eAppendix for "Mediation Analysis with an Ordinal Outcome" by Tyler J. VanderWeele, Yun Zhang, and Pilar Lim

eAppendix 1: Derivations and Proofs.

We will use the notation $A \perp\!\!\!\perp B|C$ to denote that A is independent of B conditional on C. Total effects are identified if, conditional on some set of measured covariates C, the effect of exposure A on outcome Y is unconfounded given C; in counterfactual notation, this is $Y_a \perp\!\!\!\perp A|C$. Controlled direct effects are identified if control is made for a set of covariates C that includes all confounders of not only the exposure-outcome relationship but also the mediator-outcome relationship. In counterfactual notation, we require that for all a and m,

$$Y_{am} \perp\!\!\!\perp A|C \tag{A1}$$

$$Y_{am} \perp\!\!\!\perp M|\{A, C\}. \tag{A2}$$

If this is the case then the controlled direct effect is identified by

$$\frac{P(Y_{am} > j|c)}{P(Y_{am} \leq j|c)} \Big/ \frac{P(Y_{a^*m} > j|c)}{P(Y_{a^*m} \leq j|c)} = \frac{P(Y > j|a, m, c)}{P(Y \leq j|a, m, c)} \Big/ \frac{P(Y > j|a, m, c)}{P(Y \leq j|a, m, c)}$$

since

$$\begin{aligned}
\frac{P(Y_{am} > j|c)}{P(Y_{am} \leq j|c)} \Big/ \frac{P(Y_{a^*m} > j|c)}{P(Y_{a^*m} \leq j|c)} &= \frac{P(Y_{am} > j|a, c)}{P(Y_{am} \leq j|a, c)} \Big/ \frac{P(Y_{a^*m} > j|a^*, c)}{P(Y_{a^*m} \leq j|a^*, c)} \text{ by (A1)} \\
&= \frac{P(Y_{am} > j|a, m, c)}{P(Y_{am} \leq j|a, m, c)} \Big/ \frac{P(Y_{a^*m} > j|a^*, m, c)}{P(Y_{a^*m} \leq j|a^*, m, c)} \text{ by (A2)} \\
&= \frac{P(Y > j|a, m, c)}{P(Y \leq j|a, m, c)} \Big/ \frac{P(Y > j|a^*, m, c)}{P(Y \leq j|a^*, m, c)} \text{ by consistency.}
\end{aligned}$$

Natural direct and indirect effects will be identified if, in addition to assumptions (A1) and (A2), the following two assumptions hold, that for all a, a^* and m,

$$M_a \perp\!\!\!\perp A|C \tag{A3}$$

$$Y_{am} \perp\!\!\!\perp M_{a^*}|C. \tag{A4}$$

Assumption (A3) can be interpreted as: conditional on C, there is no unmeasured confounding of the exposure-mediator relationship. On a causal diagram interpreted as a set of non-parametric structural equations[11], if assumption (A2) holds, then assumption (A4) will hold if there is no variable L that is affected by the exposure A and that itself affects both M and Y.

If assumptions (A1)-(A4) hold, then we have

$$\begin{aligned}
\frac{P(Y_{aM_{a^*}} > j|c)}{P(Y_{aM_{a^*}} \le j|c)} &= \frac{\sum_m P(Y_{am} > j|c, M_{a^*} = m)P(M_{a^*} = m|c)}{\sum_m P(Y_{am} \le j|c, M_{a^*} = m)P(M_{a^*} = m|c)} \text{ by iterated expectations} \\
&= \frac{\sum_m P(Y_{am} > j|c)P(M_{a^*} = m|a^*, c)}{\sum_m P(Y_{am} \le j|c)P(M_{a^*} = m|a^*, c)} \text{ by (A4) and (A3)} \\
&= \frac{\sum_m P(Y_{am} > j|a, c)P(M = m|a^*, c)}{\sum_m P(Y_{am} \le j|, ac)P(M = m|a^*, c)} \text{ by (A1) and consistency} \\
&= \frac{\sum_m P(Y_{am} > j|a, m, c)P(M = m|a^*, c)}{\sum_m P(Y_{am} \le j|, a, m, c)P(M = m|a^*, c)} \text{ by (A2)} \\
&= \frac{\sum_m P(Y > j|a, m, c)P(M = m|a^*, c)}{\sum_m P(Y \le j|a, m, c)P(M = m|a^*, c)} \text{ by consistency.}
\end{aligned}$$

