Mediation and interaction analysis
Introduction and overview of Stata commands

Andrea Bellavia

Utrecht Exposome Hub, Utrecht University
Departments of Environmental Health, Harvard T.H. Chan School of Public Health

abellavi@hsph.harvard.edu

2021 Stata Biostatistics and Epidemiology Virtual Symposium

Outline

1. Principles of mediation and pathway analysis
2. Overview of current approaches and STATA environments
   2.1. Classical (Baron & Kenny) approach
   2.2. Structural Equation Modeling framework
   2.3. Counterfactual approach to mediation analysis
   2.2. General approach (Imai)
3. Extensions and current gaps
4. The `med4way` command (Rino Bellocco)

1. Principles of mediation and pathway analysis

Mediation and pathway analysis attempts to elucidate the mechanisms underlaying an observed $X - Y$ association.

- How is the association generated?
- Does the association only exist for specific subgroups?

Mediation and pathway analysis address the how question. Interaction analysis addressed the for whom question

Mediation analysis - formal definition

A mediator is a covariate that mediates the association between X and Y.

$$X \longrightarrow Y$$
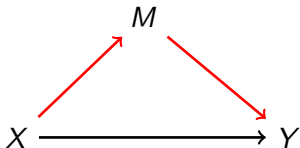
with M above, arrows from X to M and from M to Y.
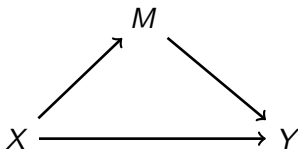
Part of the effect of X on Y is due to the fact that X causes M, which in turn causes Y

We aim to disentangle the total effect of X on Y into a direct effect that goes through all possible pathways but M . . .
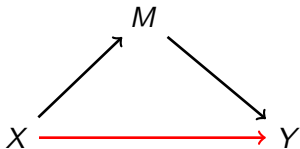
. . . and an indirect effect that goes through M

2.1 Classical approach to mediation analysis

A very basic approach for mediation analysis is the difference method, based on fitting two regression models. Let X be the exposure of interest, Y a continuous outcome, M a potential mediators, and C a set of confounders:

$$E[Y|x, c] = \alpha_0 + \alpha_1 \cdot x + \alpha_2^T \cdot c$$
$$E[Y|x, m, c] = \beta_0 + \beta_1 \cdot x + \beta_2 \cdot m + \beta_3^T \cdot c$$

If the effect of the exposure is considerably reduced when adjusting for the mediator, this is taken as an indicator of mediation.

- In this context, the effect of the exposure in the first model is interpreted as the total effect of X on Y, while the effect of the exposure in the model that also adjusts for the mediator is interpreted as the direct effect.

- Since the total effect is the sum of direct and indirect, the latter is therefore derived by taking the difference of the two.

$$\text{Total Effect (TE)} = \alpha_1$$

$$\text{Direct Effect (DE)} = \beta_1$$

$$\text{Indirect Effect (IE)} = \text{TE - DE} = \alpha_1 - \beta_1$$

Product method

The traditional approach for mediation analysis, introduced by Baron and Kenny in 1986, is the product method. Two regression models are still required, one for the effect of the exposure on the outcome while adjusting for the mediator, and one for the effect of the exposure on the mediator.

$$E[Y|x, m, c] = \beta_0 + \beta_1 \cdot x + \beta_2 \cdot m + \beta_3^T \cdot c$$

$$E[M|x, c] = \gamma_0 + \gamma_1 \cdot x + \gamma_2^T \cdot c$$

The indirect effect is calculated by taking the product of the effect of the exposure on the mediators time the effect of the mediator on the outcome.

$$\text{Direct Effect} = \beta_1$$

$$\text{Indirect Effect} = \gamma_1 \cdot \beta_2$$

$$\text{Total Effect} = \text{DE} + \text{IE}$$

- Baron and Kenny (1986) showed that the two approaches compare when linear models are fitted (continuous outcome and mediator).
- The product method however became the common choice as it can be extended to binary as well as time-to-event outcomes

- Stata does not have an official command for the Baron & Kenny approach, which simply consists in estimating regression models and combining coefficients (easily done with `lincom` and `nlcom`).

2.2 Structural Equation Modeling framework

- Structural equation modeling (SEM) is a flexible framework to evaluate the relationship between several covariates, including latent variables, generally based on generalized linear regression modeling.

- By linking several covariates, SEM has been widely used to evaluate direct dependencies among covariates (pathway analysis)

- This approach became very popular in economics and social sciences, but its use in epidemiology is limited by important assumption including linearity, normality for all evaluated variables, and absence of unmeasured confounding between all sets of evaluated covariates (Vanderweele 2012)

SEM in Stata

The `sem` command allows conducting mediation analysis as long as both the dependent variable and the mediator variable are continuous variables (and all assumptions are met)

The basic command is built on fitting the two regression models presented before

```
sem (M <- X C1 C2)(Y <- M X C1 C2)
```

2.3 Counterfactual approach to mediation analysis

The traditional and SEM approach to mediation analysis share important limitations:

1. First, if control is not made for the mediator-outcome confounders then results from the traditional approach to mediation can be highly biased

2. Exposure-mediator interaction cannot be incorporated in the traditional framework. On the other hand, if an interaction is really present and ignored, direct and indirect effects will be biased.

3. In addition, all continuous exposure and mediators are assumed to have a linear effect. Incorporating non-linearities in the traditional approach is not straigthforward.

