Mediation Analysis in the Context of Program and Policy Evaluation

Amanda Bennett, PhD
AMCHP Pre-Conference Training
CityMatCH / MCH Epidemiology Conference
September 16, 2014
Outline

• Mediation Theory
• Historic Approaches to Mediation
• Counterfactual Approach to Mediation
• Mediation in an Evaluation Context
• Implementing Mediation Analysis (SAS Code, etc.)
Mediation Theory:
What is mediation?
Why is it valuable to epidemiology?
What is Mediation?

- Mediation describes the “totality of processes that explain an observed relationship between exposure and disease”
- The effect of an exposure (A) on an outcome (Y) is carried out through an intermediary variable (M), called a mediator or “intervening variable”
- An overall effect of A on Y is seen because A changes M, and the change in M causes a change in Y

Hafeman, 2009; MacKinnon 2007
What is Mediation?

• Deals with questions of causal processes and a chain of events

• What if the effect of

\[ A \rightarrow M \rightarrow Y \]

is actually due to A’s effect on a variable M, which later affects Y?
What is Mediation?

The overall relationship of two variables (the “Total Effect”)...

...can be decomposed into two pathways: **indirect** (thru $M$) and **direct** (NOT thru $M$)
Mediation vs. Confounding

- Confounding: a 3\textsuperscript{rd} variable is correlated with (or causes) both A and Y
- Mediation has specific, directed relationships

MacKinnon 2000; MacKinnon 2007
Mediation vs. Interaction

- Interaction: the relationship between A and Y changes across values of a third variable, Z

MacKinnon 2000; MacKinnon 2007
Mediation Theory

- Before conducting a mediation analysis, it is crucial to develop a theory about the causal processes at work.
- There is no way to empirically test whether a variable is a mediator vs. a confounder... so a strong theory for a causal relationship is needed.
- Directed Acyclic Graphs (DAGs) are an important tool for explicating your theory.

Glymour 2008
Directed Acyclic Graphs (DAGs)

- **Graphs**: visual representation of relationships between variables
- **Directed**: a chain of events or processes is marked by directed arrows
- **Acyclic**: the graph does not go in a circle... there is no way to start at variable X and follow the arrows in a way that leads back to X (no feedback loops)
Complexity of Mediation Pathways

- There are theoretically many, many intermediate variables between an exposure and outcome of interest

MacKinnon 2007; Hayes 2009
Complexity of Mediation Pathways

• There may be multiple mediation pathways and some mediators may lie in more than one pathway

MacKinnon 2007; Hayes 2009
Mediation Theory in Epidemiology

However, all the variables in the true causal chain do not need to be specified (which would result in infinite pathways with multiple mediators) because

For any given pathway, the arrow represents the sum of all pathways occurring between those variables
If the true pathway is:

But you only specified:

Your A → M₃ path would represent the sum of:
- direct path of A → M₃
- indirect path of A → M₁ → M₃
- indirect path of A → M₁ → M₂ → M₃
Mediation Theory in Epidemiology

• Epidemiology has been criticized for:
  • “Risk Factorology”: individual risk factors are identified or disproved, but little effort is made to explore why these risk factors impact disease
  • “Black Box” Approach: large number of factors in causal chain between exposure and outcome are ignored

• Without an explanatory theory, development of effective interventions is limited

Hafeman, 2009; McKinlay 2000; Weed 1998
Mediation Theory in Epidemiology

• To distinguish mediation from confounding, you must have reasonable evidence that you are dealing with a specific directed sequence

• Put another way, you must reasonably believe that the relationships are **causal**
  • Bradford-Hill criteria?
    • Temporality!
Mediation Theory in Epidemiology

• Mediation Analysis can be useful in epidemiology for:
  • Strengthening main effect causal hypothesis (show *how* an exposure causes a disease)
  • Testing pathway specific hypotheses (is a specific mechanism responsible for the association of exposure and disease)
  • Evaluating and improving interventions (is a program/treatment operating the way we expect it to?)

Hafeman, 2009
Mediation Theory in Epidemiology

- Biologic relationships that epidemiologists study may not be direct, but may contain many pathways within them

- What are the mediators of...
  - Gestational Diabetes $\rightarrow$ $\uparrow$ Macrosomia
  - Breastfeeding $\rightarrow$ $\downarrow$ Infant Infections
  - Child Overweight $\rightarrow$ $\uparrow$ Asthma

- Identifying biologic mediators may require clinical knowledge
- From a biologic standpoint, there may be infinite mediators, moderators, and pathways
Example: Biologic Mediation

How much of the increased risk of perinatal mortality for women with a placental abruption is caused by their increased likelihood of preterm birth?

