Regression models for mediation analysis and their application in Stata

Rino Bellocco, Sc.D. & Alessandra Grotta, Ph.D.

Department of Statistics and Quantitative Methods
University of Milano–Bicocca
&
Department of Medical Epidemiology and Biostatistics
Karolinska Institutet

Congresso Nazionale SISMEC - Torino, 16 September 2015

Mediation analysis

IDEAS

PAPERS | ARTICLES | AUTHORS | INSTITUTIONS | DATA (FRED®) |

EAS is a RePEc
vice hosted by the
nomic Research
ision of the Federal
serve Bank of St. Louis

you know? IDEAS uses
data collected within
RePEc project, the
gest online bibliographic
abase in Economics.

owse for material

Papers
Articles
Software
Books
Chapters
Authors
Institutions
Data (FRED®)

d material

JEL Classification
NEP reports

IDEAS Search

Search page with instructions

Search for: mediation Search! Advanced Search

Search for **mediation**. Search results: **mediation : 14, mediate : 1, mediative : 0, mediations : 0, mediately : 0, mediating : 1, mediated : 1**
mediates : 0, mediateness : 0.

1. **MEDIATION**: Stata module for causal **mediation** analysis and sensitivity analysis [**61.175%**]
Raymond Hicks & Dustin Tingley (2011)
Downloadable! **mediation** estimates the role of particular causal mechanisms that **mediate** a relationship between treatment an
mediation effects and direct effects for models with continuous or binary dependent variables using methods presented in Imai
analyses for **mediation** effects that are necessary due to non-random assignment of **mediating** variable. Package replaces earli
method and "Sobel test" for the case of continuous mediator and outcome variables, producing identical results as these earlier m
inference framework and with sensitivity analyses to the key identification assumption. For models with binary mediators or outco
effects are implemented that take into account the use of non-linear models such as probit.

2. PARAMED: Stata module to perform causal **mediation** analysis using parametric regression models [**59.877%**]
Richard Emsley & Hanhua Liu (2013)
Downloadable! paramed performs causal **mediation** analysis using parametric regression models. Two models are estimated: a mo
treatment (exposure) and covariates (if specified), and a model for the outcome conditional on treatment (exposure), the mediat
extends statistical **mediation** analysis (widely known as Baron and Kenny procedure) to allow for the presence of treatment (exp
outcome regression model using counterfactual definitions of direct and indirect effects. paramed allows continuous, binary or co
binary mediators, and requires the user to specify an appropriate form for the regression models. paramed provides estimates of t
direct effect, the natural indirect effect and the total effect with standard errors and confidence intervals derived using the delt

Mediation analysis

Search for: mediation [Search!] Advanced Search

Search for **mediation**. Search results: **mediation : 14, mediate : 1, mediative : 0, mediations : 0, mediately : 0, mediating : 1, mediated : 1, mediates : 0, mediateness : 0.** Results **1-4** of **4**. Search took **0.959** seconds

1. **MEDIATION**: Stata module for causal **mediation** analysis and sensitivity analysis [**61.175%**]
   Raymond Hicks & Dustin Tingley (2011)
   Downloadable! **mediation** estimates the role of particular causal mechanisms that **mediate** a relationship between treatment and outcome variables. Calculates causal **mediation** effects and direct effects for models with continuous or binary dependent variables using methods presented in Imai et al 2010. Also calculates sensitivity analyses for **mediation** effects that are necessary due to non-random assignment of **mediating** variable. Package replaces earlier approaches like the "Baron-Kenny" method and "Sobel test" for the case of continuous mediator and outcome variables, producing identical results as these earlier methods but not put in a causal inference framework and with sensitivity analyses to the key identification assumption. For models with binary mediators or outcomes, correct calculation of **mediation** effects are implemented that take into account the use of non-linear models such as probit.

2. PARAMED: Stata module to perform causal **mediation** analysis using parametric regression models [**59.877%**]
   Richard Emsley & Hanhua Liu (2013)
   Downloadable! paramed performs causal **mediation** analysis using parametric regression models. Two models are estimated: a model for the mediator conditional on treatment (exposure) and covariates (if specified), and a model for the outcome conditional on treatment (exposure), the mediator and covariates (if specified). It extends statistical **mediation** analysis (widely known as Baron and Kenny procedure) to allow for the presence of treatment (exposure)-mediator interactions in the outcome regression model using counterfactual definitions of direct and indirect effects. paramed allows continuous, binary or count outcomes, and continuous or binary mediators, and requires the user to specify an appropriate form for the regression models. paramed provides estimates of the controlled direct effect, the natural direct effect, the natural indirect effect and the total effect with standard errors and confidence intervals derived using the delta method by default, with a bootstrap option also available.

3. GFORMULA: Stata module to implement the g-computation formula for estimating causal effects in the presence of time-varying conf [**47.580%**]
   Rhian Daniel (2010)
   Downloadable! gformula is an implementation of the g-computation procedure, used to estimate the causal effect of time-varying exposure(s) (A) on an outcome (Y) in the presence of time-varying confounders (L) that are themselves also affected by the exposure(s). The procedure can also be used to address the related problem of estimating controlled direct effects and natural direct/indirect effects when the causal effect of the exposure(s) on an outcome is **mediated** by intermediate variables, and in particular when confounders of the mediator-outcome relationships are themselves affected by the exposure(s).

