eAppendix for "A unification of mediation and interaction: a four-way decomposition" by Tyler J. VanderWeele

1. Continuous Outcomes and Linear Regression Models

1.1 Continuous Outcome, Continuous Mediator

For Y and M continuous, under assumptions (i)-(iv) and correct specification of the regression models for Y and M:

$$\begin{aligned} E[Y|a,m,c] &= \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_4' c \\ E[M|a,c] &= \beta_0 + \beta_1 a + \beta_2' c, \end{aligned}$$

VanderWeele and Vansteelandt[4] and VanderWeele[34] showed that the average controlled direct effect, the pure indirect effect, and the mediated interaction conditional on covariates $C = c$ were given by:

$$\begin{aligned} E[CDE(m^*)|c] &= (\theta_1 + \theta_3 m^*)(a - a^*) \\ E[PIE|c] &= (\theta_2\beta_1 + \theta_3\beta_1 a^*)(a - a^*) \\ E[INT_{med}|c] &= \theta_3\beta_1(a - a^*)(a - a^*). \end{aligned}$$

They also showed that the pure direct effect was given by $E[PDE|c] = \{\theta_1 + \theta_3(\beta_0 + \beta_1 a^* + \beta_2' c)\}(a - a^*)$. The reference interaction is then given by difference between the the pure direct effect and the controlled direct effect:

$$\begin{aligned} E[INT_{ref}(m^*)|c] &= \{\theta_1 + \theta_3(\beta_0 + \beta_1 a^* + \beta_2' c)\}(a - a^*) - (\theta_1 + \theta_3 m^*)(a - a^*) \\ &= \theta_3(\beta_0 + \beta_1 a^* + \beta_2' c - m^*)\}(a - a^*). \end{aligned}$$

Standard errors for these expressions could be derived using the delta method along the lines of the derivations in VanderWeele and Vansteelandt[4] or by using bootstrapping.

1.2 Continuous Outcome, Binary Mediator

For Y continuous and M binary, under assumptions (i)-(iv) and correct specification of the regression models for Y and M:

$$\begin{aligned} E[Y|a,m,c] &= \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_4' c \\ logit\{P(M=1|a,c)\} &= \beta_0 + \beta_1 a + \beta_2' c. \end{aligned}$$

Valeri and VanderWeele[16] show that the average controlled direct effect and the average pure indirect effect are given by:

$$\begin{aligned} E[CDE(m^*)|c] &= (\theta_1 + \theta_3 m^*)(a - a^*) \\ E[PIE|c] &= (\theta_2 + \theta_3 a^*)\{\frac{exp[\beta_0 + \beta_1 a + \beta_2^{'} c]}{1 + exp[\beta_0 + \beta_1 a + \beta_2^{'} c]} - \frac{exp[\beta_0 + \beta_1 a^* + \beta_2^{'} c]}{1 + exp[\beta_0 + \beta_1 a^* + \beta_2^{'} c]}\}. \end{aligned}$$

The reference interaction is given by the difference between the pure direct effect and the controlled direct effect, which were both given by Valeri and VanderWeele[16]:

$$\begin{aligned} E[INT_{ref}(m^*)|c] &= \{\theta_1(a-a^*)\} + \{\theta_3(a-a^*)\}\frac{exp[\beta_0+\beta_1 a^* + \beta_2^{'}c]}{1+exp[\beta_0+\beta_1 a^* + \beta_2^{'}c]} - (\theta_1+\theta_3 m^*)(a-a^*) \\ &= \theta_3(a-a^*)\left(\frac{exp[\beta_0+\beta_1 a^* + \beta_2^{'}c]}{1+exp[\beta_0+\beta_1 a^* + \beta_2^{'}c]} - m^*\right) \end{aligned}$$

The mediated interaction is given by the difference between the total indirect effect and the pure indirect effect, which were also both given by Valeri and VanderWeele[16]:

$$\begin{aligned} E[INT_{med}|c] &= (\theta_2+\theta_3 a)\{\frac{exp[\beta_0+\beta_1 a + \beta_2^{'}c]}{1+exp[\beta_0+\beta_1 a + \beta_2^{'}c]} - \frac{exp[\beta_0+\beta_1 a^* + \beta_2^{'}c]}{1+exp[\beta_0+\beta_1 a^* + \beta_2^{'}c]}\}. \\ &\quad -(\theta_2+\theta_3 a^*)\{\frac{exp[\beta_0+\beta_1 a + \beta_2^{'}c]}{1+exp[\beta_0+\beta_1 a + \beta_2^{'}c]} - \frac{exp[\beta_0+\beta_1 a^* + \beta_2^{'}c]}{1+exp[\beta_0+\beta_1 a^* + \beta_2^{'}c]}\} \\ &= \theta_3(a-a^*)\{\frac{exp[\beta_0+\beta_1 a + \beta_2^{'}c]}{1+exp[\beta_0+\beta_1 a + \beta_2^{'}c]} - \frac{exp[\beta_0+\beta_1 a^* + \beta_2^{'}c]}{1+exp[\beta_0+\beta_1 a^* + \beta_2^{'}c]}\}. \end{aligned}$$

2. Decomposition on a Ratio Scale and Logistic Regression Models

2.1. Four-way Decomposition on a Ratio Scale

From Proposition 1 in the text we have $Y_a - Y_{a^*}$

$$\begin{aligned} &= (Y_{am^*} - Y_{a^*m^*}) + \sum\nolimits_m (Y_{am} - Y_{a^*m} - Y_{am^*} + Y_{a^*m^*})1(M_{a^*} = m) \\ &\quad + \sum\nolimits_m (Y_{am} - Y_{a^*m})\{1(M_a = m) - 1(M_{a^*} = m)\} + (Y_{a^*M_a} - Y_{a^*M_{a^*}}). \end{aligned}$$

Taking expectations conditional on $C = c$ gives: $E(Y_a - Y_{a^*}|c)$

$$\begin{aligned} &= E(Y_{am^*} - Y_{a^*m^*}|c) + \sum\nolimits_m E[(Y_{am} - Y_{a^*m} - Y_{am^*} + Y_{a^*m^*})1(M_{a^*} = m)|c] \\ &\quad + \sum\nolimits_m E[(Y_{am} - Y_{a^*m})\{1(M_a = m) - 1(M_{a^*} = m)\}|c] + E(Y_{a^*M_a} - Y_{a^*M_{a^*}}|c). \end{aligned}$$