Ananth, 2011
Mediation Theory in Epidemiology

• When we study disparities, the relationships examined are not direct pathways, but involve unnamed mediators

• What are the mediators of...
  - Race/Ethnicity → Infant Mortality
  - Income Level → Childhood asthma

• By ignoring the mediators, we only describe disparities rather than understanding why they occur or what leads to the difference

• To change disparities, we need to find manipulable variables in the causal chain
Example: Mediation in Social Epi

Among CSHCN...

How much of the racial/ethnic disparity in unmet needs among CSHCN is caused by black children’s lower likelihood of having a medical home?

Bennett, 2012
Example: Mediation in Social Epi

Maternal Education → Inter-pregnancy Interval → Preterm Delivery

How much of the education disparity in preterm delivery is caused by women with low education having shorter inter-pregnancy intervals?

Naimi, 2014
Mediation Theory in Epidemiology

• In program/policy evaluation, we often only focus on the impact of the program on long-term, distal outcomes

• What are the mediators of...
  • Prenatal Care $\rightarrow$ ↓ Low Birth Weight
  • Safe Sleep Education $\rightarrow$ ↓ SIDS
  • Medical Home $\rightarrow$ ↓ Asthma Severity

• Identifying mediators of these relationships depends upon program theory
  • Logic models are very helpful here!
Example: Mediation in Program Eval

How much of increase in infant mortality for women getting super-adequate PNC is caused by their increased risk of medically-induced preterm birth?

VanderWeele, 2013
Exercise 1: Mediation Theory
Historic Approaches to Mediation Analysis
Traditional Epi Approaches

To quantify the relationship between an exposure and outcome, control for all variables associated with exposure and outcome, except those in the causal pathway.

Questions

Is it always wrong/bad to do this?

Why might we want to control for factors in the causal pathway?

How do our results interpretations change if we choose to control for such factors?
Traditional Epi Approach to Mediation

Compare regression models with and without the mediator

Model 1: \( Y = \beta_0 + \beta_1 A + \text{covar} \)
\( \beta_1 \) represents Total Effect

Model 2: \( Y = \beta_0 + \beta_1 A + \beta_2 M + \text{covar} \)
\( \beta_1 \) represents Direct Effect not occurring through \( M \)

If \( \beta_1 \) differs in two models, mediation is considered to be present (similar to assessment of confounding, but have strong theory for directed relationship)

VanderWeele 2010
Traditional Epi Approach Weaknesses

- Focuses only on direct effect; does not quantify indirect effect
- Does not establish significance level for mediation
- No control for confounders of the mediator-outcome \( (M \rightarrow Y) \) relationship
  - Leads to biased results
- Cannot accommodate with A-M interaction
  - If \( A \rightarrow Y \) is modified by \( M \), there is not a single adjusted estimate of the direct effect

VanderWeele 2010
Exposure-Mediator Interaction

(Dotted lines represent effect modification)

- A single estimate of the \( A \rightarrow Y \) (and \( M \rightarrow Y \)) relationship is not appropriate if \( A-M \) interaction is present

Valeri 2013; VanderWeele 2010
Baron & Kenny (1986) Mediation Model

Total Effect = c
OR c' + a*b

Direct Effect = c'

Indirect Effect = a*b
OR c - c'

MacKinnon 2007; Valeri 2013; Hayes 2009
Baron & Kenny (1986) Mediation Model

Regression Equations

\[ Y = \text{i}_1 + c^*A + e_1 \]
\[ Y = \text{i}_2 + c'^*A + b^*M + e_2 \]
\[ M = \text{i}_3 + a^*A + e_3 \]

MacKinnon 2007; Valeri 2013; Hayes 2009
“Baron & Kenny” Causal Steps

• For mediation to be present, four criteria must be met:
  • Significant $A \rightarrow Y$, without controlling for $M$
    • significant $c$
  • Significant $A \rightarrow M$
    • significant $a$
  • Significant $M \rightarrow Y$
    • significant $b$
  • Total Effect larger than Direct Effect
    • $|c| > |c'|$

MacKinnon 2007; Hayes 2009
Limitations of Causal Steps Criteria

• Assumes “consistent” mediation
  • Consistent mediation is when the DE and IE operate in the same direction
  • Inconsistent mediation, aka “suppression”, is when DE & IE operate in different directions

• Does not quantify magnitude or significance of indirect effect

Baron & Kenny Methods: Other Problems for Epidemiologists

- Cannot accommodate non-continuous variables
  - Indirect effects via $c-c'$ and $ab$ are not equivalent for non-linear models
  - Method assumes uncorrelated errors of regression equations, which is not true for non-linear models

