RACE611 CLINICAL EPIDEMIOLOGY AND EVIDENCE-BASED MEDICINE

Risk study

Assoc.Prof.Dr.Patarawan Woratanarat

Master of Science Program in Medical Epidemiology and Doctor of Philosophy Program in Clinical Epidemiology
Section for Clinical Epidemiology & Biostatistics
Faculty of Medicine Ramathibodi Hospital Mahidol University

http://www.ceb-rama.org/Academic Year 2020 Semester 1

CONTENTS

OBJECTIVES

Students should be able to

1. Define absolute, relative, attributable and population attributable risk and be able to calculate these from given data.
2. Explain the term odds ratio and describe its relationship to relative risk.
3. Assemble in a logical sequence evidence indicating a certain exposure is causally related to a specific outcome.
4. Define prognosis
5. List the factors that might bias the assessment of prognosis of a disease.
6. Describe how bias may be treated.
7. Formulate clinical question about harm.
8. Appraise article about harm.
9. Create critical appraisal topic (CAT) of harm study.

REFERENCES

1. Fletcher RH, Fletcher SW, Wagner EH. Clinical Epidemiology. The Essentials, 3rd Edition. Williams & Wilkins, Baltimore, 1996, pp 94-135.

2. Greenberg RS, Daniels SR, Flanders WD, Ely JW, Boring JR. Medical Epidemiology. 2nd Edition. Appleton & Lange, Stamford, 1996, pp 105-44.

3. Sackette DL. Bias in analytic research. J Chron Dis 1979; 32:51-63.

4. Risk, Cause and Bias. Epidemiology A – Basic methods. Centre for Clinical Epidemiology & Biostatistics, Faculty of Medicine & Health Sciences. The University of Newcastle, Australia

5. Natural History and Prognosis. Clinical Epidemiology. Centre for Clinical Epidemiology & Biostatistics, Faculty of Medicine & Health Sciences. The University of Newcastle, Australia.

6. Grimes DA, Schulz KF. Cohort studies: marching towards outcome. Lancet 2002;359:341-5.

SUGGESTED READING

1. Grimes DA, Schulz KF. Bias and causal associations in observational research. Lancet 2002; 359:248-52.

2. Grimes DA, Schulz KF. Cohort studies: marching towards outcome. Lancet 2002; 359:341-5.

3. Schulz KF, Grimes DA. Case-control studies: research in reverse. Lancet 2002; 359:431-4.

READING SECTIONS AND QUESTIONS

Reading *p. 9, 17*

Question *p. 5, 10, 15*

Assignments *p. 14*, 17

RISK

Scenario

At your clinic

A 65-year-old woman came to your clinic for follow up diabetes mellitus. She went to the restroom and suddenly fell down on the floor. She had the right hip pain and could not walk anymore. You examined her and told her that she may have fracture around the hip. She was sent for x-ray and it revealed fracture neck of femur. Now she asks you that *what factors are associated with this fracture.*

Is diabetes mellitus associated with hip fracture?

What are the consequences of having this injury?

QUESTION

What evidence would you seek for answering these questions?

Hip fracture in elderly caused by osteoporosis. There are many risk factors related to it.

The prognosis depends on the prefracture status and type of treatment

Age, sex, BMI

Femoral axis length, BMD

Falling, neuromuscular diseases

Cigarette smoking, alcohol → Hip fracture

DM, thyroid disease

Calcium intake, etc

INTRODUCTION

When the patients got some diseases or injuries, the most frequently asked questions are about the cause/risk of diseases, treatment options, and the prognosis. How can you tell them? This module will lead you to understand the risk and prognosis. The answers about these can improve medical care in prevention and treatment as well.

DEFINITION

Risk = The probability of some untoward event

The likelihood that people who are exposed to certain factors (risk factors) will subsequently develop a particular disease

The proportion of unaffected individuals who, on average, will contact the disease of interest over a specified period of time.

Risk factors = Characteristics that are associated with an increased risk of becoming diseased

Risk factors compose of :-

1. Inherited: e.g. HLA-B27
2. Infection: e.g. HIV
3. Drugs, toxin: e.g. Aspirin overdose
4. Social, environment: e.g. crowding
5. Behavior: e.g. smoking, alcohol abuse, driving without seat belts

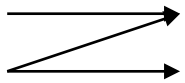
Exposure to risk factors = a person has, before becoming ill, come in contact with or has manifested the factor in question. It can take place in a single point of time (e.g. Radiation, diarrhea) or over a period of time (e.g. sun exposure, smoking).

Characteristic of exposure: ever exposed, current dose, largest dose, total cumulative dose, duration of exposure, latest contact, etc.

Recognizing risk

Large risks associated with effects that occur rapidly after exposures are easy to recognize. But most morbidity and mortality is caused by chronic diseases. The relationships between exposure and disease are far less obvious. It becomes virtually impossible for individual clinicians to develop estimates of risk based on their own experiences with patients as discussed below.

1. Long latency: Too long to remember as a risk.
   e.g. radiation and CA thyroid
2. Frequent exposure to risk factors: many risk factors are so common in society to recognize as dangerous thing.
   e.g. junk food and coronary heart disease
3. Low incidence of disease: it is difficult to draw conclusions about the risk in rare disease.
   e.g. heavy smoker and lung cancer (incidence < 2/1000)
4. Small risk: too small to detect the significant risk.
   e.g. birth control pills and breast cancer.
5. Common disease: the risk factors are already known. It becomes difficult to find a new risk factors.
6. Multiple causes and effects:



Use of risks

1. Prediction the occurrence of disease
2. Cause
   "marker" = the risk factor that is not cause of disease but increases probability of disease.
3. Diagnosis: rule in / rule out by using risk factors
4. Prevention: risk removal