Under assumption (iv) this is: $E(Y_a - Y_{a^*}|c)$

$$\begin{aligned} &= E(Y_{am^*} - Y_{a^*m^*}|c) + \sum\nolimits_m E(Y_{am} - Y_{a^*m} - Y_{am^*} + Y_{a^*m^*}|c)P(M_{a^*} = m|c) \\ &\quad + \sum\nolimits_m E(Y_{am} - Y_{a^*m}|c)\{P(M_a = m|c) - P(M_{a^*} = m|c)\} + E(Y_{a^*M_a} - Y_{a^*M_{a^*}}|c). \end{aligned}$$

and dividing by $E(Y_{a^*}|c)$ gives:

$$RR_c^{TE} - 1 = \kappa\left[RR_c^{CDE}(m^*) - 1\right] + \kappa RR_c^{INT_{ref}}(m^*) + \kappa RR_c^{INT_{med}} + (RR_c^{PIE} - 1)$$

where $RR_c^{TE} = \frac{E(Y_a|c)}{E(Y_{a^*}|c)}$, $\kappa = \frac{E(Y_{a^*m^*}|c)}{E(Y_{a^*}|c)}$, and

$$\begin{aligned}
RR_c^{CDE}(m^*) &= \frac{E(Y_{am^*}|c)}{E(Y_{a^*m^*}|c)} \\
RR_c^{INT_{ref}}(m^*) &= \sum\nolimits_m RERI(a^*, m^*)P(M_{a^*} = m|c) \\
RR_c^{INT_{med}} &= \sum\nolimits_m RERI(a^*, m^*)\{P(M_a = m|c) - P(M_{a^*} = m|c)\} \\
RR_c^{PIE} &= \frac{E(Y_{a^*M_a}|c)}{E(Y_{a^*M_{a^*}}|c)}
\end{aligned}$$

with $RERI(a^*, m^*) = \left(\frac{E(Y_{am}|c)}{E(Y_{a^*m^*}|c)} - \frac{E(Y_{a^*m}|c)}{E(Y_{a^*m^*}|c)} - \frac{E(Y_{am^*}|c)}{E(Y_{a^*m^*}|c)} + 1\right)$. Under assumptions (i)-(iii) we also have $E(Y_a|c) = E(Y|a, c)$, $E(Y_{am}|c) = \sum_m E[Y|a, m, c]P(m|a, c)$ and thus and $P(M_a = m|c) = P(M = m|a, c)$ and thus the right hand side of the equalities above would be identified from the data. VanderWeele[34] also showed that $\kappa RR_c^{INT_{med}} = \kappa \sum_m RERI(a^*, m^*)\{P(M_a = m|c) - P(M_{a^*} = m|c)\} = \left(\frac{E[Y_{aM_a}|c]}{E[Y_{a^*M_{a^*}}|c]} - \frac{E[Y_{aM_{a^*}}|c]}{E[Y_{a^*M_{a^*}}|c]} - \frac{E[Y_{a^*M_a}|c]}{E[Y_{a^*M_{a^*}}|c]} + 1\right)$ and called this latter term $RERI_{mediated}$.

Note also under assumption (iv), $(RR_c^{PIE} - 1)$ can be rewritten as

$$\begin{aligned}
(RR_c^{PIE} - 1) &= \left(\frac{E(Y_{a^*M_a}|c)}{E(Y_{a^*}|c)} - \frac{E(Y_{a^*}|c)}{E(Y_{a^*}|c)}\right) = \frac{\kappa}{E(Y_{a^*m^*}|c)}\{E(Y_{a^*M_a}|c) - E(Y_{a^*}|c)\} \\
&= \frac{\kappa}{E(Y_{a^*m^*}|c)} \sum\nolimits_m \{E[Y_{a^*m}|c] - E[Y_{a^*m^*}|c]\}\{P(M_a = m|c) - P(M_{a^*} = m|c)\} \\
&= \kappa \sum\nolimits_m \left(\frac{E(Y_{a^*m}|c)}{E(Y_{a^*m^*}|c)} - 1\right)\{P(M_a = m|c) - P(M_{a^*} = m|c)\} \\
&= \kappa \sum\nolimits_m \frac{E(Y_{a^*m}|c)}{E(Y_{a^*m^*}|c)}\{P(M_a = m|c) - P(M_{a^*} = m|c)\}
\end{aligned}$$

The proportion attributable to each of the four components is then obtained by simply dividing each of the four components in the display equation above by their sum as in Table 2. A similar decomposition could likewise be carried out on an additive scale using hazard ratios.

By similar arguments to those above but applied to Propositions 2 and 4, if assumption (iv) did not hold but assumptions (i)-(iii) all did hold, we would have that $(RR_c^{TE} - 1)$

decomposed into the product of κ and the sum of:

$$RR_c^{CDE}(m^*) - 1 = \frac{E[Y|a, m^*, c]}{E[Y|a^*, m^*, c]} - 1$$

$$\int RERI(a^*, m^*) dP(M_{a^*}|c)$$

$$= \int \{\frac{E[Y|a, m, c]}{E[Y|a^*, m^*, c]} - \frac{E[Y|a^*, m, c]}{E[Y|a^*, m^*, c]} - \frac{E[Y|a, m^*, c]}{E[Y|a^*, m^*, c]} + 1\} dP(m|a^*, c)$$

$$\int RERI(a^*, m^*)\{dP(M_a|c) - dP(M_{a^*}|c)$$

$$= \int \{\frac{E[Y|a, m, c]}{E[Y|a^*, m^*, c]} - \frac{E[Y|a^*, m, c]}{E[Y|a^*, m^*, c]}\}\{dP(m|a, c) - dP(m|a^*, c)\}$$

$$\int \frac{E[Y_{a^*m}|c]}{E[Y_{a^*m^*}|c]}\{dP(M_a|c) - dP(M_{a^*}|c)\} = \int \frac{E[Y|a^*, m, c]}{E[Y|a^*, m^*, c]}\{dP(m|a, c) - dP(m|a^*, c)\}.$$

2.2 Binary Outcome, Continuous Mediator

Suppose Y were binary and M continuous, that assumptions (i)-(iv) held, that the outcome is rare, and that the following regressions were correctly specified:

$$\begin{aligned} logit(P(Y = 1|a, m, c)) &= \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_4' c \\ E[M|a, c] &= \beta_0 + \beta_1 a + \beta_2' c. \end{aligned}$$

with M normally distribution conditional on (A, C) with variance σ^2. Suppose that the outcome is rare so that odds ratios approximate risk ratios. VanderWeele and Vansteelandt[5] derived expressions for the controlled direct effect, the pure indirect effect, and the pure direct effect, all on the risk ratio scale. The total effect, controlled direct effect, and pure indirect effect were given approximately by:

$$\begin{aligned} RR_c^{TE} &\approx \exp[\theta_1 + \theta_2\beta_1 + \theta_3(\beta_0 + \beta_1 a^* + \beta_1 a + \beta_2' c + \theta_2\sigma^2)\}(a - a^*) + \frac{1}{2}\theta_3^2\sigma^2(a^2 - a^{*2})] \\ RR_c^{CDE}(m^*) &\approx \exp[(\theta_1 + \theta_3 m^*)(a - a^*)] \\ RR_c^{PIE} &\approx \exp[(\theta_2\beta_1 + \theta_3\beta_1 a^*)(a - a^*)] \end{aligned}$$

where the approximations (here and below) hold to the extent that the outcome is rare. We

have that $\kappa = \frac{E(Y_{a^*m^*}|c)}{E(Y_{a^*}|c)}$ is given by:

$$\begin{aligned}
\kappa &= \frac{E(Y_{a^*m^*}|c)}{E(Y_{a^*}|c)} = \frac{E[Y|a^*,m^*,c]}{\int E[Y|a^*,m,c]dP(m|a^*,c)} \\
&\approx \frac{\exp(\theta_0+\theta_1 a^*+\theta_2 m^*+\theta_3 a^* m^*+\theta_4' c)}{\exp\{\theta_0+\theta_1 a^*+\theta_4' c\}\int \exp\{(\theta_2+\theta_3 a^*)m\}dP(m|a^*,c)} \\
&= \frac{\exp(\theta_2 m^*+\theta_3 a^* m^*)}{\exp\{(\theta_2+\theta_3 a^*)(\beta_0+\beta_1 a^*+\beta_2' c)+\frac{1}{2}(\theta_2+\theta_3 a^*)^2\sigma^2\}} \\
&= e^{\theta_2 m^*+\theta_3 a^* m^*-(\theta_2+\theta_3 a^*)(\beta_0+\beta_1 a^*+\beta_2' c)-\frac{1}{2}(\theta_2+\theta_3 a^*)^2\sigma^2}.
\end{aligned}$$

