EPI 202 Lab 2 Practice Problem Solutions

1. **In a study of malnutrition and diarrhea in children under age five in the Sudan, 154 episodes of diarrhea were observed in 38.51 person-years of observation among children who were 75% or less of their expected weight for age, and 287 episodes of diarrhea were observed in 88.83 person-years of observation among children who were 90% or more of their expected weight for age (El Samani et al.; Am J Epidemiol, 1988). Suppose the data from this study were broken down by sex as follows:**

Males

	≤ 75% WFA	≥ 90% WFA
Cases	75	150
PY	20.5	48

Females

	≤ 75% WFA	≥ 90% WFA
Cases	79	137
PY	18.01	40.83

From this data, we can see that, among unexposed children (those with >90% weight for age), the IRR of diarrhea comparing females & males is given by:

- IRR=IR(Females|≥90% WFA)/IR(Males|≥90% WFA)=(150/48)/(137/40.83)=0.931

Additionally, we can see that the OR comparing the odds of exposure between the female and male study bases is given as:

- OR=Odds(Exposed|Female)/Odds(Exposed|Male)=(18.01/40.83)/(20.5/48)=1.03

The crude IRR = IR(≤ 75% WFA)/IR(≥ 90% WFA) = (154/38.51)/(287/88.83) = 1.24

a. Calculate the summary incidence rate ratio for the association between weight for age and rate of diarrhea. Interpret your answer.

	a_i	b_i	N_{0i}	N_{1i}	T_i
i=1 (male)	75	150	48	20.5	69
i=2 (female)	79	137	40.83	18.01	59

Summary Incidence Rate Ratio

$$IRR_{MH} = \frac{\sum_{i=1}^{I} \frac{a_i N_{0i}}{T_i}}{\sum_{i=1}^{I} \frac{b_i N_{1i}}{T_i}}$$

$$= \frac{\left[\left(\frac{75 * 48}{69}\right) + \left(\frac{79 * 40.83}{59}\right)\right]}{\left[\left(\frac{150 * 20.5}{69}\right) + \left(\frac{137 * 18.01}{59}\right)\right]} = \frac{106.884}{86.385} = 1.237$$

After adjusting for sex, the incidence rate of diarrhea is 1.24 times higher in children who are less than 75% of their expected weight for age compared to children who are 90% or more of their expected weight for age, assuming no residual confounding, confounding by other variables, selection bias, or information bias.

b. Calculate the summary incidence rate difference for the association between weight for age and rate of diarrhea. Interpret your answer.

Summary Incidence Rate Difference:
Where the MH summary IRD is:

$$\frac{\sum w_i \left[\frac{a_i}{N_{1i}} - \frac{b_i}{N_{0i}} \right]}{\sum w_i}$$

and the weights are:

$$\frac{N_{1i} N_{0i}}{T_i}$$

Values for a, b, and N can be found in the previous table.
*Note: the weight is MH weight of the strata (difference)

Weights for males: (48*20.5)/69 = 14.26087

Weights for females: (40.83*18.01)/59 = 12.46353

Sum of weights: 14.26087 + 12.46353= 26.7244

IRD for males: (75/20.5) - (150/48) = 0.5335
IRD for females: (79/18.01) - (137/40.83) = 1.03107

Numerator of summary IRD: 14.26087 (0.5335) + 12.46353 (1.03107)= 20.45895
Denominator of summary IRD= sum of the weights= 26.7244

20.45895/26.7244= 0.766 cases/person-year =76.6/100 PY

After adjusting for sex, there were 76.6 more cases of diarrhea per 100 person-years in children who are less than 75% of their expected weight for age compared to children who are 90% or more of their expected weight for age, assuming no residual confounding, confounding by other variables, selection bias, or information bias.

c. Test the hypothesis that weight for age has no association with rate of diarrhea, after stratifying by sex. Interpret the numeric results.

H_0: The rate of diarrhea does not differ between those children who are less than 75% of their expected weight for age compared to children who are 90% or more of their expected weight for age, after stratifying by sex.

$$IRR_{MH} = 1 \leftrightarrow IRD_{Summ} = 0$$

H_A: The rate of diarrhea does differ between those children who are less than 75% of their expected weight for age compared to children who are 90% or more of their expected weight for age, after stratifying by sex.