- These limitations can be addressed by evaluating mediation analysis within a counterfactual framework (causal mediation analysis).
- This approach defines direct and indirect effects in terms of the counterfactual intervention [i.e. fixing exposure and mediator to a predefined value (controlled), or fixing the exposure to a predefined value and the mediator to the value that naturally follows (natural)].
- The total effect decomposes into the sum of natural direct and indirect effect
- Broadly speaking, natural effects provide information on mechanisms, while controlled effects can be interpreted in terms of interventions

Identification

- Counterfactual objects can not be identified at the individual level (they would require observing an individual in both the real and counterfactual world), but we are able to estimate such effects at the population levels and by making a set of assumptions.

  1. No unmeasured exposure-outcome confounding given C
  2. No unmeasured mediator-outcome confounding given C
  3. No unmeasured exposure-mediator confounding given C
  4. No effect of exposure that confounds the mediator-outcome relationship

- Note that assumptions (1) and (3) are satisfied automatically if the exposure is randomized, but not (2) and (4).

- Estimating the CDE only requires assumptions (1) and (2) to be satisfied.

Mediation analysis with exposure-mediator interaction

When exposure-mediator interaction is present one of the three needed model must be modified as follows:

$$E[Y|x] = \alpha_0 + \alpha_1 \cdot x$$

$$E[Y|x, m] = \beta_0 + \beta_1 \cdot x + \beta_2 \cdot m + \beta_3 \cdot x \cdot m$$

$$E[M|x] = \gamma_0 + \gamma_1 \cdot x$$

- In this situation the formulas for direct and indirect effects previously presented do not yield valid estimates.
- By using the counterfactual definition of direct and indirect effects, interactions can instead be taken into account, and formulas to estimate effects are available.

Four-way decomposition of the total effect

- Vanderweele (2014) showed that, by using the counterfactual approach, the total effect can be decomposed into four different components:
  - ▶ a direct effect (controlled)
  - ▶ a proportion due to mediation alone (natural indirect effect)
  - ▶ a proportion due to interaction alone (reference interaction)
  - ▶ a proportion due to both mediation and interaction (mediated interaction).
- This 4-way decomposition provides the maximum insight on clarifying the contribution of interactive and mediating mechanisms to a given observed total effect

Component	Interpretation
CDE	Treatment effect neither due to mediation nor interaction
INTref	Treatment effect only due to interaction
INTmed	Treatment effect due to both mediation and interaction
NIE	Treatment effect only due to mediation

The paramed command

- paramed (Emsley et al., 2012) was the first Stata command to be developed for conducting causal mediation analysis allowing for exposure-mediator interaction
- By using the option nointer, the command produces results from that correspond to the ones obtained with the traditional method.

```
ssc install paramed

paramed varname, avar(varname) mvar(varname) a0(real)
       a1(real) m(real) yreg(string) mreg(string)
       [cvars(varlist)  nointeraction casecontrol fulloutput
       c(numlist) bootstrap reps(integer 1000) level(cilevel)
       seed(passthru)]
```

2.4 General approach to mediation analysis

- A more flexible simulation-based approach has been developed by Imai et al (2010).

- This allows to specify much more flexible models for the outcome and the mediator and then to estimate the direct and indirect effects by simulation

- The approach makes the same confounding assumptions previously described but allows for more flexible modeling

The `medeff` command

`medeff` (Hicks and Tingley, 2011) is the Stata command for implementing this approach by Imai. It is based on the R package developed by the Authors, and is computationally intensive.

```
ssc install mediation

medeff
(regress m x c)
(logit y m x mx c),
mediate(m) treat(x)
    [interact(mx) sims(#) seed(#) vce() level()]
```

The `mediation` package also includes a `medsens` command that allows conducting sensitivity analysis

Recent extensions and current software gaps

- Several extensions have been developed over the last decade, primarily based on the counterfactual approach to mediation analysis but not only

- To our knowledge, while most of these extensions are available in R, current Stata commands do not cover most of these recent extensions. Because of the increasing interest in mediation analysis, there is a need to fill this gap

- We recently developed the `med4way command` (focus of the next presentation), which integrates the analysis of survival outcomes and the four-way decomposition of the total effects

Some critical topics that could/should be integrated include:

- Multiple mediators and interactions (Vanderweele & Vansteelandt 2014, Bellavia & Valeri 2019)
- High-dimensional mediation analysis (Blum et al. 2020)
- Time-varying exposures, mediators, and confounders (VanderWeele et al. (2017)
- Sensitivity analyses for the counterfactual approach (Vanderweele 2010)
- Multiple imputations

References

- Bellavia A, Valeri L. Decomposition of the total effect in the presence of multiple mediators and interactions. American journal of epidemiology. 2018
- Blum MG et al. Challenges Raised by Mediation Analysis in a High-Dimension Setting. Environmental health perspectives. 2020
- Discacciati A, et al. Med4way: a Stata command to investigate mediating and interactive mechanisms using the four-way effect decomposition. Int. J. Epi 2019
- Emsley R, Liu H. PARAMED: Stata module to perform causal mediation analysis using parametric regression models.
- Hicks R, Tingley D. Causal mediation analysis. The Stata Journal. 2011
- Imai K, et al. A general approach to causal mediation analysis. Psyc. methods. 2010
- Valente MJ, et al. Causal Mediation Programs in R, M plus, SAS, SPSS, and Stata. Structural Equation Modeling: A Multidisciplinary Journal. 2020
- VanderWeele TJ. Mediation analysis: a practitioner's guide. Annual review of public health. 2016
- VanderWeele TJ. Invited commentary: structural equation models and epidemiologic analysis. American journal of epidemiology. 2012
- VanderWeele TJ. Bias formulas for sensitivity analysis for direct and indirect effects. Epidemiology. 2010
- VanderWeele TJ, Tchetgen EJ. Mediation analysis with time varying exposures and mediators. Journal of the Royal Statistical Society. Series B. 2017
- VanderWeele T, Vansteelandt S. Mediation analysis with multiple mediators. Epidemiologic methods. 2014