


**Controlling Bias and
Confounding and
Determining Causality**

**Bioterrorism Epidemiology
Module 13**

Missouri Department of Health
And Senior Services


This is the final module for epidemiological methods. Subsequent modules will provide information on the six primary biological pathogens that might be used as weapons by terrorists. In this module, I will discuss some of the common biases associated with performing an epidemiological study.



Major Sources of Error in Conducting an Epidemiological Investigation

- Use of nonrandom samples of the study population
- Nonparticipation of members of the study group
- Observer variation

This slide lists some of the major sources of error in conducting a study to determine the cause of a disease outbreak. If you are not able to study the entire study population, then the sample of the population must be random or bias can be introduced. Bias might also be introduced if some members of the study sample do not participate because it is possible that the nonparticipants might differ from the participants in the proportion with disease or the amount of exposure to the toxin of interest. Another potential source of error relates to the person collecting data on people in the study and control groups. This observer might collect data differently in people in one group compared to the other which might bias the data. For example, if the observer knows that one group has disease and the other group does not, the observer might question the diseased group more thoroughly than the non-diseased group about potential exposures to toxins. This bias can be minimized by keeping the observer unaware of the disease or exposure status of the person being interviewed.




Confounding

- Factor must be associated with disease
- Factor must be associated with exposure in the study base
- Must not be an intermediate step in causal pathway
- Can be controlled in study design, analysis, or both

A special type of bias called “confounding bias” occurs when some factor that is associated with both the exposure and the disease is not part of the causal pathway. For example, if a study shows that there is a higher rate of pancreatic cancer in coffee drinkers than in non coffee drinkers, an investigator might conclude that there is a causal relation between coffee consumption and pancreatic cancer. However, if on further analysis, the investigator finds more cigarette smokers in the coffee drinking group than the non coffee drinking group, cigarette smoking might be confounding the relation between coffee consumption and pancreatic cancer since it is known that cigarette smoking is associated with pancreatic cancer. Cigarettes satisfy the requirement that a confounder must be associated with disease and the factor must be associated with exposure. In this situation, people who smoke cigarettes tend to drink coffee more than non smokers.

Confounding can be controlled in the study design and/or through statistical analysis. The most effective way of controlling bias is through the study design.




Confounding

- Control in study design
 - Random selection
 - Restrict to narrow range of values
 - Matching on potential confounders
- Control in analysis involves stratifying data, calculating effect estimate, then summarizing
- Can also use advanced statistical methods

This control through the study design can be accomplished in three primary ways. Random selection of the study and control groups is the most efficient means of control, because random selection not only controls confounding bias, but also other forms of bias. In the pancreatic example above, random sampling might have assured equal numbers of cigarette smokers in both groups. Another possibility might be to limit the study to non smokers or to match smoking history which would assure that the same proportion of cigarette smokers were in both groups.

The most common statistical control for confounding is stratification. In stratification, the investigator would look at the pancreatic cancer rate in cigarette smokers and non smokers separately and then if there is a difference, do a weighted average across the strata. Advanced statistical methods such as logistic regression can also be used.




Odds Ratio for Matched Pairs

OR (matched pair) = B/C

		Controls	
		Exposed	Not Exposed
Cases	Exposed	A	B
	Non-Exposed	C	D

If you choose to use matching to control for confounding, you will need to use an alternative formula for calculating the OR. The formula is given in this slide.




Odds Ratio for Matched Pairs

OR (matched pair) = B/C

		Controls	
		Exposed	Not Exposed
Cases	Exposed	100	300
	Non-Exposed	100	300

Use this formula to calculate the odds ratio for this data.




Odds Ratio for Matched Pairs

OR (matched pair) = B/C

		Controls	
		Exposed	Not Exposed
Cases	Exposed	100	300
	Non-Exposed	100	300


Since the data is matched, we are only interested in the discordant pairs, cases that are exposed but controls that are not exposed and cases that are not exposed and controls that are exposed. The cells where both groups are exposed or both groups are non exposed do not provide any additional information.



Effect Modification

- Level of association (RR and OR) between exposure and outcome is different in different subgroups of a population
- Usually less common than confounding

In some situations, confounding might be confused with effect modification. While confounding is an artifact that is common in epidemiologic studies, effect modification is usually a biological phenomenon and is therefore less common. For example, if older people have a weaker immune system than younger people, then older people might be less likely to fight off an infectious disease. Age in this situation is an effect modifier and the study group should be stratified by age. But in seeking a confounder, you would not adjust for age but would present your results separately for each age group.




