



Hughes
Undergraduate
Biological
Science
Education
Initiative



**Tracking the Source of Disease:
Koch's Postulates, Causality, and Contemporary Epidemiology**

Koch's Postulates

In the late 1800's, the German physician Robert Koch conducted experiments on the cause of anthrax, a disease of cattle, sheep, and humans. His experiments demonstrated that this disease was caused by a bacterium, *Bacillus anthracis*, and also established the experimental requirements for determining whether a particular pathogen is the cause of a particular disease. These requirements, known as Koch's postulates, are as follows:

- 1.) The same pathogen must be present in every case of the disease.
- 2.) The pathogen must be isolated from the diseased host and grown in pure culture.
- 3.) The pathogen from the pure culture must cause the disease when it is inoculated into a healthy, susceptible laboratory animal.
- 4.) The pathogen must be isolated from the inoculated animal and must be shown to be the original organism.

from Tortora et al., 1998

While Koch's postulates have provided a solid framework for infectious disease research for more than a century, they are not universally applicable. Classes of pathogens such as viruses and prions had not even been discovered in the late 1800's when Koch was formulating his ideas. Many of these more newly discovered pathogens as well as many bacterial pathogens are not culturable outside of a host organism. Furthermore, ethical considerations prevent researchers from fulfilling the third postulate in serious diseases affecting only human hosts. Additional complications arise from situations in which the same organism causes different diseases under different circumstances or diseases that may be caused by a community of microorganisms rather than a single pathogen. Hence, exceptions to Koch's postulates must often be made. This is not to say that there should not be strict requirements for demonstrating that a pathogen is the cause of a disease.

Causality

If one must make exceptions to Koch's Postulates, what type of evidence would one need to show that a pathogen is the **cause** of a disease? Remember that simply because the presence of a potential pathogen is strongly associated with a disease, that doesn't mean it is the cause of the disease. It may in fact be present more frequently in individuals with the disease because the disease causes changes in the relationship between the host and the potential pathogen which are favorable to the potential pathogen (a weakened immune system, for example). It may also be possible that the factor, which is causing the disease, also causes an increase in the presence of the potential pathogen.

Assignment 1: Take a few minutes to work with a partner and write down how you would test whether the following pathogens are the **cause** of the disease described. *NOTE: These are hypothetical scenarios involving made-up pathogens and diseases, although they may resemble known diseases.*

Case One: Chronic pre-weekend febrile fatigue syndrome – An unusually large number of employees at a sporting goods manufacturing company have started calling in sick on Fridays. While originally attributed to low employee morale, the similarity of symptoms has captured attention by public health authorities. The vast majority of these employees report fever, aches, extreme fatigue, and headache. An inspection of the ventilation systems reveals that the employees who have called in sick most often are in the older part of the building with an outdated air-conditioning system. A colleague of yours suspects that the mysterious outbreak of illness is being caused by a previously assumed nonpathogenic bacterium common in cooling systems – *Frigidus slimeii*. You are unable to culture *F. slimeii* from samples taken from the sick employees' noses and throats, although it is present in the building's old cooling system. What would you do to establish whether *F. slimeii* is the cause of this disease?

Case Two: Progressive degenerative neurological disease – A strange increase in cases of fatal degenerative neurological disease has occurred in a pattern familiar to public health officials. Although not a perfect correlation, cases appear to be more prevalent among injection drug users and people with multiple sex partners. A colleague suggests to you that this pattern resembles that of a virus transmitted by exchange of body fluids. Nevertheless, attempts to isolate a virus from these patients have proved fruitless. How would you go about finding the potential pathogen and establishing whether it is the cause of the disease?

Case Three: Severe pustulous feline-scratch infection – Recently, there has been an upsurge in pus-secreting infections of wounds resulting from pet cats. The infections have been extremely difficult to treat, and several have necessitated amputation. In an effort to culture and identify the bacterium causing these infections, you run into complications. The pus-filled wounds are full of bacteria of many different kinds. You are having extreme difficulty isolating a single type of bacterium associated with all cases of the disease. How would you proceed? What if you find a whole suite of bacteria associated with the infections; how would you establish that they as a community are the cause of the disease?

Write answers on a separate sheet of paper.

- 1.) Considering these scenarios, were there any techniques common to all three?
- 2.) In each case, what facts did you find it important to establish to test whether the pathogen(s) was the **cause** of the disease?
- 3.) What role, if any, did the timing of events play in establishing whether a pathogen was the **cause** of the disease?
- 4.) In each case, how could determine whether a pattern of association between the presence of the pathogen in victims and presence of the disease was strong enough to say that the pathogen **caused** the disease?
- 5.) Develop a list of at least three criteria that you could use in a variety of situations to show whether a pathogen is the **cause** of a disease.
- 6.) AFTER WRITING DOWN YOUR OWN CRITERIA, compare your criteria with those described in the handout from *Epidemiology in Medicine* by Hennekens and Buring, 1987. Which of your criteria overlap? Which did the handout list that you did not? Which did you list that the handout did not?

Contemporary Epidemiology

The strategies used by contemporary epidemiologists to uncover the causes of disease fall into several categories of experimental design.

NOTE: The following descriptions are excerpted and/or adapted from *Epidemiology in Medicine* by Hennekens and Buring, 1987.

Analytical Studies:

Analytical studies are explicitly designed to test hypotheses about whether the risk of disease differs for individuals either exposed or not exposed to a particular factor. There are two primary types of analytical study design.