STUDIES OF RISK

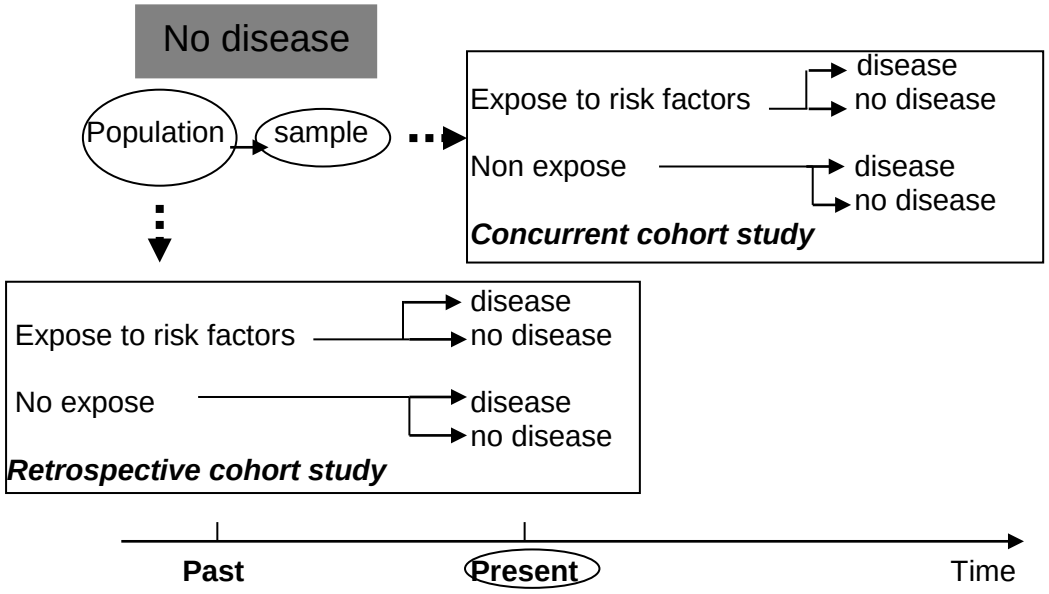
The most powerful way of determining whether exposure to a potential risk factor results in an increased risk of disease is to conduct an experiment. Unfortunately, the effects of most risk factors for humans cannot be studied with experimental studies.

Clinical studies in which the researcher gathers data by simple observing events as they happen without playing an active part (***observational studies***) are appropriate studies in less obtrusive ways.

Observational studies compose of cohort and case-control studies.

Cohorts

Cohort study begins with the exposure to cause or risk and then follow up until the outcome occurs.



Cohort studies can be conducted in two ways. It can be assembled in the present and followed into the future (*concurrent cohort study*)

Or it can be identified from past records and followed forward from that time up to the present (*historical cohort study*). The historical cohort study may not include sufficient and accurate data and may have historical bias.

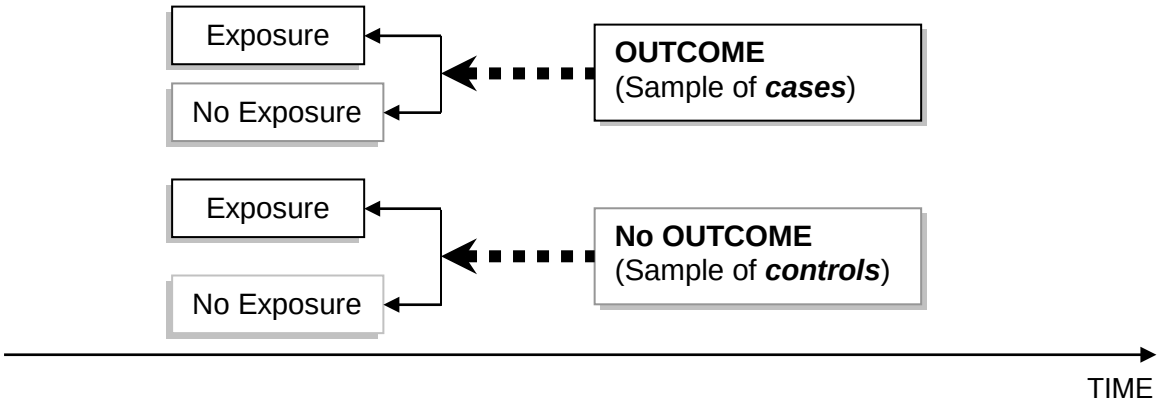
Advantages and disadvantages of cohort studies

Advantages	*Disadvantages*
The only way of establishing incidence directly	Inefficient because many more subjects must be enrolled (cannot be used for rare disease)
Follows the same logic as the clinical question	Expensive (need resources to study many people over time)
Exposure can be elicited without the bias	Results not available for a long time
Can assess the relationship between exposure and many diseases	Assess the relationship between disease and exposure to only relatively few factors

READING: Grimes DA, Schulz KF. Cohort studies: marching towards outcome. Lancet 2002; 359:341-5.

Case-control study

This type of study runs backwards. Researchers enrolled case (having diseases) and controls (having no disease) and then look back in time to ascertain each person's exposure status.



Case-control study has advantages and disadvantages differed from cohort.

Advantages	*Disadvantages*
The most efficient design in terms of time, money, and effort (recommend when incidence rate of outcome is low, long latency disease)	Cannot get the incidence rate
	Inefficient when frequency of exposure is low
	Choosing a control group is difficult to have good validity
	Obtaining exposure history

QUESTION: You try to search the risk factors for hip fracture.
What type of research you will find in medline?

Comparing risks

The basic expression of risk is incidence (a number of new cases of disease arising in a defined population during a given period of time). To compare risks, several measures of the association between exposure and disease, called measures of effect, are commonly used. They represent different concepts of risk and are used for different purposes.