- Cannot accommodate exposure-mediator interaction

MacKinnon 2007; Pearl 2012; VanderWeele 2010
Counterfactual Approach to Mediation Analysis
Counterfactual Model Review

• How would an individual’s outcome change if he/she had two different exposure statuses
  • Only one exposure status is observable, the other is hypothetical and “counter-to-fact”
  • Since not all potential outcomes are observed, true causal effects are unknowable

• Statistical solution: estimate population causal effect by comparing exposed & unexposed groups
  • To do this, groups must be “exchangeable” or equivalent on all factors except exposure

Greenland 2002; Greenland 2009; Hernan 2004; Robins 1992; Rubin 1974
Counterfactual Mediation Approach

- Robins/Greenland (1992) and Pearl (2001) used counterfactual logic to develop general equations for estimating total, direct, and indirect effects
  - Allows A-M interaction
  - Adapted to non-linear regression models (VanderWeele 2010)

- These counterfactual statements compare the outcome under different scenarios of exposure and mediator values
Dichotomous Outcomes: Total Effect (TE)

How would probability of the outcome change if exposure changed from $a$ to $a^*$?

- Compares Y across two exposure values
- Does not account/adjust for mediator
- Gives “full effect” of an exposure on an outcome occurring through all potential pathways
Dichotomous Outcomes:
Controlled Direct Effect (CDE):

How would the probability of the outcome change if exposure changed from $a$ to $a^*$ and the mediator was set to $m$?

- Compare $Y$ across two exposure values while mediator set to value $m$
- Fixes mediator while manipulating exposure
- Mediator set to fixed level (average $m$ in population)
- If A-M interaction is present, there would be a different CDE for each level of $m$
- There is no “controlled indirect effect” to parallel the CDE
Dichotomous Outcomes: Natural Direct Effect (NDE):

How would the probability of the outcome change if exposure changed from $a$ to $a^*$ and the mediator was set to its “natural” value for exposure level $a^*$?

- Compare $Y$ across two exposure values if mediator set to what would have naturally occurred at exposure $a^*$
- Fixes mediator while manipulating exposure
- Mediator set to the average $m$ among those with exposure level $a^*$
- One estimate even in presence of A-M interaction
- If no A-M interaction, $\text{NDE} = \text{CDE}$
Dichotomous Outcomes: Natural Indirect Effect (NIE):

How would the probability of the outcome change if exposure was set to $a$ and the mediator changed from what would have naturally occurred at exposure level $a$ to what would have naturally occurred at exposure level $a^*$?

- Compare $Y$ under two mediator values while setting exposure to value $a$
- Fixes the $A$ value while manipulating $M$ values
- Mediator value manipulation is based on observed $M$ values for two different exposure conditions
Counterfactual Mediation Effects: Practical Interpretations

Assuming dichotomous exposure & outcome...

- **Total Effect** = what is the change in probability of the outcome due to the exposure?
- **Controlled Direct Effect** = what is the change in probability of the outcome due to the exposure, *after controlling for the mediator*?
Counterfactual Mediation Effects: Practical Interpretations

Assuming dichotomous exposure & outcome...

• Natural Direct Effect = what is the change in probability of the outcome due to the exposure if both the exposed & unexposed had the *mediator values observed for the unexposed*?

• Natural Indirect Effect = *among the exposed*, what is the change in probability of the outcome if they kept their *observed mediator value* versus if they had experienced the *mediator level observed for the unexposed*?
Counterfactual Mediation Approach: Population Estimates in Odds Ratio Form

\textbf{Two Regression Equations:}

\begin{align*}
\text{logit} \left( P(Y = 1|a, m, c) \right) &= \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_4' c \\
\text{logit} \left( P(M = 1|a, c) \right) &= \beta_0 + \beta_1 a + \beta_2' c
\end{align*}

- Combine coefficients from these equations to produce the NDE and NIE estimates (next slide)
- Can also use Binomial or Poisson regression for first equation to get estimates on RR scale
- \textbf{Note}: \( c \) represents the SAME set of covariates for both models

VanderWeele 2010
Counterfactual Mediation Approach: Population Estimates in Odds Ratio Form

Natural Direct Effect (NDE):

\[ OR_{NDE} = \frac{\exp(\theta_1) \{1 + \exp(\theta_2 + \theta_3 + \beta_0 + \beta_2' c)\}}{1 + \exp(\theta_2 + \beta_0 + \beta_2' c)} \]

Natural Indirect Effect (NIE):

\[ OR_{NIE} = \frac{\{1 + \exp(\beta_0 + \beta_2' c)\} \{1 + \exp(\theta_2 + \theta_3 + \beta_0 + \beta_1 + \beta_2' c)\}}{\{1 + \exp(\beta_0 + \beta_1 + \beta_2' c)\} \{1 + \exp(\theta_2 + \theta_3 + \beta_0 + \beta_2' c)\}} \]

