Causal Mediation Analysis

Kristin MacDonald

Executive Director, Statistical Services
StataCorp LLC

UK Stata Conference

Outline

- Introduction
- Overview of **mediate**
- Traditional mediation
- Causal inference
- Causal mediation analysis
- Examples

Introduction

Causal mediation analysis combines:

Introduction

Causal mediation analysis combines:

- Causal inference

Introduction

Causal mediation analysis combines:

- Causal inference
- Mediation analysis

Introduction

Causal mediation analysis combines:

- Causal inference—What is the effect of a treatment on an outcome?
- Mediation analysis

Introduction

Causal mediation analysis combines:

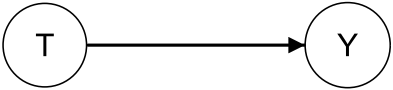
- Causal inference—What is the effect of a treatment on an outcome?
- Mediation analysis—Can the total effect of a predictor on an outcome be decomposed into a direct effect and an indirect effect through a mediating variable?

Causal inference

With causal inference, we want to answer questions about causality.

Causal inference

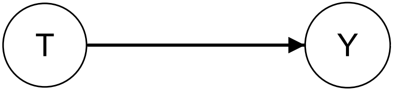
With causal inference, we want to answer questions about causality.



- What is the causal effect of **T** on **Y**?

Causal inference

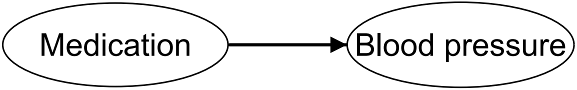
With causal inference, we want to answer questions about causality.



- What is the causal effect of **T** on **Y**?
- What is the expected difference in **Y** if the treatment **T** is applied versus if the treatment is not applied?

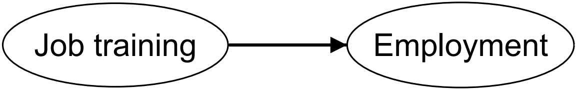
Causal inference

What is the effect of a medication on blood pressure?



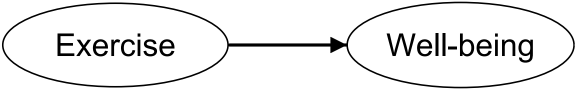
Causal inference

What is the effect of a job-training program on probability of employment?



Causal inference

What is the effect of exercise on self-perceived well-being?

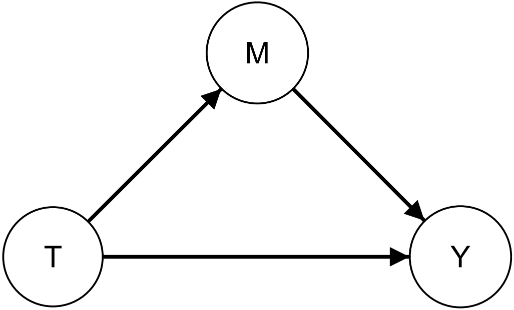


Mediation analysis

With mediation analysis, we want to better understand the effect.

Mediation analysis

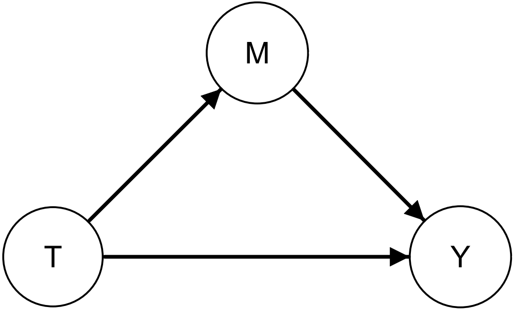
With mediation analysis, we want to better understand the effect.



- Why does **T** affect **Y**?

Mediation analysis

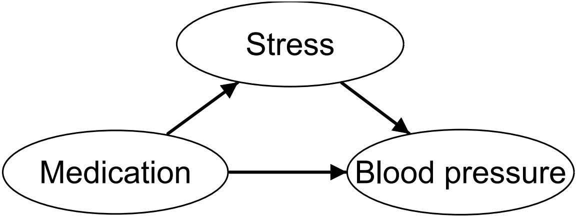
With mediation analysis, we want to better understand the effect.



- Why does **T** affect **Y**?
- Can effect of **T** on **Y** be explained either completely or partially by a change in the mediator **M**?

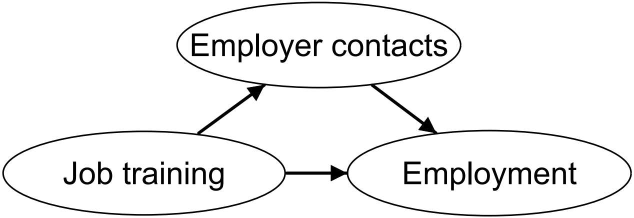
Mediation analysis

Does the medication result in lower stress levels, which in turn, results in lower blood pressure?



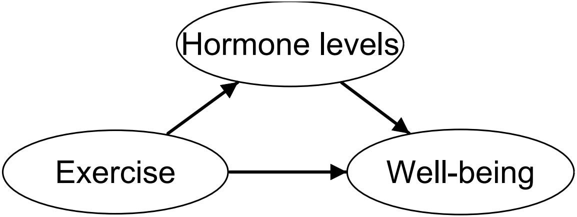
Mediation analysis

Does the job-training program put participants in contact with potential employers, which in turn, increases the probability of employment?



Mediation analysis

Does exercise change levels of some hormones, which in turn, change self-perceived well-being?



Causal mediation analysis

- With causal mediation analysis, we aim to draw causal inferences about the effect of a treatment on an outcome and to understand why the effect arises.

- To understand the why, we decompose the total effect into indirect effects through a mediator and direct effects.

Stata's **mediate** command

- In Stata 18, we introduced the **mediate** command to perform causal mediation analysis.

Stata's **mediate** command

- In Stata 18, we introduced the **mediate** command to perform causal mediation analysis.

```
mediate (ovar [omvarlist] [, omodel noconstant])
        (mvar [mmvarlist] [, mmodel noconsant])
        (tvar [, continuous(numlist)])
        [if] [in] [, stat options]
```

Stata's **mediate** command

- In Stata 18, we introduced the **mediate** command to perform causal mediation analysis.

```
mediate (ovar [omvarlist] [, omodel noconstant])
        (mvar [mmvarlist] [, mmodel noconsant])
        (tvar [, continuous(numlist)])
        [if] [in] [, stat options]
```

- *ovar* is a continuous, binary, or count outcome of interest.
- *mvar* is the mediator variable and may be continuous, binary, or count.
- *tvar* is the treatment variable and may be binary, multivalued, or continuous.

