Mediation anaalysis

Causal Inference

Muhammad Tamjid Rahman

04 September, 2021

In this study we analyse the effect of smoking on infant mortality using mediation analysis method, taking low birth weight as mediator. The 'birthweight.txt' data contains 100,000 observations with 12 variables. There are very few missing values. Since it'll not affect our result very much. So, we omitted the rows with missing values. Now we have a clean dataset with 933,299 observations and 12 variables.

It is found that 0.548% of the children in the sample died. So, the outcome can be considered to be rare.

Controlled direct effect

The controlled direct effect (CDE) odds ratios are to be estimated at the two levels of the mediator, low and regular birth weight. The estimation is under the assumptions of no unmeasured exposure-outcome confounding and mediator-outcome confounding conditional on the observed covariates. To estimate the odds ratios, two logistic regression models are fitted, one for the outcome Y and the other for the mediator M. The model for outcome Y is

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

The model for outcome M is

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

The CDE odds ratio at M=m is calculated as

$$OR^{CDE}(m) = exp[(\theta_1 + \theta_3 m)(a - a^*)]$$

Here, the treatment is binary. So, a=1 and a^*. m=1 if the birth weight is low and m=0 if the birth weight is regular.

For the confidence intervals, the standard errors are calculated as

$$se = \sqrt{\Gamma \Sigma \Gamma'}, \;\; where \quad \Sigma = \begin{pmatrix} \Sigma_\beta & 0 \\ 0 & \Sigma_\theta \end{pmatrix}$$

Σ_β is the estimator's covariance matrix of the model for outcome m and Σ_θ is the estimator's covariance matrix of the model for outcome y. The vector Γ is a vector with 24 elements, where the 12th and 14th elements are set to 1 and remaining elements are 0 for the low weight group. The 12th element is set to 1 and remaining elements are 0 for the regular weight group.

The estimated CDE odds ratios and 95 % confidence intervals are presented in Table 1.

Table 1: Odds ratios and 95% confidence intervals for the CDE

Birth weight	Odds ratio	CI
low	0.933	(0.864, 1.006)
Regular	1.627	(1.487, 1.780)

From the controlled direct effect, we can interpret that the odds for infant mortality for children whose mothers smoked is .93 times the odds of children of non-smoking mothers, given low birth weight.

On the other hand the odds for infant mortality for children whose mothers smoked is over 1.6 times the odds of children of non-smoking mothers, given regular birth weight.

These effects give us concept of the effect of exposure on the outcome when the mediator was fixed at a certain level. If the mediator can be interfered,we can know the effect of exposure on outcome by the controlled direct effect.

Sensitivity analysis

A sensitivity analysis of the natural direct and indirect effect and total effect odds ratios is performed using sensitivity analysis parameters corresponding to $\gamma = 3$ for the effect of an unmeasured binary confounding variable on the outcome and $\pi_{1m=1} = 0.04$, $\pi_{1m=0} = 0.03$ $\pi_{0m=1} = 0.15$, $\pi_{0m=0} = 0.02$ for the prevalences of the unmeasured confounder in each of the smoking by low birth weight strata. The calculated estimates are biased for the presence of unmeasured confounder. A factor bias can be used to correct the estimates. The factors are,

$$B_0 = \frac{1 + (\gamma - 1)P(U = 1 \mid a, M = 0, c)}{1 + (\gamma - 1)P(U = 1 \mid a^*, M = 0, c)}$$

$$B_1 = \frac{1 + (\gamma - 1)P(U = 1 \mid a, M = 1, c)}{1 + (\gamma - 1)P(U = 1 \mid a^*, M = 1, c)}$$

$$Corrected\ OR^{CDE}(0) = \frac{OR^{CDE}(0)}{B_0}$$

$$Corrected\ OR^{CDE}(1) = \frac{OR^{CDE}(1)}{B_1}$$

The corrected CDE odds ratios and confidence intervals are presented in Table 2

Table 2: Corrected odds ratios and 95% confidence intervals for the CDE

Birth weight	**Odds ratio**	**CI**
Low	1.122	(1.040, 1.211)
Regular	1.596	(1.458, 1.746)

In the sensitivity analysis, both adjusted estimates are over 1. So, the odds of infant mortality are higher if the mother smoke, regardless of the birth weight.