$$IRR_{MH} \neq 1 \leftrightarrow IRD_{Summ} \neq 0$$

$$Z^2 = \frac{\left[\sum_{i=1}^{I} X_i - \sum_{i=1}^{I} E_i(X_i|H_0)\right]^2}{\sum_{i=1}^{I} Var_i(X_i|H_0)} \sim \chi_1^2$$

where $X = \sum a_i$,

$E(X|H_0) = \sum \frac{N_{1i}M_{1i}}{T_i}$,

$Var(X|H_0) = \sum \frac{N_{1i}N_{0i}M_{1i}}{T_i^2}$

Values for a, b, and N can be found in the previous table.
X= 75+79= 154
$E(X|H_0)$= 20.5(225)/68.8+18.01(216)/58.84= 133.4499744451
$Var(X|H_0)$= $20.5(48)(225)/68.8^2+18.01(40.83)(216)/58.84^2$= 93.0618738684

Z^2= $(154-133.450)^2/93.062$= 4.54

The epi calculator gets a Z-square= 4.58
p-value= 0.0324

These data are not consistent with the state of nature described by the null.
We can reject the null hypothesis at the alpha=0.05 level and conclude that there is statistically significant evidence that after stratifying by sex, the rate of diarrhea does differ between those children who are less than 75% of their expected weight for age compared to children who are 90% or more of their expected weight for age, assuming no residual confounding, confounding from other variables, selection bias, or information bias.

d. Do you think sex is a confounder based on the above analyses?

In order for sex to be a confounder of the relationship between WFA and diarrheal disease, there would need to be an open backdoor path through sex. A simple version of this is if sex is a common cause of WFA and diarrheal disease.

Confounders will also have three properties: 1) they are associated with the exposure in the study base, 2) they are associated with the outcome in the unexposed, and 3) they are not a downstream consequence of the exposure or outcome. We can empirically check conditions 1) and 2), and must use substantive knowledge to decide about 3). Importantly, not all variables with these properties will be confounders, so the 'backdoor path' (causal) definition described above should be used to get an understanding of the structure of the relationship.

From the information provided above, we can see that sex is weakly associated with diarrhea in children ≥90% WFA (IRR=0.93) and weakly associated with weight for age (IRR=1.03). Additionally, sex is not a downstream consequence of weight for age or diarrhea. However, the associations between sex with the exposure and outcome are both very weak and may be the result of sampling variability. From this information, sex is not likely to be a strong confounder.

Finally, we can note that the crude and adjusted analyses produced almost identical results. This is indirect, but not irrefutable, evidence that sex is unlikely to be a strong confounder.

2. A cigarette smoking history was obtained from 6,690 Japanese-American men examined from 1965 through 1968. During the 22-year follow-up period, 37 incident cases of oral or bladder cancer were observed among the 2,344 never smokers, and 165 incident cases or oral or bladder cancer were observed among the 4,346 past and current smokers, as shown in the table below (Chyou PH, Nomura AMY, Stemmermann GN. Am J Public Health 1992;83:37-40). Assume no loss to follow-up and no competing causes of death for this and all remaining questions about these data.

The crude table and cumulative incidence for oral or bladder cancer (as calculated last week) are shown below:

	Ever smoked	Never smoked	Total
Cases	165	37	202
Non-cases	4,181	2,307	6,488
Total	4,346	2,344	6,690

$CI_{exposed}$ = 165 / 4,346 = 0.038 over the 22-year study period
$CI_{unexposed}$ = 37 / 2,344 = 0.016 over the 22-year study period

Last week, we concluded there was statistically significant evidence of an association between ever smoking compared to never smoking and the 22-year cumulative incidence of oral or bladder cancer (assuming no confounding, selection bias, information bias).

We also calculated the following crude measures of association comparing ever smokers to never smokers (reference group):
CIR = 2.41 over the 22-year study period, 95% confidence interval = (1.69, 1.43) over the 22-year study period
CID = 0.0222 over the 22-year study period, 95% confidence interval = (0.014, 0.030) over the 22-year study period

Suppose the data from this study were broken down by age as follows:

	≤55 years old			>55 years old		
	Ever smoked	Never smoked	Total	Ever smoked	Never smoked	Total
Cases	93	19	112	72	18	90
Non-cases	2,860	1,085	3,945	1,321	1,222	2,543
Total	2,953	1,104	4,057	1,393	1,240	2,633

a. Calculate the cumulative incidence ratio for the association between ever smoking compared to never smoking and oral or bladder cancer within each stratum of age (≤55, >55 years).

$CIR_{\leq 55}$ = (93/2953) / (19/1104) = 1.83 over the 22-year study period
$CIR_{>55}$ = (72/1393) / (18/1240) = 3.56 over the 22-year study period

b. Calculate the cumulative incidence difference for the association between ever smoking compared to never smoking and oral or bladder cancer within each stratum of age (≤55, >55 years).

