PHS2000B Lab 3

Sensitivity Analysis and Measurement Error

02/04/2023

To gain familiarity with some of the equations presented during the sensitivity analysis and measurement lab, let's consider a few examples.

1 Sensitivity Analysis

Let's say we are interested in the effect of low educational attainment on risk of dementia. In a recent study, you estimated an adjusted risk ratio of 1.8 (95% CI: 1.4, 2.2) for the association between low educational attainment and risk of dementia conditional on a set of covariates. You are concerned that there is another variable U, household income, that may confound the association you observed. A recent meta-analysis of the effect of low household income on dementia suggests that it increases the risk of dementia by a factor of 2.3.

> How large must the association between low household income and educational attainment be in order for this unmeasured confounder to i) completely explain away the observed association and ii) shift the observed 95% CI to include the null?

For the point estimate, we can use our Bias Factor equation to estimate the following:

$$\begin{aligned}
BF &\leq \frac{RR_{UY} \times RR_{AU}}{RR_{UY} + RR_{AU} - 1} \\
1.8 &\leq \frac{2.3RR_{AU}}{RR_{AU} + 2.3 - 1} \\
1.8 &\leq \frac{2.3RR_{AU}}{RR_{AU} + 1.3} \\
1.8(RR_{AU} + 1.3) &\leq 2.3RR_{AU} \\
1.8RR_{AU} + 2.34 &\leq 2.3RR_{AU} \\
2.34 &\leq 0.5RR_{AU} \\
\boldsymbol{RR_{AU}} &\boldsymbol{\geq} \mathbf{4.68}
\end{aligned}$$

To completely explain away the effect, $\boldsymbol{RR_{AU}}$ must be <u>at least</u> 4.68.

For the confidence interval, we can do the same:

$$\begin{aligned}
BF &\leq \frac{RR_{UY} \times RR_{AU}}{RR_{UY} + RR_{AU} - 1} \\
1.4 &\leq \frac{2.3RR_{AU}}{RR_{AU} + 2.3 - 1} \\
1.4 &\leq \frac{2.3RR_{AU}}{RR_{AU} + 1.3} \\
1.4(RR_{AU} + 1.3) &\leq 2.3RR_{AU} \\
1.4RR_{AU} + 1.82 &\leq 2.3RR_{AU} \\
1.82 &\leq 0.9RR_{AU} \\
\boldsymbol{RR_{AU}} &\boldsymbol{\geq} \mathbf{2.02}
\end{aligned}$$

To shift the 95% CI to include the null, $\boldsymbol{RR_{AU}}$ must be at least 2.02.

Calculate the E-value using the risk ratio and 95% confidence interval.

For the point estimate:

$$\begin{aligned}\text{E-value} &= RR_{obs} + \sqrt{RR_{obs}(RR_{obs}-1)}\\ &= 1.8 + \sqrt{1.8(1.8-1)}\\ &= \mathbf{3}\end{aligned}$$

For the confidence interval:

$$\begin{aligned}\text{E-value} &= RR_{obs} + \sqrt{RR_{obs}(RR_{obs}-1)}\\ &= 1.4 + \sqrt{1.4(1.4-1)}\\ &= \mathbf{2.15}\end{aligned}$$

Interpret the E-value for the point estimate and the E-value for the 95% confidence interval.

An unmeasured confounder U (household income) that is associated with both the outcome (dementia) and the exposure (low educational attainment) by a risk ratio of 3, above and beyond the measured confounders, could explain away the estimated effect, but weaker confounding could not. To move the confidence interval to include the null, such a confounder would need to be associated with the outcome and exposure by risk ratios of at least 2.15 each.

A colleague boasts that, in their recent study of educational attainment and cognitive impairment, they obtained an E-value of 6 for the point estimate! Can the magnitude of this E-value be directly compared to the E-value we calculated in our study? Why or why not?

No - for a few reasons! An E-value is specific to the measurement of the exposure, outcome, and measured confounders - all values that went into the calculation of the observed risk ratio. For example, in your colleague's study, they looked at cognitive impairment as an outcome, whereas in your study, you looked at risk of dementia. These two E-values are therefore not comparable.