We have $\int \frac{E[Y|a,m,c]}{E[Y|a^*,m^*,c]}dP(m|a^\dagger,c)$

$$\begin{aligned}
&\approx \int \exp(\theta_1 a+\theta_2 m+\theta_3 am-\theta_1 a^*-\theta_2 m^*-\theta_3 a^* m^*)dP(m|a^\dagger,c) \\
&= \exp\{\theta_1(a-a^*)-\theta_2 m^*-\theta_3 a^* m^*\}\int \exp\{(\theta_2+\theta_3 a)m\}dP(m|a^\dagger,c) \\
&= \exp\{\theta_1(a-a^*)-\theta_2 m^*-\theta_3 a^* m^*\}\exp\{(\theta_2+\theta_3 a)(\beta_0+\beta_1 a^\dagger+\beta_2' c)+\frac{1}{2}(\theta_2+\theta_3 a)^2\sigma^2\} \\
&= e^{\theta_1(a-a^*)-\theta_2 m^*-\theta_3 a^* m^*+(\theta_2+\theta_3 a)(\beta_0+\beta_1 a^\dagger+\beta_2' c)+\frac{1}{2}(\theta_2+\theta_3 a)^2\sigma^2}.
\end{aligned}$$

The reference interaction is thus given by:

$$\begin{aligned}
RR_c^{INT_{ref}}(m^*) &= \int\{\frac{E[Y|a,m,c]}{E[Y|a^*,m^*,c]}-\frac{E[Y|a^*,m,c]}{E[Y|a^*,m^*,c]}-\frac{E[Y|a,m^*,c]}{E[Y|a^*,m^*,c]}+1\}dP(m|a^*,c) \\
&= e^{\theta_1(a-a^*)-\theta_2 m^*-\theta_3 a^* m^*+(\theta_2+\theta_3 a)(\beta_0+\beta_1 a^*+\beta_2' c)+\frac{1}{2}(\theta_2+\theta_3 a)^2\sigma^2} \\
&\quad -e^{-\theta_2 m^*-\theta_3 a^* m^*+(\theta_2+\theta_3 a^*)(\beta_0+\beta_1 a^*+\beta_2' c)+\frac{1}{2}(\theta_2+\theta_3 a^*)^2\sigma^2}-e^{(\theta_1+\theta_3 m^*)(a-a^*)}+1
\end{aligned}$$

and the component due to the reference interaction $\kappa RR_c^{INT_{ref}}(m^*)$ by:

$$\begin{aligned}
&e^{\{\theta_1+\theta_3(\beta_0+\beta_1 a^*+\beta_2' c+\theta_2\sigma^2)\}(a-a^*)+\frac{1}{2}\theta_3^2\sigma^2(a^2-a^{*2})}-1 \\
&-e^{\theta_1(a-a^*)+\theta_2 m^*+\theta_3 a m^*-(\theta_2+\theta_3 a^*)(\beta_0+\beta_1 a^*+\beta_2' c)-\frac{1}{2}(\theta_2+\theta_3 a^*)^2\sigma^2} \\
&+e^{\theta_2 m^*+\theta_3 a^* m^*-(\theta_2+\theta_3 a^*)(\beta_0+\beta_1 a^*+\beta_2' c)-\frac{1}{2}(\theta_2+\theta_3 a^*)^2\sigma^2}
\end{aligned}$$

The mediated interaction is given by:

$$\begin{aligned} RR_c^{INT_{med}} &= \int \{\frac{E[Y|a,m,c]}{E[Y|a^*,m^*,c]} - \frac{E[Y|a^*,m,c]}{E[Y|a^*,m^*,c]}\}\{dP(m|a,c) - dP(m|a^*,c)\} \\ &\approx e^{\theta_1(a-a^*)-\theta_2 m^*-\theta_3 a^* m^*+(\theta_2+\theta_3 a)(\beta_0+\beta_1 a+\beta_2' c)+\frac{1}{2}(\theta_2+\theta_3 a)^2\sigma^2} \\ &\quad -e^{-\theta_2 m^*-\theta_3 a^* m^*+(\theta_2+\theta_3 a^*)(\beta_0+\beta_1 a+\beta_2' c)+\frac{1}{2}(\theta_2+\theta_3 a^*)^2\sigma^2} \\ &\quad -e^{\theta_1(a-a^*)-\theta_2 m^*-\theta_3 a^* m^*+(\theta_2+\theta_3 a)(\beta_0+\beta_1 a^*+\beta_2' c)+\frac{1}{2}(\theta_2+\theta_3 a)^2\sigma^2} \\ &\quad +e^{-\theta_2 m^*-\theta_3 a^* m^*+(\theta_2+\theta_3 a^*)(\beta_0+\beta_1 a^*+\beta_2' c)+\frac{1}{2}(\theta_2+\theta_3 a^*)^2\sigma^2}. \end{aligned}$$

and the component due to the mediated interaction $\kappa RR_c^{INT_{med}}$ by:

$$\begin{aligned} &e^{\{\theta_1+\theta_2\beta_1+\theta_3(\beta_0+\beta_1 a^*+\beta_1 a+\beta_2' c+\theta_2\sigma^2)\}(a-a^*)+\frac{1}{2}\theta_3^2\sigma^2(a^2-a^{*2})} \\ &-e^{(\theta_2\beta_1+\theta_3\beta_1 a^*)(a-a^*)} - e^{\{\theta_1+\theta_3(\beta_0+\beta_1 a^*+\beta_2' c+\theta_2\sigma^2)\}(a-a^*)+\frac{1}{2}\theta_3^2\sigma^2(a^2-a^{*2})} + 1. \end{aligned}$$

We also have that the component due to controlled direct effect is:

$$\begin{aligned} \kappa\left[RR_c^{CDE}(m^*) - 1\right] &= \kappa[e^{(\theta_1+\theta_3 m^*)(a-a^*)} - 1] \\ &= e^{\theta_1(a-a^*)+\theta_2 m^*+\theta_3 a m^*-(\theta_2+\theta_3 a^*)(\beta_0+\beta_1 a^*+\beta_2' c)-\frac{1}{2}(\theta_2+\theta_3 a^*)^2\sigma^2} \\ &\quad -e^{\theta_2 m^*+\theta_3 a^* m^*-(\theta_2+\theta_3 a^*)(\beta_0+\beta_1 a^*+\beta_2' c)-\frac{1}{2}(\theta_2+\theta_3 a^*)^2\sigma^2} \end{aligned}$$

and the component due to the pure indirect effect is:

$$\begin{aligned} (RR_c^{PIE} - 1) &= \kappa\int_m \frac{E(Y_{a^* m}|c)}{E(Y_{a^* m^*}|c)}\{dP(m|a,c) - dP(m|a^*,c)\} \\ &= \kappa\{e^{-\theta_2 m^*-\theta_3 a^* m^*+(\theta_2+\theta_3 a^*)(\beta_0+\beta_1 a+\beta_2' c)+\frac{1}{2}(\theta_2+\theta_3 a^*)^2\sigma^2} \\ &\quad -e^{-\theta_2 m^*-\theta_3 a^* m^*+(\theta_2+\theta_3 a^*)(\beta_0+\beta_1 a^*+\beta_2' c)+\frac{1}{2}(\theta_2+\theta_3 a^*)^2\sigma^2}\} \\ &= e^{(\theta_2\beta_1+\theta_3\beta_1 a^*)(a-a^*)} - 1. \end{aligned}$$

Standard errors for these various expressions could be derived using the delta method along the lines of the derivations in the Online Appendix of VanderWeele and Vansteelandt[5] or by using bootstrapping.

