that were illustrated in the main text. I use the symptom improvement example, keeping my presentation brief based on the assumption that you have read and understood the logic of ordinal regression from the main text.

Question 1: The Effects of the Intervention on the Outcome

To estimate the total effect of the intervention, I again resort to LISEM that regresses the outcome onto the treatment condition dummy variable and the covariates using ordinal regression. Table A.1 presents the syntax for the analysis.

Table A.1: Mplus Code for LISEM Total Effect Analysis For Alternative Parameterization

```
1.  TITLE: LISEM total effect analysis ;
2.  DATA: FILE IS c:\mplus\symptom.dat ;
3.  VARIABLE:
4.    NAMES ARE ID GA2 TA2 BD2 CE1 CIS1 T IMP3 ;
5.    USEVARIABLES ARE CE1 CIS1 T IMP3 ;
6.    CATEGORICAL ARE IMP3 ;
7.    MISSING ARE ALL (-9999) ;
8.  ANALYSIS:
9.    ESTIMATOR = ML ; LINK = PROBIT ;
10. MODEL:
11.   IMP3 on T CE1 CIS1 (p1 b1 b2 )  ;
12.   [IMP3$1] (t1) ;  [IMP3$2] (t2) ; [IMP3$3] (t3) ;
13. MODEL CONSTRAINT:
14.   NEW (c1probc c1probt c1diff
15.   c2probc c2probt c2diff
16.   c3probc c3probt c3diff
17.   c4probc c4probt c4diff ) ;
18. !Contrast 1 : Total effect for category 1 of the outcome measure
19.   c1probc = 1-phi(-t1 + p1*0 + b1*3.036 + b2*7.044) ;
20.   c1probt = 1-phi(-t1 + p1*1 + b1*3.036 + b2*7.044) ;
21.   c1diff = c1probt - c1probc ;
22. !Contrast 2 : Total effect for category 2 of the outcome measure
23.   c2probc = (1-phi(-t2 + p1*0 + b1*3.036 + b2*7.044))-c1probc ;
24.   c2probt = (1-phi(-t2 + p1*1 + b1*3.036 + b2*7.044))-c1probt ;
25.   c2diff = c2probt - c2probc ;
26. !Contrast 3 : Total effect for category 3 of the outcome measure
27.   c3probc = (1-phi(-t3 + p1*0 + b1*3.036 + b2*7.044))-c2probc-c1probc ;
```

```
28.   c3probt = (1-phi(-t3 + p1*1 + b1*3.036 + b2*7.044))-c2probt-c1probt ;
29.   c3diff = c3probt - c3probc ;
30. !Contrast 4 : Total effect for category 4 of the outcome measure
31.   c4probc = phi(-t3 + p1*0 + b1*3.036 + b2*7.044) ;
32.   c4probt = phi(-t3 + p1*1 + b1*3.036 + b2*7.044) ;
33.   c4diff = c4probt - c4probc ;
34. OUTPUT: SAMP RESIDUAL STANDARDIZED(STDY) CINTERVAL TECH4  ;
```

The syntax follows the same logic as Table 13.4 except the `MODEL CONSTRAINT` commands differ to reflect different contrasts. For each contrast, I assign specific values for the covariates at which to hold the covariates constant and, in this instance, I do so using their sample mean or "typical" values, which is 3.036 for `CE1` and 7.044 for `CIS1`.

The first contrast follows the same logic of the first contrast in the break point analysis. Line 19 calculates the control group's probability of being in category 1 of the outcome ("no change or got worse"). Line 20 calculates the corresponding probability for the intervention/treatment group. Line 21 calculates the difference between these two probabilities and is of primary interest.

I calculate the contrast for the second category of the outcome in the same way as the break point analysis (which combines categories 1 and 2; see Lines 22 to 25) but I subtract from it the proportion of cases in category 1 to isolate the predicted proportion for just category 2 (examine the final entries of Lines 23 and 24, `c1probc` and `c1probt`).

The contrast for the third category of the outcome variable uses the Mplus internal breakpoint dichotomy of category 4 versus categories 1, 2 and 3 combined. By subtracting the phi function of the probit from 1, I obtain the proportion of cases in categories 1, 2 and 3 combined. I then subtract from this result the proportion of cases in category 1 and also in category 2, which yields the proportion of cases in category 3 (examine the entries at the end of Lines 27 and 28).