Now, let's say we have some additional information. We know that the prevalence of our unmeasured confounder, low household income, is 0.70 in the exposed. The risk ratio comparing low household income and dementia among both exposed and unexposed is $\gamma = 2.70$

What would the prevalence of low household income among the unexposed have to be to fully explain away the point estimate of 1.8?

In order to fully explain the point estimate of 1.8 from our previous example, an unmeasured confounder U would need to produce a bias factor $B_{mult}(c)$ of 1.80:

$$\begin{aligned}B_{mult}(c) &= \frac{1+(\gamma-1)P(U=1 \mid a_1,c)}{1+(\gamma-1)P(U=1 \mid a_0,c)}\\ 1.80 &= \frac{1+(2.70-1)0.70}{1+(2.70-1)P(U=1 \mid a_0,c)}\\ \boldsymbol{P(U=1 \mid a_0,c)} &= \mathbf{0.127}\end{aligned}$$

The prevalence of low household income among the unexposed would need to be **12.7%** to fully explain away the observed effect estimate.

What would the bias factor be if $\gamma = 4$ and, using subject matter knowledge, we know that the prevalence estimates for the unmeasured confounder were 0.55 among the exposed and 0.10 among the unexposed? Using this bias factor, what would the corrected point estimate be?

$$B_{mult}(c) = \frac{1 + (4-1)0.55}{1 + (4-1)0.10} = \mathbf{2.04}$$

We could then calculate a corrected point estimate:

$$RR_{obs} = RR_{obs}/B_{mult}(c) = 1.80/2.04 = \mathbf{0.88}$$

2 Measurement Error

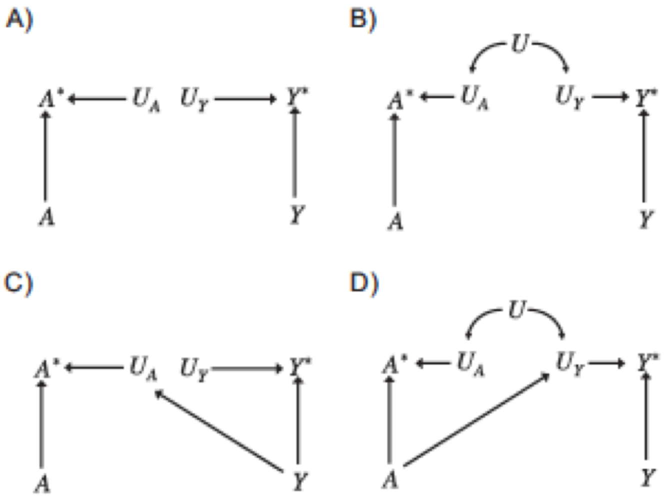

Figure 1: Types of Measurement Error.

For each of the above DAGs, identify the types of measurement error (include both independent vs dependent and differential vs non-differential in your answers).

A) Independent non-differential B) Dependent non-differential C) Independent differential D) Dependent differential[a]

---

[a] Hernán MA, Cole SR. 2009;170(8):959-962. VanderWeele TJ, Hernán MA. 2012;175(12):1303-10.

If you're testing for a rare disease, even with a highly specific test, there's a good probability that any given positive test result is a false positive. To see this in action, work through the following example

Let's say you have a test for disease X, which affects 1% of the population. This test is 97% sensitive and 98% specific. You test 1000 people. Because the prevalence of disease X is 1%, we know that 10 of these people truly have the disease; 990 do not.

What is the positive predictive value of the test?

The test is 97% sensitive, so of the 10 people who truly have the disease, $10 \times 0.97 = 9.7$ of them will have positive results (in expectation). We can round this to 10 for simplicity, and because there's no such thing as 0.7 people.

The test is 98% specific, so of the 990 people who do not have the disease, $990 \times 0.98 = 970.2$ will have negative results. Let's round this to 970. The remaining $990 - 970 = 20$ will have (false) positive results.

Therefore, there are 30 people $(10 + 20)$ with positive results, but only 10 of them actually have the disease. Positive predictive value $= 10/30 = 33\%$.

What is the negative predictive value of the test?

We know from above that 10 people who had the disease tested positive and 30 people who did not have the disease tested positive. Therefore, we had 960 people who tested negative. All of these individuals were, in fact, negative, so our NPV is 100%!

If you did not round to ten whole people, then 0.3 individuals with the disease would have tested negative. We would then have an NPV of $960/960.3 = 99.97\%$.