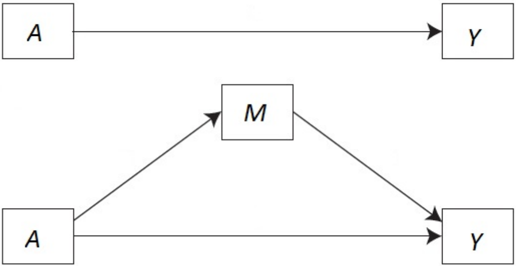
4. LDECOMP: Stata module decomposing the total effects in a logistic regression into direct and indirect effects [**33.812%**]
   Maarten L. Buis (2008)
   Downloadable! ldecomp decomposes the total effects of a categorical variable in logistic regresion into direct and indirect effects using a method method by Erikson et al. (2005) and a generalization of this method by Buis (2008). Consider an example where social class has an indirect effect on attending college through academic performance in high school. The indirect effect is obtained by comparing the proportion of lower class students that attend college with the counterfactual proportion of lower class students if they had the distribution of performance of the higher class students. This captures the association between class and attending college due to differences in performance. The direct effect of class is obtained by comparing the proportion of higher class students with the counterfactual proportion of lower

Summary

Summary

Why mediation analysis?

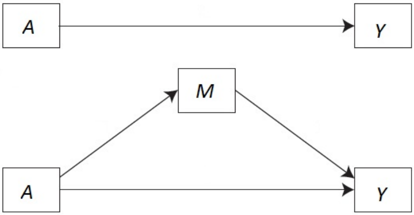
The aim is to understand *if* and *to which extent* the effect of a *treatment* variable A on an *outcome* variable Y is mediated through a variable M



Mediation analysis - example (I)

We could be interested in . . .

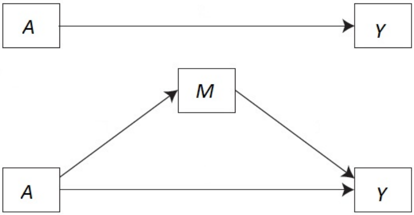
1. studying the relation between physical activity (A) and myocardial infarction (Y) (total effect)
2. evaluating the role of BMI (M) as potential mediator (direct/indirect effects)



Mediation analysis - example (II)

We could be interested in ...

1. studying the relation between maternal smoking (A) and infant mortality (Y) (total effect)
2. evaluating the role of birth weight (M) as potential mediator (direct/indirect effects)



Mediation analysis - other examples (I)

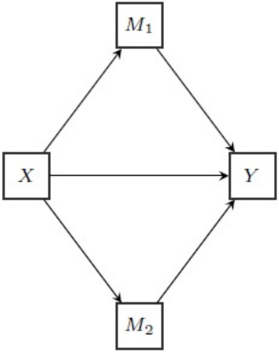


Figure 1: Rosseel, 2013

Mediation analysis - other examples (II)

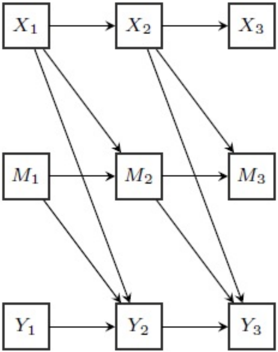


Figure 2: Rosseel, 2013

Summary

Standard approach - Baron and Kenny, 1986

- Model for the outcome (with mediator)

$$E[Y \mid a, m] = \alpha_1 + \beta_1 a + \theta m$$

- Model for the mediator

$$E[M \mid a] = \alpha_2 + \gamma a$$

- Direct effect: β_1
- Indirect effect (product method): $\theta\gamma$

Standard approach - Baron and Kenny, 1986

- linear models
- no exposure-mediator interaction
- causal interpretation?

Modern approach to mediation analysis

- Robins and Greenland (1992), Pearl (2001)
- based on conterfactuals
- more general settings
- assumptions and sensitivity analyses

Causal inference framework

- Let A be a treatment, M be a mediator, Y be an outcome,
- Let $Y(a)$ be the potential outcome Y when intervening to set A to a
- Let $M(a)$ be the potential outcome M when intervening to set A to a
- Let $Y(a, m)$ be the potential outcome Y when intervening to set A to a and M to m

Controlled direct effects

- Controlled direct effect, that compares outcomes under treatment level $A = 1$ vs. $A = 0$, fixing $M = m$:

$$CDE(m) = E(Y(1, m)) - E(Y(0, m))$$

- CDE(m) depends on M level m.
- no analougous definition of controlled indirect effect

Natural direct and indirect effects (I)

- Natural direct effect, that compares outcomes under treatment level $A = 1$ vs. $A = 0$, fixing $M = M(0)$:

$$NDE_0 = E(Y(1, M(0)) - E(Y(0, M(0))$$

- Natural indirect effect, that compares outcomes under $M = M(1)$ vs. $M = M(0)$, fixing $A = 1$:

$$NIE_1 = E(Y(1, M(1)) - E(Y(1, M(0))$$

- Total causal effect can be decomposed as:

$$TCE = E(Y(1)) - E(Y(0)) = NDE + NIE$$

Natural direct and indirect effects (II)