If we apply this result and replace a with a^* we get $\frac{P(Y_{a^*M_{a^*}} > j|c)}{P(Y_{a^*M_{a^*}} \le j|c)} = \frac{\sum_m P(Y > j|a^*, m, c)P(M = m|a^*, c)}{\sum_m P(Y \le j|a^*, m, c)P(M = m|a^*, c)}$ and from this it follows that the natural direct effect is given by:

$$NDE = \frac{\sum_m P(Y > j|a, m, c)P(M = m|a^*, c)}{\sum_m P(Y \le j|a, m, c)P(M = m|a^*, c)} / \frac{\sum_m P(Y > j|a^*, m, c)P(M = m|a^*, c)}{\sum_m P(Y \le j|a^*, m, c)P(M = m|a^*, c)}.$$

If we apply this result and replace a^* with a we get $\frac{P(Y_{aM_a} > j|c)}{P(Y_{aM_a} \le j|c)} = \frac{\sum_m P(Y > j|a, m, c)P(M = m|a, c)}{\sum_m P(Y \le j|a, m, c)P(M = m|a, c)}$ and from this it follows that the natural indirect effect is given by:

$$NIE = \frac{\sum_m P(Y > j|a, m, c)P(M = m|a, c)}{\sum_m P(Y \le j|a, m, c)P(M = m|a, c)} / \frac{\sum_m P(Y > j|a, m, c)P(M = m|a^*, c)}{\sum_m P(Y \le j|a, m, c)P(M = m|a^*, c)}.$$

Suppose that the mediator follows a normal distribution with constant conditional variance σ^2:

$$E[M|a, c] = \beta_0 + \beta_1 a + \beta_2' c$$

The ordinal logistic regression model is:

$$log\{\frac{P(Y\leq j|a,m,c)}{P(Y>j|a,m,c)}\}=\alpha_j-(\theta_1^j a+\theta_2^j m+\theta_3^j am+\theta_4^{j\prime}c)$$

We will consider potential interaction between the exposure and the mediator, but this can of course be dropped by setting $\theta_3^j=0$. The outcome regression model also implies:

$$log\{\frac{P(Y>j|a,m,c)}{P(Y\leq j|a,m,c)}\}=-\alpha_j+(\theta_1^j a+\theta_2^j m+\theta_3^j am+\theta_4^{j\prime}c).$$

Note that if the reference category $J=1$ is sufficiently common (e.g. $P(J=1|a,m,c)>90\%$) then we have the following approximation:

$$\begin{aligned}log\{P(Y>j|a,m,c)\} &\approx \\ log\{\frac{P(Y>j|a,m,c)}{P(Y\leq j|a,m,c)}\} &= -\alpha_j+(\theta_1^j a+\theta_2^j m+\theta_3^j am+\theta_4^{j\prime}c).\end{aligned}$$

Under confounding assumptions (1)-(4) we have that $log\{\frac{P(Y_{aM_{a^*}}>j|c)}{P(Y_{aM_{a^*}}\leq j|c)}\}$

$$\begin{aligned}
&\approx \log\{P(Y_{aM_{a^*}}>j|c)\}\\
&= \log\{\int P(Y_{am}>j|c,M_{a^*}=m)P(M_{a^*}=m|c)dm\}\\
&= \log\{\int P(Y_{am}>j|c)P(M_{a^*}=m|c)dm\} \text{ by (4)}\\
&= \log\{\int P(Y>j|a,m,c)P(M=m|a^*,c)dm\} \text{ by (1)-(3)}\\
&\approx \log\{\int \exp(-\alpha_j+\theta_1^j a+\theta_2^j m+\theta_3^j am+\theta_4^{j\prime}c)P(M=m|a^*,c)dm\}\\
&= \log\{\exp(-\alpha_j+\theta_1^j a+\theta_4^{j\prime}c)\int \exp\{(\theta_2^j+\theta_3^j a)m\}P(M=m|a^*,c)dm\}
\end{aligned}$$

$$\begin{aligned}
&= \log\{\exp(-\alpha_j+\theta_1^j a+\theta_4^{j\prime}c)E[e^{(\theta_2^j+\theta_3^j a)M}|a^*,c]\}\\
&= \log\{\exp(-\alpha_j+\theta_1^j a+\theta_4^{j\prime}c)\exp((\theta_2^j+\theta_3^j a)(\beta_0+\beta_1 a^*+\beta_2' c)+\frac{1}{2}(\theta_2^j+\theta_3^j a)^2\sigma^2)\}\\
&= -\alpha_j+\theta_1^j a+\theta_4^{j\prime}c+(\theta_2^j+\theta_3^j a)(\beta_0+\beta_1 a^*+\beta_2' c)+\frac{1}{2}(\theta_2^j+\theta_3^j a)^2\sigma^2.
\end{aligned}$$