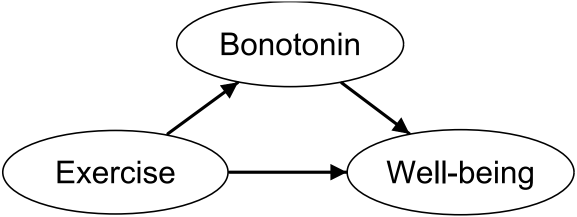
Stata's **mediate** command

Outcome \ Mediator	linear	logit	probit	Poisson	exp. mean
linear	X	X	X	X	X
logit		X	X	X	
probit	X	X	X	X	X
Poisson	X	X	X	X	X
exp. mean	X	X	X	X	X

Note: X indicates a supported model combination

A first example

For a simple example using the **mediate** command, we continue with our hypothesis that exercise affects well-being and that this may, at least in part, be because of a change in hormone levels. We will consider a fictional hormone **bonotonin**.



A first example

```
. webuse wellbeing
(Fictional well-being data)

. list wellbeing bonotonin exercise in 1/5, abbreviate(10)
```

	wellbeing	bonotonin	exercise
1.	71.73816	196.5467	Control
2.	68.66573	195.8572	Exercise
3.	71.05155	228.6035	Exercise
4.	69.44469	206.6651	Exercise
5.	75.62035	261.6855	Exercise

- Both **wellbeing** and **bonotonin** are continuous, so we use the default linear model for the outcome and the mediator.
- **exercise** is a binary treatment variable with 0 representing the control group and 1 representing exercise group.

A first example

```
. mediate (wellbeing) (bonotonin) (exercise)

Iteration 0:  EE criterion =  5.104e-27
Iteration 1:  EE criterion =  2.031e-28

Causal mediation analysis                          Number of obs = 2,000

Outcome model:     Linear
Mediator model:    Linear
Mediator variable: bonotonin
Treatment type:    Binary
-----------------------------------------------------------------------------
             |               Robust
   wellbeing | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+---------------------------------------------------------------
NIE          |
    exercise |
   (Exercise |
          vs |
    Control) |   9.799821   .3943251    24.85   0.000     9.026958    10.57268
-------------+---------------------------------------------------------------
NDE          |
    exercise |
   (Exercise |
          vs |
    Control) |   2.891453   .2304278    12.55   0.000     2.439823    3.343083
-------------+---------------------------------------------------------------
TE           |
    exercise |
   (Exercise |
          vs |
    Control) |   12.69127   .4005941    31.68   0.000     11.90612    13.47642
-----------------------------------------------------------------------------
Note: Outcome equation includes treatment-mediator interaction.
```

- We estimate the total effect of exercise on well-being is 12.7, with an indirect effect of 9.8 and a direct effect of 2.9.

A first example

- What proportion of the total effect exercise on well-being is mediated through bonotonin levels?

```
. estat proportion

Proportion mediated                                    Number of obs = 2,000

------------------------------------------------------------------------------
             |                Robust
   wellbeing | Proportion   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
    exercise |
   (Exercise |
          vs |
    Control) |     .77217   .0172979    44.64   0.000     .7382668    .8060732
------------------------------------------------------------------------------
```

Traditional mediation analysis: The formulation

- In traditional mediation analysis, we write models for the outcome and the mediator.

$$Y = \beta_0 + \beta_1 M + \beta_2 T + \epsilon$$
$$M = \alpha_0 + \alpha_1 T + \nu$$

- Then we define direct, indirect, and total effects as

$$\begin{aligned} \textit{Direct} &= \beta_2 \\ \textit{Indirect} &= \alpha_1 * \beta_1 \\ \textit{Total} &= \beta_2 + \alpha_1 * \beta_1 \end{aligned}$$

Traditional mediation analysis: Estimation

- We can fit the linear regression models using **regress** and then manually compute direct, indirect, and total effects and the corresponding standard errors.
- Alternatively, we can fit models simultaneously using, for instance, the **sem** command.

```
. sem (wellbeing <- bonotonin exercise)
      (bonotonin <- exercise)
```

After **sem**, we can use **estat teffects** to compute direct, indirect, and total effects and the corresponding standard errors.

Traditional mediation analysis: Estimation

```
. estat teffects

Direct effects
------------------------------------------------------------------------------
             |                 OIM
             | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
Structural   |
  wellbeing  |
...
    exercise |   2.996658   .2119904    14.14   0.000     2.581164    3.412151
...
------------------------------------------------------------------------------

Indirect effects
------------------------------------------------------------------------------
             |                 OIM
             | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
Structural   |
  wellbeing  |
...
    exercise |   9.694617   .3717311    26.08   0.000     8.966037     10.4232
...
------------------------------------------------------------------------------

Total effects
------------------------------------------------------------------------------
             |                 OIM
             | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
Structural   |
  wellbeing  |
...
    exercise |   12.69127   .3983777    31.86   0.000     11.91047    13.47208
...
------------------------------------------------------------------------------
```

Traditional mediation analysis

- Traditional mediation analysis began with linear models for both the outcome and the mediator.
- The model for the outcome did not include a mediator by treatment interaction term.
- What can a causal inference approach add?

Causal inference workflow

Common steps in a causal inference approach:

Causal inference workflow

Common steps in a causal inference approach:

1. **Hypothetical modeling.** Researchers make assumptions about relationships among variables based on their understanding and expertise. These assumptions may be illustrated by using a causal diagram.

Causal inference workflow

Common steps in a causal inference approach:

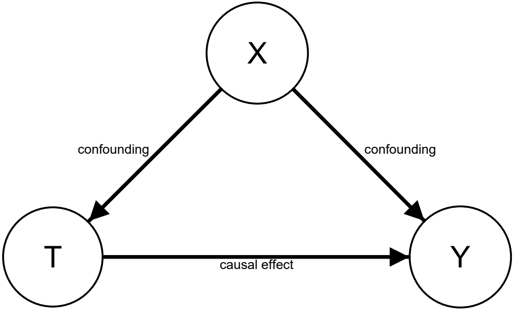
1. **Hypothetical modeling.** Researchers make assumptions about relationships among variables based on their understanding and expertise. These assumptions may be illustrated by using a causal diagram.
2. **Causal effect identification.** Based on the assumptions made in the first phase, the researcher tries to determine whether the causal effect can be identified.

