

## Causal Inference

### Introduction

Causal inference – the art and science of making a causal claim about the relationship between two factors – is in many ways the heart of epidemiologic research. Under most circumstances if we see an association between an exposure and a health outcome of interest, we would like to answer the question: is one causing the other? We care about causal inference because, ultimately, we want to intervene to improve public health, and interventions can be targeted on removing known causes of adverse health outcomes (or adding known causes of beneficial health outcomes).

Moving from measuring an association to inferring a causal link is not trivial. Before we decide whether a demonstrated association is plausibly causal, we need to first know what a cause is. To find this out requires a brief look into causal theory. This exercise will introduce you to causal theory as used by epidemiologists and will lead you through the steps that epidemiologists commonly use to assess whether an observed association is plausibly causal. Finally, this exercise will show you why it's often difficult to make a confident claim of causality from epidemiological data.

### Faculty Highlight: Dr. Mervyn Susser

Dr. Mervyn Susser is Special Lecturer of Epidemiology, Mailman School of Public Health, and Sergievsky Professor of Epidemiology Emeritus of The Gertrude H. Sergievsky Center, College of Physicians and Surgeons.

Dr. Susser received his medical degree in 1950, from the University of Witwatersrand in Johannesburg, South Africa. He began his career in community and primary health care in Alexandra, a township for Africans on the outskirts of the city. Then in England from 1956, for nearly a decade thereafter, he taught in the Department of Social and Preventive Medicine at Manchester University.

His main research, in collaboration with Zena Stein, was on epidemiological and family and cultural aspects of mental retardation and child development, on psychiatric disorders, and on reproductive health and neuro- developmental disorders.

As Professor and Head of Epidemiology at Columbia University School of Public Health (1966-1978), this joint research moved into large-scale studies of the epidemiology of nutritional effects on child development, including the landmark Dutch Famine Study.

From 1978 through 1990, Dr. Susser was Sergievsky Professor of Epidemiology and Director of the newly endowed Sergievsky Center. He continued to conduct research in collaboration with Zena Stein on neurodevelopment and reproduction, and also on HIV/AIDS.

Dr. Susser is renowned not only for his research, but also for his thinking and writing on causality and causal inference. He remains a leading thinker and inspirational figure in the Department of Epidemiology here at Columbia.

#### Read more about Dr. Susser's work:

Susser, M. Causal thinking in the health sciences: concepts and strategies of epidemiology. New York: Oxford Press, 1973. (This book is available in the library).

Susser, M. A conversation with Mervyn Susser. Interview by Nigel Paneth. Epidemiology. 2003; Nov;14(6):748-52.

Susser, M. Glossary: Causality in public health science. J. Epidemiol. Community Health. 2001;55: 376-378.

## Step 1: Learning Objectives

1. Understand the definition of a cause as it applies to epidemiology.
2. Describe the *sufficient-component-cause* model using Rothman's causal heuristic.
3. Distinguish between different elements of the *sufficient-component-cause* model: necessary and sufficient causes, and neither necessary nor sufficient component causes.
4. Describe some criticisms of the *sufficient-component-cause* model.
5. Apply Bradford Hill's "causal criteria" to the assessment of exposure-outcome association.
6. Discover the limitations inherent in using "causal criteria" for causal inference.

## Step 2: Student Role - Your Plan of Action

You have just begun your internship at the [Epiville Department of Health \(DOH\)](#) and are looking forward to getting some "real world" Epi experience. Your supervisor, head of the Epiville Department of Health, Dr. Morissa Zapp, gives you a quick tour of the Department and apologizes for seeming hurried and distracted, but the Epiville Attorney General's Office has asked for her advice on a class action lawsuit they are thinking of pursuing. Dr. Zapp informs you that you will be doing the background research on this project. She arranges for you to meet with the Attorney General in his office.