$CID_{\leq 55}$ = (93/2953) - (19/1104) = 0.0143 over the 22-year study period
$CID_{>55}$ = (72/1393) - (18/1240) = 0.0372 over the 22-year study period

c. Do the results from 2a and 2b suggest there might effect measure modification on:

The multiplicative scale?

There appears to be effect measure modification on the multiplicative scale. The CIR among men who are older than 55 years (3.56 over the 22-year study period) is nearly double the CIR among men who are 55 years or younger (1.83 over the 22-year study period).

The additive scale?

It seems like there is effect measure modification on the additive scale. The CID among men who are older than 55 years (0.0372 over the 22-year study period) is more than double the CID among men who are 55 years or younger (0.0143 over the 22-year study period).

d. For the stratified data, perform the test of homogeneity for the cumulative incidence ratio. State the null and alternative hypotheses. Interpret your findings.

H_0: There is no effect measure modification of the association between ever versus never smoking and oral or bladder cancer across the strata of age on the multiplicative scale ($CIR_{\leq 55} = CIR_{>55}$)

H_1: There is effect measure modification of the association between ever versus never smoking and oral or bladder cancer across the strata of age on the multiplicative scale ($CIR_{\leq 55} \neq CIR_{>55}$)

$$H = \sum_{i=1}^{I} \frac{[\hat{X}_i - \hat{X}_{summary}]^2}{\hat{Var}_i[\hat{X}_i]} \sim \chi^2_{I-1}$$

$X_i = \ln(CIR_i)$,
$X_{summ} = \ln(CIR_{MH})$
$Var(X_i) = Var(\ln(CIR_i))$
Degrees of freedom = # of strata - 1 = 2-1 = 1 degree of freedom

$$\hat{CIR}_{MH} = \frac{\sum_{i=1}^{I} w_i' \hat{CIR}_i}{\sum_{i=1}^{I} w_i'} = \frac{\sum_{i=1}^{I} \frac{b_i N_{1i}}{T_i} \left(\frac{a_i}{N_{1i}}\right) \Big/ \left(\frac{b_i}{N_{0i}}\right)}{\sum_{i=1}^{I} \frac{b_i N_{1i}}{T_i}} = \frac{\sum_{i=1}^{I} \frac{a_i N_{0i}}{T_i}}{\sum_{i=1}^{I} \frac{b_i N_{1i}}{T_i}}$$

CIR_{MH} = [(93*1104/4057) + (72*1240/2633)] / [(19*2953/4057) + (18*1393/2633)]
= 2.54 over the 22-year study period

$\ln(CIR_{MH}) = \ln(2.54) = 0.932$

From question 2a:
$\ln(CIR_{\leq 55}) = \ln(1.830) = 0.604$
$\ln(CIR_{>55}) = \ln(3.561) = 1.27$

$Var(\ln(CIR_i)) = \frac{c_i}{a_i N_{1i}} + \frac{d_i}{b_i N_{0i}}$

$Var(\ln(CIR_{\leq 55})) = 2860/(93*2953) + 1085/(19*1104) = 0.06214$
$Var(\ln(CIR_{>55})) = 1321/(72*1393) + 1222/(18*1240) = 0.06792$

$$H = \sum_{i=1}^{I} \frac{[\hat{X}_i - \hat{X}_{summary}]^2}{\hat{Var}_i[\hat{X}_i]} \sim \chi^2_{I-1}$$

$Z^2 = [(0.604-0.932)^2 / [0.06214] + [(1.27-0.932)^2] / [0.06792]$
$= 1.731 + 1.682 = 3.41 \sim \chi^2$ with 1 df
p-value =0.065

At the alpha=0.05 level, we fail to reject the null hypothesis. We do not have sufficient evidence to conclude that there is effect measure modification by age on the multiplicative scale for the association between ever smoking compared to never smoking and the cumulative incidence of oral or bladder cancer (assuming no confounding, selection bias, or information bias).

However, note that the p-value is quite close to 0.05 and studies are often underpowered for tests of homogeneity.

e. For the stratified data, perform the test of homogeneity for the cumulative incidence difference. State the null and alternative hypotheses. Interpret your findings.