Causal inference workflow

Common steps in a causal inference approach:

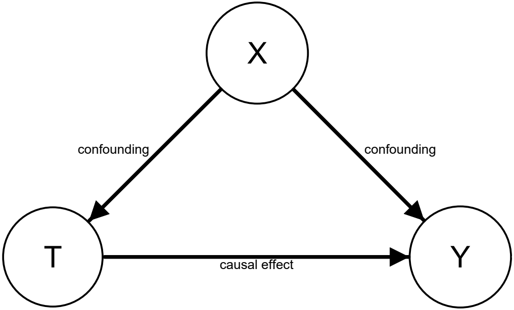
1. **Hypothetical modeling.** Researchers make assumptions about relationships among variables based on their understanding and expertise. These assumptions may be illustrated by using a causal diagram.
2. **Causal effect identification.** Based on the assumptions made in the first phase, the researcher tries to determine whether the causal effect can be identified.
3. **Parameter estimation.** If the answer to the second phase is positive, the researcher can then to estimate the causal effect. Stata provides a variety of commands such as the **teffects** suite to estimate average treatment effects (ATEs), average treatment effects on the treated (ATETs), and other causal estimands of interest.

Causal diagram



- In this very simple causal diagram, we are interested in estimating the causal effect of treatment **T** on outcome **Y**, but we believe that **X** also affects both **T** and **Y**.

Causal diagram



- In this very simple causal diagram, we are interested in estimating the causal effect of treatment **T** on outcome **Y**, but we believe that **X** also affects both **T** and **Y**.
- **X** is a counfounder, and we must somehow control for confounding to obtain an unbiased estimate of the causal effect.

Defining a causal effect

How do we define a causal effect?

Defining a causal effect

How do we define a causal effect?

- A common causal-inference approach is based on the potential-outcomes framework.

Defining a causal effect

How do we define a causal effect?

- A common causal-inference approach is based on the potential-outcomes framework.
- For a binary treatment T, we can define two potential outcomes.
  - $Y(0)$ is the potential outcome that would have been observed if treatment $T = 0$ was assigned.
  - $Y(1)$ is the potential outcome that would have been observed if treatment $T = 1$ was assigned.

Defining a causal effect

How do we define a causal effect?

- A common causal-inference approach is based on the potential-outcomes framework.
- For a binary treatment T, we can define two potential outcomes.
  - $Y(0)$ is the potential outcome that would have been observed if treatment $T = 0$ was assigned.
  - $Y(1)$ is the potential outcome that would have been observed if treatment $T = 1$ was assigned.
- The individual treatment effect is the difference in the two potential outcomes, $Y(1) - Y(0)$.

Defining a causal effect

How do we define a causal effect?

- A common causal-inference approach is based on the potential-outcomes framework.
- For a binary treatment T, we can define two potential outcomes.
  - $Y(0)$ is the potential outcome that would have been observed if treatment $T = 0$ was assigned.
  - $Y(1)$ is the potential outcome that would have been observed if treatment $T = 1$ was assigned.
- The individual treatment effect is the difference in the two potential outcomes, $Y(1) - Y(0)$.
- The average treatment effect (ATE) is $E[Y(1) - Y(0)]$.

Fundamental problem of causal inference

Subject	T	Y	Y(1)	Y(0)	Y(1) – Y(0)
1	0	2.1	?	2.1	?
2	1	3.7	3.7	?	?
3	1	4.2	4.2	?	?
4	0	6.2	?	6.2	?
…	…	…	…	…	…

For each individual, we can observe only one of $Y(1)$ or $Y(0)$.

Identifying a causal effect

Can we estimate the ATE given that $Y(1) - Y(0)$ is never observed?

Identifying a causal effect

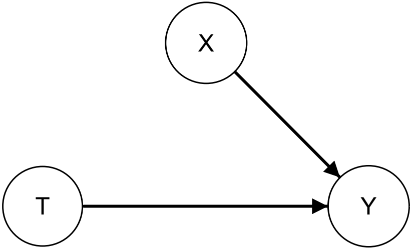
Can we estimate the ATE given that $Y(1) - Y(0)$ is never observed?

- What if we have a randomized control trial (RCT)?

Identifying a causal effect

Can we estimate the ATE given that $Y(1) - Y(0)$ is never observed?

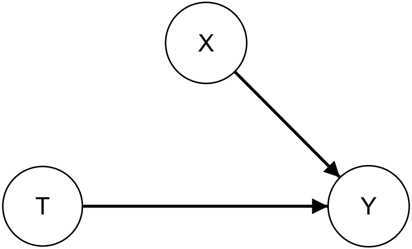
- What if we have a randomized control trial (RCT)?
- In an RCT, we randomize the treatment; therefore, T is independent of $Y(0)$, $Y(1)$, and X



Identifying a causal effect

Can we estimate the ATE given that $Y(1) - Y(0)$ is never observed?

- What if we have a randomized control trial (RCT)?
- In an RCT, we randomize the treatment; therefore, T is independent of $Y(0)$, $Y(1)$, and X

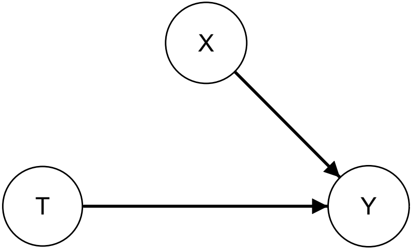


- In this case $E[Y(0)] = E[Y|T = 0]$ and $E[Y(1)] = E[Y|T = 1]$.

Identifying a causal effect

Can we estimate the ATE given that $Y(1) - Y(0)$ is never observed?

- What if we have a randomized control trial (RCT)?
- In an RCT, we randomize the treatment; therefore, T is independent of $Y(0)$, $Y(1)$, and X



- In this case $E[Y(0)] = E[Y|T = 0]$ and $E[Y(1)] = E[Y|T = 1]$.
- We can estimate the ATE as $E[Y|T = 1] - E[Y|T = 0]$.

Identifying a causal effect

Can we estimate the ATE given that $Y(1) - Y(0)$ is never observed?

Identifying a causal effect

Can we estimate the ATE given that $Y(1) - Y(0)$ is never observed?

- What if we have observational data?

Identifying a causal effect

Can we estimate the ATE given that $Y(1) - Y(0)$ is never observed?

- What if we have observational data?
- The treatment T is not independent of $Y(0)$, $Y(1)$, and X

Identifying a causal effect

Can we estimate the ATE given that $Y(1) - Y(0)$ is never observed?

- What if we have observational data?
- The treatment T is not independent of $Y(0)$, $Y(1)$, and X
- To identify the ATE, we need to make some assumptions.

Identifying a causal effect

Can we estimate the ATE given that $Y(1) - Y(0)$ is never observed?

- What if we have observational data?
- The treatment T is not independent of $Y(0)$, $Y(1)$, and X
- To identify the ATE, we need to make some assumptions.
  - Unconfoundedness or conditional independence. If we condition on confounders X, the treatment assignment is as good as random.