Indirect and direct effects

Using the same logistic regression models, the natural direct effect (NDE) and natural indirect effect (NIE) odds ratios are computed as

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

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

The total effect (TE) odds ratio is calculated as

$$OR^{TE} = OR^{NDE}.OR^{NIE}$$

The estimated indirect, direct and total effects are presented in Table 3.

Table 3: Odds ratios and 95% confidence intervals for the NDE, NIE and TE

Effect	Odds ratio	CI
Direct	1.129	(1.036, 1.231)
Indirect	1.308	(1.283, 1.333)
Total	1.477	(1.355, 1.610)

From the natural direct effect we can interpret that if there was no effect of smoking on low birth weight, children mortality would be 1.129 times whose mother smoke than the children whose mothers doesn't smoke. The natural indirect effect tells us that if everyone is smoker, the odds of dying is around 1.308 times higher for low birth weight compared to regular weight.

The total effect tells us that the infant mortality is 1.477 times if the mother smokes than the mother does not.

NDE and NIE tell the effect of exposure directly on the outcome and through the mediator.

The total effect tells us the overall effect of the exposure, both directly and through mediator.

Sensitivity analysis

A sensitivity analysis of the natural direct and indirect effect and total effect odds ratios using sensitivity analysis parameters corresponding to $\gamma = 3$ for the effect of an unmeasured binary confounding variable on the outcome and $\pi_{1m=1} = 0.04$, $\pi_{1m=0} = 0.03$ $\pi_{0m=1} = 0.15$, $\pi_{0m=0} = 0.02$ for the prevalences of the unmeasured confounder in each of the smoking by low birth weight strata. The calculated estimates are biased for the presence of unmeasured confounder. A factor bias can be used to correct the estimates. The factors are,

$$B_0 = \frac{1 + (\gamma - 1)P(U = 1 \mid a, M = 0, c)}{1 + (\gamma - 1)P(U = 1 \mid a^*, M = 0, c)}$$

$$B_1 = \frac{1 + (\gamma - 1)P(U = 1 \mid a, M = 1, c)}{1 + (\gamma - 1)P(U = 1 \mid a^*, M = 1, c)}$$

$$B_2 = \frac{1 + (\gamma - 1)P(U = 1 \mid a^*, M = 1, c)}{1 + (\gamma - 1)P(U = 1 \mid a^*, M = 0, c)}$$

The corrected point estimates from the sensitivity analysis are presented in Table 4

Table 4: Corrected odds ratios for the NDE, NIE and TE

Effect	Odds ratio
Direct	1.251
Indirect	1.300
Total	1.626

In the sensitivity analysis the odds ratio for the indirect effect is almost unchanged. But the odds ratio the direct effect and total have changed. The effect of an unmeasured confounder affected the direct effect,if the mediated effect is fixed. The infant mortality is higher if the mother is smoker.

Miscellaneous aspects on mediation analysis

The proportion of the effect mediated on the risk difference scale calculated as

$$PM = \frac{OR^{NDE}(OR^{NIE} - 1)}{OR^{NDE}.OR^{NIE} - 1} = 72.87\%$$

So, 72.87% of the effects of smoking on infant mortality is mediated through low birth weight.

To test for mediation, it would be preferable to test the natural indirect effect rather than the proportion mediated. The NIE has lower variance than the PM. Resulting in less wide confidence intervals in NIE than PM. Sometimes direct and indirect effects work in opposite directions. In that case the proportion mediated doesn't work well. The NIE is more well-off and can be tested in every cases.

In mediation analysis,the assumption of no unmeasured mediator-outcome confounder can create bias and sometimes the effect doesn't go to the right direction. This can be corrected using sensitivity analysis techniques. The assumption "Cross-world independence" can also be problematic for mediation analysis. The cross-world independence assumption is

$$Y(a,m) \perp\!\!\!\perp M(a^*)$$

Where, Y(a,m) is the outcome of Y if exposure A were set to A = a and the mediator M were set to M=m. M(a^*) is the value of M under the assignment A = a^*. This assumption states an independence between counterfactual outcome and mediator values when the exposure is set to A = a for the outcome and it is set to A = a^* for the mediator. This unrealistic occurance makes the cross-world independence assumption impossible to verify and not easy to justify based on background knowledge, even without relying on other equally problematic assumptions.