H_0: There is no effect measure modification of the association between ever versus never smoking and oral or bladder cancer across the strata of age on the additive scale ($CID_{\leq 55} = CID_{>55}$)

H_1: There is effect measure modification of the association between ever versus never smoking and oral or bladder cancer across the strata of age on the additive scale ($CID_{\leq 55} \neq CID_{>55}$)

$$H = \sum_{i=1}^{I} \frac{[\hat{X}_i - \hat{X}_{summary}]^2}{\hat{Var}_i[\hat{X}_i]} \sim \chi^2_{I-1}$$

$X_i = CID_i$
$X_{summ} = CID_{SUMM}$
$Var(X_i) = Var(CID_i)$
Degrees of freedom = # of strata - 1 = 2-1 = 1 degree of freedom
The summary CID is:

$$\frac{\sum w_i \left[\frac{a_i}{N_{1i}} - \frac{b_i}{N_{0i}}\right]}{\sum w_i}$$ where the weight is

$$\frac{N_{1i}N_{0i}}{T_i}$$

From question 2b:
$CID_{\leq 55}$ = 0.0143 over the 22-year study period
$CID_{>55}$ = 0.0372 over the 22-year study period

$w_{\leq 55}$ = (2953*1104)/4057 = 803.577
$w_{>55}$ = (1393*1240)/2633 = 656.0273

CID_{Summ} = [(803.577*0.0143) + (656.0273*0.0372)] / (803.577+ 656.0273) = 0.02459253 over the 22-year study period

$Var(CID_i) = \frac{a_i c_i}{N_{1i}^3} + \frac{b_i d_i}{N_{0i}^3}$

	≤55 years old			>55 years old		
	Ever smoked	Never smoked	Total	Ever smoked	Never smoked	Total
Cases	93	19	112	72	18	90
Non-cases	2,860	1,085	3,945	1,321	1,222	2,543
Total	2,953	1,104	4,057	1,393	1,240	2,633

$Var(CID_{\leq 55})$ = ((93*2860)/(2953^3)) + ((19*1085)/(1104^3)) = 2.564962e-05
$Var(CID_{>55})$ = ((72*1321)/(1393^3)) + ((18*1222)/(1240^3)) = 4.67236e-05

Z^2 = ((0.0143-0.0246)^2 / 2.564962e-05) + ((0.0372-0.0246)^2 / 4.67236e-05)
= 4.136124 + 3.397855 = 7.533 ~ χ^2 with 1 df
p-value =0.007

At the alpha=0.05 level, we reject the null hypothesis. We have sufficient evidence to conclude that there is effect measure modification by age on the additive scale for the association between ever smoking compared to never smoking and the cumulative incidence of oral or bladder cancer (assuming no confounding, selection bias, or information bias).

f. Given these results, would you present a MH summary CIR or summary CID?

Though we did not reject the null hypothesis of no effect measure modification on the multiplicative scale for the association between smoking and the cumulative incidence of oral or bladder cancer, we can see that the CIR among men who are older than 55 years (3.56 over the 22-year study period) is nearly double the CIR among men who are 55 years or younger (1.83 over the 22-year study period). Additionally, studies are often underpowered to detect statistically significant effect measure modification. As a result, presenting an MH summary CIR

may be inappropriate. Instead, we may want to present CIR stratified by age. Or, if we want to present a summary CIR, we could standardize by using population-based weights (which we will discuss this in a future lecture).

We would not want to present a summary CID, as we found statistically significant evidence of effect measure modification of the association between smoking and the cumulative incidence of oral or bladder cancer on the additive scale. We could present cumulative incidence differences stratified by age. Or, if we wanted to present a summary measure, we could present a standardized CID (which we will discuss in a future lecture).

3. We are interested in evaluating if the association between low birth weight and risk of autism is different among women that received at least one dose of acetaminophen during pregnancy. The unconfounded risks of autism are presented below:

	No Acetaminophen	Acetaminophen
Normal Birth Weight	1.00	1.04
Low Birth Weight	1.08	2.25

   a. Calculate RERI and interpret the meaning of this quantity in words.

RERI = 2.25 - 1.08 - 1.04 + 1 = 1.13
Since RERI > 1, we can say that there is additive effect modification by acetaminophen on the association between birth weight and autism.

For those infants with low birth weight whose mothers used at least one dose of acetaminophen during pregnancy, 1.13 represents the additional cumulative incidence of autism that is due to the interaction (above and beyond individual effects).

b. Assuming that both acetaminophen and low birth weight are harmful for all individuals with respect to their risk of autism, can you make any claims about acetaminophen and low birth weight in terms of the sufficient and component causes framework?

Yes. We can say that there is sufficient cause interaction. That is, we can state that low birth weight and acetaminophen use during pregnancy are both present in the same sufficient cause pie (i.e., there is mechanistic interaction between exposure and interaction).