Total Effect (TE) = NDE * NIE

[equations for the RR estimates are the same - use betas from binomial or poisson regression for equation 1]
Identification Assumptions

- For a causal effect to be validly identified, there must be:
  - No unmeasured A-M confounding
  - No unmeasured A-Y confounding
  - No unmeasured M-Y confounding
  - No M-Y confounders that are caused by A
Identification Assumptions

1. No uncontrolled $Z_1$
2. No uncontrolled $Z_2$
3. No uncontrolled $Z_3$
4. No $Z_3$ caused by $A$ (even if controlled)

In practice, these strong assumptions are unlikely to be met - if violated, effect estimates may not be valid and causal interpretations may not be appropriate

Pearl 2001; Valeri 2013; VanderWeele 2010
Sensitivity Analyses

- To evaluate the impact of potential unmeasured confounding on results, you can conduct a sensitivity analysis, involving:
  - Calculating bias terms based on estimations of the extent of unmeasured confounding
  - Obtaining bias-corrected estimates for NDE and NIE
  - Determining how violation of identification assumptions changes results and conclusions
Mediation Example:

From: Ananth & VanderWeele (2011). “Placental abruption and perinatal mortality with preterm delivery as a mediator...” AJE

RR-TE  = 13.8  (13.5-14.1)
RR-NDE = 10.2  (9.8-10.6)
RR-NIE  = 1.35  (1.3 -1.4)
Mediation Example (Ananth, 2011)

$$RR_{TE} = 13.8 \ (13.5\text{-}14.1)$$

**Interpretation:** Women with placental abruption have over 13 times the risk of perinatal mortality as women without abruption **[due to all causal pathways combined]**
Mediation Example (Ananth, 2011)

\[ RR_{NDE} = 10.2 \ (9.8-10.6) \]

**Interpretation:** If women with abruption had the same prevalence of PTB as women without abruption, women with abruption would have a perinatal mortality risk 10 times that of those without abruption.

**Alternate Interpretation:** Women with placental abruption have 10 times the risk for perinatal mortality as those without abruption, due to all causes other than increased preterm birth.

Ananth 2011
Mediation Example (Ananth, 2011)

$RR_{NIE} = 1.35 \ (1.3-1.4)$

Interpretation: Among those with abruption, the risk of perinatal mortality is 35% higher than it would be if they experienced PTB at the lower prevalence observed among women without abruption

Alternate Interpretation: If women with placental abruption had the same PTB risk as women without abruption, the risk of perinatal mortality among those with placental abruption would be 35% lower than what they currently experience

Ananth 2011
Mediation Example (Ananth 2011)

Overall Conclusions

• “The findings suggest that early delivery is not the primary mediator through which abruption is associated with increased mortality risk.”

• “...There were pathways for the effects of abruption [on perinatal mortality] not through preterm delivery... and these other pathways generally accounted for a majority of the effect.”

Ananth 2011
Mediation in an Evaluation Context
How Might Mediation Be Relevant to Program/Policy Evaluation?

Program as “Exposure”

- Is Program X causing a change in Y because of its effect on an intermediary variable?
- Is Program X affecting Y through a hypothesized intermediary variable?
- Why does Program X have an (unexpected) effect on the outcome?
Program as “Exposure”: Example #1

Purpose: Determine whether program is successfully impacting outcome through hypothesized intermediary variables

Example:
A peer counseling intervention aims to improve breastfeeding duration for participants by improving breastfeeding self-efficacy. Suppose you want to evaluate whether the program is working at all and working as intended.
Program as “Exposure”

\[ \text{RR}_{\text{TE}} = 1.70 \ (1.55-1.85) \]
\[ \text{RR}_{\text{NDE}} = 1.55 \ (1.42-1.68) \]
\[ \text{RR}_{\text{NIE}} = 1.10 \ (1.06-1.14) \]

How would you interpret these relative risks? Is the program “working”?

Improved Self-Efficacy

Peer Counseling

BF Dur >6 mos
Program as “Exposure”

$$RR_{TE} = 1.70 \ (1.55-1.85)$$
Interpretation: Women who receive peer counseling are 70% more likely to breastfeeding ≥6 months than women who do not receive peer counseling

$$RR_{NDE} = 1.55 \ (1.42-1.68)$$
Interpretation: If women who received peer counseling had the same self-efficacy as those not receiving peer counseling (If peer counseling did not impact self-efficacy), peer counseling recipients would still be 55% more likely to breastfeeding ≥6 months than women who do not receive peer counseling
Program as “Exposure”

\[ \text{RR}_{\text{NIE}} = 1.10 \ (1.06-1.14) \]

**Interpretation:** Among women receiving peer counseling, the rate of breastfeeding ≥6 months is 10% higher than would be expected if they had the self-efficacy of women who did not receive peer counseling.