Identifying a causal effect

Can we estimate the ATE given that $Y(1) - Y(0)$ is never observed?

- What if we have observational data?
- The treatment T is not independent of $Y(0)$, $Y(1)$, and X
- To identify the ATE, we need to make some assumptions.
  - Unconfoundedness or conditional independence. If we condition on confounders X, the treatment assignment is as good as random.
  - Stable unit treatment value assumption (SUTVA). The treatment of each individual is unrelated to the outcome of the treatment of all the other individuals in the population.

Identifying a causal effect

Can we estimate the ATE given that $Y(1) - Y(0)$ is never observed?

- What if we have observational data?
- The treatment T is not independent of $Y(0)$, $Y(1)$, and X
- To identify the ATE, we need to make some assumptions.
  - Unconfoundedness or conditional independence. If we condition on confounders X, the treatment assignment is as good as random.
  - Stable unit treatment value assumption (SUTVA). The treatment of each individual is unrelated to the outcome of the treatment of all the other individuals in the population.
  - Overlap. Each individual has a positive probability of receiving each treatment level.

Identifying a causal effect

Can we estimate the ATE given that $Y(1) - Y(0)$ is never observed?

- What if we have observational data?
- The treatment T is not independent of $Y(0)$, $Y(1)$, and X
- To identify the ATE, we need to make some assumptions.
  - Unconfoundedness or conditional independence. If we condition on confounders X, the treatment assignment is as good as random.
  - Stable unit treatment value assumption (SUTVA). The treatment of each individual is unrelated to the outcome of the treatment of all the other individuals in the population.
  - Overlap. Each individual has a positive probability of receiving each treatment level.
- Now, the causal effect is identified:

$$E[Y(1)] - E[Y(0)] = E_x[E[Y|T=1,X] - E[Y|T=0,X]]$$

Identifying a causal effect

Can we estimate the ATE given that $Y(1) - Y(0)$ is never observed?

- What if we have observational data?
- The treatment T is not independent of $Y(0)$, $Y(1)$, and X
- To identify the ATE, we need to make some assumptions.
  - Unconfoundedness or conditional independence. If we condition on confounders X, the treatment assignment is as good as random.
  - Stable unit treatment value assumption (SUTVA). The treatment of each individual is unrelated to the outcome of the treatment of all the other individuals in the population.
  - Overlap. Each individual has a positive probability of receiving each treatment level.
- Now, the causal effect is identified:

  $$E[Y(1)] - E[Y(0)] = E_x[E[Y|T = 1, X] - E[Y|T = 0, X]]$$
- We can use commands such as **teffects** to estimate the ATE.

Estimating the causal effect

```
. teffects ra (wellbeing age) (exercise)
Iteration 0:   EE criterion =  1.261e-27
Iteration 1:   EE criterion =  1.707e-29
Treatment-effects estimation                    Number of obs     =      2,000
Estimator      : regression adjustment
Outcome model  : linear
Treatment model: none
------------------------------------------------------------------------------
             |               Robust
   wellbeing | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
ATE          |
    exercise |
  (Exercise  |
         vs  |
   Control)  |   12.76801   .3961873    32.23   0.000      11.9915    13.54452
-------------+----------------------------------------------------------------
POmean       |
    exercise |
     Control |   57.06904   .2738341   208.41   0.000     56.53234    57.60575
------------------------------------------------------------------------------
```

Mediation analysis via potential outcomes

Mediation analysis via potential outcomes

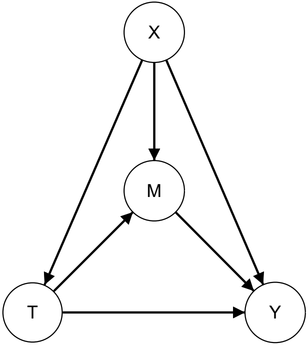
- We can extend the potential-outcomes framework to mediation analysis.

Mediation analysis via potential outcomes

- We can extend the potential-outcomes framework to mediation analysis.
- We can define a total average treatment effect as well as direct and indirect effects in terms of potential outcomes.

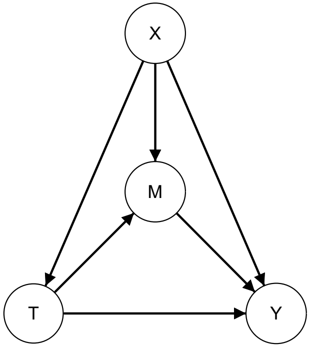
Mediation analysis: Causal inference workflow

- A causal diagram



Mediation analysis: Causal inference workflow

- A causal diagram



- As before, we will make assumptions that allow us to get unbiased estimates of the causal effects, even in the presence of confounders.

Potential outcomes

- We now have potential outcomes for the the mediator M and for the outcome Y.

Potential outcomes

- We now have potential outcomes for the the mediator M and for the outcome Y.
- For the mediator, we have
  - $M(0)$ is the potential outcome of the mediator that would have been observed if treatment $T = 0$ was assigned.
  - $M(1)$ is the potential outcome of the mediator that would have been observed if treatment $T = 1$ was assigned.

Potential outcomes

- Formally, let t be the treatment level with respect to the outcome, and let t' be the treatment level with respect to the mediator, the potential outcomes become $Y[t, M(t')]$.

Potential outcomes

- Formally, let t be the treatment level with respect to the outcome, and let t' be the treatment level with respect to the mediator, the potential outcomes become $Y[t, M(t')]$.
- This leads to four types of potential outcomes:

Potential outcomes

- Formally, let t be the treatment level with respect to the outcome, and let t' be the treatment level with respect to the mediator, the potential outcomes become $Y[t, M(t')]$.
- This leads to four types of potential outcomes:
  - $Y[0, M(0)]$ is the potential outcome that would be observed if treatment $T = 0$ was assigned.

Potential outcomes

- Formally, let t be the treatment level with respect to the outcome, and let t' be the treatment level with respect to the mediator, the potential outcomes become $Y[t, M(t')]$.
- This leads to four types of potential outcomes:
  - $Y[0, M(0)]$ is the potential outcome that would be observed if treatment $T = 0$ was assigned.
  - $Y[1, M(1)]$ is the potential outcome that would be observed if treatment $T = 1$ was assigned.