2.3 Binary Outcome, Binary Mediator

Suppose both Y and M were binary, that assumptions (i)-(iv) held, that the outcome was rare and that the following regressions were correctly specified:

$$\begin{aligned} logit\{P(Y=1|a,m,c)\} &= \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_4' c \\ logit\{P(M=1|a,c)\} &= \beta_0 + \beta_1 a + \beta_2' c. \end{aligned}$$

Valeri and VanderWeele[16] show that the average total effect, controlled direct effect and the

average pure indirect effect conditional on $C = c$ are given approximately by:

$$\begin{aligned} RR_c^{TE} &\approx \frac{\exp(\theta_1 a)\{1+\exp(\beta_0+\beta_1 a^*+\beta_2^{'}c)\}\{1+\exp(\beta_0+\beta_1 a+\beta_2^{'}c+\theta_2+\theta_3 a)\}}{\exp(\theta_1 a^*)\{1+\exp(\beta_0+\beta_1 a+\beta_2^{'}c)\}\{1+\exp(\beta_0+\beta_1 a^*+\beta_2^{'}c+\theta_2+\theta_3 a^*)\}} \\ RR_c^{CDE}(m^*) &\approx \exp\{(\theta_1+\theta_3 m)(a-a^*)\} \\ RR_c^{PIE} &\approx \frac{\{1+\exp(\beta_0+\beta_1 a^*+\beta_2^{'}c)\}\{1+\exp(\beta_0+\beta_1 a+\beta_2^{'}c+\theta_2+\theta_3 a^*)\}}{\{1+\exp(\beta_0+\beta_1 a+\beta_2^{'}c)\}\{1+\exp(\beta_0+\beta_1 a^*+\beta_2^{'}c+\theta_2+\theta_3 a^*)\}} \end{aligned}$$

where the approximations (here and below) hold to the extent that the outcome is rare. We have that $\kappa = \frac{E(Y_{a^*m^*}|c)}{E(Y_{a^*}|c)}$ is given by:

$$\begin{aligned} \kappa &= \frac{E(Y_{a^*m^*}|c)}{E(Y_{a^*}|c)} = \frac{E[Y|a^*,m^*,c]}{\int E[Y|a^*,m,c]dP(m|a^*,c)} \\ &\approx \frac{\exp(\theta_0+\theta_1 a^*+\theta_2 m^*+\theta_3 a^* m^*+\theta_4^{'}c)}{\exp\{\theta_0+\theta_1 a^*+\theta_4^{'}c\}\int \exp\{(\theta_2+\theta_3 a^*)m\}dP(m|a^*,c)} \\ &= \frac{\exp(\theta_2 m^*+\theta_3 a^* m^*)}{\frac{1+\exp(\beta_0+\beta_1 a^*+\beta_2^{'}c+\theta_2+\theta_3 a^*)}{1+\exp(\beta_0+\beta_1 a^*+\beta_2^{'}c)}} \\ &= \frac{e^{\theta_2 m^*+\theta_3 a^* m^*}\{1+e^{\beta_0+\beta_1 a^*+\beta_2^{'}c}\}}{1+e^{\beta_0+\beta_1 a^*+\beta_2^{'}c+\theta_2+\theta_3 a^*}}. \end{aligned}$$

We also have $\int \frac{E[Y|a,m,c]}{E[Y|a^*,m^*,c]}dP(m|a^\dagger,c)$

$$\begin{aligned} &\approx \int \exp(\theta_1 a+\theta_2 m+\theta_3 am-\theta_1 a^*-\theta_2 m^*-\theta_3 a^* m^*)dP(m|a^\dagger,c) \\ &= \exp\{\theta_1(a-a^*)-\theta_2 m^*-\theta_3 a^* m^*\}\int \exp\{(\theta_2+\theta_3 a)m\}dP(m|a^\dagger,c) \\ &= \frac{e^{\theta_1(a-a^*)-\theta_2 m^*-\theta_3 a^* m^*}}{1+e^{\beta_0+\beta_1 a^\dagger+\beta_2^{'}c}}(1+e^{\beta_0+\beta_1 a^\dagger+\beta_2^{'}c+\theta_2+\theta_3 a}) \\ &\quad \frac{e^{\theta_1(a-a^*)-\theta_2 m^*-\theta_3 a^* m^*}(1+e^{\beta_0+\beta_1 a^\dagger+\beta_2^{'}c+\theta_2+\theta_3 a})}{1+e^{\beta_0+\beta_1 a^\dagger+\beta_2^{'}c}}. \end{aligned}$$

The reference interaction is thus given by: $RR_c^{INT_{ref}}(m^*) =$

$$\begin{aligned} &\int\{\frac{E[Y|a,m,c]}{E[Y|a^*,m^*,c]}-\frac{E[Y|a^*,m,c]}{E[Y|a^*,m^*,c]}-\frac{E[Y|a,m^*,c]}{E[Y|a^*,m^*,c]}+1\}dP(m|a^*,c) \\ = &\frac{e^{\theta_1(a-a^*)-\theta_2 m^*-\theta_3 a^* m^*}(1+e^{\beta_0+\beta_1 a^*+\beta_2^{'}c+\theta_2+\theta_3 a})}{1+e^{\beta_0+\beta_1 a^*+\beta_2^{'}c}}-\frac{e^{-\theta_2 m^*-\theta_3 a^* m^*}(1+e^{\beta_0+\beta_1 a^*+\beta_2^{'}c+\theta_2+\theta_3 a^*})}{1+e^{\beta_0+\beta_1 a^*+\beta_2^{'}c}} \\ &-e^{(\theta_1+\theta_3 m^*)(a-a^*)}+1 \end{aligned}$$

and the component due to the reference interaction $\kappa RR_c^{INT_{ref}}(m^*)$ by:

$$
\begin{aligned}
&= \frac{e^{\theta_1(a-a^*)}(1+e^{\beta_0+\beta_1 a^*+\beta_2' c+\theta_2+\theta_3 a})}{1+e^{\beta_0+\beta_1 a^*+\beta_2' c+\theta_2+\theta_3 a^*}} - 1 \\
&\quad - \frac{e^{\theta_1(a-a^*)+\theta_2 m^*+\theta_3 a m^*}(1+e^{\beta_0+\beta_1 a^*+\beta_2' c})}{1+e^{\beta_0+\beta_1 a^*+\beta_2' c+\theta_2+\theta_3 a^*}} + \frac{e^{\theta_2 m^*+\theta_3 a^* m^*}(1+e^{\beta_0+\beta_1 a^*+\beta_2' c})}{1+e^{\beta_0+\beta_1 a^*+\beta_2' c+\theta_2+\theta_3 a^*}}
\end{aligned}
$$

