Health Effects of Temperature and Air Pollution Using Causal Mediation Analysis

Joel Schwartz
Harvard School of Public Health
- Introduction
- Causal inference
- Mediation analysis
- Mixed-effects mediation analysis for longitudinal data
- Applications to the Normative Aging Study (NAS)
- Conclusions
Air Pollution and Temperature

• Are associated with Deaths from cardiovascular and respiratory disease
• Reductions in lung function, increases in inflammatory markers, atherosclerosis, etc.
• It is commonly assumed that understanding the pathways behind these associations is a job for toxicologists and biologists
Mediation Analysis

• Is one way to address the question in the species of interest and at the exposure levels of interest
Causal inference has become popular in epidemiology

- I will briefly discuss how to compute causal effects
- Then Mediation analysis, which allows us:
  - To take into account intermediary variables
  - To investigate biological mechanisms
  - To assess their importance
- Present a method for mediation analysis for longitudinal data in
- the presence of heterogeneity.
- Applications:
  - Environmental exposures: air pollution and temperature
  - Mediator: ICAM-1 DNA methylation
  - Outcome: ICAM-1 protein levels
Causal Modeling

- Potential outcomes (also called counterfactuals): $Y_i^a$ and $Y_i^{a*}$
  - $Y_i^a$: what would be the value of the outcome $Y_i$ if participant $i$ were exposed to $a$?
  - $Y_i^{a*}$: what would be the value of the outcome $Y_i$ if participant $i$ were exposed to $a^*$?
  - For each participant $i$, only $Y_i^a$ or $Y_i^{a*}$ is observed.

- $E(Y_i^a - Y_i^{a*})$: Average causal effect of changing the exposure from $a^*$ to $a$
Important assumptions to conclude that "association = causation"

- Temporality
  - Can be achieved with prospective design.

- $Y_i^a \prod A_i \mid C_i = c$
  - No unmeasured exposure-outcome confounding given the confounding variables.
  - The group of people exposed to $a^*$ is exchangeable to the group of people exposed to $a$ given the confounding variables.
  - The counterfactual outcome does not depend on the actual observed exposure.
  - Can be achieved with randomization.
**Figure**: Directed Acyclic Diagram (DAG)

- $A_i$: exposure
- $Y_i$: outcome
- $C_{1i}$: confounding variables
- $C_{2i}$: risk factors of the outcome
We often hear people talk about confounding in terms of correlation

- E.g. a confounder is correlated with exposure and outcome
- But correlation does not have arrowheads
- This can cause problems, that visualization can clarify
What about these graphs?
What about this case?
Example

• I am interested in the effects of Flavonoids
• Is Vitamin C a confounder?
• Does eating an orange make you eat berries?
• What is going on?
Food Preference Cluster

Fruit Pref → Berries → Oranges

[Diagram showing relationships between fruit preferences and specific fruits like berries and oranges]
$E(Y_i^a - Y_i^{a*} | C_i = c)$: Average causal effect of changing the exposure from $a^*$ to $a$

- Methods to Compute Causal Effects:
  - Inverse Probability Weighting (IPW)
  - G Estimation
  - Propensity Scores
  - Marginal Structural Models
Mediation analysis allows us to answer interesting questions:

1. Does the exposure $A_i$ have a direct effect on $Y_i$?
2. Is the effect of the exposure $A_i$ mediated by a mediator $M_i$?
3. How much of the effect of the exposure $A_i$ on $Y_i$ mediated by $M_i$?

Three important quantities in mediation analysis:

1. Natural direct effect (NDE)
2. Natural indirect effect (NIE)
3. Proportion mediated (PM)
- $A_i$: exposure
- $M_i$: mediator
- $Y_i$: outcome
- $C_{1i}$: confounding variables related to the exposure
- $C_{2i}$: confounding variables related to the mediator

**Figure**: Directed Acyclic Diagram (DAG)
Counterfactuals for the $i^{th}$ person exposed to $a$: $M_i^a$ and $Y_i^{a,m}$