The contrast for category 4 of the outcome metric follows the identical logic of the last contrast in the break point analysis.

The output for the contrasts in proportion form appears in the Mplus output section `New/Additional Parameters:`

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

New/Additional Parameters
    C1PROBC            0.243      0.025      9.730      0.000
    C1PROBT            0.018      0.005      3.807      0.000
    C1DIFF            -0.225      0.023     -9.655      0.000
    C2PROBC            0.496      0.027     18.246      0.000
    C2PROBT            0.207      0.021      9.903      0.000
    C2DIFF            -0.289      0.028    -10.169      0.000
```

```
    C3PROBC            0.243      0.023     10.750      0.000
    C3PROBT            0.537      0.027     20.168      0.000
    C3DIFF             0.294      0.029     10.229      0.000
    C4PROBC            0.017      0.005      3.818      0.000
    C4PROBT            0.237      0.025      9.667      0.000
    C4DIFF             0.220      0.023      9.574      0.000
```

I summarize the results in Table A.2 using percents instead of proportions, i.e. the proportions are multiplied by 100.

Table A.2: LISEM Total Effects based on Profile Analysis

Contrast Categories	Treatment Percent	Control Percent	Difference
C1: No change or worse	1.8 ±0.8	24.3 ±4.8	-22.5 ±4.4*
C2: Minimal improvement	20.7 ±4.2	49.6 ±5.4	-28.9 ±5.6*
C3: Much improved	53.7 ±5.4	24.3 ±4.6	29.4 ±5.8*
C3: Very much improved	23.7 ±5.0	1.7 ±0.8	22.0 ±4.6*

As noted in the main text, the percentage differences in Table A.2 can change depending on the values at which the covariates are held constant. In the above analysis, I set the covariates equal to their mean values, but I invariably want to explore variants of the syntax in which I hold the covariates constant at different values. In the main text for the break point analysis, I did so for the 25th and 75th quantiles of the covariates. I do not do so here to save space but you will want to pursue similar analyses for the present case.

To calculate average marginal effects that map onto the contrasts, I use the results for the probit equation from the analysis in Table 13.3, which I repeat here for reference:

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

 IMP3       ON
    T                1.394      0.098     14.292      0.000
    CE1              0.498      0.053      9.358      0.000
    CIS1             0.725      0.094      7.731      0.000

Thresholds
    IMP3$1           5.923      0.650      9.119      0.000
    IMP3$2           7.261      0.666     10.901      0.000
    IMP3$3           8.729      0.682     12.800      0.000
```

These results yield the following three probit subequations:

$$\text{Probit}(IMP3_{d1}) = -5.923 + 1.394\ T + .498\ CE1 + .725\ CIS1$$

$$\text{Probit}(IMP3_{d2}) = -7.261 + 1.394\ T + .498\ CE1 + .725\ CIS1$$

$$\text{Probit}(IMP3_{d3}) = -8.729 + 1.394\ T + .498\ CE1 + .725\ CIS1$$

I use these equations to calculate AMEs for the contrasts using the syntax in Table A.3.