Example of Effect Modification

		Diseased	Not Diseased
Attended Soccer Game	YES	40	160
	NO	12	240
		Diseased	Not Diseased
Lived within one mile of stadium	YES	32	160
	NO	20	240

Calculate the OR, 95% CI and p values for these two tables

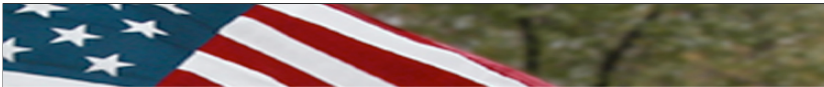
In this example, an epidemiologist determined that people who had attended a soccer game became ill. However, he also observed that some people living within a one mile radius of the soccer stadium also became ill. It is hypothesized that a toxic agent was released during the soccer game but that the agent also drifted to the surrounding community. Calculate the OR and 95% confidence interval for this data.



Example of Effect Modification

		Diseased	Not Diseased	OR (CI)
Attended Soccer Game	YES	40	160	5.0 (2.5 - 9.8) P < .001
	NO	12	240	
		Diseased	Not Diseased	OR
Lived within one mile of stadium	YES	32	160	2.4 (1.3 - 4.3) P < .004
	NO	20	240	

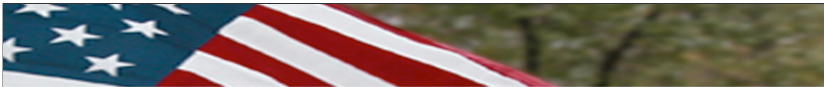
From these calculations it is evident that there was more disease in people who attended the soccer game than in those that did not, approximately a five fold increase in disease. However, there was also an almost two and a half fold increase in disease in those people that lived within one mile of the stadium. What we do not know from these calculations is what is the disease rate in those living outside the one mile area.



Effect Modification

	Diseased	Not Diseased
Attended Soccer Game and lived within one mile	24	80
Attended Soccer Game and did not live within one mile	16	80
Did not attend Soccer Game and lived within one mile	8	80
Did not attend Soccer Game and did not live within one mile	4	160


This table shows a stratification that answers this question. Calculate the OR and 95% confidence interval for each strata using those who did not attend the soccer game and who do not live within one mile of the soccer arena as your reference group.



Effect Modification

	Diseased	Not Diseased	OR (CI)
Attended Soccer Game and lived within one mile	24	80	12.0 (4.0 – 35.7)
Attended Soccer Game and did not live within one mile	16	80	8.0 (2.6 – 24.7)
Did not attended Soccer Game and lived within one mile	8	80	4.0 (1.2 – 13.7)
Did not attended Soccer Game and did not live within one mile	4	160	reference


These calculations show that people who both attended the soccer game and lived within one mile of the arena had the highest risk of disease. In this situation, living within one mile of the stadium modified the effect of attending the soccer game. This is probably because people who both attended the soccer game and lived within one mile of the area had more opportunity for exposure. This also might represent a dose response in that those attending the game and living within the area had more exposure than those just attending the game.



Key Factors to Consider in Determining Causality

- **Temporality:** disease occurs within a biologically reasonable time frame after exposure
- **Chance:** how likely that results are due to chance
- **Consistency:** studies should demonstrate similar associations

The last two slides of this module discuss the overall purpose of epidemiology: the evaluation of cause and effect. From previous modules we learned that an elevated risk ratio measured by a RR or an OR is an indication that there is an association between an exposure and an outcome, usually disease or death. But the risk measure does not necessarily indicate that the exposure caused the disease. A number of factors must be considered in making this determination. First, the exposure must precede the disease. This criterion must be met before considering any other factors. Latency must be considered to determine if exposure preceded disease, because if the minimum latency period after exposure to disease is two weeks and a person comes down with disease two days after exposure, it is highly unlikely that the exposure caused the disease. Once this criterion has been determined, then we determine the degree of likelihood that the association is due to chance. The effect of chance is measured by the p value or the confidence interval discussed in previous modules. If there is more than a 5% probability that the observed risk is the result of chance, we are less likely to consider the association causal. Investigators will also look at other available studies to see if similar associations have been found by others using different groups of people.



Key Factors to Consider in Determining Causality

- Strength: greater estimate of risk and more precise
- Dose Response: increase in effect correlated with exposure
- Biological Plausibility: makes sense in terms of biological knowledge

The strength of the association is also important. A risk ratio might not be statistically significant yet be quite large. In most situations, a large risk ratio in a well conducted study is important even if it is not statistically significant. If you are able to measure a dose for exposure, then if there is a causal relation between exposure and disease, we would expect to see the risk ratio increase with greater exposure. The final factor to consider is biological plausibility. Is there a biological mechanism that explains how the toxins associated with the exposure cause the disease? While this factor increases our confidence in a causal relation, it is not a necessary factor because a relation might exist but scientists have not yet found it.

You have now completed all the modules on epidemiological methods. I encourage you to periodically review the slides to reinforce your understanding of the methods that you will be using throughout your career and to refer to epidemiological and statistical texts for more explanation of the methods discussed in these modules.

The next set of modules on the six main biological agents that might be used by terrorists was created by Steven Lawrence, M.D., a faculty member in the Center for the Study of Bioterrorism and Emerging Infections.