- Natural direct effect, that compares outcomes under treatment level $A = 1$ vs. $A = 0$, fixing $M = M(1)$:

$$NDE_1 = E(Y(1, M(1)) - E(Y(0, M(1))$$

- Natural indirect effect, that compares outcomes under $M = M(1)$ vs. $M = M(0)$, fixing $A = 0$:

$$NIE_0 = E(Y(0, M(1)) - E(Y(0, M(0))$$

- Total causal effect can be decomposed as:

$$TCE = E(Y(1)) - E(Y(0)) = NDE + NIE$$

Natural direct and indirect effects (II)

- Natural direct effect, that compares outcomes under treatment level $A = 1$ vs. $A = 0$, fixing $M = M(1)$:

$$NDE_1 = E(Y(1, M(1)) - E(Y(0, M(1))$$

- Natural indirect effect, that compares outcomes under $M = M(1)$ vs. $M = M(0)$, fixing $A = 0$:

$$NIE_0 = E(Y(0, M(1)) - E(Y(0, M(0))$$

- Proportion of mediation can be computed as:

$$PM = \frac{NIE}{NIE + NDE} = \frac{NIE}{TCE}$$

Decomposition for dichotomous outcomes

- Natural direct effect

$$OR_0^{NDE} = \frac{P(Y_{1M_0} = 1)/P(Y_{1M_0} = 0)}{P(Y_{0M_0} = 1)/P(Y_{0M_0} = 0)}$$

- Natural indirect effect

$$OR_1^{NIE} = \frac{P(Y_{1M_1} = 1)/P(Y_{1M_1} = 0)}{P(Y_{1M_0} = 1)/P(Y_{1M_0} = 0)}$$

- Total causal effect

$$OR^{TE} = OR^{NIE} \times OR^{NDE}$$

Assumptions (I)

- no unmeasured exposure-outcome confounding
- no unmeasured exposure-mediator confounding
- no unmeasured mediator-outcome confounding

Assumptions (I)

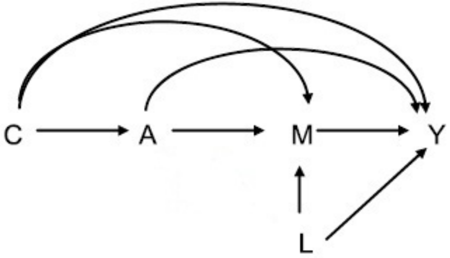


Figure 3: VanderWeele, 2009

Assumptions (cont'd)

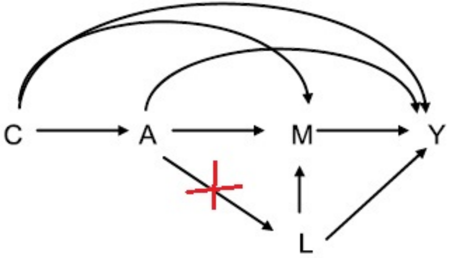


Figure 4: VanderWeele, 2009

Summary

Mediation analysis in Stata

- paramed (Emsley et al., 2012)
- medeff (Hicks and Tingley, 2011)
- ldecomp (Buis, 2010)
- gformula (Daniel, De Stavola and Cousens, 2012)

Mediation analysis in Stata

- paramed (Emsley et al., 2012)
- medeff (Hicks and Tingley, 2011)
- ldecomp (Buis, 2010)
- gformula (Daniel, De Stavola and Cousens, 2012)

paramed

- Valeri, L. and VanderWeele, T.J. (2013). *Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros*
- macro is available also in SAS and SPSS
- parametric approach

paramed - models

- Outcome model - logistic

$$logit\{P(Y = 1 \mid a, m, c)\} = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_4' c$$

- Mediator model - linear

$$E(M \mid a, c) = \beta_0 + \beta_1 a + \beta_2' c$$

where **c** is the set of **known confounders** of the relationships:

- exposure-outcome
- exposure-mediator
- mediator-outcome

paramed - estimation

If the outcome Y is rare, then:

$$log(OR^{NDE} \mid c) \approx \left\{\theta_1 + \theta_3(\beta_0 + \beta_1 a^* + \beta_2^{'} c + \theta_2\sigma^2)\right\}(a - a^*) + 0.5\theta_3^2\sigma^2(a^2 - a^{*2})$$

$$log(OR^{NIE} \mid c) \approx (\theta_2\beta_1 + \theta_3\beta_1 a)(a - a^*)$$

where a, a^* are the treatment levels and σ^2 is the variance of the gaussian error term in the mediator model

paramed - models

- Outcome model - logistic

$$logit\{P(Y = 1 \mid a, m, c)\} = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_4' c$$