The Attorney General, Mike Broomberg, wants your help. According to the EDOH, the prevalence of smoking in Epiville among youths has been increasing in the past few years. He has long suspected that exposure to images in television and the movies that portray smoking as "cool" may be part of the reason for this increase. Additionally, he has just come across a study purporting to show an association between television watching and the initiation of smoking among children ([click here for the study](#)). Based on this study, Mr. Broomberg is thinking of bringing a class action lawsuit against the television networks to recover medical costs to the city associated with teenage smoking. He wants your opinion about whether the evidence presented by the article is sufficient to make the claim that television watching is a cause of smoking initiation.

## Step 3: Study Design

### Part A: Review Causal Theory

After your meeting with Mr. Broomberg, you start outlining the steps you'll need to take to provide an informed opinion concerning whether the study ([link to article](#)) provides sufficient evidence of a causal link between watching television and early initiation of smoking. While reading the article and jotting down some extremely informative notes, you begin asking yourself a deceptively simple question: *exactly what is a 'cause'?*

While gestating on this perplexing question, you decide to do a little office cleaning. Your "office" is actually a cubicle in an old storage room deep in the bowels of the Epiville Department of Health, and it is cluttered with old papers and newspaper articles. You decide it's time to move these boxes of files when one headline from an old *Epiville Press* newspaper catches your attention. You pause in your cleaning and read.

1. Which of the following statements best explains the Mayor's false belief that increasing the stork population would increase the birth rate?
- The Mayor implied there was an association between the stork population and the birth rate that was not there.  
**Incorrect.** Looking at the figure provided in the article, there is clearly an association between the stork population and the birth rate. His mistake was not in ascribing an association between these two, but rather in speculating on the reason for such an association.
  - The Mayor mistakenly thought that 2 factors which are statistically associated (the stork population and the birth rate) must be causally related.  
**Correct.** One of the central tenets of epidemiologic study is that association does not necessarily imply causation – there may be alternate explanations for this association (such that a third factor, in this case the Industrial Revolution, is causing both a decrease in storks and a decrease in the birth rate).
  - The Mayor neglected to look at the statistical significance of the association. Only if the results were statistically significant could he conclude that a causal relationship exists.  
**Incorrect.** "Statistical significance" only tells you the probability of observing the data that you observed (or something more extreme) given the hypothesis that there is no association between the exposure and outcome. Statistical significance does not indicate that an association is causal. It is merely concerned with whether an association could be due to chance given sampling variability in your data.

Since this news report has got you thinking about causal theory again, you decide that your work environment is "clean enough." You're ready to get back to work when you notice today's *EpiVille Press*, so you decide to check the baseball scores from last night.

2. You are struck with the manager's claim that I-Pod did not cause the team to lose the game. Which of the following (A only, B only, or both A and B) could also have caused the Riskfactors to lose the game?
- The Riskfactors failed to score 2 runs in the 7th, 8th, or 9th inning.  
**Incorrect.** While this is a potential cause of the loss, there is a more complete answer available.
  - Riskfactor pitcher, Vinnie Virchow, gave up a home run and the Riskfactors did not score any runs.  
**Incorrect.** While this is a potential cause of the loss, there is a more complete answer available.
  - Both A and B  
**Correct.** Note that there are multiple causes of the event operating here. In epidemiology, we speak of a "web of causality" and look to explain whether certain "risk factors" are "A cause" (not THE cause) of a particular outcome.

### Part B: Review Rothman's heuristic and the Sufficient-Component Cause model

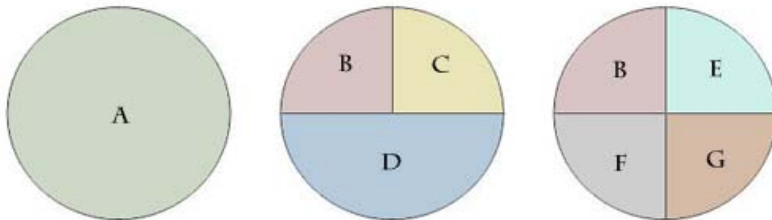
The two newspaper articles highlight some important concepts about causality. In epidemiology, some of these concepts have been coalesced into a theory of disease causation, based on the premise that there are multiple causes for most given diseases. This theory was made "famous" (for epidemiologists, at least) by Kenneth Rothman and his heuristic showing causes of disease as distinct pies (Aschengrau, pp 391-393).