Natural direct effect (NDE) $= \mathbb{E}(Y_i^{a,M_i^a} - Y_i^{a*,M_i^a*})$

Natural indirect effect (NIE) $= \mathbb{E}(Y_i^{a,M_i^a} - Y_i^{a*,M_i^a*})$

Proportion mediated $= \frac{\text{NIE}}{\text{NIE} + \text{NDE}}$
Assumptions

1) $Y_{i}^{a,m} \perp A_{i} \mid C_{i} = c$ (No unmeasured exposure-outcome confounding)

2) $Y_{i}^{a,m} \perp M_{i} \mid A_{i} = a, C_{i} = c$ (No unmeasured mediator-outcome confounding)

3) $M_{i}^{a} \perp A_{i} \mid C_{i} = c$ (No unmeasured exposure-mediator confounding)

4) $Y_{i}^{a,m} \perp M_{i}^{a*} \mid C_{i} = c$ (No mediator-outcome confounder affected by the exposure)
Start without Interactions: We fit two Regressions

1. \[ E[M_i|C_i = c, A_i = a] = \beta_0 + \beta_1 \ a + \beta_c^T c \]

2. \[ E[Y_i|C_i = c, A_i = a, M_i = m] = \gamma_0 + \gamma_1 \ a + \gamma_2 \ m + \gamma_c^T c \]

\[
\begin{align*}
A_i & \xrightarrow{\beta_1} M_i & \xrightarrow{\gamma_2} Y_i
\end{align*}
\]

Natural Direct Effect (NDE) = \((a - a^*) \ \gamma_1\)

Natural Indirect Effect (NIE) = \((a - a^*) \ \gamma_2 \ \beta_1\)
Mediator and Modifier?

Mediation analysis can include interactions as well:

1. \( E[M_i | C_i = c, A_i = a] = \beta_0 + \beta_1 a + \beta_c^T c \)

2. \( E[Y_i | C_i = c, A_i = a, M_i = m] = \gamma_0 + \gamma_1 a + \gamma_2 m + \gamma_3 am + \gamma_c^T c \)

Natural Direct Effect (NDE) = \( (a - a^*)[\gamma_1 + \gamma_3\beta_0 + a^*\gamma_3\beta_1 + \gamma_3 \beta_c^T c] \)

Natural Indirect Effect (NIE) = \( (a - a^*) \gamma_2\beta_1 + a (a - a^*) \gamma_3\beta_1 \)
Standard Methods for Mediation Analysis

• Do not take into account Longitudinal data

• Do not allow for Heterogeneity in Response
  – For the Mediator response to Exposure
  – For the Outcome response to the Mediator

• We derived the direct and indirect causal effects in the presence of longitudinal data and heterogeneity.
A Key Concept

• Suppose there is heterogeneity in the effect of exposure on the mediator, and in the effect of the mediator on the outcome
• If there is correlation in those, then the mediation formula must take it into account
• If the high responders to A->M are the high responders to M->Y then there is a larger mediated effect
• We define two mixed models in the presence of heterogeneity

\[ E[M_i|C_i = c, A_i = a, b_i] = (\beta_0 + b_0i) + (\beta_1 + b_1i) a + \beta_c^T c \]

\[ E[Y_i|C_i = c, A_i = a, M_i = m, g_i] = (\gamma_0 + g_0i) + (\gamma_1 + g_1i) a + (\gamma_2 + g_2i) m + (\gamma_3 + g_3i) am + \gamma_c^T c \]

• \( M_i \) and \( Y_i \) are vectors of length \( n_i \) (repeated measures)
• \( A_i \) : exposure
• \( C_i \) : confounding variables or risk factors
• \( b_i \) and \( g_i \) are vectors of random effects
Assumptions

For $i=1,...,N$ and $j=1,...,n_i$:

1) $Y_{ij}^{a,m} \perp A_{ij} \mid C_{ij} = c, b_i, g_i$ (No unmeasured exposure-outcome confounding)

2) $Y_{ij}^{a,m} \perp M_{ij} \mid A_{ij} = a, C_{ij} = c, b_i, g_i$ (No unmeasured mediator-outcome confounding)

3) $M_{ij}^a \perp A_{ij} \mid C_{ij} = c, b_i, g_i$ (No unmeasured exposure-mediator confounding)

4) $Y_{ij}^{a,m} \perp M_{ij}^a \mid C_{ij} = c, b_i, g_i$ (No mediator-outcome confounder affected by the exposure)
Figure: DAG at visits $j=J$ and $j=J+1$
For $j=1,...,J$:

**Natural Direct Effect (NDE)**

$$= (a - a^*) \left[ \gamma_1 + \gamma_3 \beta_0 + \sigma_{g3,b0} + a^* \gamma_3 \beta_1 + \sigma_{g3,b1} + \gamma_3 \beta_c^T c \right]$$

$$= (a - a^*) \gamma_1 \text{ (if no interaction)}$$

**Natural Indirect Effect (NIE)**

$$= (a - a^*) \left[ \gamma_2 \beta_1 + \sigma_{g2,b1} \right] + a \left( a - a^* \right) \left[ \gamma_3 \beta_1 + \sigma_{g3,b1} \right]$$

$$= (a - a^*) \left[ \gamma_2 \beta_1 + \sigma_{g2,b1} \right] \text{ (if no interaction)}$$
The correlation in the random slopes matters

Natural Indirect Effect \( (NIE) = (a - a^*) \left[ \gamma_2 \beta_1 + \sigma_{g_2,b_1} \right] \)
The Normative Aging Study

• Is a longitudinal study of men recruited in 1965
• Medical examinations every four years
• These include inflammatory markers and DNA methylation
- We investigate whether PM$_{2.5}$ has an effect on ICAM-1 protein through a change in ICAM-1 DNA methylation.

\[
\text{PM}_{2.5} \rightarrow \text{ICAM-1 methylation} \rightarrow \text{ICAM-1 protein}
\]

- We also examine whether temperature has an effect on ICAM-1 protein through a change in ICAM-1 DNA methylation.

\[
\text{Temperature} \rightarrow \text{ICAM-1 methylation} \rightarrow \text{ICAM-1 protein}
\]
• We measured ICAM-1 DNA methylation on blood samples collected after an overnight fast and smoking abstinence (1999-2009).

• DNA methylation was measured with highly quantitative methods based on bisulfite PCR pyrosequencing.

• A total of 777 participants had their DNA methylation assessed one to five times within a period of three to five years in between.

• ICAM-1 protein level was quantified by ELISA assay.
We fit two linear mixed-effects models simultaneously in SAS:

$$E[M_i|C_i = c, AP_i = a, b_i] = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i}) a + \beta_c^T c$$

$$E[Y_i|C_i = c, AP_i = a, M_i = m, g_i] = (\gamma_0 + g_{0i}) + (\gamma_1 + g_{1i}) a + (\gamma_2 + g_{2i}) m + \gamma_c^T c$$

- Covariates: age, batch, bmi, temperature, % of neutrophils, smoking, seasonal sine and cosine, diabetes
Air pollution direct effect $\hat{\gamma}_1 = 0.058$ (95%CI: 0.005 to 0.110)

Air pollution indirect effect

$= \hat{\gamma}_2 \times \hat{\beta}_1 + \hat{\sigma}_{g_2,b_1} = (-0.06706)(-0.08193) - 0.00062$

$= 0.005$ (95%CI: -0.016 to 0.026)

Note: Test of $\sigma_{g_2,b_1}=0$ gives a p-value=0.95

Proportion mediated (PM)=7.8%
Temperature direct effect $\hat{\gamma}_1 = -0.207$ (95% CI: -0.415 to 0.002)

Temperature indirect effect

$= \hat{\gamma}_2 \times \hat{\beta}_1 + \hat{\sigma}_{\hat{g}_2, \hat{b}_1} = (-0.07211)(0.4174) - 0.01109$

$= -0.041$ (95% CI: -0.076 to -0.006)

Note: Test of $\sigma_{g_2, b_1} = 0$ gives a p-value = 0.33

Proportion mediated (PM) = 16.6%
• We obtained a significant positive direct effect of PM2.5 on ICAM-1 protein levels.
• We found a significant indirect negative effect of temperature on ICAM-1 protein levels through DNA methylation. The proportion mediated was equal to 16.6%.
• The correlation between the random slopes were non-negative. This suggests that individuals experiencing an effect of exposure on the mediator are not different individuals experiencing an effect of the mediator on the outcome, on average.
Next Steps

• Other mediation analyses
  – MI with BP as a mediator, etc
  – Methylation in pathways (e.g. MAPK)

• How much of overall population heterogeneity is through the mediator

• Multiple M