- Mediator model - logistic

$$logit\{P(M = 1 \mid a, c)\} = \beta_0 + \beta_1 a + \beta_2' c$$

- exposure-outcome
- exposure-mediator
- mediator-outcome

paramed - estimation

If the outcome Y is rare, then:

$$log(OR^{NDE} \mid c) \approx \frac{exp[\theta_1 a](1 + exp[\theta_2 + \theta_3 a + \beta_0 + \beta_1 a^* + \beta_2^{'} c])}{exp[\theta_1 a^*](1 + exp[\theta_2 + \theta_3 a^* + \beta_0 + \beta_1 a^* + \beta_2^{'} c])}$$

$$log(OR^{NIE} \mid c) \approx \frac{[1 + exp(\beta_0 + \beta_1 a^* + \beta_2^{'} c)][1 + exp(\theta_2 + \theta_3 a + \beta_0 + \beta_1 a + \beta_2^{'} c]}{[1 + exp(\beta_0 + \beta_1 a + \beta_2^{'} c)][1 + exp(\theta_2 + \theta_3 a + \beta_0 + \beta_1 a^* + \beta_2^{'} c]}$$

paramed

- advantages: continuous, binary or count Y, and continuous or binary M; A-M interaction; bootstrap; case-control
- limitations: no sensitivity analyses routines

Example I

The Swedish National March Cohort - 1997

- 43,863 subjects, men and women

Exposure, mediator and outcome

- past physical activity: PA, binary (87% physical active)
- body mass index in 1997: BMI, continuous (mean 24.7; std.dev. 3.5)
- myocardial infarction in 1997-2007: MI, binary (1,200 events)

paramed - syntax

```
paramed mi_dic,

avar(tpa_c) mvar(bmi)

a0(0) a1(1) m(0)

yreg(logistic) mreg(linear)

nointer cvars(agein)
```

paramed - output

```
Logistic regression                               Number of obs   =      37635
                                                  LR chi2(3)      =    1647.96
                                                  Prob > chi2     =     0.0000
Log likelihood = -4491.4153                       Pseudo R2       =     0.1550

------------------------------------------------------------------------------
      mi_dic |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       tpa_c |  -.2617198   .0867014    -3.02   0.003    -.4316514   -.0917881
         bmi |   .0540316   .0084465     6.40   0.000     .0374768    .0705864
       agein |   .1011044   .0030058    33.64   0.000     .0952131    .1069958
       _cons |  -10.72708   .3231284   -33.20   0.000     -11.3604   -10.09376
------------------------------------------------------------------------------
```

paramed - output

bmi	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
tpa_c	-1.06481	.0533235	-19.97	0.000	-1.169325	-.960294
agein	.0303716	.0012058	25.19	0.000	.0280082	.0327349
_cons	24.03829	.079328	303.02	0.000	23.88281	24.19378

paramed - output

	Estimate	Std Err	P>|z|	[95% Conf	Interval]
nde = cde	.76972671	.0867014	0.003	.64943371	.91230128
nie	.94409039	.00944413	0.000	.92677556	.96172871
mte	.72669159	.08615045	0.000	.61378659	.86036528

Note

- exposure-mediator interaction is not allowed, then controlled direct effects are equal to the natural direct effect

paramed - output

	Estimate	Std Err	P>|z|	[95% Conf	Interval]
nde = cde	.76972671	.0867014	0.003	.64943371	.91230128
nie	.94409039	.00944413	0.000	.92677556	.96172871
mte	.72669159	.08615045	0.000	.61378659	.86036528

Total effect $= 0.73$

- odds for MI if everyone had been active vs. odds for MI if everyone had been inactive

paramed - output

```
             |   Estimate     Std Err    P>|z|    [95% Conf    Interval]
-------------+----------------------------------------------------------
             |
   nde = cde |  .76972671     .0867014   0.003    .64943371    .91230128
         nie |  .94409039    .00944413   0.000    .92677556    .96172871
             |
         mte |  .72669159    .08615045   0.000    .61378659    .86036528
```

Natural direct effect = 0.77

- we fix BMI to the value that it would have taken without physical activity
- we compare odds for MI if everyone had been active vs. odds for MI if everyone had been inactive

paramed - output

```
             |   Estimate      Std Err    P>|z|    [95% Conf    Interval]
-------------+----------------------------------------------------------
             |
   nde = cde |  .76972671     .0867014    0.003    .64943371    .91230128
         nie |  .94409039    .00944413    0.000    .92677556    .96172871
             |
         mte |  .72669159    .08615045    0.000    .61378659    .86036528
```

Natural indirect effect = 0.94

- we assume that everyone is physically active
- we compare the odds for MI, when BMI changes from the value with physical activity to the one without physical activity

Example II

The birth weight paradox

- maternal smoking appears to have a protective effect for low birth weight infants

Data

- 2003 U.S. linked birth certificate-infant mortality files from the National Center for Health Statistics
- random sample of 1,000,000 records

Example II

Exposure, mediator and outcome

- low birth weight (infant was born < 2500 grams)
- smoking (maternal smoking during pregnancy)
- death (infant died within 1 year of birth)

Other variables

- maternal drinking, education, race
- marital status

Example II

Exposure, mediator and outcome

- low birth weight (infant was born < 2500 grams) binary
- smoking (maternal smoking during pregnancy) binary
- death (infant died within 1 year of birth) binary