A cause of an outcome is defined as "something that makes a difference" (Susser, 1973), or, more formally, as an "event, condition, or, characteristic that preceded the outcome of interest such that, had it not occurred, the outcome would not have occurred when and how it did" (Rothman and Greenland, pg. 8).

Looking at the newspaper article on last night's baseball game and the answers to Question 2 above, you hypothesize that the following factors were "component causes" in last night's loss to the Biostatown Frequentists:

1. I-pod striking out with the bases loaded.
2. Vinnie Virchow giving up a solo home run.
3. The Epiville Riskfactors not scoring any runs in the 7th, 8th, or 9th inning.
4. I-pod getting caught stealing second base to end the game.

However, there are likely other "component causes" to last night's loss. Ken Rothman proposed a "sufficient-component cause" model of causation to explain the idea that there are often many contributing causes to a single event. Consider the following heuristic which represents the potential "causes" of last night's loss:



**Legend**

- A** = The Epiville RiskFactors scoring fewer runs than the Biostatown Frequentists
- B** = Vinnie Virchow giving up a home run
- C** = I-Pod striking out with the bases loaded in the 6th inning
- D** = RiskFactors failing to score in the 7th, 8th, or 9th inning
- E** = RiskFactors failing to score through the 8th inning
- F** = I-Pod hitting a single with 2 outs in the 9th inning
- G** = I-Pod getting caught stealing in the 9th inning

Each "pie" represents one sufficient cause of the outcome of interest, in this case, last night's loss. Thus, the "cause" of the loss could be viewed as operating through cause A alone, or through the combination of causes B, C, and D, or through the combination of causes B, E, F, and G.

In this heuristic, if the three "sufficient causes" represent all of the potential causes of last night's loss, then cause A represents a **sufficient** cause of disease (if A occurred, then the RiskFactors would inevitably have lost).

Causes B-G are "component" causes, since they are neither **necessary** for the outcome of interest (since there is another sufficient-cause pie which does not include them) nor **sufficient** (since, for example, having component B would "cause" the loss only if components C and D also occurred).

Rothman's schematic has proven remarkably useful in explaining the multicausal nature of diseases. This schematic is useful at explaining why, for example, we consider smoking to be a cause of lung cancer even though (a) not everyone who smokes gets lung cancer and (b) not everyone who has lung cancer, smoked. Rothman's schematic is also useful at demonstrating which "causes" of a particular outcome, if removed, would lead to the greatest reduction in the incidence of that outcome.

However, there is a limitation to this model. Namely, his heuristic says nothing about the mechanisms through which the causes operate. In fact, “cause A” can be thought of as containing causes B, C, and D (or B, E, F, and G). Under this model, all “causes,” no matter how or when they operate, are given essentially the same importance. The sufficient-component cause model’s inattention to causal pathways has led to the criticism that utilizing it leads to “black box” epidemiology, identifying “risk factors” without gaining any insight into the mechanisms through which they operate.

Now that you’ve seen the difficulty inherent in making a causal claim about something that has been observed, you begin to look at the Attorney General’s case. You start by reading the article and answering the following questions:

3. Attorney General Mike Broomberg asks you if you think that watching television is “a cause” of early smoking initiation. You think back to your Principles of Epidemiology class and recall that epidemiologists have a fairly distinct definition of “cause.” Which of the following statements best describes the hypothesized causal link between television viewing and smoking initiation?

a. All individuals **with long hours of television viewing** during childhood will go on to start smoking early in adolescence.