Similarly, we have that $log\{\frac{P(Y_{aM_a}>j|c)}{P(Y_{aM}\leq j|c)}\}$

$$\approx -\alpha_j + \theta_1^j a + \theta_4^{j\prime} c + (\theta_2^j + \theta_3^j a)(\beta_0 + \beta_1 a + \beta_2' c) + \frac{1}{2}(\theta_2^j + \theta_3^j a)^2\sigma^2$$

and $log\{\frac{P(Y_{a^*M_{a^*}}>j|c)}{P(Y_{a^*M_{a^*}}\leq j|c)}\}$

$$\approx -\alpha_j + \theta_1^j a^* + \theta_4^{j\prime} c + (\theta_2^j + \theta_3^j a^*)(\beta_0 + \beta_1 a^* + \beta_2' c) + \frac{1}{2}(\theta_2^j + \theta_3^j a^*)^2\sigma^2$$

Thus for the natural indirect effect odds ratio we have

$$\begin{aligned}
\log\{NIE\} &= \log[\frac{P(Y_{aM_a} > j|c)}{P(Y_{aM_a} \leq j|c)} / \frac{P(Y_{aM_{a^*}} > j|c)}{P(Y_{aM_{a^*}} \leq j|c)}] \\
&= \log[\frac{P(Y_{aM_a} > j|c)}{P(Y_{aM_a} \leq j|c)}] - \log[\frac{P(Y_{aM_{a^*}} > j|c)}{P(Y_{aM_{a^*}} \leq j|c)}] \\
&\approx -\alpha_j + \theta_1^j a + \theta_4^{j\prime} c + (\theta_2^j + \theta_3^j a)(\beta_0 + \beta_1 a + \beta_2' c) + \frac{1}{2}(\theta_2^j + \theta_3^j a)^2\sigma^2 \\
&\quad -\{-\alpha_j + \theta_1^j a + \theta_4^{j\prime} c + (\theta_2^j + \theta_3^j a)(\beta_0 + \beta_1 a^* + \beta_2' c) + \frac{1}{2}(\theta_2^j + \theta_3^j a)^2\sigma^2\} \\
&= (\theta_2^j \beta_1 + \theta_3^j \beta_1 a)(a - a^*).
\end{aligned}$$

Exponentiating the equalities, $NIE \approx \exp\{(\theta_2^j\beta_1 + \theta_3^j\beta_1 a)(a - a^*)\}$.

For the natural direct effect odds ratio, we have that

$$\begin{aligned}
\log\{NDE\} &= \log[\frac{P(Y_{aM_{a^*}} > j|c)}{P(Y_{aM_{a^*}} \leq j|c)} / \frac{P(Y_{a^*M_{a^*}} > j|c)}{P(Y_{a^*M_{a^*}} \leq j|c)}] \\
&= \log[\frac{P(Y_{aM_{a^*}} > j|c)}{P(Y_{aM_{a^*}} \leq j|c)}] - \log[\frac{P(Y_{a^*M_{a^*}} > j|c)}{P(Y_{a^*M_{a^*}} \leq j|c)}] \\
&\approx -\alpha_j + \theta_1^j a + \theta_4^{j\prime} c + (\theta_2^j + \theta_3^j a)(\beta_0 + \beta_1 a^* + \beta_2' c) + \frac{1}{2}(\theta_2^j + \theta_3^j a)^2\sigma^2 \\
&\quad -\{-\alpha_j + \theta_1^j a^* + \theta_4^{j\prime} c + (\theta_2^j + \theta_3^j a^*)(\beta_0 + \beta_1 a^* + \beta_2' c) + \frac{1}{2}(\theta_2^j + \theta_3^j a^*)^2\sigma^2\} \\
&= \{\theta_1^j + \theta_3^j(\beta_0 + \beta_1 a^* + \beta_2' c + \theta_2^j\sigma^2)\}(a - a^*) + 0.5\theta_3^{j2}\sigma^2(a^2 - a^{*2}).
\end{aligned}$$

Exponentiating gives

$$NDE \approx \exp[\{\theta_1^j + \theta_3^j(\beta_0 + \beta_1 a^* + \beta_2' c + \theta_2^j\sigma^2)\}(a - a^*) + 0.5\theta_3^{j2}\sigma^2(a^2 - a^{*2})]. \qquad \text{(A1)}$$

Now suppose that $\theta_3^j = 0$ then the natural indirect effect odds ratio reduces to $NIE \approx \exp\{\theta_2^j\beta_1(a-$

$a^*)\}$ and the natural direct effect odds ratio reduces to $NDE \approx \exp\{\theta_1^j(a-a^*)\}$.