Other variables

- maternal drinking, education, race
- marital status

Example II

Question 1

- Calculate estimates of what the controlled direct effect odds ratios would be for the low birth weight group and the regular birth weight group if it were the case that the assumptions of no unmeasured exposure-outcome confounding and mediator-outcome confounding conditional on the observed covariates held.

paramed - syntax

```
paramed death,

avar(smoking) mvar(lbw)

a0(0) a1(1) m(0)

yreg(logistic) mreg(logistic)

cvars(mar drinking hispanic black

nativeamerican asian agebelow20

ageabove35 somecollege)
```

paramed - syntax

```
paramed death,

avar(smoking) mvar(lbw)

a0(0) a1(1) m(0)

yreg(logistic) mreg(logistic) cvars(...)
```

Now...

exposure-mediator interaction is allowed!

paramed - outcome model

```
Logistic regression                              Number of obs     =     933299
                                                 LR chi2(12)       =   13323.16
                                                 Prob > chi2       =     0.0000
Log likelihood = -25070.873                      Pseudo R2         =     0.2099

------------------------------------------------------------------------------
       death |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     smoking |   .4865302   .0670969     7.25   0.000     .3550228    .6180376
         lbw |   3.517508   .0373642    94.14   0.000     3.444275     3.59074
_smoking_X~w |  -.5563947   .0714215    -7.79   0.000    -.6963783   -.4164112
         mar |   .0996677   .0343631     2.90   0.004     .0323172    .1670182
    drinking |  -.2189616    .060874    -3.60   0.000    -.3382724   -.0996508
    hispanic |  -.1100233   .0423943    -2.60   0.009    -.1931147    -.026932
       black |   .2594655   .0376343     6.89   0.000     .1857036    .3332274
 nativeamer~n |   .3509367   .1274273     2.75   0.006     .1011838    .6006896
       asian |  -.2243916   .0789109    -2.84   0.004    -.3790542    -.069729
   agebelow20 |   .1141238   .0433233     2.63   0.008     .0292117    .1990358
   ageabove35 |  -.0399296   .0430927    -0.93   0.354    -.1243898    .0445306
  somecollege |  -.2047545   .0336721    -6.08   0.000    -.2707505   -.1387585
       _cons |  -6.560523   .0610743  -107.42   0.000    -6.680227    -6.44082
------------------------------------------------------------------------------
```

paramed - mediator model

```
Logistic regression                               Number of obs   =     933299
                                                  LR chi2(10)     =    9203.33
                                                  Prob > chi2     =     0.0000
Log likelihood =  -257131.4                       Pseudo R2       =     0.0176

------------------------------------------------------------------------------
         lbw |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     smoking |   .4794296   .0121403    39.49   0.000     .4556351    .5032241
         mar |   .1878066   .0093212    20.15   0.000     .1695373    .2060759
    drinking |  -.5303817   .0162099   -32.72   0.000    -.5621525    -.498611
    hispanic |  -.0567797   .0111899    -5.07   0.000    -.0787116   -.0348478
       black |   .6478243   .0104637    61.91   0.000     .6273159    .6683327
nativeamer~n |  -.0443687    .040648    -1.09   0.275    -.1240373    .0352999
       asian |   .1867056   .0180545    10.34   0.000     .1513196    .2220917
  agebelow20 |   .0893198    .012603     7.09   0.000     .0646184    .1140212
  ageabove35 |   .2919824   .0108425    26.93   0.000     .2707315    .3132333
 somecollege |  -.0751146    .008927    -8.41   0.000    -.0926111    -.057618
       _cons |  -2.872066   .0150736  -190.54   0.000     -2.90161   -2.842522
------------------------------------------------------------------------------
```

For normal birth weight children...

```
paramed death,
avar(smoking) mvar(lbw)
a0(0) a1(1) m(0)
yreg(logistic) mreg(logistic)
cvars(mar drinking hispanic black
nativeamerican asian agebelow20
ageabove35 somecollege)
```

For normal birth weight children...

	Estimate	Std Err	P>|z|	[95% Conf	Interval]
cde	1.6266622	.06709686	0.000	1.4262097	1.8552882
nde	1.1294185	.04397595	0.006	1.0361481	1.2310848
nie	1.3077558	.00964651	0.000	1.2832621	1.332717
mte	1.4770036	.04396609	0.000	1.3550548	1.6099272

Controlled direct effect $= 1.63$

- smoking is a risk factor

For low birth weight children...

```
paramed death,
avar(smoking) mvar(lbw)
a0(0) a1(1) m(1)
yreg(logistic) mreg(logistic)
cvars(mar drinking hispanic black
nativeamerican asian agebelow20
ageabove35 somecollege)
```

For low birth weight children...

