E2040 Introduction to Epidemiology and Environmental Health
Measures of association and effect —Practice Exercises

Exercise 1

In their study of the relationship between smoking and lung cancer, Wynder and Graham (1950) identified 100 men with cancer of the lungs (cases) and 186 men who had other chest diseases but no cancer. After identifying these men, the researchers asked them about their lifetime smoking habits. Individuals who smoked pipes or cigars were included in this dataset (1 cigar was equivalent to 5 cigarettes and 1 pipeful as 2.5 cigarettes). The researchers then divided their study participants into 6 categories, as follows:

0 = nonsmokers (less than 1 cigarette per day for over 20 years)
1 = light smokers (1-9 cigarettes per day for more than 20 years)
2 = moderately heavy smokers (10-15 cigarettes per day for more than 20 years)
3 = heavy smokers (16-20 cigarettes per day for more than 20 years)
4 = excessive smokers (21-34 cigarettes per day for more than 20 years)
5 = chain smokers (35+ cigarettes per day for more than 20 years)

This type of study is called case-control study (we will learn more about this in the coming weeks), and the disease being observed is *prevalent* disease.

Smoking intensity	Cancer status		Total	Prevalence by smoking intensity
	Cancer	**No cancer**		
None	0	26	**26**	
Light	5	20	**25**	
Moderately heavy	7	21	**28**	
Heavy	35	76	**111**	
Excessive	30	27	**57**	
Chain	23	16	**39**	
Total	**100**	**186**	**286**	

1. Describe the observed prevalence of disease depending on smoking intensity.

2. What level of disease would you expect in these patients? Use the above table to calculate the expected cancer prevalence for each cell.

3. Calculate the X^2 statistics to understand if there is an association between cancer status and smoking in the population.

4. Based on the correct DF and a 2-sided alpha of 0.05, would you say that smoking intensity is related or unrelated to lung cancer?

Exercise 2

The hypothetical dataset below contains data on 30 adults from Brno who developed bronchitis during the week of October 10, 2023. Information on sex (1=male, 2=female), smoking (0=non-smoker, 1=smoker) and bronchitis (0=no, 1=yes) is available for each person.

Case number	Sex	Smoking status	Bronchitis
1	1	1	0
2	1	1	1
3	2	0	0
4	1	0	0
5	1	1	1
6	2	1	1
7	2	0	0
8	2	0	0
9	2	0	1
10	1	1	1
11	1	0	0
12	1	1	0
13	2	1	0
14	1	0	0
15	2	0	1
16	2	0	0
17	1	1	1
18	1	1	1
19	1	0	0
20	2	0	0
21	2	0	0
22	1	1	0
23	1	1	1
24	1	0	0
25	2	1	1
26	1	0	0
27	1	0	1
28	2	1	0
29	1	1	1
30	2	1	1

1. What is the prevalence of smoking in this sample:
   a. Overall?
   b. In men and women separately?

2. Create the following 2x2 tables:
   a. Bronchitis incidence by sex
   b. Bronchitis incidence by smoking status

3. Calculate the following:
    a. Incidence of bronchitis in men and women
    b. Incidence of bronchitis in smokers and non-smokers

4. Do you think the incidence of bronchitis is different among smokers vs. non-smokers?
    a. State your null and alternative hypothesis about this relationship.
    b. Select a test of statistical significance to carry out your hypothesis test.
    c. Perform the calculations.
    d. Interpret the results and state your conclusions.

Exercise 3

One study followed 3,600 employees of London Underground for 20 years. The employees completed a smoking questionnaire at the start of the study and investigators tracked the incidence of lung cancer throughout the duration of the study.

	Lung cancer	No cancer	Total
Smoker	300	1000	
Non-smoker	100	2200	
Total			

a. Complete all parts of the table above. Calculate the absolute risk of developing lung cancer among smokers and non-smokers, as well as overall.

b. Calculate the relative risk (RR) of lung cancer by smoking status. Interpret the RR, assuming that it is statistically significant (its 95% CI does not cross 1).

c. What do you conclude from your calculations? Which statistics is more informative with respect to the impact of smoking on lung cancer risk in the population? Explain your answer.

Exercise 4

In a study of 1000 retired policemen in Austria, researchers tracked the development of liver cancer over a 25-year period. Half of the policemen consumed alcohol regularly during their retirement and 20 cases of liver cancer occurred in this group. In the rest of the study population, 10 cases were diagnosed.

a. Create a table to illustrate the data from this study. What was the incidence of cancer in each group?

Alcohol consumption	Liver cancer	No cancer	Total	Incidence
Regular consumer				
Non-consumer				
Total				

b. What is the relative risk of liver cancer among regular consumers vs. non-consumers of alcohol?

c. Calculate the odds ratio for the same association. What do you conclude from the results of (b) and (c)?

d. What other information would you like to have about these individuals to better understand the relationship between drinking and risk of liver cancer? List the variables/type of information you wish were available and explain how it might be a risk factor for liver cancer.

Exercise 5

Ignac Semmelweis (1818-1865) began his medical career in 1844 in obstetrics and midwifery at the Vienna General Hospital (Allgemeines Krankenhaus). There were two maternity wards (divisions) in the hospital: patients in the first division were examined by doctors and medical students, while midwives cared for patients in the second division. Semmelweis noticed that there were more maternal deaths in the first division than in the second. In this exercise, you will follow Semmelweis' steps in investigating a problem. Table 1 below lists Semmelweis' first observations.

Year	First division			Second division		
	Births	Deaths	Mortality rate	Births	Deaths	Mortality rate
1841	3,036	237		2,442	86	
1842	3,287	518		2,659	202	
1843	3,060	274		2,739	169	
1844	3,157	260		2,956	68	
1845	3,492	241		3,241	66	
1846	4,010	459		3,754	105	
Total						

a. Calculate the maternal mortality rate for each year and overall for the two maternity wards. Complete the table above.

Maternal mortality rate = ($number\ of\ maternal\ deaths/number\ of\ live\ births$) x 100,000

b. Do you agree with the conclusion Semmelweis reached that there were more deaths in the first division?

c. Is it necessary to calculate the death rate in each year to compare the two divisions?

Semmelweiss observed that the medical students and doctors who attended the births often did so soon after performing autopsies on deceased patients. On the other hand, the midwives were not involved in performing any autopsies. Based on this observation, Semmelweiss recommended that the medical students and doctors clean their hands with a chlorine solution after performing an autopsy. This measure began in May of 1847 and again, Semmelweiss recorded the number of maternal deaths. Table 2 below shows maternal deaths before and after the intervention.

Year	Births	Deaths	Mortality rate
Jan-Apr 1846	1193	194	
May-Aug 1846	1039	140	
Sep-Dec 1846	1120	125	
Jan-Apr 1847	1240	84	
Total			
INTERVENTION			
May-Aug 1847	1076	50	
Sep-Dec 1847	1059	42	
Jan-Apr 1848	1155	14	
May-Aug 1848	1107	7	
Total			

d. Calculate maternal mortality rates for each period prior to and after the intervention. Complete the above table.

e. Was the implemented intervention successful? How would you test this hypothesis?

f. Briefly comment on the significance and implications of this research from the point of view of epidemiology and clinical practice.