

Lab

Mediation analysis in the presence of exposure-mediator interaction

In this lab we will analyze simulated data based on the case-control lung cancer study that was described in the lecture [data was simulated so that the distribution of outcomes, exposures and covariates were roughly similar to those in the original case control study]. We will be estimating direct and indirect effects. Note that these simulated data are restricted to Caucasians and so population stratification is not likely an issue in the analyses you will run. The assumption that the outcome is rare is very reasonable here. The .csv dataset has been sent to you. The data consists of the following variables:

Variable in Dataset	Description
case	lung cancer
snp	0 vs. 1/2 C alleles at rs8034191
smoking	1=ever smoker; 0=never smoker
sex	1=male 2=female
colgrad	1=college graduate 0=not college graduate
age	in years

Questions

1. Let the exposure be the genetic variant, the mediator smoking status and the outcome lung cancer status.

(a) We start by employing traditional approaches to mediation analysis and estimate total effect, direct effect, and difference method and product method indirect effect.

(b) We proceed calculating an estimate and a 95% confidence interval for the multiplicative interaction between the SNP and smoking on lung cancer on the odds ratio scale using logistic regression, adjusted for the covariates.

(c) What the controlled direct effect odds ratios would be for the smoker group and non-smoker group? We can evaluate these and all other effects at the mean level of the covariates

2.

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

(b) Give the estimate of the proportion of the effect mediated on the risk difference scale under these assumptions.

3.

What about possible unmeasured confounding?

(a) Which assumption(s) might be less plausible in this context and what could be an unmeasured confounder?

(b) What could be the impact of violation of the assumptions?

(c) What could be the impact of measurement error in this study?

(d) What could be an approach that might provide a solution to unmeasured confounding and measurement error simultaneously?