- 1.) Case Control Studies – a case group or series of patients who have a disease and a control group without the disease are selected and researchers compare the proportion of each group with exposure to the factor in question
- 2.) Cohort Studies – subjects are classified on the basis of the presence or absence of exposure to a particular factor and followed for a specific period of time to determine the development of disease in each group

Assignment 2:

Researchers are now recognizing the strong association of certain autoimmune conditions with prior exposure to specific pathogens. One theory about the development of Type I diabetes is that it is triggered by infection with Coxsackievirus B. Design an experiment to investigate the possibility of exposure to Coxsackievirus B as a causal factor in the development of Type I (insulin dependent) diabetes mellitus – an autoimmune disorder. With your partner, write a hypothesis regarding the role of Coxsackievirus B in Type I diabetes. Then write an outline of how you would design a study to determine whether Coxsackievirus B is a causal factor in the development of Type I diabetes. Use one of the types of analytical studies outlined above and identify what type of study yours is (case-control or cohort).

Present your study design to the rest of the class.

Assignment 3:

Examine the simulated data provided. How might you analyze this data to determine whether Coxsackievirus B is a causal factor in the development of Type I diabetes?

1.) Decide whether or not you will include all individuals observed or exclude some. If you decide to exclude some individuals, what is your rationale? Cross out the rows containing individuals you choose to exclude.

2.) A common tool for analyzing data from case-control and cohort studies is the two by two table comparing exposure and presence of the disease. Using the blank two by two table as a guide, construct a similar table for the IDDM/Coxsackievirus B study and fill in the appropriate count values.

	Disease +	Disease -	Total
Exposure +	a	b	a + b
Exposure -	c	d	c + d
Total	a + c	b + d	a + b + c + d

3.) For two by two tables with count data, one usually calculates an odds ratio (OR)

$$OR = (ad) \div (bc)$$

The odds ratio is the ratio of odds of exposure among cases to the odds of exposure among controls. For example, an odds ratio of 3.4 would indicate that those individuals with exposure to the factor in this study were 3.4 times more likely to get the disease than individuals without exposure.

Calculate the odds ratio for your data.

4.) In order to calculate statistical significance of the difference in odds between cases and controls, epidemiologists typically use the Chi-square test. Using the following instructions, calculate a Chi-square for your data, and determine whether the pattern observed is statistically significant.

The Chi-square tests for differences between observed and expected values for frequencies, so first we must calculate the expected values for each frequency in the two by two table. Expected values are calculated from row, column, and overall totals of the values in the table. The expected values represent what the values would be if the distribution were random.

	Disease +	Disease -	Total
Exposure +	45	13	45 + 13 = 58
Exposure -	5	37	5 + 37 = 42
Total	45 + 5 = 50	13 + 37 = 50	TOTAL = 100

For each cell in the table, the expected value is calculated by the formula

$$(\text{Column total} \times \text{Row total}) / \text{Total}$$

For instance, the expected value for the cell with the observed value 45, would be

$$(50 \times 58) / 100 = 29$$

The Chi-square statistic itself, denoted by the symbol χ^2 is described by the following formula:

$$\chi^2 = \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

For the data above, the Chi-square statistic is,

$$\chi^2 = \frac{(\text{observed} - \text{expected})^2}{\text{expected}} = \frac{(45 - 29)^2}{29} + \frac{(13 - 29)^2}{29} + \frac{(5 - 21)^2}{21} + \frac{(37 - 21)^2}{21}$$

$$\chi^2 = 42.04$$

Calculate the Chi-square statistic for your data:

Now that you have calculated the Chi-square statistic, you still haven't answered the question about how probable it is that the pattern we are observing could have occurred by chance alone.

In order to answer that question, we will need four things:

- Our Chi-square value calculated from the table
- The **degrees of freedom** for our experiment
- A **probability level**
- A statistical table

Degrees of freedom correspond roughly to the sample size or number of replicates for the experiment. In general, the greater the degrees of freedom, the more likely we will be able to show that the pattern observed is not likely to be due to chance alone.

For a Chi-square test, the degrees of freedom or "df" are calculated as follows:

Chi-square: $df = (\text{number of rows} - 1)(\text{number of columns} - 1)$

NOTE: In some tables, the symbol for df may also be represented by the Greek letter ν .

The **probability level** describes the probability that the pattern we observe could have occurred by chance alone.

The probability level can be set at whatever value one wants, but the most widely accepted value is 0.05. A statistical analysis that meets a probability level of 0.05 demonstrates that there is only a 0.05 or 1/20 chance that the observed pattern is due to chance alone.

The probability level is often referred to as the **p-value** or **-value**.

Once one has calculated the value of the Chi-square statistic and the degrees of freedom, and decided the probability level, it is time to refer to a statistical table. Use the table to find the value of the statistic for the df () and probability level (p or) that you have calculated.

If the value of the Chi-square statistic from the table is less than the Chi-square value you calculated, then your results are said to be **statistically significant**.

= 0.95	0.50	0.25	0.10	0.05	0.025	0.01	0.005	0.001	<i>(from Zar, 1984)</i>
0.004	0.455	1.323	2.706	3.841	5.024	6.635	7.879	10.828	

Using the part of the table for Chi-square values above, which only describes values for 1 degree of freedom, determine whether the pattern observed in your data is statistically significant.

Chi-square (from table) = Chi-square(calculated) =

Significant? yes/no

5.) What conclusions can you make based on your data and statistical analyses? How well do these data meet the criteria for establishing causality previously discussed?