Lab 6. Practice Problem Solutions

The tables below show data from a hypothetical cohort study of the relationship between aspirin use (yes versus no) and 5-year risk of stroke (assume there was no loss to follow-up or competing risks). The tables are stratified by sex, a suspected confounder.

Male

	Aspirin use	No aspirin use	Total
Stroke	60	40	100
No stroke	90	10	100
Total	150	50	200

Female

	Aspirin use	No aspirin use	Total
Stroke	75	60	150
No stroke	75	40	100
Total	150	100	250

1. Using standardization, compute the cumulative incidence ratio for the total study population. Briefly interpret your results.

Let A represent aspirin intake (1=yes, 0=no), Y represent stroke (1=yes, 0=no), and L represent gender (1=male, 0=female).

$\Pr[Y^{a=1}=1] = \Sigma_l \Pr[Y=1 \mid A=1, L=l]*\Pr[L=l]$
$= \Pr[Y=1 \mid A=1, L=1]*\Pr[L=1] + \Pr[Y=1 \mid A=1, L=0]*\Pr[L=0]$
$= (60/150)*(200/450) + (75/150)*(250/450)$
$= 0.1778 + 0.2778 = 0.4556$

$\Pr[Y^{a=0}=1] = \Sigma_l \Pr[Y=1 \mid A=0, L=l]*\Pr[L=l]$
$= \Pr[Y=1 \mid A=0, L=1]*\Pr[L=1] + \Pr[Y=1 \mid A=0, L=0]*\Pr[L=0]$
$= (40/50)*(200/450) + (60/100)*(250/450)$
$= 0.3556 + 0.3333 = 0.6889$

$\text{CIR} = [Y^{a=1}=1] / [Y^{a=0}=1]$

= 0.4556/0.6889 = 0.6613 over 5 years

Alternative approach:

$$\frac{\widehat{R_1}}{\hat{R}_0} = \frac{\sum_{i=1}^{I} \frac{a_i T_i}{N_{1i}}}{\sum_{i=1}^{I} \frac{b_i T_i}{N_{0i}}} = \frac{\left[\frac{60 * (200)}{150}\right] + \left[\frac{75 * (250)}{150}\right]}{\left[\frac{40 * (200)}{50}\right] + \left[\frac{60 * (250)}{100}\right]} = \frac{205}{310} = 0.6613 \ over\ 5\ years$$

If everyone in the study had taken aspirin, we would have observed 0.66 times the 5-year cumulative incidence of stroke compared to if no one in the study had taken aspirin (assuming no residual confounding by sex, no confounding by other variables, no selection bias, and no information bias).

2. Using standardization, compute the cumulative incidence ratio for the exposed. Briefly interpret your results.

$\Pr[Y^{a=1}=1 \mid A=1] = \Sigma_l \Pr[Y=1 \mid A=1, L=l]*\Pr[L=l \mid A=1]$
$= \Pr[Y=1 \mid A=1, L=1]*\Pr[L=1 \mid A=1] + \Pr[Y=1 \mid A=1, L=0]*\Pr[L=0 \mid A=1]$
= (60/150)*(150/300) + (75/150)*(150/300)
= 0.2 + 0.25 = 0.45

$\Pr[Y^{a=0}=1 \mid A=1] = \Sigma_l \Pr[Y=1 \mid A=0, L=l]*\Pr[L=l \mid A=1]$
$= \Pr[Y=1 \mid A=0, L=1]*\Pr[L=1 \mid A=1] + \Pr[Y=1 \mid A=0, L=0]*\Pr[L=0 \mid A=1]$
= (40/50)*(150/300) + (60/100)*(150/300)
= 0.4 + 0.3 = 0.7

CIR in exposed = $[Y^{a=1}=1 \mid A=1] / [Y^{a=0}=1 \mid A=1]$
= 0.45/0.7 = 0.6429 over 5 years

Alternative approach:

$$\widehat{SMR} = \frac{\sum_{i=1}^{I} a_i}{\sum_{i=1}^{I} \frac{b_i N_{1i}}{N_{0i}}} = \frac{60 + 75}{\left[\frac{40 * (150)}{50}\right] + \left[\frac{60 * (150)}{100}\right]} = \frac{135}{210} = 0.6429 \ over\ 5\ years$$

If all the exposed had taken aspirin, we would have observed 0.64 times the 5-year cumulative incidence of stroke compared to if none of the exposed had taken aspirin (assuming no residual confounding by sex, no confounding by other variables, no selection bias, and no information bias).