Potential outcomes

- Formally, let t be the treatment level with respect to the outcome, and let t' be the treatment level with respect to the mediator, the potential outcomes become $Y[t, M(t')]$.
- This leads to four types of potential outcomes:
  - $Y[0, M(0)]$ is the potential outcome that would be observed if treatment $T = 0$ was assigned.
  - $Y[1, M(1)]$ is the potential outcome that would be observed if treatment $T = 1$ was assigned.
  - $Y[1, M(0)]$ is the potential outcome that would be observed if treatment $T = 1$ was assigned, but the mediator is held at its value that would be observed if if $T = 0$ was assigned.

Potential outcomes

- Formally, let t be the treatment level with respect to the outcome, and let t' be the treatment level with respect to the mediator, the potential outcomes become $Y[t, M(t')]$.
- This leads to four types of potential outcomes:
  - $Y[0, M(0)]$ is the potential outcome that would be observed if treatment $T = 0$ was assigned.
  - $Y[1, M(1)]$ is the potential outcome that would be observed if treatment $T = 1$ was assigned.
  - $Y[1, M(0)]$ is the potential outcome that would be observed if treatment $T = 1$ was assigned, but the mediator is held at its value that would be observed if if $T = 0$ was assigned.
  - $Y[0, M(1)]$ is the potential outcome that would be observed if treatment $T = 0$ was assigned, but the mediator is held at its value that would be observed if if $T = 1$ was assigned.

Potential outcomes

```
. mediate (wellbeing) (bonotonin) (exercise), pomeans
Iteration 0:  EE criterion =  5.104e-27
Iteration 1:  EE criterion =  2.023e-28
Causal mediation analysis                            Number of obs = 2,000
Outcome model:      Linear
Mediator model:     Linear
Mediator variable: bonotonin
Treatment type:     Binary
```

wellbeing	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
POmeans						
Y0M0	57.11317	.2753201	207.44	0.000	56.57355	57.65278
Y1M0	60.00462	.3157888	190.02	0.000	59.38569	60.62356
Y0M1	66.68199	.3258477	204.64	0.000	66.04334	67.32064
Y1M1	69.80444	.2898927	240.79	0.000	69.23626	70.37262

Note: Outcome equation includes treatment-mediator interaction.

Direct, indirect, and total treatment effects

Direct, indirect, and total treatment effects

- Average direct, indirect, and total treatment effects are contrasts between potential-outcome means.

Direct, indirect, and total treatment effects

- Average direct, indirect, and total treatment effects are contrasts between potential-outcome means.
- The total effect is:

$$\tau \equiv E[Y(1)] - E[Y(0)] = E[Y(1, M(1))] - E[Y(0, M(0))]$$

Direct, indirect, and total treatment effects

- Average direct, indirect, and total treatment effects are contrasts between potential-outcome means.
- The total effect is:
$$\tau \equiv E[Y(1)] - E[Y(0)] = E[Y(1, M(1))] - E[Y(0, M(0))]$$
- The effect of the treatment on the outcome through the mediator is the indirect effect:
$$\delta(t) \equiv E[Y(t, M(1))] - E[Y(t, M(0))], \quad t \in \{0, 1\}$$

Direct, indirect, and total treatment effects

- Average direct, indirect, and total treatment effects are contrasts between potential-outcome means.
- The total effect is:

$$\tau \equiv E[Y(1)] - E[Y(0)] = E[Y(1, M(1))] - E[Y(0, M(0))]$$

- The effect of the treatment on the outcome through the mediator is the indirect effect:

$$\delta(t) \equiv E[Y(t, M(1))] - E[Y(t, M(0))], \quad t \in \{0, 1\}$$

- The direct effect of the treatment is:

$$\zeta(t) \equiv E[Y(1, M(t))] - E[Y(0, M(t))], \quad t \in \{0, 1\}$$

Direct, indirect, and total treatment effects

- Average direct, indirect, and total treatment effects are contrasts between potential-outcome means.
- The total effect is:

$$\tau \equiv E[Y(1)] - E[Y(0)] = E[Y(1, M(1))] - E[Y(0, M(0))]$$

- The effect of the treatment on the outcome through the mediator is the indirect effect:

$$\delta(t) \equiv E[Y(t, M(1))] - E[Y(t, M(0))], \quad t \in \{0, 1\}$$

- The direct effect of the treatment is:

$$\zeta(t) \equiv E[Y(1, M(t))] - E[Y(0, M(t))], \quad t \in \{0, 1\}$$

- Notice that the total effect is the sum of direct and indirect effects

$$\tau = \delta(0) + \zeta(1)$$
$$\tau = \delta(1) + \zeta(0)$$

Estimands

- Denoting $E[Y(t, M(t'))]$ as $Y_{tM_{t'}}$, we define the following treatment effects of interest

(Total) natural indirect effect (**NIE**)	$Y_{1M_1} - Y_{1M_0}$
(Pure) natural direct effect (**NDE**)	$Y_{1M_0} - Y_{0M_0}$
(Pure) natural indirect effect (**PNIE**)	$Y_{0M_1} - Y_{0M_0}$
(Total) natural direct effect (**TNDE**)	$Y_{1M_1} - Y_{0M_1}$
Total effect (**TE**)	$Y_{1M_1} - Y_{0M_0}$

Decomposition 1

```
. mediate (wellbeing) (bonotonin) (exercise)

Iteration 0:  EE criterion =  5.104e-27
Iteration 1:  EE criterion =  2.031e-28

Causal mediation analysis                               Number of obs = 2,000

Outcome model:      Linear
Mediator model:     Linear
Mediator variable: bonotonin
Treatment type:     Binary
------------------------------------------------------------------------------
             |               Robust
   wellbeing | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
NIE          |
    exercise |
   (Exercise |
          vs |
    Control) |   9.799821   .3943251    24.85   0.000     9.026958    10.57268
-------------+----------------------------------------------------------------
NDE          |
    exercise |
   (Exercise |
          vs |
    Control) |   2.891453   .2304278    12.55   0.000     2.439823    3.343083
-------------+----------------------------------------------------------------
TE           |
    exercise |
   (Exercise |
          vs |
    Control) |   12.69127   .4005941    31.68   0.000     11.90612    13.47642
------------------------------------------------------------------------------
Note: Outcome equation includes treatment-mediator interaction.
```

Decomposition 2

```
. mediate (wellbeing) (bonotonin) (exercise), pnie tnde te

Iteration 0:  EE criterion =  5.104e-27
Iteration 1:  EE criterion =  3.672e-28

Causal mediation analysis                              Number of obs = 2,000

Outcome model:      Linear
Mediator model:     Linear
Mediator variable:  bonotonin
Treatment type:     Binary
------------------------------------------------------------------------------
             |               Robust
   wellbeing | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
PNIE         |
    exercise |
   (Exercise |
          vs |
    Control) |   9.568827   .3884522    24.63   0.000     8.807475    10.33018
-------------+----------------------------------------------------------------
TNDE         |
    exercise |
   (Exercise |
          vs |
    Control) |   3.122447   .2418591    12.91   0.000     2.648412    3.596482
-------------+----------------------------------------------------------------
TE           |
    exercise |
   (Exercise |
          vs |
    Control) |   12.69127   .4005941    31.68   0.000     11.90612    13.47642
------------------------------------------------------------------------------
Note: Outcome equation includes treatment-mediator interaction
```

Chosing a decomposition

Which decomposition do we want?