**Incorrect.** While this is a description of a sufficient cause of disease, the article cited by the Attorney General does not claim that everyone who watches television will initiate smoking.

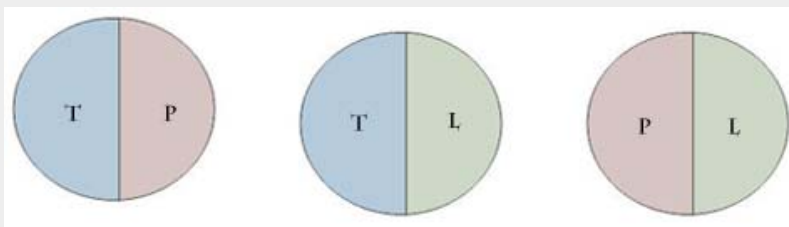
b. All individuals who initiate smoking early on in childhood would have been exposed to substantial television viewing during childhood.

**Incorrect.** This is a description of a necessary cause of disease. The article cited by the Attorney General does not claim that television viewing is required for individuals to initiate smoking.

c. Individuals who watch substantial amounts of television early on in childhood are more likely to initiate smoking early, compared with those individuals who watch television less frequently during childhood.

**Correct.** This statement allows for a “multicausal” theory of causation. While television watching is neither necessary nor sufficient, it is still hypothesized to be a cause.

4. You decide to draw some “Rothman Causal Pies” to help you think about how watching television might act as a cause of smoking initiation. Based on the causal pies given below and assuming that they are true, which of the following is a true statement?



**T** = Television Watching

**P** = Peer Pressure

**L** = Low Self-Esteem

- a. Television watching is a necessary cause of smoking initiation.  
**Incorrect.** In the above model, it is possible to initiate smoking without watching television, due to a combination of peer pressure and low self-esteem. Since television watching is not present in every causal pie it is not a necessary cause.
- b. Television watching is a sufficient cause of smoking initiation.  
**Incorrect.** In the above model, watching television alone cannot cause smoking initiation, but instead must be accompanied by peer pressure or low self-esteem. Since television watching does not have its “own pie”, it is not a sufficient cause.
- c. Television watching is a cause of smoking initiation that is neither necessary nor sufficient.  
**Correct.** In the above model, watching television in the context of peer pressure or low self-esteem is a cause of initiating smoking.

## Step 4: Data Analysis

Being in the Attorney General's Office surrounded by all of those legal books, led you to think about how Hill's causal "criteria" ([click here](#) for a link to the original article or see Aschengrau, pp 384-391) can be used to assess evidence for causality in much the same way that a prosecutor attempts to establish guilt in a court of law. Basically, epidemiologists have looked to lists of 'causal criteria' as **inductive** ways of building an argument to support the notion that a given association is causal. Your job is to use Hill's criteria to give the Attorney General guidance about whether the Gidwani et al article shows that television viewing is a cause of early initiation of smoking.

- 5. The authors state in the results, “Controlling for baseline characteristics, youth who watched >5 hours of television per day were 5.99 times more likely to initiate smoking behaviors than those youth who watched 0-2 hours per day.” What does this piece of information tell us?
  - a. Watching >5 hours of television per day causes youths to initiate smoking.  
**Incorrect.** They are simply stating the statistical association that they found in their data; this information does not tell us if this association is causal.
  - b. There is a statistical association between watching >5 hours of television per day and initiation of smoking after taking into account other baseline characteristics of youth who watched television that were measured in this study.  
**Correct.** They are stating that there is an association between the exposure and outcome, but cannot determine on the basis of these data alone whether this association is causal. They also report that their finding is “adjusted for” baseline differences between children who watch television and those who do not. Adjustment is a way to control for possible differences in the two comparison groups which could be independently associated with differences in television viewing patterns ([confounding](#)).
  - c. If kids watch less TV, they will be less likely to initiate smoking.