**General Conclusions:**
While peer counseling does significantly impact breastfeeding duration by improving self-efficacy, the main effect of this program operates through other mechanisms.
Program as “Exposure”

Purpose #2: Explore reasons for an unexpected or unanticipated program effect

Example:
Compared to other women, women receiving adequate+ PNC (on the APNCU index) have the highest rates of infant mortality, even after adjusting for demographics and medical conditions

Hafeman 2009; VanderWeele 2013
Program as “Exposure”

Hypothesis:
The IM increase for adequate+ PNC women is due to an increased risk of medically-induced preterm birth.

From: VanderWeele 2013 “Medically induced preterm birth and the associations between prenatal care and infant mortality” Annals of Epidemiology
Program as “Exposure”

\[ OR_{TE} = 1.67 \ (1.62-1.72) \]
\[ OR_{NDE} = 0.98 \ (0.90-1.04) \]
\[ OR_{NIE} = 1.70 \ (1.68-1.72) \]

How would you interpret these odds ratios?
Program as “Exposure”

\[
\text{OR}_{\text{TE}} = 1.67 \ (1.62-1.72)
\]

Interpretation: Women receiving adequate+ PNC have 67% higher risk of infant mortality than women receiving adequate PNC

\[
\text{OR}_{\text{NDE}} = 0.98 \ (0.90-1.04)
\]

Interpretation: If women receiving adequate+ PNC were to have the same prevalence of medically-induced preterm birth as those with adequate PNC, adequate+ PNC would no longer be associated with an increased risk of infant mortality over adequate PNC

VanderWeele 2013
Program as “Exposure”

\[ \text{OR}_{\text{NIE}} = 1.70 \ (1.68-1.72) \]

**Interpretation:** Among women receiving adequate+ PNC, their observed risk of infant mortality is approximately 70% higher than what would be expected if they experienced medically-induced PTB at the same rate as women getting adequate PNC.

**General Conclusions:**
After accounting for medically-induced PTB, there is no longer a significant difference in IM related to adequate+ vs. adequate PNC. An increased risk of medically-induced PTB fully accounts for the observed increased risk of IM among adequate+ PNC recipients.
How Might Mediation Be Relevant to Program/Policy Evaluation?

Scenario #2: Program as “Mediator”

• Is an observed difference between the outcomes of two groups caused by their differential access to a program?

• How would equalizing the receipt of a service/program change the observed disparity between two groups?
Program as “Mediator”

**Example:**
Black CSHCN are more likely to have unmet healthcare needs than white CSHCN, but they are also less likely to have a medical home. How much of the observed racial disparity in unmet needs is due to a disparity in the medical home?

*Bennett 2012*
Program as “Mediator”

\[
\begin{align*}
\text{OR}_{\text{TE}} &= 1.9 \ (1.3-2.9) \\
\text{OR}_{\text{NDE}} &= 1.6 \ (1.1-2.4) \\
\text{OR}_{\text{NIE}} &= 1.2 \ (1.1-1.3)
\end{align*}
\]

How would you interpret these odds ratios?
Program as “Mediator”

\( \text{OR}_{\text{TE}} = 1.9 \ (1.2-2.9) \)

Interpretation: Black CSHCN have 90% higher odds of having unmet healthcare needs than white CSHCN.

\( \text{OR}_{\text{NDE}} = 1.6 \ (1.1-2.4) \)

Interpretation: If Black CSHCN had the same level of access to a medical home as white CSHCN, they would have 60% higher odds of unmet healthcare needs than white children. Assuring equal access to the medical home, therefore, would reduce the observed total disparity in unmet healthcare needs.
Program as “Mediator”

$\text{OR}_{\text{NIE}} = 1.2 \ (1.1-1.3)$

**Interpretation:** Among Black CSHCN, the observed odds of unmet healthcare needs are approximately 20% higher than what would be expected if they accessed the medical home at the same rate as white CSHCN.

**General Conclusions:**
Reducing disparities in the medical home serves as one strategy to reduce disparities in unmet healthcare needs among CSHCN... but other strategies will also be needed to fully eliminate the disparity.
Program as “Mediator”

**Example:**
Even among breastfeeding initiators, black women tend to have shorter duration of exclusive breastfeeding than white women.

Baby-friendly practices, like breastfeeding in the first hour after delivery, are likely to improve exclusive breastfeeding.

To what extent might the disparity in exclusive breastfeeding be caused by disparities in experience of Baby-Friendly hospital practices?