If what is of interest is the controlled direct effect, then if assumptions 1 and 2 hold then we have that $CDE(m) =$

$$\frac{P(Y_{am}>j|c)}{P(Y_{am}\le j|c)}/\frac{P(Y_{a^*m}>j|c)}{P(Y_{a^*m}\le j|c)} = \frac{P(Y_{am}>j|a,m,c)}{P(Y_{am}\le j|a,m,c)}/\frac{P(Y_{a^*m}>j|a,m,c)}{P(Y_{a^*m}\le j|a,m,c)}$$

Under the ordinal logistic regression model we have that $CDE(m) =$

$$\begin{aligned} & \frac{P(Y_{am}>j|a,m,c)}{P(Y_{am}\le j|a,m,c)}/\frac{P(Y_{a^*m}>j|a,m,c)}{P(Y_{a^*m}\le j|a,m,c)} \\ = & \frac{\exp\{-\alpha_j+\theta_1^j a+\theta_2^j m+\theta_3^j am+\theta_4^{j\prime}c\}}{\exp\{-\alpha_j+\theta_1^j a^*+\theta_2^j m+\theta_3^j a^*m+\theta_4^{j\prime}c\}} \\ = & \exp\{(\theta_1^j+\theta_3^j m)(a-a^*)\}. \end{aligned}$$

Now suppose that we have a proportional odds model with $\theta_1^j, \theta_2^j, \theta_3^j, \theta_4^j$ all constant across j so that

$$log\{\frac{P(Y\le j|a,m,c)}{P(Y>j|a,m,c)}\} = \alpha_j - (\theta_1 a+\theta_2 m+\theta_3 am+\theta_4' c).$$

The expressions above then simplify to:

$$\begin{aligned} NIE &\approx \exp\{(\theta_2\beta_1+\theta_3\beta_1 a)(a-a^*)\} \\ NDE &\approx \exp[\{\theta_1+\theta_3(\beta_0+\beta_1 a^*+\beta_2' c+\theta_2\sigma^2)\}(a-a^*)+0.5\theta_3^2\sigma^2(a^2-a^{*2})] \\ CDE(m) &= \exp\{(\theta_1+\theta_3 m)(a-a^*)\} \end{aligned}$$

We now consider standard errors for the controlled direct effect and natural direct and indirect effect log odds ratios. Suppose that the ordinal logistic regression model has been fit using standard software and that the linear regression model has likewise been fit using standard software. Let $\theta_0 = (\alpha_1, ..., \alpha_J)$ Suppose furthermore that the resulting estimates $\hat{\beta}$ of $\beta \equiv (\beta_0, \beta_1, \beta_2')'$, $\hat{\theta}$ of

$\theta \equiv (\theta_0, \theta_1, \theta_2, \theta_3, \theta_4')'$ and $\hat{\sigma}^2$ of σ^2 have covariance matrices

$$\Sigma_\beta = \begin{pmatrix} \sigma_{00}^\beta & \sigma_{01}^\beta & \sigma_{02}^\beta \\ \sigma_{10}^\beta & \sigma_{11}^\beta & \sigma_{12}^\beta \\ \sigma_{20}^\beta & \sigma_{21}^\beta & \sigma_{22}^\beta \end{pmatrix},$$

$$\Sigma_\theta = \begin{pmatrix} \sigma_{00}^\theta & \sigma_{01}^\theta & \sigma_{02}^\theta & \sigma_{03}^\theta & \sigma_{04}^\theta \\ \sigma_{10}^\theta & \sigma_{11}^\theta & \sigma_{12}^\theta & \sigma_{13}^\theta & \sigma_{14}^\theta \\ \sigma_{20}^\theta & \sigma_{21}^\theta & \sigma_{22}^\theta & \sigma_{23}^\theta & \sigma_{24}^\theta \\ \sigma_{30}^\theta & \sigma_{31}^\theta & \sigma_{32}^\theta & \sigma_{33}^\theta & \sigma_{34}^\theta \\ \sigma_{40}^\theta & \sigma_{41}^\theta & \sigma_{42}^\theta & \sigma_{43}^\theta & \sigma_{44}^\theta \end{pmatrix}$$