**The following quotes provide evidence to support Hill's Criteria. Choose the criterion which correctly matches each quote.**

- 1. “We examined the relationship between television viewing and initiation of smoking and found a strong dose-response relationship with increasing hours.”

**Dose Response**

2.	“In this study, television viewing was measured 2 years before smoking initiation.”	<b>Temporality</b>
3.	“The association was substantial, with youth who watched >5 hours per day being 5.99 times as likely to initiate smoking than youth who watched 0 to 2 hours per day.”	<b>Strength of Association</b>
4.	“...Television provides adolescents with role models, including movie and television stars and athletes, who portray smoking as a personally and socially rewarding behavior.”	<b>Biological Plausibility</b>
5.	“The findings are consistent with social learning theory.”	<b>Coherence</b>
6.	“A similar association between television viewing and the onset of alcohol use has been reported...”	<b>Analogy</b>

**7. Which of the criteria that you reviewed above is the one that is essential for TV watching to be a cause of smoking initiation?**

- a. Biological plausibility  
**Incorrect.** It is possible for a factor to be causal even if we do not know the exact biological mechanism by which it operates.
- b. Dose-response  
**Incorrect.** It is possible that there is a threshold effect of TV viewing, so that only after a certain length of viewing it is harmful.
- c. Temporality  
**Correct.** By the definition of cause, the cause must precede the effect. Temporality is the only ‘required’ criterion in the evaluation of causality.

**Intellectually curious?**

[Learn more about the limitations of causal criteria](#)

Hill’s ‘criteria’ are often used by epidemiologists in an attempt to rule out alternate explanations for an association (other than a causal explanation). Epidemiologists also look to other explanations, those not explicitly covered in Hill’s guidelines, when assessing whether an association is plausibly causal. You will learn more about these “alternative explanations” in the lectures on bias and confounding. Being the vigorous young epidemiologist that you are, you know that you should consider potential alternate explanations for the association between television viewing and smoking initiation before giving your report to the Attorney General.

- 8. Which of the following alternate explanations could have possibly caused the authors to find that youth who watched >5 hours of television per day were 5.99 time more likely to initiate smoking behaviors, if watching TV did not truly cause smoking initiation?**

- a. If a true cause of smoking initiation was poor parental monitoring, and youth who watch a lot of TV are less likely to be monitored by parents, then the association between TV watching and smoking initiation would really be due to their shared association with parental monitoring. This is an example of confounding.  
**Correct.** If the true cause was being home alone after school, and youth who watch TV are more likely to be home alone after school, then the association between TV watching and smoking initiation would really be due to their shared association with being home alone after school. This is an example of confounding ([link to dictionary](#)).
- b. Youth who watch TV are more likely to under-report true smoking habits on a survey compared to youth who do not watch TV.  
**Incorrect.** If that were the case, then it would appear that TV-watchers are LESS likely to initiate smoking.
- c. Youth who watch a lot of TV are more likely to have participated in the follow-up survey since they have more free time.  
**Incorrect.** For this type of phenomenon to explain the results, the youth who watched a lot of television AND started smoking early would have to be disproportionately followed-up, compared to children who watched a lot of television and did not initiate smoking early.

## Step 5: Discussion Questions

1. Do you believe that television viewing is a cause of early smoking initiation? Why or why not? What evidence in addition to what is presented in the article and the module would you need to definitively be convinced that television watching is a cause of smoking initiation?
2. Do you think the epidemiologic evidence is conclusive enough to merit a public health intervention? What would be the general rules to devise a public health intervention?
3. Think of two examples of exposure/outcome relationships that you believe are causal, and describe why you believe that the relationship is a causal one. What points of evidence were necessary in your evaluation of a causal relationship?

### Intellectually curious?

Can statistical tests be used to distinguish between association and causation? Why or why not?