Bennett 2014
Program as “Mediator”

$$OR_{TE} = 1.62 \ (1.32-1.98)$$
$$OR_{NDE} = 1.54 \ (1.25-1.88)$$
$$OR_{NIE} = 1.05 \ (1.02-1.09)$$

How would you interpret these odds ratios?
Program as “Mediator”

\( \text{OR}_{\text{TE}} = 1.62 \ (1.32-1.98) \)

**Interpretation:** Black women have 60% higher odds of stopping exclusive breastfeeding before 2 weeks.

\( \text{OR}_{\text{NDE}} = 1.54 \ (1.25-1.88) \)

**Interpretation:** If Black women had the same prevalence of breastfeeding in the first hour as White women, they would still have over 50% higher odds of stopping exclusive breastfeeding before 2 weeks.
Program as “Mediator”

\[ \text{OR}_{\text{NIE}} = 1.05 (1.02-1.09) \]

**Interpretation:** Among Black women, the odds of stopping exclusive breastfeeding before 2 weeks are approximately 5% higher than what would be expected if they breastfed in the first hour at the same rate as white women.

**General Conclusions:**

The disparity in breastfeeding in the first hour accounts for a very small portion of the disparity in exclusive BF at 2 weeks. Ensuring that black women breastfeed in the first hour at the same prevalence as white women would not meaningfully change the disparity in exclusive BF.
Implementing Mediation Analysis in Statistical Software
Preliminary Steps

Before jumping into mediation analysis:

• Develop your mediation theory / model

• Examine individual pathways to understand the direction and magnitude of the relationships between variables
  \[ A \rightarrow Y \quad A \rightarrow M \quad M \rightarrow Y \]

• Examine exposure-mediator interaction to determine if significant
  • Will you want to include the interaction term in your mediation analysis?
Macros for Statistical Software

- Valeri and VanderWeele (2013) developed a macro for SAS and SPSS
  - Calculates the CDE, NDE, NIE, and TE
  - Calculates 95% CIs for effects
  - Flexible for different types of outcome and mediator variables

- Macro adapted for Stata by Lu & Emsley (2013) into command called: paramed
SAS Mediation Macro

Basic Macro From
(all elements required)

%mediation
(data=, yvar=, avar=, mvar=, cvar=, a0=, a1=, m=, nc=, yreg=, mreg=, interaction=)
RUN;
SAS Mediation Macro

\%
mediation
(data= libname.dataset, 
 yvar= outcome, 
 avar= exposure, 
 mvar= mediator, 
 cvar= var1 var2 var3, 
 a0= 0, 
 a1= 1, 
 ...)

Parts of code to edit are in *italic pink*

List a single variable name for each of these

List all covariates: must be continuous or dichotomous (convert categorical to dummy)

Values for two exposure levels of interest: 0 and 1 if exposure is dichot.
SAS Mediation Macro (continued)

... Mediator level at which CDE is calculated
m = 1,
nc = 3, Number of covariates listed in cvar list

yreg= logistic, Specify model types (for Y & M vars):
yreg options = linear, logistic,
mreg= logistic, loglinear, negbin, poisson
mreg options = linear or logistic

interaction= true or false) “false” option will not
RUN; include A-M interaction
in model, $\beta_3$ set to zero
SAS Mediation Macro

There are also optional keywords in the macro to specify:

• A case-control design for rare outcomes
• Method of standard error calculation
  • Default is delta method, but can specify bootstrapping
• Covariates values to obtain specific conditional effect estimates
SAS Mediation Macro

Special Considerations for Macro Use

• Subset data to complete cases only
  (remove any observations missing on any of the variables and covariates in analysis)

• Manually edit macro to accommodate complex sample surveys (PRAMS, NSCH, BRFSS, etc.)
### SAS Macro Output: Odds Ratio Estimates

*interaction = false option*

<table>
<thead>
<tr>
<th>Obs</th>
<th>Effect</th>
<th>Estimate</th>
<th>p_value</th>
<th>_95__CI_lower</th>
<th>_95__CI_upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>cde</td>
<td>1.53673</td>
<td>.000033765</td>
<td>1.25428</td>
<td>1.88278</td>
</tr>
<tr>
<td>2</td>
<td>nde</td>
<td>1.53673</td>
<td>.000033765</td>
<td>1.25428</td>
<td>1.88278</td>
</tr>
<tr>
<td>3</td>
<td>nie</td>
<td>1.03427</td>
<td>.000039993</td>
<td>1.01777</td>
<td>1.05104</td>
</tr>
<tr>
<td>4</td>
<td>marginal</td>
<td>1.58939</td>
<td>.000007938</td>
<td>1.29698</td>
<td>1.94773</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Effect</th>
<th>_95__CI</th>
<th>_95__CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>total effect</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CDE and NDE will be the same with interaction term is not included
### SAS Macro Output: Odds Ratio Estimates