The mediated interaction is given by: $RR_c^{INT_{med}} =$

$$
\begin{aligned}
&\int \{\frac{E[Y|a,m,c]}{E[Y|a^*,m^*,c]} - \frac{E[Y|a^*,m,c]}{E[Y|a^*,m^*,c]}\}\{dP(m|a,c) - dP(m|a^*,c)\} \\
&= \frac{e^{\theta_1(a-a^*)-\theta_2 m^*-\theta_3 a^* m^*}(1+e^{\beta_0+\beta_1 a+\beta_2' c+\theta_2+\theta_3 a})}{1+e^{\beta_0+\beta_1 a+\beta_2' c}} - \frac{e^{-\theta_2 m^*-\theta_3 a^* m^*}(1+e^{\beta_0+\beta_1 a+\beta_2' c+\theta_2+\theta_3 a^*})}{1+e^{\beta_0+\beta_1 a+\beta_2' c}} \\
&\quad - \frac{e^{\theta_1(a-a^*)-\theta_2 m^*-\theta_3 a^* m^*}(1+e^{\beta_0+\beta_1 a^*+\beta_2' c+\theta_2+\theta_3 a})}{1+e^{\beta_0+\beta_1 a^*+\beta_2' c}} + \frac{e^{-\theta_2 m^*-\theta_3 a^* m^*}(1+e^{\beta_0+\beta_1 a^*+\beta_2' c+\theta_2+\theta_3 a^*})}{1+e^{\beta_0+\beta_1 a^*+\beta_2' c}}
\end{aligned}
$$

and the component due to the mediated interaction $\kappa RR_c^{INT_{med}}$ by:

$$
\begin{aligned}
&= \frac{e^{\theta_1(a-a^*)}(1+e^{\beta_0+\beta_1 a+\beta_2' c+\theta_2+\theta_3 a})(1+e^{\beta_0+\beta_1 a^*+\beta_2' c})}{(1+e^{\beta_0+\beta_1 a^*+\beta_2' c+\theta_2+\theta_3 a^*})(1+e^{\beta_0+\beta_1 a+\beta_2' c})} - \frac{(1+e^{\beta_0+\beta_1 a+\beta_2' c+\theta_2+\theta_3 a^*})(1+e^{\beta_0+\beta_1 a^*+\beta_2' c})}{(1+e^{\beta_0+\beta_1 a^*+\beta_2' c+\theta_2+\theta_3 a^*})(1+e^{\beta_0+\beta_1 a+\beta_2' c})} \\
&\quad - \frac{e^{\theta_1(a-a^*)}(1+e^{\beta_0+\beta_1 a^*+\beta_2' c+\theta_2+\theta_3 a})}{(1+e^{\beta_0+\beta_1 a^*+\beta_2' c+\theta_2+\theta_3 a^*})} + 1
\end{aligned}
$$

We also have that the component due to controlled direct effect is:

$$
\begin{aligned}
\kappa\left[RR_c^{CDE}(m^*) - 1\right] &= \kappa[e^{(\theta_1+\theta_3 m^*)(a-a^*)} - 1] \\
&= \frac{e^{\theta_1(a-a^*)+\theta_2 m^*+\theta_3 a m^*}(1+e^{\beta_0+\beta_1 a^*+\beta_2' c})}{1+e^{\beta_0+\beta_1 a^*+\beta_2' c+\theta_2+\theta_3 a^*}} - \frac{e^{\theta_2 m^*+\theta_3 a^* m^*}(1+e^{\beta_0+\beta_1 a^*+\beta_2' c})}{1+e^{\beta_0+\beta_1 a^*+\beta_2' c+\theta_2+\theta_3 a^*}}
\end{aligned}
$$

and the component due to the pure indirect effect is:

$$
\begin{aligned}
&\kappa \int_m \frac{E(Y_{a^* m}|c)}{E(Y_{a^* m^*}|c)}\{dP(m|a,c) - dP(m|a^*,c)\} \\
&= \kappa\left(\frac{e^{-\theta_2 m^*-\theta_3 a^* m^*}(1+e^{\beta_0+\beta_1 a+\beta_2' c+\theta_2+\theta_3 a^*})}{1+e^{\beta_0+\beta_1 a+\beta_2' c}} - \frac{e^{-\theta_2 m^*-\theta_3 a^* m^*}(1+e^{\beta_0+\beta_1 a^*+\beta_2' c+\theta_2+\theta_3 a^*})}{1+e^{\beta_0+\beta_1 a^*+\beta_2' c}}\right) \\
&= \frac{\{1+\exp(\beta_0+\beta_1 a^*+\beta_2' c)\}\{1+\exp(\beta_0+\beta_1 a+\beta_2' c+\theta_2+\theta_3 a^*)\}}{\{1+\exp(\beta_0+\beta_1 a+\beta_2' c)\}\{1+\exp(\beta_0+\beta_1 a^*+\beta_2' c+\theta_2+\theta_3 a^*)\}} - 1.
\end{aligned}
$$

Standard errors for these expressions could be derived using the delta method along the lines of the derivations in the Online Appendix of Valeri and VanderWeele[16] or by using bootstrapping.

3. SAS Code for the 4-Way Decomposition

3.1. Continuous Outcome, Continuous Mediator

To estimate the components of the 4-way decomposition for the effect of exposure A on a continuous outcome Y with continuous mediator M under the regression models in Section 1.1, one can use the code below. Suppose we have a dataset named 'mydata' with outcome variable 'y', exposure variables 'a' and mediator 'm' and three covariates 'c1', 'c2' and 'c3'. If there were more or fewer covariates the user would have to modify the second, third, fourth, fifth and tenth lines of the code below to include these covariates.

The user must input in the third line of code the two levels of A ('a1=' and 'a0=') that are being compared (these are exposure levels 1 and 0 in the code below but this could be modified for an ordinal or continuous exposure) and the level of $M = m^*$ ('mstar=') at which to compute the controlled direct effect and the remainder of the decomposition (it is assumed in the code below that the mediator is fixed to the value $M = m^* = 0$ but this could be modified). The user must also input in the third line of the code the value of the covariates C at which the effects are to be calculated ('cc1=', 'cc2' and 'cc3='). Alternatively the mean value of these covariates in the sample could be inputted on this line as a summary measure. The code below on line 3 specifies these as 10, 10, and 20 which should be altered according to the covariate values in the application of interest.

The output will include estimates and confidence intervals for the total effect as well as the four components of the total effect, i.e. the controlled direct effect, the reference interaction, the mediated interaction, and the pure indirect effect; the output will also include estimates and confidence intervals for the proportion of the total effect due to each of the four components; and estimates and confidence intervals for the overall proportion mediated, the overall proportion due to interaction, and the overall proportion of the effect that would be eliminated if the mediator M were fixed to the value m^*, specified by the user.

```
proc nlmixed data=mydata;
parms t0=0 t1=0 t2=0 t3=0 tc1=0 tc2=0 tc3=0 b0=0 b1=0 bc1=0 bc2=0 bc3=0 ss_m=1 ss_y=1;
a1=1; a0=0; mstar=0; cc1=10; cc2=10; cc3=20;
mu_y=t0 + t1*A + t2*M + t3*A*M + tc1*C1 + tc2*C2 + tc3*C3;
mu_m =b0 + b1*A + bc1*C1 + bc2*C2 + bc3*C3;
ll_y= -((y-mu_y)**2)/(2*ss_y)-0.5*log(ss_y);
ll_m= -((m-mu_m)**2)/(2*ss_m)-0.5*log(ss_m);
ll_o= ll_m + ll_y;
model Y ~general(ll_o);
bcc = bc1*cc1 + bc2*cc2 + bc3*cc3;
cde = (t1 + t3*mstar)*(a1-a0);
intref = t3*(b0 + b1*a0 + bcc - mstar)*(a1-a0);
intmed = t3*b1*(a1-a0)*(a1-a0);
pie = (t2*b1 + t3*b1*a0)*(a1-a0);
te = cde + intref + intmed + pie;
estimate 'Total Effect' te;
estimate 'CDE' cde;
estimate 'INTref' intref;
estimate 'INTmed' intmed;
estimate 'PIE' pie;
```

```
estimate 'Proportion CDE' cde/te;
estimate 'Proportion INTref' intref/te;
estimate 'Proportion INTmed' intmed/te;
estimate 'Proportion PIE' pie/te;
estimate 'Overall Proportion Mediated' (pie+intmed)/te;
estimate 'Overall Proportion Attributable to Interaction' (intref+intmed)/te;
estimate 'Overall Proportion Eliminated' (intref+intmed+pie)/te;
run;
```