3. Using standardization, compute the cumulative incidence ratio for the unexposed. Briefly interpret your results.

$\Pr[Y^{a=1}=1 \mid A=0] = \Sigma_l \Pr[Y=1 \mid A=0, L=l]*\Pr[L=l \mid A=0]$
$= \Pr[Y=1 \mid A=1, L=1]*\Pr[L=1 \mid A=0] + \Pr[Y=1 \mid A=1, L=0]*\Pr[L=0 \mid A=0]$
$= (60/150)*(50/150) + (75/150)*(100/150)$
$= 0.1333 + 0.3333 = 0.4667$

$\Pr[Y^{a=0}=1 \mid A=0] = \Sigma_l \Pr[Y=1 \mid A=0, L=l]*\Pr[L=l \mid A=0]$
$= \Pr[Y=1 \mid A=0, L=1]*\Pr[L=1 \mid A=0] + \Pr[Y=1 \mid A=0, L=0]*\Pr[L=0 \mid A=0]$
$= (40/50)*(50/150) + (60/100)*(100/150)$
$= 0.2667 + 0.4000 = 0.6667$

CIR in exposed $= [Y^{a=1}=1 \mid A=0] / [Y^{a=0}=1 \mid A=0]$
$= 0.4667/0.6667 = 0.7000$ over 5 years

Alternative approach:

$$\widehat{SRR} = \frac{\sum_{i=1}^{I} \frac{a_i N_{0i}}{N_{1i}}}{\sum_{i=1}^{I} b_i} = \frac{\left[\frac{60*(50)}{150}\right] + \left[\frac{75*(100)}{150}\right]}{40+60} = \frac{70}{100} = \boldsymbol{0.700\ over\ the\ 5\ years}$$

If all the unexposed had taken aspirin, we would have observed 0.70 times the 5-year cumulative incidence of stroke compared to if none of the unexposed had taken aspirin (assuming no residual confounding by sex, no confounding by other variables, no selection bias, and no information bias).

4. Using inverse probability weighting, compute the cumulative incidence ratio for the total study population. Briefly interpret your results. Compare your results with those from question 1.

L	**A**	**Y**	**N(L,A,Y)**	**f[A|L]**	**W(L,A,Y) (1/f[A|L])**	**PS (W*N)**
0	0	1	60	100/250 = 0.4	2.5	150
0	0	0	40	100/250 = 0.4	2.5	100
0	1	1	75	150/250 = 0.6	1.67	125
0	1	0	75	150/250 = 0.6	1.67	125
1	0	1	40	50/200 = 0.25	4	160
1	0	0	10	50/200 = 0.25	4	40
1	1	1	60	150/200 = 0.75	1.33	80
1	1	0	90	150/200 = 0.75	1.33	120

Pseudopopulation Collapsed over L

	A=1	A=0	Total
Y=1	125+80 = 205	150+160 = 310	515
Y=0	125+120 = 245	100+40 = 140	385
	450	450	900

Causal risk ratio
= $\Pr[Y^{a=1}=1] / \Pr[Y^{a=0}=1]$
= [205/450]/[310/450]
= 0.4556/0.6889
= 0.6613 over 5 years

If everyone in the study had taken aspirin, we would have observed 0.66 times the 5-year cumulative incidence of stroke compared to if no one in the study had taken aspirin (assuming no residual confounding by sex, no confounding by other variables, no selection bias, and no information bias).

The causal CIRs calculated in questions 1 and 4 are identical. This is because both standardization and IPW are non-parametric methods to adjust for confounding.

5. Would be appropriate to present a Mantel-Haenszel summary estimate for the data above? Why or why not?

No, if you do the Test of Homogeneity (using the EPI202 calculator), there is significant evidence to reject the null that all strata are equal on both the additive ($p=0.0014$) and multiplicative scales ($p = 0.0021$). Thus, stratum-specific estimates should be presented instead of Mantel-Haenszel summary estimates. It would also be appropriate to report the standardized or inverse probability-weighted effect measures, since these weights reflect the population structure, in contrast with the MH weights which are derived from the data.