**interaction = true option**

<table>
<thead>
<tr>
<th>Obs</th>
<th>Effect</th>
<th>Estimate</th>
<th>p_value</th>
<th>_95__CI_lower</th>
<th>_95__CI_upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>cde</td>
<td>1.07974</td>
<td>0.81027</td>
<td>0.57718</td>
<td>2.01986</td>
</tr>
<tr>
<td>2</td>
<td>nde</td>
<td>1.54967</td>
<td>0.00003</td>
<td>1.26293</td>
<td>1.90151</td>
</tr>
<tr>
<td>3</td>
<td>nie</td>
<td>1.04658</td>
<td>0.00062</td>
<td>1.01966</td>
<td>1.07421</td>
</tr>
<tr>
<td>4</td>
<td>marginal t</td>
<td>1.62186</td>
<td>0.00000</td>
<td>1.31821</td>
<td>1.99545</td>
</tr>
</tbody>
</table>

**CDE ≠ NDE when A-M interaction is included;**

- **CDE** = effect of A on Y when M = 0
- **NDE** = effect of A on Y when M = avg. m value among the unexposed
Exercise #2: Mediation
Planning and Interpretation
Final Questions?

Contact Information:
Amanda Bennett
acaven3@uic.edu
References - Mediation Methods & Theory


References - Mediation Methods & Theory


References - Examples of Mediation in MCH


More Details about Mediation Equations
Counterfactual Mediation Notation

\[ Y = \text{outcome value for each individual} \]
\[ A = \text{exposure value for each individual} \]
\[ M = \text{mediator value for each individual} \]
\[ C = \text{set of covariates for each individual} \]

\[ Y_a = \text{potential outcome for each individual when} \]
\[ A = a \]
\[ M_a = \text{potential mediator value for each individual} \]
\[ \text{when} \ A = a \]
\[ Y_{am} = \text{potential outcome for each individual when} \]
\[ A = a \text{ and } M = m \]

Pearl 2001; Pearl 2012; VanderWeele 2010
Individual-Level Counterfactual Effect Estimates
*(Hypothetical)*

Total Effect (TE) = \( Y_a - Y_{a^*} \)

Controlled Direct Effect (CDE) = \( Y_{am} - Y_{a^*m} \)

Natural Direct Effect (NDE) = \( Y_{aM_{a^*}} - Y_{a^*M_{a^*}} \)

Natural Indirect Effect (NIE) = \( Y_{aM_{a^*}} - Y_{a^*M_{a^*}} \)

Pearl 2001; Pearl 2012; VanderWeele 2010
General Effect Estimates: Total Effect (TE)

- General form:
  \[
  TE_{a^*,a} = E(Y|A = a) - E(Y|A = a^*)
  \]

Risk Ratio form (dichotomous Y):

\[
RR_{a,a^*|c}^{TE} = \frac{P(Y_a=1|c)}{P(Y_{a^*}=1|c)}
\]

Pearl 2001; Pearl 2012; VanderWeele 2010
General Effect Estimates: Controlled Direct Effect (CDE)

- General form:
  \[ CDE_{a^*,a|m}(Y) = E(Y|A = a, M = m) - E(Y|A = a^*, M = m) \]

Risk Ratio Form (dichotomous Y):

\[ RR^{CDE}_{a,a^*|c}(m) = \frac{P(Y_{am} = 1|c)}{P(Y_{a^*m} = 1|c)} \]

Pearl 2001; Pearl 2012; VanderWeele 2010
General Effect Estimates: Natural Direct Effect (CDE)

\[ NDE_{a^*, a}(Y) = \sum_z [E(Y|a, m) - E(Y|a^*, m)]P(m|a^*) \]

Risk Ratio Form (dichotomous Y):

\[ RR_{a, a^*|c}(a^*) = \frac{P(Y_{aM_{a^*}} = 1|c)}{P(Y_{a^*M_{a^*}} = 1|c)} \]

Pearl 2001; Pearl 2012; VanderWeele 2010
General Effect Estimates:  
Natural Indirect Effect (NIE)

General form:

\[ NIE_{a^*, a}(Y) = \sum_{z} \{E(Y|a^*, m) \times [P(m|a) - P(m|a^*)]\} \]

Risk Ratio Form (dichotomous Y):

\[ RR_{a, a^*|c}(\alpha) = \frac{P(Y_{am\alpha}=1|c)}{P(Y_{am\alpha^*}=1|c)} \]