3.2. Continuous Outcome, Binary Mediator

To estimate the components of the 4-way decomposition for the effect of exposure A on a continuous outcome Y with binary mediator M under the regression models in Section 1.2, one can use the code below. The explanation of the code follows that presented in Section 3.1 above.

```
proc nlmixed data=mydata;
parms t0=0 t1=0 t2=0 t3=0 tc1=0 tc2=0 tc3=0 b0=1 b1=0 bc1=0 bc2=0 bc3=0 ss_y=1;
a1=1; a0=0; mstar=0; cc1=10; cc2=10; cc3=20;
mu_y=t0 + t1*A + t2*M + t3*A*M + tc1*C1 + tc2*C2 + tc3*C3;
p_m=(1+exp(-(b0 + b1*A + bc1*C1 + bc2*C2 + bc3*C3)))**-1;
ll_y= -((y-mu_y)**2)/(2*ss_y)-0.5*log(ss_y);
ll_m= m*log (p_m)+(1-m)*log(1-p_m);
ll_o= ll_m + ll_y;
model Y ~general(ll_o);
bcc = bc1*cc1 + bc2*cc2 + bc3*cc3;
cde = (t1 + t3*mstar)*(a1-a0);
intref = t3*(a1-a0)*(exp(b0+b1*a0+bcc)/(1+exp(b0+b1*a0+bcc)) - mstar);
intmed = t3*(a1-a0)*(exp(b0+b1*a1+bcc)/(1+exp(b0+b1*a1+bcc))-exp(b0+b1*a0+bcc)/(1+exp(b0+b1*a0+bcc)));
pie = (t2 + t3*a0)*(exp(b0+b1*a1+bcc)/(1+exp(b0+b1*a1+bcc))-exp(b0+b1*a0+bcc)/(1+exp(b0+b1*a0+bcc)));
te = cde + intref + intmed + pie;
estimate 'Total Effect' te;
estimate 'CDE' cde;
estimate 'INTref' intref;
estimate 'INTmed' intmed;
estimate 'PIE' pie;
estimate 'Proportion CDE' cde/te;
estimate 'Proportion INTref' intref/te;
estimate 'Proportion INTmed' intmed/te;
estimate 'Proportion PIE' pie/te;
estimate 'Overall Proportion Mediated' (pie+intmed)/te;
estimate 'Overall Proportion Attributable to Interaction' (intref+intmed)/te;
estimate 'Overall Proportion Eliminated' (intref+intmed+pie)/te;
run;
```

3.3. Binary Outcome, Continuous Mediator

To estimate the components of the 4-way decomposition on the ratio scale for the effect of exposure A on a binary outcome Y with continuous mediator M under the regression models

in Section 2.2, one can use the code below. Suppose we have a dataset named 'mydata' with outcome variable 'y', exposure variables 'a' and mediator 'm' and three covariates 'c1', 'c2' and 'c3'. If there were more or fewer covariates the user would have to modify the second, third, fourth, fifth and tenth lines of the code below to include these covariates.

The user must input in the third line of code the two levels of A ('a1=' and 'a0=') that are being compared (these are exposure levels 1 and 0 in the code below but this could be modified for an ordinal or continuous exposure) and the level of $M = m^*$ ('mstar=') at which to compute the controlled direct effect and the remainder of the decomposition (it is assumed in the code below that the mediator is fixed to the value $M = m^* = 0$ but this could be modified). The user must also input in the third line of the code the value of the covariates C at which the effects are to be calculated ('cc1=', 'cc2' and 'cc3='). Alternatively the mean value of these covariates in the sample could be inputted on this line as a summary measure. The code below on line 3 specifies these as 58.57, 1.44, and 0.34 which should be altered according to the covariate values in the application of interest.

The output will include estimates and confidence intervals for the total effect risk ratio, the excess relative risk (i.e. the relative risk minus 1) as well as the four components of the excess relative risk, i.e. the excess relative risks due to the controlled direct effect, to the reference interaction, to the mediated interaction, and to the pure indirect effect; the output will also include estimates and confidence intervals for the proportion of the excess relative risk due to each of the four components; and estimates and confidence intervals for the overall proportion mediated, the overall proportion due to interaction, and the overall proportion of the effect that would be eliminated if the mediator M were fixed to the value m^*, specified by the user.

```
proc nlmixed data=mydata;
parms t0=1 t1=0 t2=0 t3=0 tc1=0 tc2=0 tc3=0 b0=0 b1=0 bc1=0 bc2=0 bc3=0 ss_m=1;
a1=1; a0=0; mstar=0; cc1=58.57; cc2=1.44; cc3=0.34;
p_y=(1+exp(-(t0 + t1*A + t2*M + t3*A*M + tc1*C1 + tc2*C2 + tc3*C3)))**-1;
mu_m =b0 + b1*A + bc1*C1 + bc2*C2 + bc3*C3;
ll_m= -((m-mu_m)**2)/(2*ss_m)-0.5*log(ss_m);
ll_y= y*log (p_y)+(1-y)*log(1-p_y);
ll_o= ll_m + ll_y;
model Y ~general(ll_o);
bcc = bc1*cc1 + bc2*cc2 + bc3*cc3;
CDE_comp = exp( t1*(a1-a0)+t2*mstar + t3*a1*mstar - (t2+t3*a0)*(b0+b1*a0+bcc)
   -  (1/2)*(t2+t3*a0)*(t2+t3*a0)*ss_m )
   - exp(t2*mstar + t3*a0*mstar - (t2+t3*a0)*(b0+b1*a0+bcc) -  (1/2)*(t2+t3*a0)*(t2+t3*a0)*ss_m );
INTref_comp = exp((t1+t3*(b0+b1*a0+bcc+t2*ss_m))*(a1-a0) + (1/2)*t3*t3*ss_m*(a1*a1-a0*a0)) - (1.0)
   -exp(t1*(a1-a0)+t2*mstar+t3*a1*mstar-(t2+t3*a0)*(b0+b1*a0+bcc)- (1/2)*(t2+t3*a0)*(t2+t3*a0)*ss_m)
   +exp(t2*mstar+t3*a0*mstar-(t2+t3*a0)*(b0+b1*a0+bcc)- (1/2)*(t2+t3*a0)*(t2+t3*a0)*ss_m);
INTmed_comp = exp( (t1+t2*b1+t3*(b0+b1*a0+b1*a1+bcc+t2*ss_m))*(a1-a0)
    + (1/2)*t3*t3*ss_m*(a1*a1-a0*a0) )
   -exp( (t2*b1+t3*b1*a0)*(a1-a0) ) -exp( (t1+t3*(b0+b1*a0+bcc+t2*ss_m ))*(a1-a0)
    + (1/2)*t3*t3*ss_m*(a1*a1-a0*a0) ) + (1);
PIE_comp = exp( (t2*b1+t3*b1*a0)*(a1-a0) ) - (1);
terr=cde_comp+intref_comp+intmed_comp+pie_comp;
total = exp((t1 + t3*(b0+b1*a0+bcc + t2*ss_m))*(a1-a0)+(1/2)*t3*t3*ss_m*(a1*a1-a0*a0))
    *exp((t2*b1+t3*b1*a1)*(a1-a0));
estimate 'Total Effect Risk Ratio' total;
```

```
estimate 'Total Excess Relative Risk' total-1;
estimate 'Excess Relative Risk due to CDE' cde_comp*(total-1)/terr;
estimate 'Excess Relative Risk due to INTref' intref_comp*(total-1)/terr;
estimate 'Excess Relative Risk due to INTmed' intmed_comp*(total-1)/terr;
estimate 'Excess Relative Risk due to PIE' pie_comp*(total-1)/terr;
estimate 'Proportion CDE' cde_comp/terr;
estimate 'Proportion INTref' intref_comp/terr;
estimate 'Proportion INTmed' intmed_comp/terr;
estimate 'Proportion PIE' pie_comp/terr;
estimate 'Overall Proportion Mediated' (pie_comp+intmed_comp)/terr;
estimate 'Overall Proportion Attributable to Interaction' (intref_comp+intmed_comp)/terr;
estimate 'Overall Proportion Eliminated' (intref_comp+intmed_comp+pie_comp)/terr;
run;
```