Nguyen, Schmid, and Stuart (2021) give suggestions for three scenarios:

Chosing a decomposition

Which decomposition do we want?

Nguyen, Schmid, and Stuart (2021) give suggestions for three scenarios:

1. Want know whether there is a mediation effect? Use NIE and NDE.

Chosing a decomposition

Which decomposition do we want?

Nguyen, Schmid, and Stuart (2021) give suggestions for three scenarios:

1. Want know whether there is a mediation effect? Use NIE and NDE.

   We are assuming there is some direct effect and we want to know whether any mediating effect also exists.

Chosing a decomposition

Which decomposition do we want?

Nguyen, Schmid, and Stuart (2021) give suggestions for three scenarios:

1. Want know whether there is a mediation effect? Use NIE and NDE.

   We are assuming there is some direct effect and we want to know whether any mediating effect also exists.

2. Want to know whether any direct effect exists in addition to a mediation effect? Use PNIE and TNDE.

Chosing a decomposition

Which decomposition do we want?

Nguyen, Schmid, and Stuart (2021) give suggestions for three scenarios:

1. Want know whether there is a mediation effect? Use NIE and NDE.

   We are assuming there is some direct effect and we want to know whether any mediating effect also exists.

2. Want to know whether any direct effect exists in addition to a mediation effect? Use PNIE and TNDE.

   We are assuming that some mediating effect exists and want to know whether there is an affect through any other mechanisms.

Chosing a decomposition

Which decomposition do we want?

Nguyen, Schmid, and Stuart (2021) give suggestions for three scenarios:

1. Want know whether there is a mediation effect? Use NIE and NDE.

   We are assuming there is some direct effect and we want to know whether any mediating effect also exists.

2. Want to know whether any direct effect exists in addition to a mediation effect? Use PNIE and TNDE.

   We are assuming that some mediating effect exists and want to know whether there is an affect through any other mechanisms.

3. Have no prior assumption about whether direct or indirect effects exist? Use both decompositions.

Chosing a decomposition

Which decomposition do we want?

Nguyen, Schmid, and Stuart (2021) give suggestions for three scenarios:

1. Want know whether there is a mediation effect? Use NIE and NDE.

   We are assuming there is some direct effect and we want to know whether any mediating effect also exists.

2. Want to know whether any direct effect exists in addition to a mediation effect? Use PNIE and TNDE.

   We are assuming that some mediating effect exists and want to know whether there is an affect through any other mechanisms.

3. Have no prior assumption about whether direct or indirect effects exist? Use both decompositions.

   We describe try to learn all we can from all decompositions.

Behind the scenes

- **mediate** estimates all effects parameters, auxiliary parameters, and their variance–covariance matrix via generalized method of moments.
- We can specify **aequations** option to see estimated auxiliary parameters—the parameters estimated for the outcome and treatment models.

Auxiliary parameter estimates

```
. mediate (wellbeing) (bonotonin) (exercise), aequations

Iteration 0:  EE criterion =  5.104e-27
Iteration 1:  EE criterion =  2.031e-28

Causal mediation analysis                              Number of obs = 2,000

Outcome model:      Linear
Mediator model:     Linear
Mediator variable: bonotonin
Treatment type:     Binary
```

wellbeing	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
NIE						
exercise (Exercise vs Control)	9.799821	.3943251	24.85	0.000	9.026958	10.57268
NDE						
exercise (Exercise vs Control)	2.891453	.2304278	12.55	0.000	2.439823	3.343083
TE						
exercise (Exercise vs Control)	12.69127	.4005941	31.68	0.000	11.90612	13.47642

Auxiliary parameter estimates

wellbeing						
exercise						
Exercise	2.065871	.8723559	2.37	0.018	.3560846	3.775657
bonotonin	.2130222	.0034547	61.66	0.000	.2062512	.2197932
exercise#						
c.bonotonin						
Exercise	.0051424	.0046954	1.10	0.273	-.0040604	.0143452
_cons	22.91374	.5633648	40.67	0.000	21.80956	24.01791
bonotonin						
exercise						
Exercise	44.91939	1.641668	27.36	0.000	41.70178	48.137
_cons	160.544	1.142508	140.52	0.000	158.3047	162.7832

Note: Outcome equation includes treatment-mediator interaction.

Causal identification and assumptions

- We defined our effects of interest in terms of potential-outcome means.

Causal identification and assumptions

- We defined our effects of interest in terms of potential-outcome means.
- We need to consider what causal assumptions are required to identify those effects.

Causal identification and assumptions

- We defined our effects of interest in terms of potential-outcome means.
- We need to consider what causal assumptions are required to identify those effects.
- The potential-outcome means are the result of an integral of the conditional expectation of the outcome with respect to the conditional distribution of the mediator:

$$f[Y(t, M(t'))|\boldsymbol{X} = \boldsymbol{x}] = \int f[Y|M = m, T = t, \boldsymbol{X} = \boldsymbol{x}]\, dF[m|T = t', \boldsymbol{X} = \boldsymbol{x}]$$

Causal identification and assumptions

- We defined our effects of interest in terms of potential-outcome means.
- We need to consider what causal assumptions are required to identify those effects.
- The potential-outcome means are the result of an integral of the conditional expectation of the outcome with respect to the conditional distribution of the mediator:

$$f[Y(t, M(t'))|\boldsymbol{X} = \boldsymbol{x}] = \int f[Y|M = m, T = t, \boldsymbol{X} = \boldsymbol{x}]\, dF[m|T = t', \boldsymbol{X} = \boldsymbol{x}]$$

- This is sometimes referred to as the "mediation formula".

Causal identification and assumptions

- We defined our effects of interest in terms of potential-outcome means.
- We need to consider what causal assumptions are required to identify those effects.
- The potential-outcome means are the result of an integral of the conditional expectation of the outcome with respect to the conditional distribution of the mediator:

$$f[Y(t, M(t'))|\boldsymbol{X} = \boldsymbol{x}] = \int f[Y|M = m, T = t, \boldsymbol{X} = \boldsymbol{x}]\, dF[m|T = t', \boldsymbol{X} = \boldsymbol{x}]$$

- This is sometimes referred to as the "mediation formula".
- Notice that this is a nonparametric identification result.