Expression	*Question*	*Definition*
Attributable risk (risk difference)	What is the incidence of disease attributable to exposure?	AR = Ie – Io
Relative risk (risk ratio)	How many times more likely are exposed persons to become diseased, relative to nonexposed persons?	RR = Ie / Io
Population attributable risk	What is the incidence of disease in a population, associated with the occurrence of a risk factor?	ARp = AR x P
Population attributable fraction	What fraction of disease in a population is attributable to exposure to a risk factor?	AFp = ARp / I_T
Odds ratio	How many times more likely are diseased persons having been exposed to a given risk factor relative to non-diseased persons?	OR = ad/ bc

Ie = incidence in exposed persons

Io = incidence in nonexposed persons

P = prevalence of exposure to a risk factor

I_T = total incidence of disease in a population

Example:

A cohort study of 100 women who had been accidentally exposed to persticides was followed up overtime and their outcome compared to 100 women who had not been exposed to pesticides. Thirty women who had been exposed suffered miscarriages. Compared to10 women who had not been exposed.

	Miscarriage	*No miscarriage*	*Total*
Exposed	30 a	70 b	100
Not exposed	10 c	90 d	100
	40	160	200

What is the incidence of miscarriage attributable to pesticide exposure?

AR = Ie – Io

= (a / (a+b)) – (c /(c+d))

= (30 / 100) – (10 / 100)

= 20/100

How many times more likely are exposed persons to become miscarriage, relative to nonexposed persons?

RR = Ie / Io

= (a / (a+b)) / (c /(c+d))

= (30 / 100) / (10 / 100)

= 3.0

What is the incidence of miscarriage in a population, associated with the occurrence of a pesticide exposure?

ARp = AR x P

= (20/100) x (100/200)

= 0.1 or 10/100

What fraction of miscarriage in a population is attributable to exposure to pesticide?

AFp = ARp / I_T

= 0.1 / (40 / 200)

= 0.02 or 2/100

To interpret the OR results, you should consider the 95% confidence interval (CI). This module is not mentioned about the calculation of 95% CI. However, the significant of being risk should have 95%CI, which is not included 1. When 95%CI included it inferred that risk of having disease in exposed group is equal to risk of having disease in no exposed group. For example RR = 4 with 95% CI =1.7-8.9 is a significant risk factor. While OR = 3 with 95%CI = 0.7 – 6.8 which has a range of 95% CI include 1 is no significant.

Example:

A case-control study of 100 men with stroke compared to 100 controls. Twenty cases had high blood pressure while 5 controls had high blood pressure.

	Stroke	*Controls*	*Total*
High BP	20 a	5 b	25
Normal BP	80 c	95 d	175
	100	100	200

$$OR = \frac{\text{Odds that diseased person has been exposed}}{\text{Odds that non-diseased person has been exposed}}$$

$$= \{(a/a+b) / (b/a+b)\} / \{(c/c+d)/(d/c+d)\}$$

$$= (a/b) / (c/d)$$

$$= ad/bc$$

$$= (20 \times 95) / (5 \times 80)$$

$$= 19/4$$

$$= 4.75$$

To interpret the OR results, you should consider the 95% confidence interval (CI). This module is not mentioned about the calculation of 95% CI. However, the significant of being risk should have 95%CI which is not included 1 as mentioned in RR before. For example OR = 3 with 95% CI =1.2-5.4 is a significant risk factor. While OR = 3 with 95%CI = 0.2 – 9.8 is not significant.

It is important to remember that attributable risk (Ie-Io) as opposed to attributable risk fraction or percent cannot be calculated for case-control studies even if outcome is rare as Ie-Io is based solely on incident values.

In a case-control study we cannot calculate ARp%. However, we can calculate as an equivalent.

$$ARp\ \% = \frac{Po \times (RR - 1) \times 100}{[Po \times (RR - 1) + 1]}$$

Po = prevalence of exposure in the control group

The odds ratio is approximately equal to the relative risk only when the incidence of disease is low. Because of assumptions that must be made in the calculations. In general, distortion of the relative risk becomes large enough to matter at disease rates in unexposed people of greater than about 1/100.

Example:

A cohort study of lung cancer where of a population of 12,000 men, 000 were exposed to chemical fumes from smelter operations. After 9 years of follow-up 10 cases of lung cancer were observed in the exposed group and 25 cases in the non-exposed group. These data may be arranged in a fourfold table.

	Cancer	*Not cancer*	*Total*
Exposed	10 a	1990 b	2000
Not exposed	25 c	9975 d	10000
	35	11965	12000

RR = Ie / Io

= (10/2000) / (25/10000)

= 5 / 2.5

= 2.0

OR = ad / bc

= (10 x 9975) / (1990 x 25)

= 9975 / 49750

= 2.0

RISK AND PREVENTION

If the exposure is prevention, so that Ie (incidence among exposed) is less than Io (incidence among unexposed), the attributable risk is meaningless. The prevention fraction (PF) can be defined.

$$PE = \frac{Io - Ie}{Io}$$

ASSIGNMENTS

A cohort study is conducted to evaluate the relationship between dietary calcium supplementation and the occurrence of hip fractures in post-menopausal women. A total of 100 women who are taking calcium supplements and 100 women who are not taking the supplements are followed over 3 years.

During the follow up period, there are 5 women with hip fractures in the calcium group and 10 women with hip fractures in the group not taking calcium

1. What is the risk of hip fractures in the calcium group?
2. What is the risk of hip fractures in the group not talking calcium supplement?
3. What is the risk ratio for the occurrence of hip fractures?
4. The correct interpretation of these study results is that the point estimate for the risk ratio indicates that calcium supplementation with respect to hip fractures is
   a. protective
   b. deleterious
   c. neutral
   d. cannot be determined.

CAUSE

Does it really cause a disease?

Is this exposure really causing disease or ill-health in the population at risk?

This is the common question to answer the risk. As mentioned before, marker is a risk factor that does not cause a disease. How can we tell a cause of a disease?

Criteria for judgment of causal

- **Temporal sequence**

  Did exposure precede outcome?

- **Strength of association**

  How strong is the effect, measured as relative risk or odds ratio?

- **Consistency of association**

  Has effect been seen by others?