The code given above is applicable to cohort data. For case-control studies in which sampling is done on the outcome Y, if the outcome is rare, then the code above can be adapted by fitting the mediator regression only among the controls. This can be done by replacing the sixth line of code by: ll_m= -(((m-mu_m)**2)/(2*ss_m)-0.5*log(ss_m))*(1-y);

3.4. Binary Outcome, Binary Mediator

To estimate the components of the 4-way decomposition for the effect of exposure A on a binary outcome Y with binary mediator M under the regression models in Section 2.3, one can use the code below. The explanation of the code follows that presented in Section 3.3 above.

```
proc nlmixed data=mydata;
parms t0=1 t1=0 t2=0 t3=0 tc1=0 tc2=0 tc3=0 b0=0 b1=0 bc1=0 bc2=0 bc3=0;
a1=1; a0=0; mstar=0; cc1=58.57; cc2=1.44; cc3=0.34;
p_y=(1+exp(-(t0 + t1*A + t2*M + t3*A*M + tc1*C1 + tc2*C2 + tc3*C3)))**-1;
p_m =(1+exp(-(b0 + b1*A + bc1*C1 + bc2*C2 + bc3*C3)))**-1;
ll_y= y*log (p_y)+(1-y)*log(1-p_y);
ll_m= m*log (p_m)+(1-m)*log(1-p_m);
ll_o= ll_m + ll_y;
model Y ~general(ll_o);
bcc = bc1*cc1 + bc2*cc2 + bc3*cc3;
CDE_comp = exp(t1*(a1-a0)+t2*mstar+t3*a1*mstar)*(1+exp(b0+b1*a0+bcc))/(1+exp(b0+b1*a0+bcc+t2+t3*a0))
   - exp(t2*mstar+t3*a0*mstar)*(1+exp(b0+b1*a0+bcc))/(1+exp(b0+b1*a0+bcc+t2+t3*a0));
INTref_comp = exp(t1*(a1-a0))*(1+exp(b0+b1*a0+bcc+t2+t3*a1))/(1+exp(b0+b1*a0+bcc+t2+t3*a0)) - (1)
   -exp(t1*(a1-a0)+t2*mstar+t3*a1*mstar)*(1+exp(b0+b1*a0+bcc))
    /(1+exp(b0+b1*a0+bcc+t2+t3*a0))
   + exp(t2*mstar+t3*a0*mstar)*(1+exp(b0+b1*a0+bcc))/(1+exp(b0+b1*a0+bcc+t2+t3*a0));
INTmed_comp = exp(t1*(a1-a0))*(1+exp(b0+b1*a1+bcc+t2+t3*a1))*(1+exp(b0+b1*a0+bcc))
    /( (1+exp(b0+b1*a0+bcc+t2+t3*a0))*(1+exp(b0+b1*a1+bcc))  )
   - (1+exp(b0+b1*a1+bcc+t2+t3*a0))*(1+exp(b0+b1*a0+bcc)) / ( (1+exp(b0+b1*a0+bcc+t2+t3*a0))
    *(1+exp(b0+b1*a1+bcc))  )
   - exp(t1*(a1-a0))*(1+exp(b0+b1*a0+bcc+t2+t3*a1))/(1+exp(b0+b1*a0+bcc+t2+t3*a0)) + (1);
PIE_comp = (1+exp(b0+b1*a0+bcc))*(1+exp(b0+b1*a1+bcc+t2+t3*a0)) / ( (1 + exp(b0+b1*a1+bcc))
    *(1+exp(b0+b1*a0+bcc+t2+t3*a0)) ) -(1);
```

```
terr=cde_comp+intref_comp+intmed_comp+pie_comp;
total = exp(t1*a1)*(1+exp(b0+b1*a0+bcc))*(1+exp(b0+b1*a1+bcc+t2+t3*a1))
    / ( exp(t1*a0)*(1 + exp(b0+b1*a1+bcc))*(1+exp(b0+b1*a0+bcc+t2+t3*a0)) );
estimate 'Total Effect Risk Ratio' total;
estimate 'Total Excess Relative Risk' total-1;
estimate 'Excess Relative Risk due to CDE' cde_comp*(total-1)/terr;
estimate 'Excess Relative Risk due to INTref' intref_comp*(total-1)/terr;
estimate 'Excess Relative Risk due to INTmed' intmed_comp*(total-1)/terr;
estimate 'Excess Relative Risk due to PIE' pie_comp*(total-1)/terr;
estimate 'Proportion CDE' cde_comp/terr;
estimate 'Proportion INTref' intref_comp/terr;
estimate 'Proportion INTmed' intmed_comp/terr;
estimate 'Proportion PIE' pie_comp/terr;
estimate 'Overall Proportion Mediated' (pie_comp+intmed_comp)/terr;
estimate 'Overall Proportion Attributable to Interaction' (intref_comp+intmed_comp)/terr;
estimate 'Overall Proportion Eliminated' (intref_comp+intmed_comp+pie_comp)/terr;
run;
```

The code given above is applicable to cohort data. For case-control studies in which sampling is done on the outcome Y, if the outcome is rare, then the code above can be adapted by fitting the mediator regression only among the controls. This can be done by replacing the sixth line of code by: ll_m= (m*log (p_m)+(1-m)*log(1-p_m))*(1-y);

Decomposition in the Presence of an Exposure-Induced Mediator-Outcome Confounder

Consider a setting in which there is a variable L that is affected by exposure A and in turn affects both M and Y as in Figure 4. Although several of the components of the four-way decomposition are not identified in this setting, alternative effects which randomly set M to a value chosen from the distribution of a particular exposure level can be identified. The discussion here will give a randomized interventional interpretation to Proposition 4 in the text and extend that result to settings such as Figure 4 in which there is a mediator-outcome confounder affected by the exposure.