and

$$\Sigma_{\sigma^2} = (\sigma_{11}^{\sigma^2}),$$

respectively, where σ_{ij}^β is the covariance between $\hat{\beta}_i$ and $\hat{\beta}_j$ and σ_{ij}^θ is the covariance between $\hat{\theta}_i$ and $\hat{\theta}_j$. These matrices can be obtained from standard statistical software packages. Under a linear regression for the mediator, let RSS denote the residual sum of square; an unbiased estimate of $\hat{\sigma}^2$ is given by $RSS/(n-p)$ where n is the sample size and p is the number of parameters in the regression model; the variance of $\hat{\sigma}^2$ can be estimated by $\frac{2\hat{\sigma}^4}{n-p}$. Then standard errors of the log of the controlled direct effect odds ratios and natural direct and indirect effect effects can be obtained (using the Delta method) as

$$\sqrt{\Gamma\Sigma\Gamma'}|a - a^*| \tag{A2}$$

with

$$\Sigma \equiv \begin{pmatrix} \Sigma_\beta & 0 & 0 \\ 0 & \Sigma_\theta & 0 \\ 0 & 0 & \Sigma_{\sigma^2} \end{pmatrix}$$

and with $\Gamma \equiv (0, 0, 0', 0', 1, 0, m, 0', 0)$ for the log of the controlled direct effect odds ratio, $\Gamma \equiv (0, \theta_2 + \theta_3 a, 0', 0', 0, \beta_1, \beta_1 a, 0', 0)$ for the log of the natural indirect effect odds ratio, and $\Gamma \equiv (\theta_3, \theta_3 a^*, \theta_3 c, 0', 1, \theta_3\sigma^2, \beta_0 + \beta_1 a^* + \beta_2' c + \theta_2\sigma^2 + \theta_3\sigma^2(a + a^*), 0', \theta_3\theta_2 + 0.5\theta_3^2(a + a^*))$ for the log of the natural direct effect odds ratio. In these expressions, $0'$ denotes a row vector of the dimension of c or $\theta_0 = (\alpha_1, ..., \alpha_J)$ containing zeroes only. Once a confidence interval has been obtained for the log controlled direct effect odds ratio and the log natural direct/indirect effect odds ratio as given, then confidence intervals for the controlled direct effect odds ratio and the natural indirect and direct effects can be obtained by simply exponentiating the confidence intervals in for the log of these quantities.

The effects above are sometimes referred to as "pure" (natural) direct effects and "total" natural indirect effects.[9] These are $\frac{P(Y_{aM_{a^*}}>j|c)}{P(Y_{aM_{a^*}}\leq j|c)}/\frac{P(Y_{a^*M_{a^*}}>j|c)}{P(Y_{a^*M_{a^*}}\leq j|c)}$ and $\frac{P(Y_{aM_a}>j|c)}{P(Y_{aM}\leq j|c)}/\frac{P(Y_{aM_{a^*}}>j|c)}{P(Y_{aM_{a^*}}\leq j|c)}$. Instead we could consider the effects $\frac{P(Y_{aM_a}>j|c)}{P(Y_{aM_a}\leq j|c)}/\frac{P(Y_{a^*M_a}>j|c)}{P(Y_{a^*M_a}\leq j|c)}$ and $\frac{P(Y_{a^*M_a}>j|c)}{P(Y_{a^*M_a}\leq j|c)}/\frac{P(Y_{a^*M_{a^*}}>j|c)}{P(Y_{a^*M_{a^*}}\leq j|c)}$. These are what Robins and Greenland[9] refer to as the "total" (natural) direct effects and "pure" natural indirect effects. The total direct effect and pure indirect effect likewise multiply to the total effect. These two effects differ from the effects primarily considered here in the way that they account for mediated interaction[1,8.13], with "total" denoting which of the effects (direct or indirect) picks up the mediated interaction and "pure" denoting the effect that does not. Likewise the natural direct and indirect effects on the difference scale that were considered in the text: $P(Y_{aM_{a^*}}=j|c)-P(Y_{a^*M_{a^*}}=j|c)$ for the natural direct effect and $P(Y_{aM_a}=j|c)-P(Y_{aM_{a^*}}=j|c)$ for the natural indirect effect could be referred to as pure direct effect and the total indirect effect. Once again we could alternatively consider a "total direct effect", $P(Y_{aM_a}=j|c)-P(Y_{a^*M_a}=j|c)$, and a pure indirect effect, $P(Y_{a^*M_a}=j|c)-P(Y_{a^*M_{a^*}}=j|c)$ and the two of these effect sum to the total effect, $P(Y_a=j|c)-P(Y_{a^*}=j|c)$. This total direct effect and pure indirect effect on the difference are also reported by Imai et al.[5,6] software in addition to the pure direct effect and total indirect effect.