- **Biological gradient (dose-response relationship)**

  Does increased exposure result in more of the outcome?

- **Specificity of association**

  Does exposure lead only to outcome?

- **Biological plausibility**

  Does the association make sense?

- **Coherence with existing knowledge**

  Is the association consistent with available evidence?

- **Experimental evidence**

  Has a randomized controlled trial been done?

- **Analogy**

  Is the association similar to others?

QUESTION: What evidence would you seek to decide whether a causal relationship exists between steroid abuse and adrenal insufficiency?

Critical appraisal for risk study

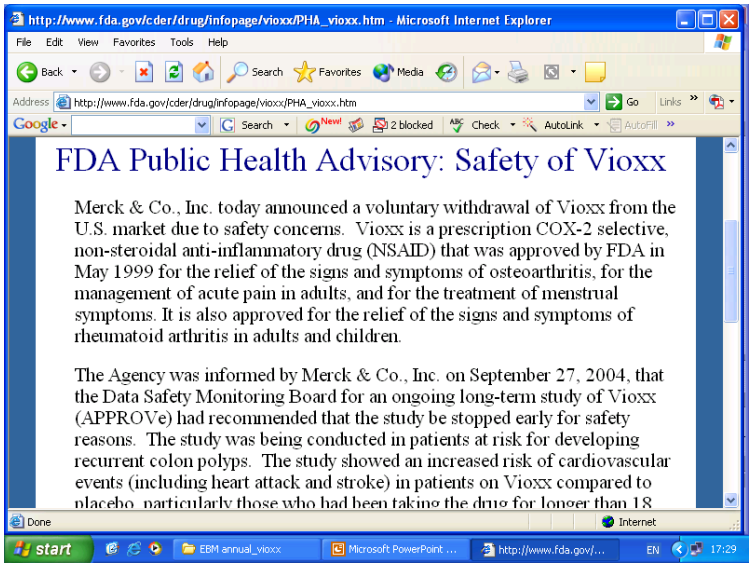
APPRAISING AN ARTICLE ABOUT HARM

Please choose your own scenario for your critical appraise.

The scenarios below demonstrate only as examples.

Example: Scenario 1

A 70-year-old woman has underlying of triple vessels disease, diabetes mellitus, and renal insufficiency. She had sustained hip fracture and underwent open reduction with dynamic hip screw fixation 2 weeks ago. After surgery, she recently receives 200 mg of celecoxib once daily for postoperative pain control. You heard about removal rofecoxib by FDA because of serious cardiovascular events, and begin thinking about celecoxib might have the same serious side effects since they are selective-cox-2 inhibitors.



This prompt you to search the medical literature on risk of myocardial infarction after taking celecoxib compared with rofecoxib and/or other NSAIDs in patients who already have cardiovascular disease.

(Hippisley-Cox J, Coupland C. Risk of myocardial infarction in patients taking population based nested case-control analysis non-steroidal anti-inflammatory drugs: cyclo-oxygenase-2 inhibitors or conventional *BMJ* 2005;330;1366-9. in Appendix 5).

Example: Scenario 2

A 54-year-old woman has been menopause for 2 years. She is quite skinny but healthy. Today she visits her mother who had hip fracture and underwent dynamic screw fixation. You think she has a risk of hip fracture since she is thin and has history of maternal hip fracture. But her bone mineral density, T-score is –0.9 SD, which is in normal limit. You are curious about how high the risk of this patient.

This problem leads you search PubMed about risks of hip fracture in postmenopausal women. Finally you got a paper talking about bone mineral density and risk of hip fracture in women (Cauley JA, Lui L, Ensrud KE, et al. Bone mineral density and the risk of incident nonspinal fractures in black and white women. JAMA 2005;293:2102-8. in Appendix 6)

ASSIGNMENTS

After selecting **your own** scenario, please answer the following questions and perform critical appraisal as shown below.

A. What is your clinical question?

   P:

   I:

   C:

   O:

B. What are your search terms?

C. Read the article and critically appraise its validity using the Appraisal Guides for an Article on harm (Appendix 1&2).

D. Appraise the results of the study, discussing the rationale for each in worksheet for harm study (Appendix 3).

E. Create critical appraisal topic (CAT) from this study (Appendix 4).

READING: Tugwell P, Haynes B. Assessing claims of causation. Haynes RB, Sackett DL, Guyatt GH, Tugwell P. Clinical epidemiology. How to do clinical practice research, 3rd Ed. Philadephia: Lippincott Williams & Wilkins, 2006:356-87.

Levine M, Haslan D, Walter S, et al. Harm. Guyatt G, Rennie D. Users' guides to the medical literature. Essentials of evidence-based clinical practice.Chicago: AMA Press, 2002:121-54.

APPENDIX (included articles assigned for reading)

Appendix 1: ASSESSING CLAIMS OF CAUSATION

Tugwell P, Haynes B. Assessing claims of causation. Haynes RB, Sackett DL, Guyatt GH, Tugwell P. Clinical epidemiology. How to do clinical practice research, 3rd Ed. Philadephia: Lippincott Williams & Wilkins, 2006:356-87.

CLINICAL RESEARCH SCENARIO: CAUSE OR COINCIDENCE?

Do silicone breast implants *cause* rheumatologic diseases? Or if women with breast implants experience disorders such as rheumatoid arthritis, is this merely a *coincidence*? More than a million women have undergone surgical implantation of silicone breast implants.