Pearl 2001; Pearl 2012; VanderWeele 2010
General Effect Estimates

- For Linear Regression: \( TE = NDE + NIE \)
- For Binomial/Logistic Regression: \( TE = NDE \times NIE \)

Proof for Linear Regression

\[
\begin{align*}
TE &= Y_1 - Y_0 \\
CDE &= Y_{1m} - Y_{0m} \\
NDE &= Y_{1M_0} - Y_{0M_0} \\
NIE &= Y_{1M_1} - Y_{1M_0}
\end{align*}
\]
Impact of Misclassification

Exposure and Outcome Misclassification

• Independent and Non-differential:
  • TE and DE biased towards null
• Dependent and/or Non-Differential
  • TE and DE bias may occur in either direction

Mediator Misclassification

• Non-differential
  • IE biased towards null, DE away from null
  • No TE bias
• Differential
  • IE/DE bias may occur in either direction
  • No TE bias

Ogburn 2012; VanderWeele 2012a; VanderWeele 2012b
References - Misclassification in Mediation


Choice of Reference Groups for Dichotomous Variables
Choosing Reference Groups

• By changing the reference group for each variable, you can change the signs of the individual paths, as well as the direction, magnitude and interpretation of the effects.

• It is important to carefully consider how your coding of dichotomous variables will impact your effect estimates and interpretability.

• Based on your research question at hand, you may want the NDE and NIE to reflect one specific comparison over another.
Choosing Reference Groups

Example:
Among Children with Asthma...
  • Exposure = Medical Home
  • Mediator = Unmet Needs
  • Outcome = ER Visit in Last Year

• Each of these variables is dichotomous and you have a choice about whether “yes” or “no” will be the reference group in your mediation analysis
Choosing Reference Groups: Scenario #1

Comparison = Yes vs. No for all

Children in a medical home are considered the “exposed”

Expected effect values:
NDE is negative: OR < 1
NIE is negative: OR < 1
Choosing Reference Groups: Scenario #1

NDE Interpretation: The odds ratio of an ER visit for children in a medical home vs. not in a medical home, if all kids had unmet needs at the same prevalence as those not in a medical home.

NIE Interpretation: Among children in a medical home, the ratio of the odds of an ER visit under their current prevalence of unmet needs versus what their prevalence of unmet needs would be if they were not in a medical home.
Choosing Reference Groups: **Scenario #2**

Comparison = no vs. yes for all

Children **NOT** in a medical home are considered “exposed”

**Expected effect values:**
- NDE is negative: OR < 1
- NIE is negative: OR < 1

Changing the reference level for all three vars keeps the relationship directionality, but changes interpretations
Choosing Reference Groups: Scenario #2

**NDE Interpretation:**
The odds ratio of **no ER visit** for children **NOT in a medical home** vs. **in a medical home**, if all kids had **no unmet needs** at the same prevalence as those **in a medical home**

**NIE Interpretation:**
Among children **NOT in a medical home**, the ratio of the odds of **no ER visit** under their current prevalence of **NO unmet needs** versus what their prevalence of **NO unmet needs** would be if they were **in a medical home**
Choosing Reference Groups: Scenario #3

Reference values = mix and match based on “desired” value
Med home: yes vs. no
Unmet needs: no vs. yes
1+ ER Visit: no vs. yes

Expected effect values:
NDE is positive: OR > 1
NIE is positive: OR > 1

Changing the reference level for exposure or outcome alters the direction & interpretation of the relationships
Choosing Reference Groups: **Scenario #3**

**NDE Interpretation:**
The odds ratio of **NO ER visit** for children in a **medical home** vs. **NOT in a med home**, if all kids had **no unmet needs** at the same prevalence as those **NOT in a med home**.

**NIE Interpretation:**
Among children **in a medical home**, the ratio of the odds of **no ER visit** under their current prevalence of **no unmet needs** versus what their prevalence of **no unmet needs** would be if they were **NOT in a medical home**.
Choosing Reference Groups: Scenario #4

Reference values = mix and match
Med home: yes vs. no
Unmet needs: yes vs. no
1+ ER Visit: no vs. yes

Expected effect values:
NDE is positive: OR > 1
NIE is positive: OR > 1

Changing the reference level for a dichotomous mediator does not change overall effect direction or interpretation
**Choosing Reference Groups: Scenario #4**

**NDE Interpretation:**
The odds ratio of no ER visit for children in a medical home vs. NOT in a medical home, if all kids had unmet needs at the same prevalence as those NOT in a medical home.

**NIE Interpretation:**
Among children in a medical home, the ratio of the odds of no ER visit under their current prevalence of unmet needs versus what their prevalence of unmet needs would be if they were NOT in a medical home.