We note that with a nominal categorical outcome ($j=1,...K$) under a multinomial logistic regression:

$$log\{\frac{P(Y=j|a,m,c)}{P(Y=1|a,m,c)}\}=\theta_0^j+\theta_1^j a+\theta_2^j m+\theta_3^j am+\theta_4^{j\prime}c$$

with a reference category ($J=1$) that is relatively common (e.g. $>90\%$) and a normally distributed mediator with constant conditional variance σ^2, it is straightforward to show that the analytic expressions for natural direct and indirect effect odds ratios given in VanderWeele and Vansteelandt[3] are applicable also to the multinomial logistic regression but one would have a different natural direct and indirect effect odds ratio, $\frac{P(Y_{aM_{a^*}}=j|c)}{P(Y_{aM_{a^*}}=1|c)}/\frac{P(Y_{a^*M_{a^*}}=j|c)}{P(Y_{a^*M_{a^*}}=1|c)}$ and $\frac{P(Y_{aM_a}=j|c)}{P(Y_{aM_a}=1|c)}/\frac{P(Y_{aM_{a^*}}=j|c)}{P(Y_{aM_{a^*}}=1|c)}$, for each of the $j=2,...,K$ categories compared with the reference category $j=1$.

eAppendix 2. Illustration

The methods in this paper were developed to assess the direct and indirect effects of an actual trial on a suicide risk reduction intervention; the parameters in the simulated illustration were set to approximately correspond to anticipated distributions and effect sizes. Treatment of suicidal ideation with major depressive disorder (MDD) represents an unmet medical need. Finding treatments to reduce the imminent risk of suicide in patients with MDD is important for clinical practice. In conducting clinical research in this patient population, there are challenges to appropriately measuring improvement in suicidal ideation risk. It is difficult to differentiate how much of the improvement is due to the nonspecific improvement in symptoms of depression and how much of the improvement is an independent effect of treatment on suicide risk. There is then interest in examining the mediating role of depressive symptoms on assessment of suicide risk.

A simulated illustration was performed using the following specifications. Subjects were randomly assigned to receive an active treatment or placebo. The subject's suicide risk and depressive symptoms were measured before dosing and after dosing, using a 5-item assessment of suicide risk rating scale and a depression continuous scale, respectively. Both scales were scored such that the higher the score, the worse the symptoms. A mediation analysis was then conducted to assess the impact of the effect of treatment on the change from baseline in the assessment of suicide risk score above and beyond the effect of treatment on the change from baseline to end point in the depression scale total score, the primary efficacy endpoint. Since the change from baseline in the assessment of suicide risk is an ordinal variable, we propose a mediation model that assesses the degree of treatment effect upon an ordinal response variable in the presence of another variable (i.e. the mediating variable).

Parameters for the simulated illustration have been set to correspond to what we might expect from the suicide study: distribution of baseline assessment of suicide risk, change from baseline in depression rating scale, and change from baseline in assessment of suicide risk. A sample size of 300 was employed. The distribution of baseline assessment of suicide risk corresponding to levels $1-5$ were set at 0.0526, 0.0526, 0.0526, 0.474, and 0.368, respectively. The frequency of the outcome (i.e. change in assessment of suicide risk) was set for levels $-4, ..., 0$ as $14, 30, 21, 20, 215$ respectively. The simulation proceeded as follows:

Step 1: Generate C (baseline assessment of suicide risk as covariate) from the multinomial distribution with the probability shown above.

Step 2: Randomly assign subjects to treatment group (A =1 or A=0), then generate mediator (change in depression rating scale) from the model $E[M|a,c] = \beta_0 + \beta_1 a + \beta_2' c$ with $\beta_0 = -9, \beta_1 = -6, \beta_2 = -0.1$ with standard deviation or the residual error equal to $\sigma = 10$

Step 3: Generate the outcome from the model $log\{P(Y \leq j|a,m,c)/P(Y > j|a,m,c)\} = \alpha_j - (\theta_1 a + \theta_2 m + \theta_3 am + \theta_4' c)$ with $\theta_1 = -0.5, \theta_2 = 0.1, \theta_3 = 0, \theta_4 = -0.1, \alpha_1 = -5.1, \alpha_2 = -3.8, \alpha_3 = -3.3, \alpha_4 = -2.9$.

The proportional odds model without an interaction term was fit to the simulated data and direct and indirect effects were estimated using the simulation based approach described above. The results for the estimates of the natural direct and indirect effects and the total effect on the difference scale, along with 95% confidence intervals, for each category are given in Table 1. The omnibus tests for any mediation had p-value < 0.01. The simulated results would suggest that active treatment has a clinically meaningful independent effect on the assessment of suicide risk that extends beyond its effect on depressive symptoms, but that some of the effect is indeed mediated by depressive symptoms. For all five outcome levels roughly a half (and always between a third and two third) of the effect appears to be mediated (Table 1).