It is undisputed that the implants leak silicone into the surrounding tissues and cause local fibrosis. Several hundreds of breast-implant recipients have subsequently developed clinically significant symptoms of connective tissue diseases (CTDs), including Raynaud's phenomenon, fibromyalgia, and classic rheumatologic diseases such as scleroderma, lupus, and rheumatoid arthritis. Courts in the United States have awarded up to $14 million in damages to single cases. Missing in these early legal decisions was systematic consideration of whether, in such a large population, these conditions would be expected to occur anyway, unrelated to silicone breast implants, that is, by coincidence. How does one assess whether silicone breast implants have caused an increase in these conditions? After a number of these court decisions had been made in favor of the plaintiffs, a US federal court convened a National Science Panel, including one of us (Peter Tugwell) as a member (1). The panel was asked to assist in evaluating expert testimony and scientific evidence presented in lawsuits brought against silicone breast-implant manufacturers. We were to assess whether existing studies provide scientific evidence of an association between silicone breast implants and systemic classic/accepted CTD, atypical connective disease, and certain signs and symptoms identified by plaintiffs in the lawsuits. To do so, we performed a systematic review of published studies, using principles of causation to marshal the evidence for the court's decision. As Peter Tugwell put it to the court in his deposition, "If we take rheumatoid arthritis, without implants the frequency in the population is 1% (1 woman in 100 women), so in 1

million women without implants, 10,000 (1% of 1 million) of these women will have rheumatoid arthritis (‘expected number’). So the question we need to answer is: Were women with breast implants more likely to develop rheumatoid arthritis than women who had no implants?” Studying whether one thing causes another is a challenging task. The best scientific test of this putative relationship would be a randomized controlled trial (RCT) in which women, initially free of connective tissue complaints, who consent to be part of this trial are randomly allocated to receive or not receive silicone breast implants, and then followed to assess the incidence of such complaints. A single RCT might settle such a matter if it were large enough to detect an important difference in the risk for CTDs, and if silicone breast implants were homogeneous enough in their nature to generalize from a single study. But both the difficulty and expense of mounting a convincing study and the diversity of medical devices (with different brands, construction, and continual changes) render definitive RCTs improbable.

Further, as is often the case in causal questions, RCTs can be infeasible or unethical, especially if the potential cause is likely to be noxious, as for, say, smoking or asbestos or breast implants. If RCTs are not possible, more types of evidence are needed, although none by themselves will be close to compelling. To make matters worse, those who have a vested interest in avoiding a causal claim (e.g., that smoking is bad for health) frequently insist on “absolute truth,” something that neither clinical epidemiology nor any other scientific approach can offer. Thus, we cannot provide in this chapter a recipe for a definitive study that you can conduct to assess a causal claim. Nevertheless, principles and procedures for testing claims for causation have been widely accepted for more than half a century. These are based on the evidence accumulated from many investigations, each assessed for relative scientific merit and collectively weighed for the strength, consistency, and temporality of findings. We will explore these in this chapter and attempt to tie down a causal claim with several lines of evidence. We’ll begin with the basic ground rules that have been established for studying causation before returning to the scenario and how it played out.

10.1 BASIC PRINCIPLES OF ASSESSING CAUSATION

Because multiple studies will be needed to assess a causal claim, and because each of these studies will have limitations of both method and execution, the general procedure for assessing causation will follow the principles set out in Chapter 2 for systematic reviews. We

will begin with a review of these principles in light of assessing causation, and later in this chapter, after considering the special principles for *settling* questions of causation, we will return to this approach. Although there are many variants of the principles for systematic reviews, those set out by Sir Austin Bradford Hill many decades ago are still both simple and powerful. These guides are summarized in Table 10–1 and can be used to organize the evidence that is to be retrieved and reviewed. It is important to bear in mind that these are not "criteria" or "rules" and that following the guides will lead to an assembly of evidence, usually with shades of gray, rather than a black-and-white conclusion. Thus, a decision about causation is best based on the weight of the evidence at the time of decision and, especially if the evidence is not strong, the decision may be later overthrown by better research. That said, weighing the evidence according to its strengths and weaknesses can often get us convincingly past the paralysis of insisting on absolute truth. The best ("weightiest") evidence for causation comes from rigorous experiments in humans (i.e., RCTs). If experimental evidence is lacking, then

TABLE 10–1 **Austin Bradford Hill's Guides for Assessing Causation (2), in Descending Order of Importance***a*

1. Experimental evidence: Is there evidence from true experiments in humans?
2. Strength of association: How strongly associated is the putative risk with the outcome of interest?
3. Consistency: Have the results been replicated by different studies, in different settings, by different investigators, and under different conditions?
4. Temporality: Did the exposure precede the disease?
5. Biological gradient: Are increasing exposures (i.e., dose and duration) associated with increasing risks of disease?
6. Coherence: Is the association consistent with the natural history and epidemiology of the disease?
7. Specificity: Is the exposure associated with a very specific disease rather than a wide range of diseases?
8. Plausibility: Is there a credible biological or physical mechanism that can explain the association?
9. Analogy: Is there a known relation between a similar putative cause and effect?

*a*In our view!

From Hill AB. *Principles of medical statistics*, 9th ed. London: Lancet, 1971, with permission.

TABLE 10–2 **Organization and Analysis of Evidence to Assess Claims of Causation**

Evidence from the hierarchy of research designs true experiments

- cohort studies
- case–control studies
- analytic surveys

Strength of association Consistency—especially among studies of higher quality

Temporal sequence—from prospective studies

Gradient—by dose or duration of exposure

Sense—from epidemiology, biology, and analogy

strength of association from "lesser" studies becomes particularly important, and the quality of the studies pertaining to strength of association becomes paramount. Thus, prospective cohort studies with comparable controls and careful and independent (blinded) assessment of exposure and outcomes outweigh case–control studies and surveys, no matter how well the latter are done, provided the cohort study is competently done (e.g., successfully following a high proportion of its cohort). This hierarchy of evidence is taken into account in the reorganization of Hill's guides shown in Table 10–2. In this, we have amalgamated the lesser guides into "sense."