Causal identification and assumptions

- In addition to assumptions for standard SUTVA and overlap assumptions, we need to make an assumption of sequential ignorability.

Causal identification and assumptions

- In addition to assumptions for standard SUTVA and overlap assumptions, we need to make an assumption of sequential ignorability.
- Sequential ignorability essentially means

Causal identification and assumptions

- In addition to assumptions for standard SUTVA and overlap assumptions, we need to make an assumption of sequential ignorability.
- Sequential ignorability essentially means
  - No unobserved confounding in the treatment-outcome relationship.

Causal identification and assumptions

- In addition to assumptions for standard SUTVA and overlap assumptions, we need to make an assumption of sequential ignorability.
- Sequential ignorability essentially means
  - No unobserved confounding in the treatment-outcome relationship.
  - No unobserved confounding in the mediator-outcome relationship.

Causal identification and assumptions

- In addition to assumptions for standard SUTVA and overlap assumptions, we need to make an assumption of sequential ignorability.
- Sequential ignorability essentially means
  - No unobserved confounding in the treatment-outcome relationship.
  - No unobserved confounding in the mediator-outcome relationship.
  - No unmeasured confounding in the treatment-mediator relationship.

Causal identification and assumptions

- In addition to assumptions for standard SUTVA and overlap assumptions, we need to make an assumption of sequential ignorability.
- Sequential ignorability essentially means
  - No unobserved confounding in the treatment-outcome relationship.
  - No unobserved confounding in the mediator-outcome relationship.
  - No unmeasured confounding in the treatment-mediator relationship.
  - There are no (observed) confounders in the mediator-outcome relationship that are caused by the treatment.

Causal identification and assumptions

- To meet these assumptions, we may need to add covariates to the model for the outcome, the model for the mediator, or both.
- Here, we adjust for **age** in both models before estimating potential-outcome means and the effects.

Including covariates

```
. mediate (wellbeing age) (bonotonin age) (exercise)

Iteration 0:   EE criterion =  6.163e-27
Iteration 1:   EE criterion =  4.924e-29

Causal mediation analysis                              Number of obs = 2,000

Outcome model:       Linear
Mediator model:      Linear
Mediator variable: bonotonin
Treatment type:      Binary
------------------------------------------------------------------------------
             |                Robust
   wellbeing | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
NIE          |
    exercise |
   (Exercise |
          vs |
    Control) |   9.851525   .3905733    25.22   0.000     9.086015    10.61703
-------------+----------------------------------------------------------------
NDE          |
    exercise |
   (Exercise |
          vs |
    Control) |   2.915712   .2327821    12.53   0.000     2.459468    3.371957
-------------+----------------------------------------------------------------
TE           |
    exercise |
   (Exercise |
          vs |
    Control) |   12.76724   .3964534    32.20   0.000      11.9902    13.54427
------------------------------------------------------------------------------
```

Controlled direct effects

- What would be the causal effect if we could set the mediator to a specific value? To explore this, we estimate a controlled direct effect (CDE).

Controlled direct effects

- What would be the causal effect if we could set the mediator to a specific value? To explore this, we estimate a controlled direct effect (CDE).
- To estimate CDEs, we use only the results of the outcome equation.

Controlled direct effects

- What would be the causal effect if we could set the mediator to a specific value? To explore this, we estimate a controlled direct effect (CDE).
- To estimate CDEs, we use only the results of the outcome equation.
- For a binary treatment, we now have potential outcomes of the form $Y(0|M = m)$. and $Y(1|M = m)$ where m is the specified value of the mediator.

Controlled direct effects

- What would be the causal effect if we could set the mediator to a specific value? To explore this, we estimate a controlled direct effect (CDE).
- To estimate CDEs, we use only the results of the outcome equation.
- For a binary treatment, we now have potential outcomes of the form $Y(0|M = m)$. and $Y(1|M = m)$ where m is the specified value of the mediator.
- CDE(m) is then the average of the differences between potential outcomes.

Controlled direct effects

- What would be the causal effect if we could set the mediator to a specific value? To explore this, we estimate a controlled direct effect (CDE).
- To estimate CDEs, we use only the results of the outcome equation.
- For a binary treatment, we now have potential outcomes of the form $Y(0|M=m)$. and $Y(1|M=m)$ where m is the specified value of the mediator.
- CDE(m) is then the average of the differences between potential outcomes.
- For binary treatment, CDE(m) is defined as $Y(1|M=m) - Y(0|M=m)$.

Controlled direct effects

- What would be the causal effect if we could set the mediator to a specific value? To explore this, we estimate a controlled direct effect (CDE).
- To estimate CDEs, we use only the results of the outcome equation.
- For a binary treatment, we now have potential outcomes of the form $Y(0|M=m)$. and $Y(1|M=m)$ where m is the specified value of the mediator.
- CDE(m) is then the average of the differences between potential outcomes.
- For binary treatment, CDE(m) is defined as $Y(1|M=m) - Y(0|M=m)$.
- Perhaps we want to know the effect of exercise on well-being if we had a medication that stabilized bonotonin levels at 200 for everyone in the population.

Controlled direct effects

```
. estat cde, mvalue(200)
Controlled direct effect                              Number of obs = 2,000
Mediator variable: bonotonin
Mediator value = 200
```

	CDE	Delta-method std. err.	z	P>|z|	[95% conf.	interval]
exercise (Exercise vs Control)	3.121577	.2315869	13.48	0.000	2.667675	3.575479

Modifying outcome and mediator models

What if we have a different type of outcome, mediator, or treatment?

Modifying outcome and mediator models

What if we have a different type of outcome, mediator, or treatment?

- **mediate** allows a continuous, binary, or count outcome. You can specify a linear, logit, probit, Poisson, or exponential mean model for the outcome.

Modifying outcome and mediator models

What if we have a different type of outcome, mediator, or treatment?

- **mediate** allows a continuous, binary, or count outcome. You can specify a linear, logit, probit, Poisson, or exponential mean model for the outcome.
- **mediate** allows a continuous, binary, or count mediator. You can specify a linear, logit, probit, Poisson, or exponential mean model for the mediator.

Modifying outcome and mediator models

What if we have a different type of outcome, mediator, or treatment?

- **mediate** allows a continuous, binary, or count outcome. You can specify a linear, logit, probit, Poisson, or exponential mean model for the outcome.
- **mediate** allows a continuous, binary, or count mediator. You can specify a linear, logit, probit, Poisson, or exponential mean model for the mediator.
- **mediate** allows a binary, multivalued, a continuous treatment.