Table A.3: Syntax for Calculating AMEs

```
TITLE: LISEM total effect analysis ;
DATA: FILE IS c:\mplus\symptom.dat ;
DEFINE:
!Contrast 1
  PROBIT1C = -5.923 + 1.394*0 + .498*CE1 + .725*CIS1 ;
  PROB1C = 1-PHI(PROBIT1C) ; !C1 prob for control group
  PROBIT1T = -5.923 + 1.394*1 + .498*CE1 + .725*CIS1 ;
  PROB1T = 1-PHI(PROBIT1T) ; !C1 prob for treat group
  IME1 = PROB1T-PROB1C ;
!Contrast 2
  PROBIT2C = -7.261 + 1.394*0 + .498*CE1 + .725*CIS1 ;
  PROB2C = 1-PHI(PROBIT2C)-PROB1C ; !C2 prob for control group
  PROBIT2T = -7.261 + 1.394*1 + .498*CE1 + .725*CIS1 ;
  PROB2T = 1-PHI(PROBIT2T)-PROB1T ; !C2 prob for treat group
  IME2 = PROB2T-PROB2C ;
!Contrast 3
  PROBIT3C = -8.729 + 1.394*0 + .498*CE1 + .725*CIS1 ;
  PROB3C = (1-PHI(PROBIT3C))-PROB1C-PROB2C ; !C3 prob for cntrl grp
  PROBIT3T = -8.729 + 1.394*1 + .498*CE1 + .725*CIS1 ;
  PROB3T = (1-PHI(PROBIT3T))-PROB1T-PROB2T ; !Prob for treat group
  IME3 = PROB3T-PROB3C ;
!Contrast 4
  PROBIT4C = -8.729 + 1.394*0 + .498*CE1 + .725*CIS1 ;
  PROB4C = PHI(PROBIT4C) ; !C3 probability for control group
  PROBIT4T = -8.729 + 1.394*1 + .498*CE1 + .725*CIS1 ;
  PROB4T = PHI(PROBIT4T) ; !Probability for treat group
  IME4 = PROB4T-PROB4C ;
VARIABLE:
  NAMES ARE ID GA2 TA2 BD2 CE1 CIS1 T IMP3 ;
  USEVARIABLES ARE IME1 IME2 IME3 IME4 PROB1C PROB1T
  PROB2C PROB2T PROB3C PROB3T PROB4C PROB4T ;
  MISSING ARE ALL (-9999) ;
ANALYSIS:
 ESTIMATOR=ML ; TYPE=BASIC ;
OUTPUT: !use defaults on output
```

The logic follows that for the total effect AME for break point analysis but adapted to the re-parameterized contrasts. I obtain the means from the output called `RESULTS FOR BASIC ANALYSIS`. The results are shown in Table A.4.

Table A.4: AMEs for Total Effect Analysis

Contrast Categories	Treatment Percent	Control Percent	Difference
C1: No change	4.8	28.7	-23.9
C2: Minimal improve	22.4	41.1	-18.7
C3: Much improved	44.7	25.6	19.1
C4: Very much improve	28.2	4.6	23.5

A disadvantage of the Mplus syntax is that it does not yield confidence intervals or significance tests for the AMEs. Because the analysis is LISEM based and the contrasts in my website program *AMEs: Ordinal-multinomial* map onto the current parameterizations, I can use my program to obtain the same AMEs but with standard errors, p values and confidence intervals. Here is the (edited) program output:

```
Average marginal effects

 Group  Term Contrast Estimate Std. Error      z Pr(>|z|)   2.5 %  97.5 %
     1 TREAT    1 - 0  -0.2387    0.01969 -12.13   <0.001 -0.2773 -0.2001
     2 TREAT    1 - 0  -0.1875    0.02024  -9.26   <0.001 -0.2272 -0.1479
     3 TREAT    1 - 0   0.1905    0.02037   9.35   <0.001  0.1506  0.2305
     4 TREAT    1 - 0   0.2357    0.01957  12.04   <0.001  0.1974  0.2741
```

I highlight in red the AME estimates, p values, and 95% confidence intervals. The `Group` column lists the metric categories of IMP3.

Question 2: Effect of the Intervention on the Mediators

The estimation of the effects of the intervention on the mediators follow the same principles and focus on the same Mplus output as described in the main text for the break analysis contrasts in ordinal regression I do not repeat this material here in the interest of space,

Question 3: Effects of the Mediators on the Outcome

The syntax for the profile analysis for characterizing the effect of the GA mediator on the outcome when holding the other mediators constant at their typical baseline values of 0 and the covariates constant at their mean values is shown in Table A.5. The first profile values are GA2= 0, TA2 = 0, BD2 = 0, T = 0, CE1 = 3.036, CIS1 = 7.044 and the second profile vales are GA2= 1, TA2 = 0, BD2 = 0, T = 0, CE1 = 3.036, CIS1 = 7.044.