	Estimate	Std Err	P>|z|	[95% Conf	Interval]
cde	.93252013	.04781394	0.144	.84909869	1.0241375
nde	1.1294185	.04397595	0.006	1.0361481	1.2310848
nie	1.3077558	.00964651	0.000	1.2832621	1.332717
mte	1.4770036	.04396609	0.000	1.3550548	1.6099272

Controlled direct effect = 0.93

- smoking is protective

Example II

Question 2

- Do you believe these correspond to the true controlled direct effects? Do you believe that the assumption of no unmeasured confounding is likely to hold?
- Decide on what you think are reasonable sensitivity analysis parameter values for (i) a scenario that would correspond to mild mediator-outcome confounding, (ii) a scenario that would correspond to moderate mediator-outcome confounding, (iii) a scenario that would correspond to severe mediator-outcome confounding.

Sensitivity analyses for controlled direct effects

Suppose:

- U is a confounding variable of the M-Y relationship
- (C,U) is sufficient to control confounding
- U increases the risk of infant mortality γ fold
- π_{im} is the prevalence of U amongst those with $A = i$ and $M = m$

Sensitivity analyses for controlled direct effects

Then, the bias factor B is equal to:

$$B = \frac{1 + (\gamma - 1)\pi_{1m}}{1 + (\gamma - 1)\pi_{0m}}$$

(Vanderweele, 2010)

In Stata

- specify sensitivity parameters
- B computation
- unbiased CDEs by dividing point and interval estimates by B

In Stata - setting I

- Suppose that U increases the risk of infant mortality 1.5 (γ) fold
- Suppose that the prevalence of U for low-birth weight infants whose mothers smoke is 0.025 (π_{11}) and the prevalence of U for low-birth weight infants whose mothers do not smoke is 0.14 (π_{01})

In Stata - setting I

```
. scalar gamma = 1.5
. scalar pi1m = 0.025
. scalar pi0m = 0.14
. scalar B = (1 + (gamma - 1)*pi1m)/(1 + (gamma -1)*pi0m)
. di B
. .94626168
```

In Stata - setting I

```
. paramed death, avar(smoking) mvar(lbw) ...
. mat eff = e(effects)

eff[4,5]
       Estimate     Std_Err        P>|z|   [95%_Conf   Interval]
cde   .93252013   .04781394   .14396739   .84909869   1.0241375
nde   1.1294185   .04397595   .00564901   1.0361481   1.2310848
nie   1.3077558   .00964651           0   1.2832621    1.332717
mte   1.4770036   .04396609           0   1.3550548   1.6099272

. scalar cde = eff[1,1]
. scalar cde_l = eff[1,4]
. scalar cde_u = eff[1,5]
. di cde/B " " cde_l/B " " cde_u/B
. 98547807 .89731911 1.0822984
```

In Stata - setting II

- Suppose that U increases the risk of infant mortality 2.5 (γ) fold
- Suppose that the prevalence of U for low-birth weight infants whose mothers smoke is 0.025 (π_{11}) and the prevalence of U for low-birth weight infants whose mothers do not smoke is 0.14 (π_{01})

In Stata - setting II

```
. scalar gamma = 2.5
. scalar pi1m = 0.025
. scalar pi0m = 0.14
. scalar B = (1 + (gamma - 1)*pi1m)/(1 + (gamma -1)*pi0m)
. di B
. .85743802
```

In Stata - setting II

```
. paramed death, avar(smoking) mvar(lbw) ...
. mat eff = e(effects)

eff[4,5]
       Estimate     Std_Err       P>|z|   [95%_Conf   Interval]
cde   .93252013   .04781394   .14396739   .84909869   1.0241375
nde   1.1294185   .04397595   .00564901   1.0361481   1.2310848
nie   1.3077558   .00964651           0   1.2832621    1.332717
mte   1.4770036   .04396609           0   1.3550548   1.6099272

. scalar cde = eff[1,1]
. scalar cde_l = eff[1,4]
. scalar cde_u = eff[1,5]
. di cde/B " " cde_l/B " " cde_u/B
. 1.0875656 .99027414 1.1944158
```

In Stata - setting III

- Suppose that U increases the risk of infant mortality 3.5 (γ) fold
- Suppose that the prevalence of U for low-birth weight infants whose mothers smoke is 0.025 (π_{11}) and the prevalence of U for low-birth weight infants whose mothers do not smoke is 0.14 (π_{01})

In Stata - setting III

```
. scalar gamma = 3.5
. scalar pi1m = 0.025
. scalar pi0m = 0.14
. scalar B = (1 + (gamma - 1)*pi1m)/(1 + (gamma -1)*pi0m)
. di B
. .78703704
```

In Stata - setting III

```
. paramed death, avar(smoking) mvar(lbw) ...
. mat eff = e(effects)

eff[4,5]
       Estimate     Std_Err       P>|z|   [95%_Conf   Interval]
cde   .93252013   .04781394   .14396739   .84909869   1.0241375
nde   1.1294185   .04397595   .00564901   1.0361481   1.2310848
nie   1.3077558   .00964651           0   1.2832621    1.332717
mte   1.4770036   .04396609           0   1.3550548   1.6099272

. scalar cde = eff[1,1]
. scalar cde_l = eff[1,4]
. scalar cde_u = eff[1,5]
. di cde/B " " cde_l/B " " cde_u/B
. 1.1848491 1.0788548 1.301257
```

Example II

Question 3

- Calculate estimates of what the natural direct and indirect effect and total effect odds ratios would be, allowing for exposure-mediator interaction, if it were the case that the assumptions of no unmeasured (i) exposure-outcome, (ii) mediator-outcome, and (iii) exposure-mediator confounding and (iv) no exposure-induced mediator-outcome confounder held conditional on the observed covariates.