Table 1. Direct, indirect and total effects on the difference scale with 95% confidence intervals for simulated suicide risk reduction example

	$P(Y = 0)$	$P(Y = -1)$	$P(Y = -2)$	$P(Y = -3)$	$P(Y = -4)$
Indirect Effect	$-.10\ (-.17, -.06)$	0.03 (.01, .05)	0.05 (.02, .07)	0.02 (.01, .03)	0.01 (.00, .02)
Direct Effect	$-.12\ (-.22, -.03)$	0.02 (.00, .05)	0.05 (.01, .09)	0.03 (.01, .06)	0.02 (.00, .05)
Total Effect	$-.22\ (-.34, -.13)$	0.05 (.02, .09)	0.09 (.05, .14)	0.05 (.02, .08)	0.03 (.01, .07)

eAppendix 3: Simulation-Based Code and Example

We generate data from a linear regression model

$$E[M|a,c] = \beta_0 + \beta_1 a + \beta_2' c.$$

with $\beta_0 = 0, \beta_1 = -0.33, \beta_2 = 0.15$ and from a proportional odds model

$$log\{\frac{P(Y \leq j|a,m,c)}{P(Y > j|a,m,c)}\} = \alpha_j - (\theta_1 a + \theta_2 m + \theta_3 am + \theta_4' c)$$

with $\alpha_1 = -0.5, \alpha_2 = 0, \alpha_3 = 0.5, \alpha_4 = 1, \alpha_5 = 1.3, \alpha_6 = 2$ and $\theta_1 = -0.5, \theta_2 = 0.2, \theta_3 = 0, \theta_4 = 0.2$. We simulate 500 observations. This can be done with the following code in R:

```
Nn <- 500
b1 <- -0.33
b2 <- 0.15
t1 <- -0.5
t2 <- 0.2
t4 <- 0.2
a1 <- -.5
a2 <- 0
a3 <- 0.5
a4 <- 1
a5 <- 1.3
a6 <- 2
a <- rbinom(Nn,1,0.5)
c <- rnorm(Nn)
m <- rnorm(Nn) + a*b1 + c*b2
p1 <- exp(a1 - (t1*a + t2*m+t4*c)) / (1 + exp(a1 - (t1*a + t2*m+t4*c)) )
p2 <- exp(a2 - (t1*a + t2*m+t4*c)) / (1 + exp(a2 - (t1*a + t2*m+t4*c)) ) - exp(a1 - (t1*a + t2*m+t4*c)) / (1 + exp(a1 - (t1*a + t2*m+t4*c)) )
p3 <- exp(a3 - (t1*a + t2*m+t4*c)) / (1 + exp(a3 - (t1*a + t2*m+t4*c)) ) - exp(a2 - (t1*a + t2*m+t4*c)) / (1 + exp(a2 - (t1*a + t2*m+t4*c)) )
p4 <- exp(a4 - (t1*a + t2*m+t4*c)) / (1 + exp(a4 - (t1*a + t2*m+t4*c)) ) - exp(a3 - (t1*a + t2*m+t4*c)) / (1 + exp(a3 - (t1*a + t2*m+t4*c)) )
p5 <- exp(a5 - (t1*a + t2*m+t4*c)) / (1 + exp(a5 - (t1*a + t2*m+t4*c)) ) - exp(a4 - (t1*a + t2*m+t4*c)) / (1 + exp(a4 - (t1*a + t2*m+t4*c)) )
p6 <- exp(a6 - (t1*a + t2*m+t4*c)) / (1 + exp(a6 - (t1*a + t2*m+t4*c)) ) - exp(a5 - (t1*a + t2*m+t4*c)) / (1 + exp(a5 - (t1*a + t2*m+t4*c)) )
p7 <- 1 - exp(a6 - (t1*a + t2*m+t4*c)) / (1 + exp(a6 - (t1*a + t2*m+t4*c)) )
rrr <- runif(Nn)
```

```
y <- 1 + (rrr > p1)
y <- y + (rrr > (p1+p2))
y <- y + (rrr > (p1+p2+p3))
y <- y + (rrr > (p1+p2+p3+p4))
y <- y + (rrr > (p1+p2+p3+p4+p5))
y <- y + (rrr > (p1+p2+p3+p4+p5+p6))
```