10.2 EVIDENCE FROM TRUE EXPERIMENTS IN HUMANS

As we've mentioned, this is the most important guide in distinguishing between coincidence and causation. Optimally, this evidence will come from RCTs. In situations in which the potential cause is "internal" (e.g., high blood pressure) or "self-inflicted" (e.g., alcohol or drugs or smoking), "reverse trials" can be done. Trials of lipid lowering, blood pressure lowering, blood sugar lowering, smoking cessation, and so on convincingly contribute to our causal understanding of harmful factors in our internal environments, particularly when lowering the suspected culprit by many means—for example, drugs that work by different mechanisms—has the same effects on the outcomes of interest. Some trials are clearly much harder to do than others. For example, a trial of blood pressure lowering is straightforward, but a trial of smoking cessation is not. Smokers could be allocated to be offered a special

smoking cessation program or no intervention (often euphemistically called “usual care”), and both groups then followed to see whether harmful effects of smoking were less in the intervention group. Even if such an RCT were done, readers who recall Chapters 4 to 7 on testing treatments will appreciate the trouble in getting a “clean” answer from a study when at least two “Cs”—low compliance (with smoking cessation) in the intervention group and contamination (quitting smoking) in the control group— are likely to make a mess of the results. In other situations, a trial of what proves to be a harmful substance can be done (or must be done!) because prior observational studies have suggested a benefit. For example, observational studies of combined estrogen–progestin hormonal replacement therapy (HRT) for postmenopausal women suggested a substantial cardiac benefit (5), but this was convincingly shown to be false by two RCTs (6,7). With the luxury of hindsight, follow-up reevaluations of observational studies purported that these really showed the same result as RCTs after adjusting for differences in baseline features (8,9). It is self-evident that observational and experimental studies can produce results that are consistent with one another, but this is most likely to occur when observational studies take special measures to avoid biases that are inherently avoided in RCTs. Observational studies are usually done before RCTs and with fewer resources, so that their ability to reduce bias is limited, even if the investigators are aware of the possible biases. In addition, no amount of resources can eliminate unknown confounders in observational studies, whereas RCTs neutralize the effect of such biases by ensuring that they are randomly distributed to the groups being compared.

It is often claimed that observational studies are needed to look for rarer adverse effects of medications, but the sample sizes of “pivotal studies” required for approval are now increasing so that less common adverse effects can be detected. For example, pivotal studies of coxibs for pain and arthritis are now required to be large enough to detect adverse effect rates as low as 1% in RCTs (10). Where individual studies are too small, metaanalyses of RCTs should be considered. For example, a meta-analysis of RCTs showed the lack of efficacy of vitamin E for lowering the risk of cardiovascular events and a small increase in risk of death and cardiovascular events with use of (11). Similarly, but less convincingly, Hemminki and McPherson’s review of small hormone replacement trials (12) raised the possibility that hormone replacement therapy (HRT) was unlikely to lower cardiovascular risk long before the definitive trials confirmed that HRT was actually harmful. Our key point is this:

RCTs provide the best evidence for causation, so don't give up on the notion of doing an RCT to settle a causal issue just because it may be difficult or contentious to do. Many precedents exist where hoards of biased observational studies have been overthrown by a single, large, well-done RCT. A brief statistical interlude: If we did an RCT of breast implants, or a reverse trial of removing them, to see if they cause musculoskeletal (MSK) complaints, Table 10–3 would be a good way to display the findings. The results would then be calculated as the risk in exposed versus nonexposed. That is: rate in exposed divided by rate in unexposed _ a>1a _ b2 _ c>1c _ d2.

Appendix 2: Guideline for Critical Appraisal on Harm study

Levine M, Haslan D, Walter S, et al. Harm. Guyatt G, Rennie D. Users' guides to the medical literature. Essentials of evidence-based clinical practice. Chicago: AMA Press, 2002:121-54.

Critical Appraisal on Harm study

EBM Elective CEU
TABLE 1B-5
Users' Guides for an Article About Harm Are the results valid? In a cohort study, aside from the exposure of interest did the exposed and control groups start and finish with the same risk for the outcome? ■ Were patients similar for prognostic factors that are known to be associated with the outcome (or did statistical adjustment level the playing field)? ■ Were the circumstances and methods for detecting the outcome similar? ■ Was the follow-up sufficiently complete? In a case-control study, did the case and control group have the same risk (chance) for being exposed in the past? ■ Were case and controls similar with respect to the indication or circumstances that would lead to exposure? ■ Were the circumstances and methods for determining exposure similar for case and controls? What are the results? ■ How strong in the association between exposure and out-come? ■ How precise is the estimated of the risk? How con I apply the results to patient care? ■ Were the study patients similar to the patient in my practice? ■ Was follow-up sufficiently complete? ■ Is the exposure similar to what might occur in my patient? ■ What is the magnitude of the risk? ■ Are there any benefits that are known to be associated with exposure?

Appendix 3: Users' guide for critical appraisal Risk & Harm

A. Are the results valid?	
• Were patients similar for prognostic factors that are known to be associated with the outcome (or did statistical adjustment level the playing field)?	
• Were the circumstances and methods for detecting the outcome similar?	
• Was the follow-up sufficiently complete?	
• Were case and controls similar with respect to the indication or circumstances that would lead to exposure?	
• Were the circumstances and methods for determining exposure similar for case and controls?	
B. What are the results?	
• How strong in the association between exposure and out-come?	
• How precise is the estimated of the ris	
C. How can I apply the results to patient care?	
• Were the study patients similar to the patient in my practice?	
• Was follow-up sufficiently complete?	
• Is the exposure similar to what might occur in my patient?	
• What is the magnitude of the risk?	
• Are there any benefits that are known to be associated with exposure?	

TUTOR'S GUIDES FOR ARTICLES ON HARM

Title: Hippisley-Cox J, Coupland C. Risk of myocardial infarction in patients taking cyclo-oxygenase-2 inhibitors or conventional non-steroidal anti-inflammatory drugs: population based nested case-control analysis. BMJ 2005;330:1366-9.

Are the results valid?

1. Did the investigators demonstrate similarity in all known determinants of outcome? Did they adjust for differences in the analysis?