Table A.5: Code for Effect of Mediator on Outcome Profile Analysis

```
1.  TITLE: Ordinal regression with probit profile analysis 1 ;
2.  DATA: FILE IS c:\mplus\symptom.dat ;
3.  VARIABLE:
4.    NAMES ARE ID GA2 TA2 BD2 CE1 CIS1 T IMP3 ;
5.    USEVARIABLES ARE GA2 TA2 BD2 CE1 CIS1 T IMP3 ;
6.    CATEGORICAL ARE IMP3 ;
7.    MISSING ARE ALL (-9999) ;
8.  ANALYSIS:
9.  ESTIMATOR = ML ; LINK = PROBIT ;
10. MODEL:
11.   GA2 on T CE1 CIS1 (p1 b1 b2)  ;
12.   TA2 on T CE1 CIS1 (p2 b3 b4)  ;
13.   BD2 on T CE1 CIS1 (p3 b5 b6)  ;
14.   IMP3 on GA2 TA2 BD2 T CE1 CIS1 (p4 p5 p6 p7 b7 b8)  ;
15.   [IMP3$1] (t1) ; [IMP3$2] (t2) ; [IMP3$3] (t3) ;
16. MODEL INDIRECT:
17. IMP3 IND T ;
18. MODEL CONSTRAINT:
19.  NEW (c1pm0 c1pm1 c1diff c2pm0 c2pm1 c2diff
20.      c3pm0 c3pm1 c3diff c4pm0 c4pm1 c4diff ) ;
21.  CONTRAST 1
22.  c1pm0 = 1-phi(-t1+p4*0+p5*0+p6*0+p7*0+b7*3.036+b8*7.044) ;
23.  c1pm1 = 1-phi(-t1+p4*1+p5*0+p6*0+p7*0+b7*3.036+b8*7.044) ;
24.  c1diff = c1pm1-c1pm0 ;   !prob difference
25.  !CONTRAST 2
26.  c2pm0 = (1-phi(-t2+p4*0+p5*0+p6*0+p7*0+b7*3.036+b8*7.044))-c1pm0 ;
27.  c2pm1 = (1-phi(-t2+p4*1+p5*0+p6*0+p7*0+b7*3.036+b8*7.044))-c1pm1 ;
28.  c2diff = c2pm1-c2pm0 ; !prob difference
29.  !CONTRAST 3
30.  c3pm0 = (1-phi(-t3+p4*0+p5*0+p6*0+p7*0+b7*3.036+b8*7.044))-c1pm0-c2pm0 ;
 31. c3pm1 = (1-phi(-t3+p4*1+p5*0+p6*0+p7*0+b7*3.036+b8*7.044))-c1pm1-c2pm1 ;
 32. c3diff = c3pm1-c3pm0 ; !prob difference
 33. !CONTRAST 4
 34. c4pm0 = phi(-t3+p4*0+p5*0+p6*0+p7*0+b7*3.036+b8*7.044);
 35. c4pm1 = phi(-t3+p4*1+p5*0+p6*0+p7*0+b7*3.036+b8*7.044);
 36. c4diff = c4pm1-c4pm0 ;    !prob difference
 37. OUTPUT: SAMP RESIDUAL STAND(STDY) CINTERVAL TECH4  ;
```

I use the same logic as for the total effect analysis to isolate the separate category probabilities by subtracting out the category probabilities that were lumped into it by virtue of the way Mplus defines IMP_{d1}, IMP_{d2} and IMP_{d3}. For example, for Contrast 2, I calculate the probability for categories 1 and 2 combined and then I subtract from this the probability of category 1 (see Line 26 and 27). For Contrast 3, I calculate the probability for categories 1, 2 and 3 combined and then I subtract from this the probabilities of categories 1 and 2 (see Line 30 and 31). When TA2 is the target (not shown in the syntax), I manipulate the value of TA2 for p5 between 0 and 1 across the two profiles for a contrast and when BD2 is the target (also not shown in the syntax), I manipulate the value of BD2 for p6. Here are the GA2 results:

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

New/Additional Parameters
         C1PM0               0.230      0.026      8.967      0.000
         C1PM1               0.062      0.017      3.554      0.000
         C1DIFF             -0.168      0.020     -8.476      0.000
         C2PM0               0.553      0.029     19.107      0.000
         C2PM1               0.430      0.040     10.677      0.000
         C2DIFF             -0.122      0.036     -3.366      0.001
         C3PM0               0.211      0.023      9.191      0.000
         C3PM1               0.461      0.042     10.867      0.000
         C3DIFF              0.250      0.035      7.063      0.000
         C4PM0               0.007      0.002      2.915      0.004
         C4PM1               0.047      0.014      3.298      0.001
         C4DIFF              0.040      0.013      3.145      0.002
```

Table A.6 summarizes the above results but expressed as percentages. I also provide results for when I ran comparable syntax for the other two mediators.

Table A.6 Effects of Mediators on Outcome for Profile Analysis

	Contrast	Percent at Score of 1	Percent at Score of 0	Difference
GA2	C1: No change	6.2 ±3.4	23.0 ±5.2	-16.8 ±4.2*
	C2: Minimal improvement	43.0 ±8.0	55.3 ±5.8	-9.1 ±6.6*
	C3: Much improved	46.1 ±8.4	21.1 ±4.6	25.0 ±7.0*
	C4: Very much improved	4.7 ±2.8	0.7 ±0.4	4.0 ±2.6*
	C1: No change	7.6 ±4.0	23.0 ±5.2	-15.4 ±4.0*

TA2	C2: Minimal improvement	45.9 ±6.8	55.3 ±5.8	-9.3 ±6.6*
	C3: Much improved	42.7 ±8.6	21.1 ±4.6	21.7 ±7.2*
	C4: Very much improved	3.7 ±2.4	0.7 ±0.4	3.1 ±2.3*
BD2	C1: No change	7.7 ±4.2	23.0 ±5.2	-15.3 ±4.2*
	C2: Minimal improvement	46.1 ±7.8	55.3 ±5.8	-9.1 ±6.6*
	C3: Much improved	42.5 ±8.8	21.1 ±4.6	21.4 ±7.2*
	C4: Very much improved	3.7 ±2.4	0.7 ±0.4	3.0 ±2.0*

As with the total effect analysis, you would explore these effects for other profile scenarios to document their generalizability across different predictor contexts.

For the average marginal effects, I again used the results for Equation 13.4 from the original ordinal regression model. Here is the relevant output from that analysis:

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

IMP3       ON
   GA2                0.801      0.107      7.485      0.000
   TA2                0.692      0.119      5.828      0.000
   BD2                0.684      0.114      6.017      0.000
   T                  0.174      0.174      0.997      0.319
   CE1                0.147      0.067      2.198      0.028
   CIS1               0.295      0.106      2.775      0.006

Thresholds
   IMP3$1             1.786      0.795      2.245      0.025
   IMP3$2             3.305      0.805      4.105      0.000
   IMP3$3             5.000      0.811      6.169      0.000
```

These results yield the following three probit subequations:

$$IMP3_{d1} = -1.786 + .801\ GA2 + .692\ TA2 + .684\ BD2 + .174\ T + .147\ CE1 + .295\ CIS1$$

$$IMP3_{d2} = -3.305 + .801\ GA2 + .692\ TA2 + .684\ BD2 + .174\ T + .147\ CE1 + .295\ CIS1$$

$$IMP3_{d3} = -5.000 + .801\ GA2 + .692\ TA2 + .684\ BD2 + .174\ T + .147\ CE1 + .295\ CIS1$$

I use these equations to calculate the AMEs for the mediator contrasts using the method of Cameron and Trivedi. Table A.6 presents the syntax for the GA2 mediator.