Modifying outcome and mediator models

What if we have a different type of outcome, mediator, or treatment?

- **mediate** allows a continuous, binary, or count outcome. You can specify a linear, logit, probit, Poisson, or exponential mean model for the outcome.
- **mediate** allows a continuous, binary, or count mediator. You can specify a linear, logit, probit, Poisson, or exponential mean model for the mediator.
- **mediate** allows a binary, multivalued, a continuous treatment.

Logit models

- To demonstrate, we model a binary mediator, **bbonotonin**, which is an indicator for at least 10% increase in bonotonin over the baseline level.
- We also have a binary outcome, **bwellbeing**, which is an indicator for at least 10% improvement in well-being over the baseline value.
- We will fit a logit model for both the outcome and mediator and estimate the effects of interest using the same definitions based on potential-outcome means.

Logit models

```
. mediate (bwellbeing basewell age, logit)
>         (bbonotonin basebono age, logit)
>         (exercise), nointeraction

Iteration 0:  EE criterion =  4.840e-18
Iteration 1:  EE criterion =  1.836e-33

Causal mediation analysis                              Number of obs = 2,000

Outcome model:      Logit
Mediator model:     Logit
Mediator variable: bbonotonin
Treatment type:     Binary
-------------------------------------------------------------------------------
             |              Robust
  bwellbeing | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+-----------------------------------------------------------------
NIE          |
    exercise |
   (Exercise |
          vs |
    Control) |    .1110896   .0142334     7.80   0.000     .0831926    .1389866
-------------+-----------------------------------------------------------------
NDE          |
    exercise |
   (Exercise |
          vs |
    Control) |     .146092   .0189224     7.72   0.000     .1090047    .1831792
-------------+-----------------------------------------------------------------
TE           |
    exercise |
   (Exercise |
          vs |
    Control) |    .2571816   .0143876    17.88   0.000     .2289824    .2853807
-------------------------------------------------------------------------------
Note: Outcome equation does not include treatment-mediator interaction.
```

Logit models

- With a binary outcome, the effects are interpreted on a probability scale or as risk differences.

Logit models

- With a binary outcome, the effects are interpreted on a probability scale or as risk differences.
- We expect the probability of better well-being to be 0.26 higher if everyone in the population exercises than if no one exercises. Of that, the probability of better well-being is 0.11 higher because of an increase in bonotonin levels which and 0.15 higher because of other factors.

Logit models

- With a binary outcome, the effects are interpreted on a probability scale or as risk differences.
- We expect the probability of better well-being to be 0.26 higher if everyone in the population exercises than if no one exercises. Of that, the probability of better well-being is 0.11 higher because of an increase in bonotonin levels which and 0.15 higher because of other factors.
- We can use **estat rr** to report risk ratios or **estat or** to report odds ratios.

Risk ratios

```
. estat rr
estat rr requires potential-outcome means; refitting model ...
Transformed treatment effects                      Number of obs = 2,000
```

bwellbeing	Risk ratio	Robust std. err.	z	P>|z|	[95% conf. interval]	
NIE exercise (Exercise vs Control)	1.245647	.0392724	6.97	0.000	1.171004	1.325047
NDE exercise (Exercise vs Control)	1.477205	.0708189	8.14	0.000	1.344724	1.622738
TE exercise (Exercise vs Control)	1.840076	.0706258	15.89	0.000	1.70673	1.983839

Odds ratios

```
. estat or
estat or requires potential-outcome means; refitting model ...
Transformed treatment effects                          Number of obs = 2,000
```

bwellbeing	Odds ratio	Robust std. err.	z	P>|z|	[95% conf. interval]	
NIE						
exercise (Exercise vs Control)	1.562536	.0898293	7.76	0.000	1.396031	1.748901
NDE						
exercise (Exercise vs Control)	1.871182	.1490494	7.87	0.000	1.600713	2.187352
TE						
exercise (Exercise vs Control)	2.92379	.1841129	17.04	0.000	2.584315	3.307858

Continuous treatment

- **mediate** supports binary, multivalued, and continuous treatments.
- When the treatment is continuous, we need to include the **continuous()** option in the treatment specification and define the values at which we want the potential-outcome means to be evaluated. The first value will be considered the control.

Continuous treatment

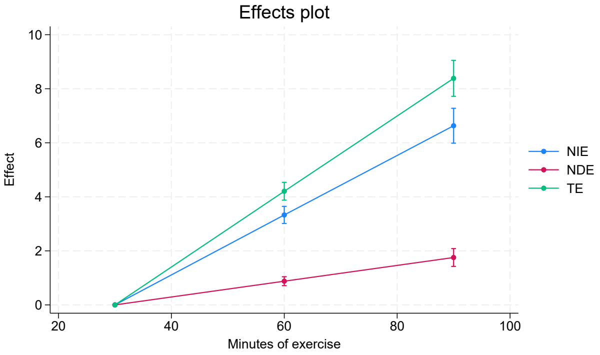
```
. mediate (wellbeing) (bonotonin) (cexercise, continuous (30 60 90))

Iteration 0:  EE criterion =  8.416e-28
Iteration 1:  EE criterion =  8.416e-28  (backed up)

Causal mediation analysis                           Number of obs = 2,000

Outcome model:      Linear
Mediator model:     Linear
Mediator variable: bonotonin
Treatment type:     Continuous

Continuous treatment levels:
  0: cexercise = 30 (control)
  1: cexercise = 60
  2: cexercise = 90
```

wellbeing	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
NIE						
cexercise						
(1 vs 0)	3.329037	.1613581	20.63	0.000	3.012781	3.645293
(2 vs 0)	6.630837	.3292353	20.14	0.000	5.985548	7.276127
NDE						
cexercise						
(1 vs 0)	.8769353	.0841601	10.42	0.000	.7119845	1.041886
(2 vs 0)	1.753871	.1683203	10.42	0.000	1.423969	2.083772
TE						
cexercise						
(1 vs 0)	4.205972	.1679266	25.05	0.000	3.876842	4.535103
(2 vs 0)	8.384708	.3394717	24.70	0.000	7.719356	9.05006

Note: Outcome equation includes treatment-mediator interaction.

Graphing effects

- When we evaluate effects at multiple points, we can use **estat effectsplot** to easily compare the effects visually.



Final remarks

- Learn more:
  https://www.stata.com/manuals/causalmediate.pdf

Thank you!

References

Nguyen, T. Q., I. Schmid, and E. A. Stuart. 2021. Clarifying causal mediation analysis for the applied researcher: Defining effects based on what we want to learn. Psychological Methods 26: 255-271.