Epi 202 Lab 3: Practice Problems Proposed Solutions

Note to students: **please read** the following guidelines on when it is appropriate to use the phrase 'over the study period'. Going forward in the last class, we will deduct points both when the phrase is necessary but left out of your interpretations as well as when it is unnecessary and inappropriately included in your interpretations.

"Over the Study Period"

1. Should not be used as a phrase in interpreting measures of association. For measures of association that are estimated using count-based data it is helpful and appropriate to say "during the X-year follow-up period"
    - **a.** CIR/CID/OR in a closed cohort with no loss to follow-up
    - **b.** OR in a cumulative-incidence sampled case-control study
    - **c.** CIR as approximated from OR in a cumulative-incidence sampled case-control study where the rare disease assumption is met
2. Not necessary and often inappropriately used for rate-based measures (IRR/IRD) because time accounted for in calculation
    - **a.** IRR/IRD in cohort study
    - **b.** OR in an incidence-density or risk-set-sampled case-control study
    - **c.** Rationale for not using "over the study period" with rates: The rate (and rate ratio or difference) does not apply to all individuals over the entirety of the study period, because not everyone contributed follow-up time throughout the entirety of the study period. Rate only applies to individuals over the person-time actually contributed.
    - **d.** Ex: We are studying depression among sleep-deprived vs. non-deprived college freshmen in an open cohort active from 2000-2010. The study runs for 10 years, and each freshman contributes a maximum of one year to the cohort before they "age out".
    - **e.** IRR can be interpreted as the rate ratio of depression among exposed v. unexposed freshmen participating in this study.
    - **f.** IRR CANNOT be interpreted as the 10-year IRR of depression, or the IRR of depression "over the study period".
        - **i.** Interpreting it as such implies that you followed everyone for 10 years, when in fact, everyone contributed a maximum of one year.
        - **ii.** If you would like to give context about the study/time scale in which the measure was estimated, you could instead say something like: "Among college freshmen between 2000-2010, the IR of depression among the exposed was X times the IR of depression among the unexposed"

The data from an incidence density-sampled case-control study of the association between caffeine intake and Parkinson's disease is below. The exposure of interest is caffeine intake (high vs. low), and the outcome is Parkinson's disease. Cases were appropriately matched to controls (1:1 matching) by age and smoking status.

Pair #	Case exposure	Control exposure	Pair #	Case exposure	Control exposure
1	+	-	11	+	-
2	+	+	12	-	-
3	-	+	13	-	+
4	-	-	14	-	+
5	-	+	15	-	-
6	+	-	16	+	-
7	-	-	17	-	+
8	-	+	18	-	+
9	-	+	19	+	+
10	+	+	20	+	-

1. Make a 2x2 table for the matched case-control data

	Exposed controls	Non-exposed controls
Exposed cases	3	5
Non-exposed cases	8	4

2. Calculate and interpret the odds ratio for the association of coffee intake (high vs. low) and Parkinson's disease, accounting for the matching factors.

$$\hat{OR}_{MH} = \frac{f_{10}}{f_{01}}$$

= 5/8
= 0.625

The odds of Parkinson's disease for individuals with high caffeine intake is 0.625 times the odds of Parkinson's disease for individuals with low caffeine intake, after accounting for the matching factors (smoking status and age), assuming no residual confounding by smoking status and age, no other unmeasured confounding, no selection bias, and no information bias.

3. Test whether there is an association between caffeine intake (high vs. low) and the incidence of Parkinson's disease after accounting for the matching factors. State your hypotheses.

H_0: There is no association between caffeine intake (high vs. low) and the odds of Parkinson's disease, after accounting for the matching factors (smoking status and age)

H_1: There is an association between caffeine intake (high vs. low) and the odds of Parkinson's disease, after accounting for the matching factors (smoking status and age)

Mantel-Haenszel Chi-square test statistic (aka McNemar's test)

$$Z^2 = \frac{[f_{10} - f_{01}]^2}{f_{10} + f_{01}} \sim \chi_1^2$$

$= \frac{[5-8]^2}{5+8} = 9/13 = 0.692$ with 1 d.f.