Table A.6 Syntax for AME for Effects of Mediators on Outcomes

```
1.  TITLE: AME for effects of GA2 mediator on outcome ;
2.  DATA: FILE IS c:\mplus\symptom.dat ;
3.  DEFINE:
4.   DELTA = SQRT(0.538)/1000 ; !divide SD of GA2 by 1000
5.   PROBIT = -1.786+.801*GA2+.692*TA2+.684*BD2+.174*T+.147*CE1+.295*CIS1;
6.   PROB1C1 = 1-PHI(PROBIT) ;    !base prob for contrast 1
7.   PROBIT = -3.305+.801*GA2+.692*TA2+.684*BD2+.174*T+.147*CE1+.295*CIS1;
8.   PROB1C2 = (1-PHI(PROBIT))-PROB1C1 ;    !base prob for contrast 2
9.   PROBIT = -5.000+.801*GA2+.692*TA2+.684*BD2+.174*T+.147*CE1+.295*CIS1;
10.  PROB1C4 = PHI(PROBIT) ;    !base prob for contrast 4
11.  PROB1C3 = 1-(PROB1C4+PROB1C1+PROB1C2) ;
12.  !Increment for Contrasts
13.  GA2=GA2+DELTA ;
14. !Incremented probabilities for contrasts
15.  PROBIT = -1.786+.801*GA2+.692*TA2+.684*BD2+.174*T+.147*CE1+.295*CIS1;
16.  PROB2C1 = 1-PHI(PROBIT) ;    !inc prob for contrast 1
17.  PROBIT = -3.305+.801*GA2+.692*TA2+.684*BD2+.174*T+.147*CE1+.295*CIS1;
18.  PROB2C2 = (1-PHI(PROBIT))-PROB2C1 ;    !inc prob for contrast 2
19. PROBIT = -5.000+.801*GA2+.692*TA2+.684*BD2+.174*T+.147*CE1+.295*CIS1;
20. PROB2C4 = PHI(PROBIT) ;   !inc prob for contrast 4
21. PROB2C3 = 1-(PROB2C4+PROB2C1+PROB2C2) ;
22. !Calculate individual marginal effects
23. IMEC1 = (PROB2C1-PROB1C1)/DELTA ;   !individ me contrast 1
24. IMEC2 = (PROB2C2-PROB1C2)/DELTA ;   !individ me contrast 2
25. IMEC3 = (PROB2C3-PROB1C3)/DELTA ;   !individ me contrast 3
26. IMEC4 = (PROB2C4-PROB1C4)/DELTA ;   !individ me contrast 4
27. VARIABLE:
28. NAMES ARE ID GA2 TA2 BD2 CE1 CIS1 T IMP3 ;
29. USEVARIABLES ARE IMEC1 IMEC2 IMEC3 IMEC4 ;
30. MISSING ARE ALL (-9999) ;
31. ANALYSIS:
32. ESTIMATOR = ML ;  TYPE = BASIC ;
33. OUTPUT: !use defaults on output
```

For GA2, the average marginal effects for the four contrasts were -0.12, -0.07, 0.07, and 0.12, respectively. The code modifications to analyze TA2 would change Line 4 to use the standard deviation of T2:

```
DELTA = SQRT(0.529)/1000 ;
```

and Line 13 to

```
TA2 = TA2 + delta ;
```

The analysis for BD2 makes comparable changes substituting BD2 for TA2.

Here are the (edited) AME results for GA2, TA2, BD2 and the treatment dummy

variable for the direct effect of treatment condition independent of the mediators when I used the program on my website *AMEs: Ordinal-multinomial* as applied using LISEM for the equation:

Probit(IMP3) = a_4 + p_4 GA2 + p_5 TA2 + p_6 BD2 + p_7 T + b_7 CE1 + b_8 CIS1

```
Average marginal effects

Group  Term Contrast Estimate Std. Error      z Pr(>|z|)    2.5 %   97.5 %
     1 BD2      dY/dX  -0.1046    0.01720 -6.083  < 0.001 -0.13830 -0.07089
     2 BD2      dY/dX  -0.0599    0.01038 -5.767  < 0.001 -0.08022 -0.03953
     3 BD2      dY/dX   0.0632    0.01110  5.697  < 0.001  0.04147  0.08497
     4 BD2      dY/dX   0.1012    0.01629  6.214  < 0.001  0.06931  0.13318
     1 GA2      dY/dX  -0.1224    0.01694 -7.227  < 0.001 -0.15564 -0.08923
     2 GA2      dY/dX  -0.0701    0.01080 -6.493  < 0.001 -0.09125 -0.04893
     3 GA2      dY/dX   0.0740    0.01125  6.580  < 0.001  0.05196  0.09604
     4 GA2      dY/dX   0.1185    0.01626  7.289  < 0.001  0.08666  0.15039
     1 TA2      dY/dX  -0.1058    0.01677 -6.311  < 0.001 -0.13869 -0.07296
     2 TA2      dY/dX  -0.0606    0.01072 -5.652  < 0.001 -0.08159 -0.03957
     3 TA2      dY/dX   0.0640    0.01129  5.663  < 0.001  0.04183  0.08610
     4 TA2      dY/dX   0.1024    0.01601  6.400  < 0.001  0.07107  0.13381
     1 TREAT    1 - 0  -0.0263    0.02535 -1.039  0.29899 -0.07601  0.02336
     2 TREAT    1 - 0  -0.0168    0.01762 -0.951  0.34147 -0.05131  0.01778
     3 TREAT    1 - 0   0.0177    0.01848  0.956  0.33886 -0.01855  0.05390
     4 TREAT    1 - 0   0.0254    0.02448  1.038  0.29919 -0.02256  0.07339
```