Effect decomposition

	Estimate	Std Err	P>|z|	[95% Conf	Interval]
cde	.93252013	.04781394	0.144	.84909869	1.0241375
nde	1.1294185	.04397595	0.006	1.0361481	1.2310848
nie	1.3077558	.00964651	0.000	1.2832621	1.332717
mte	1.4770036	.04396609	0.000	1.3550548	1.6099272

Total effect = 1.48

- odds for infant death if all women had been smokers vs. odds for child death if all women had been non-smokers

Effect decomposition

	Estimate	Std Err	P>|z|	[95% Conf	Interval]
cde	.93252013	.04781394	0.144	.84909869	1.0241375
nde	1.1294185	.04397595	0.006	1.0361481	1.2310848
nie	1.3077558	.00964651	0.000	1.2832621	1.332717
mte	1.4770036	.04396609	0.000	1.3550548	1.6099272

Natural direct effect = 1.13

- we fix birth weight to the value that it would have taken without maternal smoking
- we compare odds for infant death if all women had been smokers vs. odds for child death if all women had been non-smokers

Effect decomposition

	Estimate	Std Err	P>|z|	[95% Conf	Interval]
cde	.93252013	.04781394	0.144	.84909869	1.0241375
nde	1.1294185	.04397595	0.006	1.0361481	1.2310848
nie	1.3077558	.00964651	0.000	1.2832621	1.332717
mte	1.4770036	.04396609	0.000	1.3550548	1.6099272

Natural indirect effect = 1.31

- we assume that all the women are smokers
- we compare the odds for infant death, when birth weight changes from the value it would have had under smoking assumption to the one it would have had under the non smoking assumption

Example II

Question 4

- Give the estimate of the proportion of the effect mediated on the risk difference scale under these assumptions of unmeasured (i) exposure-outcome, (ii) mediator-outcome, and (iii) exposure-mediator confounding and (iv) no exposure-induced mediator-outcome confounder held conditional on the observed covariates.
- Do you think that the proportion mediated is an overestimate or an underestimate? Why?

Example II

The proportion mediated on the risk difference scale is:

$$\frac{NDE * (NIE - 1)}{(NDE * NIE) - 1} = 0.73$$

Note

The proportion mediated is likely overestimated, because a common cause of birth weight and infant mortality is likely to explain some of the association between the mediator and the outcome.

Mediation analysis in Stata

- paramed (Emsley et al., 2012)
- medeff (Hicks and Tingley, 2011)
- ldecomp (Buis, 2010)
- gformula (Daniel, De Stavola and Cousens, 2012)

medeff

- Imai, K., Keele, L. and Tingley, D. (2010). *A general approach to causal mediation analysis*
- parametric approach
- non-parametric extension in R
- quasi-Bayesian Monte Carlo algorithm (King et al., 2000)

medeff

- advantages: continuous and binary Y and M; A-M interaction; sensitivity analyses (medsens)
- limitations: no decomposition in terms of OR; computationally intensive

Example I

The Swedish National March Cohort - 1997

- 43,863 subjects, men and women

Exposure, mediator and outcome

- past physical activity: PA, binary (87% physical active)
- body mass index in 1997: BMI, continuous (mean 24.7; std.dev. 3.5)
- myocardial infarction in 1997-2007: MI, binary (1,200 events)

medeff - syntax

```
medeff
(regress bmi pa age)
(logit mi pa bmi age),
mediate(bmi) treat(pa)
sims(1000)
```

medeff - output

Effect	Mean	[95% Conf.	Interval]
ACME1	-.0015834	-.0022292	-.0010069
ACME0	-.002114	-.0030084	-.0013764
Direct Effect 1	-.0114884	-.0183685	-.0051785
Direct Effect 0	-.012019	-.0192047	-.0054538
Total Effect	-.0136024	-.0208373	-.0071637

Note that ...

- A.C.M.E. is the Average Causal Mediated Effect
- effects are marginal

medeff - output

Effect	Mean	[95% Conf.	Interval]
Average Mediation	-.0018487	-.0026075	-.0012015
Average Direct Effect	-.0117537	-.018795	-.0053152
% of Tot Eff mediated	.1367193	.0887204	.2580612

On average ...