$\Pr[\chi_1^2 > 0.69] = 0.4054$

These data are consistent with the state of nature described by the null. There is insufficient evidence in these data to reject the null hypothesis at alpha=0.05; we fail to reject the null hypothesis and conclude that, after accounting for the matching factors (smoking status and age), there is no statistically significant association between high caffeine intake compared to low caffeine intake and the odds of Parkinson's disease, assuming no residual confounding by smoking status and age, no other unmeasured confounding, no selection bias, and no information bias.

4) Compute and interpret the 95% confidence interval for the odds ratio that you computed.

RGB variance for $\ln(OR_{MH})$:

$$\text{Var}[ln(\hat{OR}_{MH})] = \frac{1}{f_{10}} + \frac{1}{f_{01}}$$

$= 1/5 + 1/8$
$= 0.325$

Recall that we calculated $OR_{MH}=0.625$
$\ln(OR_{MH}) = -0.47$

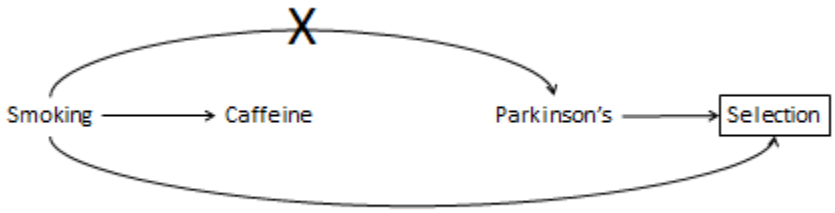
$\ln(OR_{MH}) \pm 1.96\sqrt{Var\ln[(OR_{MH})]}$
$= -0.47 \pm 1.96\sqrt{0.325}$

= -0.47 ± 1.96(0.57)
= -1.587, 0.647)
= e^(-1.587, 0.647)
= (0.20, 1.91)

(Note that the 95% CI is wide, indicating poor precision and lack of power, as there were only 13 matched pairs (5+8) that contributed information to the OR)

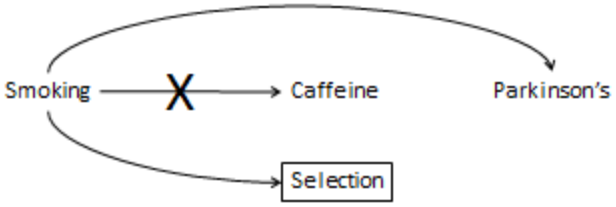
With 95% confidence, these data are consistent with odds ratios ranging from 0.20 to 1.91 for the association between high caffeine intake compared to low caffeine intake and the odds of Parkinson's disease, after accounting for the matching factors (smoking status and age), assuming no residual confounding by smoking status and age, no other unmeasured confounding, no selection bias, and no information bias.

5) Suppose the investigators were interested in whether smoking modifies the association between caffeine intake and the incidence of Parkinson's disease. Could this be evaluated in these case-control data? Could the investigators evaluate smoking status as an independent predictor of the incidence of Parkinson's disease in these data? Draw a DAG including smoking status, caffeine intake, and Parkinson's disease to support your conclusions.



The investigators cannot evaluate the main effect of smoking on the incidence of Parkinson's disease in their matched case-control study as, by design, the distribution of smoking is the same among the cases and controls. There is no arrow between smoking status and Parkinson's disease after matching, so the main effect of smoking cannot be evaluated. The investigators can still evaluate whether smoking status modifies the association between caffeine intake and the incidence of Parkinson's disease.

6) Imagine another group of investigators were examining the association between caffeine intake and the incidence rate of Parkinson's disease in a prospective cohort where individuals with low and high caffeine intake were matched on age and smoking status. Could these investigators evaluate effect measure modification by smoking status in their matched cohort? Could the investigators evaluate the overall association between smoking status and the incidence rate of Parkinson's disease in their cohort? Draw a DAG including smoking status, caffeine intake, and Parkinson's disease to support your conclusions.



The investigators can evaluate both the main effect of smoking status on the incidence rate of Parkinson's disease as well as possible effect modification by smoking status for the association between caffeine intake and Parkinson's disease. The investigators matched individuals with high and low caffeine intake on smoking status, so the distribution of smoking is the same within levels of caffeine. This removes the arrow from smoking status to caffeine intake. There is still an arrow from smoking status to the incidence of Parkinson's disease so the main effect of smoking can still be evaluated in this cohort.