To carry out the simulation-based approach, Imai et al.'s software can be loaded using the commands:

```
require(MASS)
install.packages("mediation")
library(mediation)
library(sandwich)
```

The following code will then carry out the mediation analysis for each outcome value j: the total effect, $P(Y_a = j|c) - P(Y_{a^*} = j|c)$; the natural direct effect $P(Y_{aM_{a^*}} = j|c) - P(Y_{a^*M_{a^*}} = j|c)$; and the natural indirect effect $P(Y_{aM_a} = j|c) - P(Y_{aM_{a^*}} = j|c)$.

```
fy <- factor(y)
mydata <- data.frame(fy, y, a, m)
med.fit <- lm(m ~a, data = mydata)
out.fit <- polr(fy ~m + a, data = mydata)
med.out <- mediate(med.fit, out.fit, treat = "a", mediator = "m", boot=TRUE, sims=1000)
summary(med.out)
```

The relevant output is then "ACME (treated)" as this is what we have been referring to as the natural indirect effects and "ADE (control)" as this is what we have been referring to as the natural direct effect, and the "Total effect", as discussed further in eAppendix 1 above. In this example, this output for one random sample of size 500 is:

```
                Pr(Y=1) Pr(Y=2) Pr(Y=3) Pr(Y=4) Pr(Y=5) Pr(Y=6) Pr(Y=7)
ACME (treated) 0.01924 -1.27e-03 -0.002938 -0.004168 -0.002036 -0.003594 -0.005231
```

```
2.5%           0.00378 -2.97e-03 -0.005747 -0.008402 -0.004226 -0.007548 -0.010679
97.5%          0.03671 -5.85e-05 -0.000524 -0.000833 -0.000396 -0.000751 -0.000956
p-value        0.01200 3.60e-02 0.012000 0.012000 0.012000 0.012000 0.012000
```

```
              Pr(Y=1) Pr(Y=2) Pr(Y=3) Pr(Y=4) Pr(Y=5) Pr(Y=6) Pr(Y=7)
ADE (control) 0.1189 0.000654 -0.0130 -0.0247 -0.0136 -0.02612 -0.0422
2.5%          0.0459 -0.005040 -0.0250 -0.0428 -0.0244 -0.04705 -0.0750
97.5%         0.1949 0.007077 -0.0046 -0.0094 -0.0049 -0.00928 -0.0156
p-value       0.0020 0.790000 0.0020 0.0020 0.0020 0.00200 0.0020
```

```
             Pr(Y=1) Pr(Y=2) Pr(Y=3) Pr(Y=4) Pr(Y=5) Pr(Y=6) Pr(Y=7)
Total Effect 0.1381 -0.000617 -0.01591 -0.0288 -0.01568 -0.0297 -0.0474
2.5%         0.0634 -0.006993 -0.02853 -0.0466 -0.02666 -0.0512 -0.0795
97.5%        0.2102 0.006110 -0.00668 -0.0132 -0.00658 -0.0124 -0.0208
p-value      0.0000 0.918000 0.00000 0.0000 0.00000 0.0000 0.0000
```

Because whether the direct and indirect effects are positive or negative often just depends on the signs of the coefficients, and since the p-values are computed by non-parametric bootstrap, many of the p-values numerically coincide.

Note that the effects themselves will vary across categories j. It may be the case that some of the direct and indirect effects for certain categories j are non-zero and some are not. It is possible to construct an omnibus test as to whether any of the direct effects are non-zero, or whether any of the indirect effects are non-zero. This can be done as follows. For example, for the indirect effect let I_j denote the estimate of the natural indirect effect $P(Y_{aM_a} = j|c) - P(Y_{aM_{a^*}} = j|c)$ for category j. Calculate an overall measure of departure from the null of no effect for any of the J outcome levels as $Q = I_1^2 + I_2^2 + ... + I_J^2$. Then create some number K (e.g. $K = 5000$) bootstrapped samples from the original data. For each bootstrapped sample k, estimate the mediated effects for this sample and call these $I_{k1}, I_{k2}, ..., I_{kJ}$. If we then consider $(I_{k1} - I_1), ..., (I_{kJ} - I_J)$ each of these will have mean 0 across repeated bootstrapped samples. Our departure measure for the kth sample is $Q_k = (I_{k1} - I_1)^2 + (I_{k2} - I_2)^2 + ... + (I_{kJ} - I_J)^2$. We can calculate Q_k for each of the $K = 5000$

samples. The p-value is then the proportion of the K samples for which Q_k is greater than the actual observed $Q = I_1^2 + I_2^2 + ... + I_J^2$ from the actual data. The same procedure could be used for the natural direct effect and the total effect.