No. Table 1 showed higher percentages of use of aspirin, statin, tricyclic antidepressant, selective serotonin reuptake inhibitor, ischaemic heart disease, diabetes, hypertension, osteoarthritis, rheumatoid arthritis, smoking, and obesity in cases than controls.

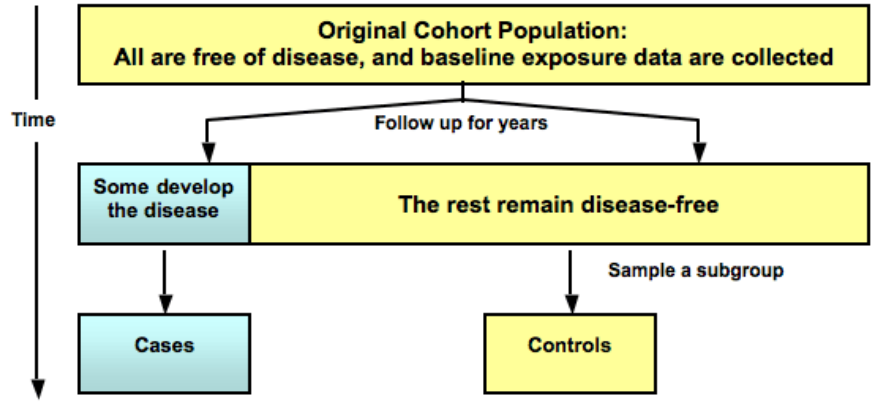
Yes. The authors adjusted for these factors in the analysis (Table 3).

Suggested follow-up questions:

a. What study design would practically ensure that all known (and unknown) determinants of outcome between the study groups would be equal?
   Answer: RCT's

b. What are the reasons for the lack of RCT's on harmful interventions?
   Answer: It would be unethical and often adverse effects occur rarely and over-prolonged periods of exposure thus making an RCT may not be feasible.

c. What other study designs may be employed when a randomized trial is not feasible?
   Answer: Cohort studies, case-control studies, case series, case reports. Also a nested case control study is the other study design. It has characteristics as shown below.



Suggest a discussion of the basic study design and the advantages and disadvantages of cohort vs. case-control vs. nested case control.

Study design	*Advantages*	*Disadvantages*
Case-control	Cheaper Valuable for rare condition Short duration	Baseline risk not measured No temporal relation Recall bias
Cohort	Baseline risk measured Temporal relation	Expensive Not valuable for rare disease, disease with long duration Long duration
Nested case control	Cheaper than cohort Baseline risk measured Temporal relation Unbiased association Decrease selection bias (cases & controls are from the same cohort)	Controls may not represent entire cohort due to die or loss of follow-up

d. Should there be differences in determinants of outcome, what may be done to have an unbiased estimate of the harmful exposure?

Answer: Statistical adjustment for prognostic factors. However, one can only adjust for known determinants of outcome.

e. Optional points regarding OR:

1.) It approximates RR when incidence is very small.

Exposure	Disease		Total
	Yes	No	
Yes	a	b	a+b
No	c	d	c+d
Total	a+c	b+d	a+b+c+d

RR = [a/(a+b)] / [c/(c+d)]

OR = [a/b] / [c/d]

When the incidence is very small,

a/b ~ a/(a+b), when a/(a+b) is small; and

c/d ~ c/(c+d), when c/(c+d) is small: therefore

OR ~ RR

2.) It is non directional:

$$\frac{a/c}{b/d} = \frac{a/d}{b/c}$$

eg- the odds of myocardial infarction patients using celecoxib is the same as the odds of using celecoxib patients having myocardial infarction.

2. Were exposed patients equally likely to be identified in the two groups?

Yes. The authors extracted and coded data on the medical history and use of prescribed drugs (selective and non-selective NSAIDs) in three years before their index date by using the QRESEARCH database (see p.1366-7).

Follow-up questions:

a. In case control studies, how is exposure ascertained?

Answer: usually through a questionnaire. But this study, the exposure ascertained by using database.

b. What biases may be introduced by the method of determining exposure?

Answer: Recall bias; the increase in the probability that a person with the outcome will recall the exposure, interviewer bias- increase in the probability of exposure because of deeper probing by the interviewer.

3. Were the outcomes measured in the same way in the groups being compared?

Yes. See p. 1367. The outcome for cases was defined as the first diagnosed of myocardial infarction recorded as the cause of death. The outcome for controls was defined as alive (and registered with the practice) at the time their matched case had myocardial infarction.

Suggested follow-up questions:

a. How would violation of this criterion lead to bias?

Answer: In cohort studies, ascertainment of outcome might become more intense among exposed patients, thus making the association seem stronger (label: surveillance bias). In case control study, ascertainment of outcome (the definition of cases and controls) might become unclear (misclassification). The association would not be presented.

4.Was follow-up sufficiently complete?

Yes. This is a nested case-control study. They retrospectively collected the data for 3 years, which was quite long enough to see the effect of NSAIDs as a risk factor of myocardial infarction. However, controls might not represent entire cohort because of death or loss of follow-up before index date

Suggested follow-up questions:

a. How would be lost to follow-up lead to bias.

Answer: In cohort studies, drop-outs might represent of patients with a higher event rate.

Re-introduce the concept of a sensitivity analysis – assuming events for drop-outs.

B. What are the results?

1. How strong is the association between exposure and outcome?

See Table 3, p. 1367. The adjusted OR was 1.21 (95% CI: 0.96,1.54) in celecoxib group, 1.32 (95% CI: 1.09, 1.61) in rofecoxib group, and 1.27 (95% CI: 1.01, 1.60) in naproxen group.

2. How precise is the estimate of the risk?

See Table 3, p. 1367 for 95% confidence interval.

Follow-up questions:

a. What is the difference between a point estimate and interval estimate?

