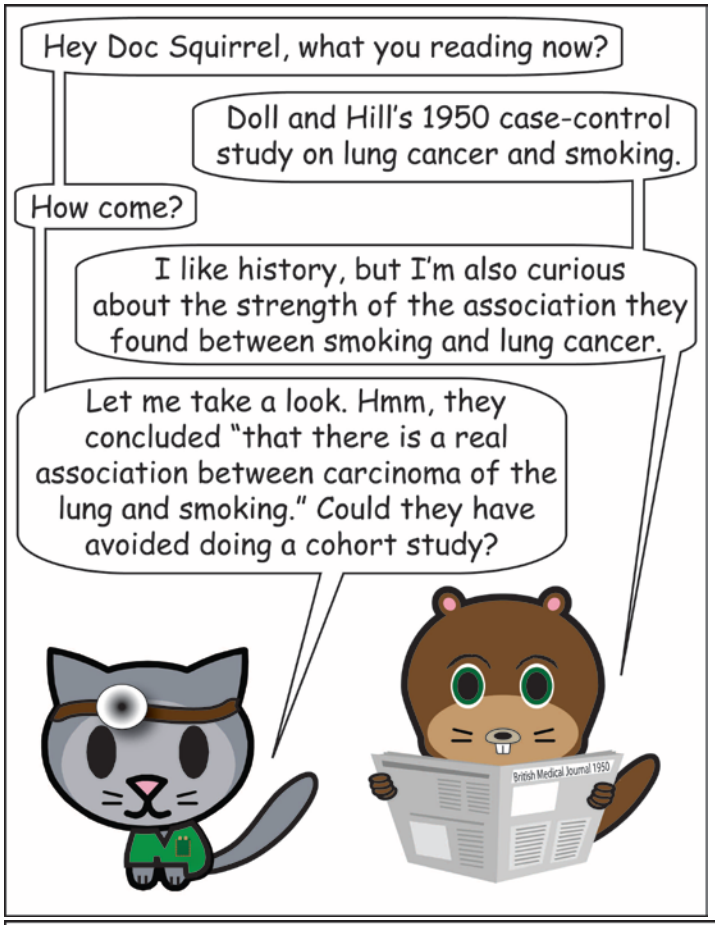
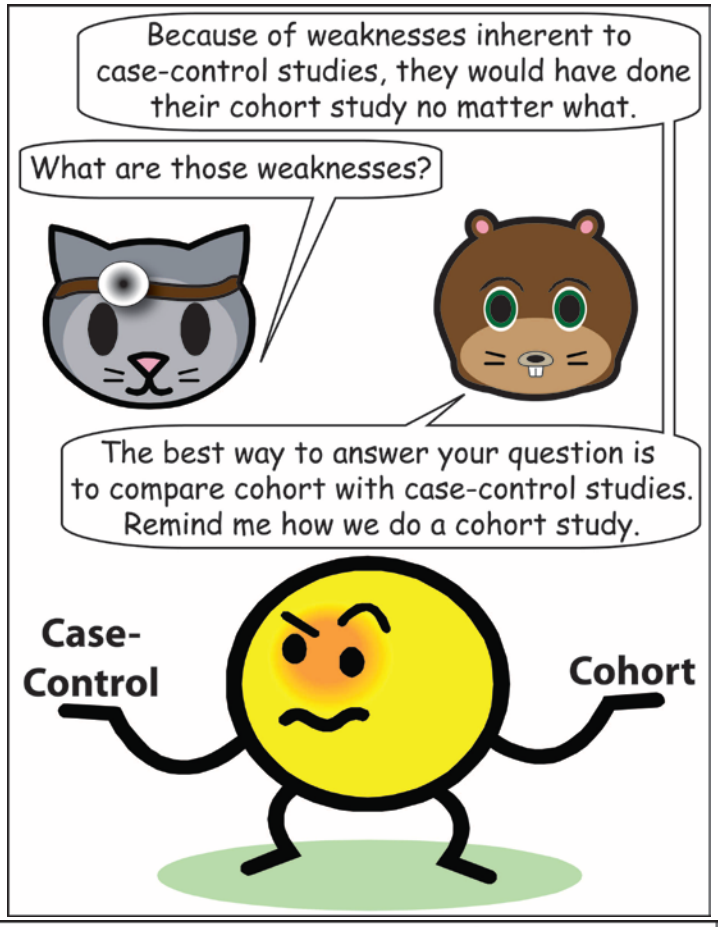
The Continuing Adventures of
"Doc Squirrel" and "Kid Cat" !!