- the indirect effect represents the 14% of the total effect

medsens

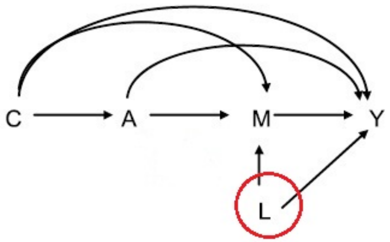


Figure 5: VanderWeele, 2009

- sensitivity analyses are based on the correlation between the error terms of the mediator and of the outcome

Mediation analysis in Stata

- paramed (Emsley et al., 2012)
- medeff (Hicks and Tingley, 2011)
- ldecomp (Buis, 2010)
- gformula (Daniel, De Stavola and Cousens, 2012)

ldecomp

- Buis, M.L. (2011). *Direct and indirect effects in a logit model*
- advantages: multiple mediators; mediators can follow any distribution; A-M interaction
- limitations: binary Y; no sensitivity analyses routines

ldecomp - syntax

```
ldecomp mi age,

direct(pa) indirect(bmi)

or rindirect

reps(1000)
```

ldecomp - output

```
------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
             | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
1/0          |
             |
       total |   .7227368   .0640098    -3.67   0.000     .6075651    .8597408
             |
   indirect1 |   .9387875   .0103436    -5.73   0.000     .9187319     .959281
     direct1 |   .7698619   .0680928    -2.96   0.003       .64733    .9155877
             |
   indirect2 |   .9386944   .0103766    -5.72   0.000     .9185754     .959254
     direct2 |   .7699383   .0680732    -2.96   0.003     .6474377    .9156171
-------------+----------------------------------------------------------------
```

ldecomp - output

```
--------------------------
             |   Observed
             | Odds Ratio
-------------+------------
1/0r         |
             |
     method1 |   .1945307
     method2 |   .1948362
             |
     average |   .1946834
--------------------------
```

On average ...

- the indirect effect represents the 19% of the total effect

Mediation analysis in Stata

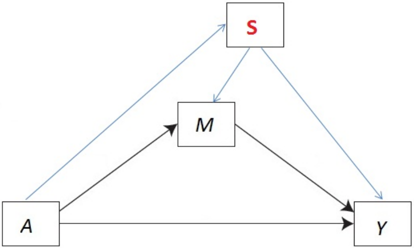
- paramed (Emsley et al., 2012)
- ldecomp (Buis, 2010)
- medeff (Hicks and Tingley, 2011)
- gformula (Daniel, De Stavola and Cousens, 2012)

gformula

- Daniel, R.M., De Stavola, B.L., and Cousens S.N. (2011). *gformula: Estimating causal effects in the presence of time-varying confounding or mediation using the g-computation formula*
- SAS macro
- implements gformula computation developed by Robin (1986)

gformula

mediator-outcome confounder: smoking status



Daniel et al. showed that this setting is methodologically related to the setting in which time-varying confounding occurs

gformula

- advantages: continuous and binary Y and M; A-M interaction; developed for settings in which there are post treatment M-Y confounders; missing values imputation option
- limitations: not suitable for settings in which there is no post-treatment confounding

gformula - syntax

```
gformula mi bmi pa age smoke,

mediation

out(mi) mediator(bmi) ex(pa) baseline(0)

base_confs(age) post_confs(smoke)

com(mi: logit, bmi: regress, smoke: logit)

eq(mi: bmi pa age smoke, bmi: pa age smoke, smoke: pa age)
```

gformula - output

	G-computation estimate	Bootstrap Std. Err.	z	P>|z|	Normal-based [95% Conf.	Interval]
TCE	-.008248	.0031653	-2.61	0.009	-.0144519	-.0020441
NDE	-.0058544	.0031253	-1.87	0.061	-.0119799	.0002712
NIE	-.0023937	.0012857	-1.86	0.063	-.0049135	.0001262

Note that ...

- effects are marginal

Summary

Further remarks

- other tools?
- causal effect measure choice
- sensitivity analyses

Summary

References (I)

- Baron RM and Kenny DA. The moderator mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6):1173-1182, 1986.
- Buis ML. Direct and indirect effects in a logit model. *The Stata Journal*, 10(1):1129, 2010.
- Daniel RM, De Stavola BL and Cousens SN. gformula: Estimating causal effects in the presence of time-varying confounding or mediation using the g-computation formula. *The Stata Journal*, 11(4):479-517, 2011.
- King G, Tomz M and Wittenberg J. Making the most of statistical analyses: Improving interpretation and presentation. American Journal of Political Science 44:347361, 2000.
- Hicks R and Tingley D. Causal Mediation Analysis. *The Stata Journal*, 11(4):609-15, 2011.
- Imai K, Keele L and Tingley D. A general approach to causal mediation analysis. *Psychological Methods*, 15:30934, 2010.
- VanderWeele TJ and Chiba Y. Sensitivity analysis for direct and indirect effects in the presence of exposure-induced mediator-outcome confounders. *Epidemiology Biostatistics and Public Health*, 11(2), 2010.

References (II)

- Pearl J. Direct and indirect effects. In *Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence*, UAI '01, pages 411-420, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers Inc.
- Robins JM. A new approach to causal inference in mortality studies with a sustained exposure period - application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7:1393-1512, 1986.
- Robins JM and Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology Cambridge Mass*, 3(2):143-155, 1992.
- Rosseel Y. Software for mediation analysis. *Symposium on Causal Mediation Analysis*, Ghent University. 2013.
- Valeri L and VanderWeele TJ. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. *Psychological Methods*, 18(2):137-50, 2013.
- VanderWeele TJ and Vansteelandt S. Conceptual issues concerning mediation, interventions and composition. *Statistics and its Interface*. 2:457-468, 2009.

Thank you