Let $G_{a|c}$ denote a random draw from the distribution of the mediator amongst those with exposure status a conditional on $C = c$. Let a and a^* be two values of the exposure e.g. for binary exposure we may have $a = 1$ and $a^* = 0$. As in VanderWeele[34], the effect $E(Y_{aG_{a|c}}|c) - E(Y_{aG_{a^*|c}}|c)$ is then the effect on the outcome of randomly assigning an individual who is given the exposure to a value of the mediator from the distribution of the mediator amongst those given exposure versus no exposure, conditional on covariates; this is a randomized interventional analogue of the pure indirect effect. Next consider the effect $E(Y_{aG_{a^*|c}}|c) - E(Y_{a^*G_{a^*|c}}|c)$; this is a direct effect comparing exposure versus no exposure with the mediator in both cases randomly drawn from the distribution of the population when given the absence of exposure, conditional on covariates; this is a randomized interventional analogue of the pure direct effect. Finally, the effect $E(Y_{aG_{a|c}}|c) - E(Y_{a^*G_{a^*|c}}|c)$ compares the expected outcome when having the exposure with the mediator randomly drawn from the distribution of the population when given the exposure, conditional on covariates to the expected outcome when not having the exposure

with the mediator randomly drawn from the distribution of the population when not exposed, conditional on covariates. With effects thus defined we have the decomposition: $E(Y_{aG_{a|c}}|c) - E(Y_{a^*G_{a^*|c}}|c) = \{E(Y_{aG_{a|c}}|c) - E(Y_{aG_{a^*|c}}|c)\} + \{E(Y_{aG_{a^*|c}}|c) - E(Y_{a^*G_{a^*|c}}|c)\}$ so that the total effect decomposes into the sum of the effect through the mediator and the direct effect. These effects arise from randomly choosing for each individual a value of the mediator from the distribution of the mediator amongst all of those with a particular exposure.

We might further decompose this as follows:

$$\begin{aligned} E(Y_{aG_{a|c}}|c) - E(Y_{a^*G_{a^*|c}}|c) &= \{E(Y_{aG_{a^*|c}}|c) - E(Y_{a^*G_{a^*|c}}|c)\} + \{E(Y_{a^*G_{a|c}}|c) - E(Y_{a^*G_{a^*|c}}|c)\} \\ &\quad + [\{E(Y_{aG_{a|c}}|c) - E(Y_{a^*G_{a|c}}|c)\} - \{E(Y_{aG_{a^*|c}}|c) - E(Y_{a^*G_{a^*|c}}|c)\}] \end{aligned}$$

where the first term in the decomposition is the randomized intervention analogue of the pure direct effect, the second is the randomized intervention analogue of the pure indirect effect, and the third is the difference between the randomized intervention analogue of the total direct effect and the pure direct effect. As shown in VanderWeele[34] this third term has the interpretation of an interaction. We have that:

$$\begin{aligned} & \{E(Y_{aG_{a|c}}|c) - E(Y_{a^*G_{a|c}}|c)\} - \{E(Y_{aG_{a^*|c}}|c) - E(Y_{a^*G_{a^*|c}}|c)\} \\ = & \sum\nolimits_m E[Y_{am} - Y_{a^*m}|G_{a|c} = m, c]P(G_{a|c} = m|c) - \sum\nolimits_m E[Y_{am} - Y_{a^*m}|G_{a^*|c} = m, c]P(G_{a^*|c} = m|c) \\ = & \sum\nolimits_m E[Y_{am} - Y_{a^*m}|c]P(M_a = m|c) - \sum\nolimits_m E[Y_{am} - Y_{a^*m}|c]P(M_{a^*} = m|c) \\ = & \sum\nolimits_m E[Y_{am} - Y_{a^*m} - Y_{am^*} + Y_{a^*m^*}|c]\{P(M_a = m|c) - P(M_{a^*} = m|c)\} \end{aligned}$$

where m^* is an arbitrary value of M. We have the three-way decomposition given in VanderWeele.[34] Moreover, for the analogue of the pure direct effect we have: $\{E(Y_{aG_{a^*|c}}|c) - E(Y_{a^*G_{a^*|c}}|c)\}$

$$\begin{aligned} = & E(Y_{am^*} - Y_{a^*m^*}|c) + \{E(Y_{aG_{a^*|c}}|c) - E(Y_{a^*G_{a^*|c}}|c) - E(Y_{am^*} - Y_{a^*m^*}|c)\} \\ = & E(Y_{am^*} - Y_{a^*m^*}|c) + \sum\nolimits_m E[Y_{am} - Y_{a^*m}|G_{a^*|c} = m, c]P(G_{a^*|c} = m|c) - E(Y_{am^*} - Y_{a^*m^*}|c) \\ = & E(Y_{am^*} - Y_{a^*m^*}|c) + \sum\nolimits_m E[Y_{am} - Y_{a^*m} - Y_{am^*} + Y_{a^*m^*}|c]P(M_{a^*} = m|c) \end{aligned}$$

i.e. the analogue of the pure direct effect is the sum of a controlled direct effect and the reference interaction term, $\sum_m E[Y_{am} - Y_{a^*m} - Y_{am^*} + Y_{a^*m^*}|c]P(M_{a^*} = m|c)$. We thus have a randomized interventional analogue of the four way decomposition.

To identify these effects the following conditions suffice: Assumptions (i) $Y_{am} \perp\!\!\!\perp A|C$ and (iii) $M_a \perp\!\!\!\perp A|C$ above, that conditional on C there is no unmeasured exposure-outcome or exposure-mediator confounding, along with an assumption (ii*) that $Y_{am} \perp\!\!\!\perp M|\{A, C, L\}$, i.e. that conditional on (A, C, L), there is no unmeasured confounding of the mediator-outcome relationship. These three assumptions would hold in the causal diagram in Figure 4. Under the three assumptions, each of these component are identified from data and it

follows from the g-formula[39] that:

$$\begin{aligned} E(Y_{am^*} - Y_{a^*m^*}|c) &= \sum\nolimits_l \{E[Y|a,l,m^*,c]P(l|a,c) - E[Y|a^*,l,m^*,c]P(l|a^*,c)\} \\ E(Y_{a^*G_{a|c}}|c) - E(Y_{a^*G_{a^*|c}}|c) &= \sum\nolimits_{l,m} E[Y|a^*,l,m,c]P(l|a^*,c)\{P(m|a,c) - P(m|a^*,c)\} \end{aligned}$$

$$\begin{aligned} & \sum\nolimits_m E[Y_{am} - Y_{a^*m} - Y_{am^*} + Y_{a^*m^*}|c]\{P(M_a = m|c) - P(M_{a^*} = m|c)\} \\ = & \sum\nolimits_{l,m} \{E[Y|a,l,m,c]P(l|a,c) - E[Y|a^*,l,m,c]P(l|a^*,c)\}\{P(m|a,c) - P(m|a^*,c)\} \end{aligned}$$

and

$$\begin{aligned} & \sum\nolimits_m E[Y_{am} - Y_{a^*m} - Y_{am^*} + Y_{a^*m^*}|c]\{P(M_{a^*} = m|c)\} \\ = & \sum\nolimits_{l,m} \{E[Y|a,l,m,c]P(l|a,c) - E[Y|a^*,l,m,c]P(l|a^*,c) - E[Y|a,l,m^*,c]P(l|a,c) \\ & + E[Y|a^*,l,m^*,c]P(l|a^*,c)\}P(m|a^*,c). \end{aligned}$$

Thus a randomized interventional analogue of the four-way decomposition holds and its components can be identified under assumptions (i), (ii*) and (iii). When Figure 3 is in fact the underlying causal diagram so the L can be chosen to be empty then assumption (ii*) simply becomes assumption (ii) in the text. And the identification results here simply reduce to those of Proposition 4 in the text. As in Proposition 4 in the text, the randomized interventional interpretation does not require the more controversial cross-world independence assumption, assumption (iv).