Supplementary Methods File

Christian A. Maino Vieytes, May A. Baydoun

Contents

1   Four-Way Decomposition Models

1.1   *Background*

Let Y denote the response, A the exposure (in this case poor oral health problems (POHP; binary–exposed/unexposed)), a its realized value (in the binary exposure case it is "1" or "0"), M the mediator (individual plasma proteomic marker levels or principal component scores), m its realized value, and $\boldsymbol{c}$ the vector of confounders. Under the assumption of no confounding, namely:

$$\text{i. } Y_{am} \coprod A|C$$

$$\text{ii. } Y_{am} \coprod M|(A, C)$$

$$\text{iii. } M_a \coprod A|C$$

$$\text{iv. } Y_{am} \coprod A*|C$$

that the effect of A on Y is unconfounded conditional on C (i), that the effect of M on Y is unconfounded conditional on (A, C), (ii), the effect of A on M is unconfounded conditional on C (iii), and that any mediator outcome confounders are not affected by the exposure (iv), we can partition the sources of total effect of the model into four components (Equation 1).[1]

$$TE = CDE + INT_{ref} + INT_{med} + PIE \tag{1}$$

We interpret this model to reflect that the total effect of the exposure, A, on the outcome, Y, is a sum of the *controlled direct effect* (CDE–i.e., the effect of A on Y not due to any interaction or mediation), *the reference interaction* (INT_{ref}– i.e., the effect of interaction only), *the mediated interaction* (INT_{med}–i.e., the

effect of interaction and mediation), and *the purely indirect effect* (PIE–i.e., the indirect effect only). Additional details on the four-way decomposition model are provided in an original publication that we refer the readers to.[1]

1.2 *Implementation*

The process of estimating the components of the partitioned model in Equation 1 requires the fitting of two regression models and then using the parameter estimates for the final computation of the components. A robust summary of the *med4way* command in Stata is described elsewhere and we refer the reviewers to that commentary though we provide a succinct summary of the implementation and estimation of the effects.[2] Under the assumption of no unmeasured confounding (as detailed above in **1.1**) we can estimate, on average, the four components of the model on a population but not the individual-level effects. The two regression models required are provided and include a model for the expectation Y conditioned on the exposure, mediator, and confounders (Equation 2) and a model for the expectation of M conditioned on the exposure and confounders (Equation 3):

$$\mathrm{E}[Y|(a, m, \boldsymbol{c})] = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 a * m + \boldsymbol{\theta}_c c \tag{2}$$

$$\mathrm{E}[M|(a, \boldsymbol{c})] = \beta_0 + \beta_1 a + \boldsymbol{\beta}_c^T \boldsymbol{c} \tag{3}$$

Estimates of these parameters therefore facilitate the direct computation of estimates of the four component sources of variation for the total effect (TE):

$$\mathrm{E}[CDE|c] = \theta_1(a - a')$$

$$\mathrm{E}[INT_{ref}|c] = \theta_3(\beta_0 + \beta_1 a' + \boldsymbol{\beta}_c^T \boldsymbol{c})(a - a')$$

$$\mathrm{E}[INT_{med}|c] = \theta_3 \beta_1 (a - a')(a - a')$$

$$\mathrm{E}[PIE|c] = (\theta_2 \beta_1 + \theta_3 \beta_1 a')(a - a')$$

where $a = 1$ and $a' = 0$ if the exposure is binary (as it is in this case).[3] The models we describe are generalizable and *Med4way* can handle outcome variables from several distributions (e.g., binomial, log-binomial, Poisson, negative binomial, Weibull, Cox, etc.).[2] In our analysis, $\mathrm{E}[Y|(a, m, c)]$ and $\mathrm{E}[M|(a, c)]$ are specified as follows:

$$\lambda(t|x, v, \boldsymbol{z}) = \lambda_0(t) + \theta_1 x + \theta_2 v + \theta_3 x * v + \boldsymbol{\theta}_z^T \boldsymbol{z} \tag{4}$$

$$\mathrm{E}[V|(x, \boldsymbol{z})] = \beta_0 + \beta_1 x + \boldsymbol{\beta}_z^T \boldsymbol{z} \tag{5}$$

where, in Equation 4, we model the log hazard at time t as a function of x, POHP, v, an individual proteomic biomarker or a principal component score, and $\boldsymbol{z}$, the vector of confounders/covariates discussed in the manuscript. In Equation 5, we model the expectation of the mediator, V, as a function of POHP and the other covariates using ordinary least squares.

References

1. VanderWeele, T. J. A unification of mediation and interaction: a four-way decomposition. *Epidemiology (Cambridge, Mass.)* **25,** 749 (2014).

2. Discacciati, A., Bellavia, A., Lee, J. J., Mazumdar, M. & Valeri, L. *Med4way: a Stata command to investigate mediating and interactive mechanisms using the four-way effect decomposition* 2019.

3. VanderWeele, T. J. & Tchetgen Tchetgen, E. J. Mediation analysis with time varying exposures and mediators. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **79,** 917–938 (2017).