Suggested exercise

1.) Ask for an estimate of the average height or weight (or some other parameter of people in the room.
2.) Ask them regarding the probability that the estimate is correct – very low.
3.) Now ask for an interval estimate, then, ask them the probability that this is correct.

Point to bring out

1.) Interval estimates are humbler because they accept a range of possibilities.
2.) Interval estimates are more likely to be correct
3.) More useful because aside from suggesting statistical significance, they convey a message, magnitude of effect, ie, the best and worst scenario.
4.) All point estimates have surrounding interval estimates.

b. What does the p-value mean when reported with point estimate of a treatment effect?

Extract: p is the probability that the observed differences are coincidental.

c. How does this form of reporting relate with interval estimates of treatment effects?

Extract: $p < 0.05$ implies the 95% CI of the RR does not contain the value 1.0

d. What are the advantages and disadvantages of reporting treatment effects as interval estimates instead of point estimates with corresponding p-values?

   Extract: 95% CI can be understood more intuitively than p-values.

C. How can I apply the results to patient care?

1.Were the study patients similar to the patients in my practice?

Yes. There are no studies indicate the differences in the biology of myocardial infarction among Asians and Caucasians. However, UK people had lower incidence of myocardial infarction when compared with south Asians (UK Prospective Diabetes Study Group, 1998).

2. Was the duration of follow-up adequate?

Yes. The authors included only patients who had completed data records 3 years before index date which was enough to see the outcome (myocardial infarction).

Follow-up questions:

a. Why is this important?

   Answer: If the follow-up is too short (eg- some diseases need a long latency period to manifest) then the author's conclusions, though valid, may not be useful.

3. What was the magnitude of the risk?

See Table 3, p. 1367. The adjusted OR was 1.21 (95% CI: 0.96,1.54) in celecoxib group, 1.32 (95% CI: 1.09, 1.61) in rofecoxib group, and 1.27 (95% CI: 1.01, 1.60) in naproxen group. In case control study, the odds ratios do not tell us how frequently the problem occurs. So we need to calculate the number needed to harm (NNH) which means number of people you would need to treat with a specific intervention for a given time to cause one additional adverse income. Large numbers are good, because they mean that adverse events are rare. Small values for NNH are bad, because they mean adverse events are common.

Number needed to harm (NNH) of having myocardial infarction in patients using celecoxib within 3 months of index date (see data on Table 2 and 3):

NNH = [(CER*(OR-1))+1] / [CER*(OR-1)*(1-CER]

= [((8988/(8988+84762))*(1.2-1))+1]

[(8988/(8988+84762))*(1.2-1)*(1-(8988/(8988+84762))]

= [(0.095*0.2) + 1] / [0.095*0.2*(1-0.095)]

= 1.019 / 0.017

= 59

Note: CER – control event rate (rate of outcome among the unexposed)

For every 59 patients treated with celecoxib within 3 months, one myocardial infarction occurs when compared with who did not use celecoxib in past 3 years.

The relative risk and the odds ratio do not tell us how frequently the problem occurs, the only tell us that the observed effect occurs more or less often in the exposed group compared to the unexposed group. (Eg – smoking may increase the risk or odds of cancer 10 fold, but what is the baseline risk in the first place?)

In a cohort, we can estimate the absolute risk reductions; in a case-control, we need to estimate it using the techniques described below.

Optional:

a. Number needed to treat (NNT) or number needed to harm (NNH) can be calculated from odds ratios using the following formula:
b. NNT or NNH can also be calculated from odds ratios using the table below

Table a: NNT and NNH calculation from odds ratios

CER	Odds Ratios																	
	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9		1.5	2	2.5	3	3.5	4	4.5	5	10
	NNTs for efficacy										NNHs for harm								
0.05	41	46	52	59	69	83	104	139	209		43	22	15	12	9	8	7	6	3
0.1	21	24	27	31	36	43	54	73	110		23	12	9	7	6	5	4	4	2
0.2	11	13	14	17	20	24	30	40	61		14	8	5	4	4	3	3	3	2
0.3	8	9	10	12	14	18	22	30	46		11	6	5	4	3	3	3	3	2
0.4	7	8	9	10	12	15	19	26	40		10	6	4	4	3	3	3	3	2
0.5	6	7	8	9	11	14	18	25	38		10	6	5	4	4	3	3	3	2
0.7	6	7	9	10	13	16	20	28	44		13	8	7	6	5	5	5	5	4
0.9	12	15	18	22	27	34	46	64	101		32	21	17	16	14	14	13	13	11

4.Should I attempt to stop the exposure?

May be. Even though NNH of celecoxib was 59, myocardial infarction is a very serious risk. All NSAIDs should be used with caution in patients with cardiovascular disease. Other drugs such as paracetamol or opioids might be better choices of treatment.

REFERENCE

1. UK Prospective Diabetes Study Group. Ethnicity and cardiovascular disease. The incidence of myocardial infarction in white, South Asian, and Afro-Caribbean patients with type 2 diabetes (U.K. Prospective Diabetes Study 32). Diabetes Care 1998;21:1271-7.

Appendix 4: Critical appraisal topic (CAT) for harm study

Clinical Question:

Citation:

A. Study Characteristics:

1. Patients included –

2. Exposure –

3. Outcome –

B. Validity Criteria:

1. Did investigators demonstrate similarity in all known determinants of outcome? Differences in analysis adjusted?

2. Were exposed patients equally likely to be identified in the 2 groups?

3. Outcomes measure in the same way in the comparison groups?

4. Was follow-up sufficiently complete?

C. Results [adjust the number of rows as needed]:

Outcome	Point Estimate (Specify if RR or OR)	95% Confidence Interval
1.		
2.		
3.		
4.		
5.		
6.		
Total		

D. Applicability:

1. Were the study patients similar to the patient in my practice?

2. Was the duration of follow-up adequate?

3. What was the magnitude of the risk?

4. Should I attempt to stop the exposure?

Author's Conclusion:

Reviewer's Conclusion:

Reviewer: **Date:**