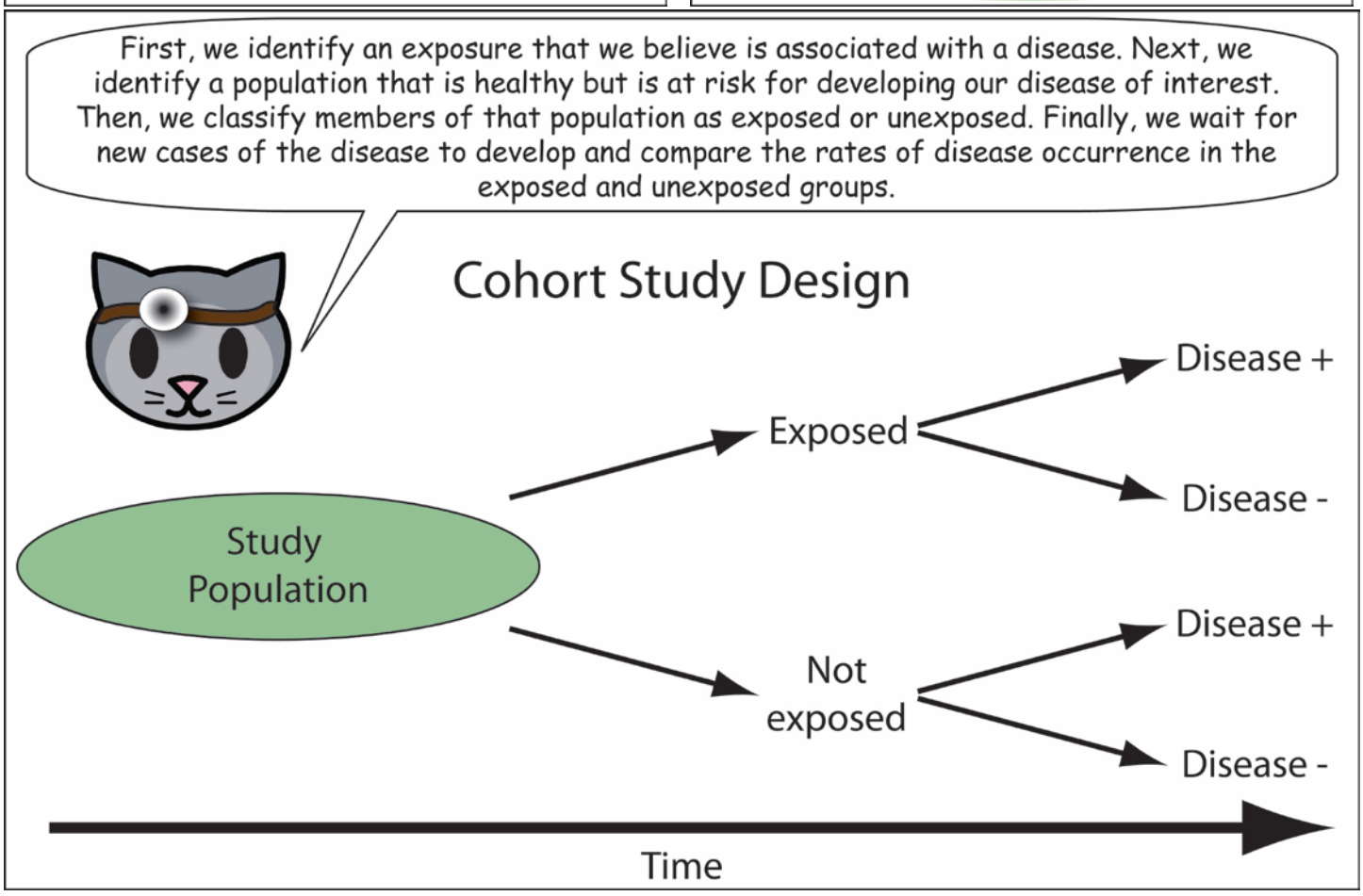
Cases and Controls

Stefan Tigges MD, MSCR


Hey Doc Squirrel, what you reading now?
Doll and Hill's 1950 case-control study on lung cancer and smoking.
How come?
I like history, but I'm also curious about the strength of the association they found between smoking and lung cancer.
Let me take a look. Hmm, they concluded "that there is a real association between carcinoma of the lung and smoking." Could they have avoided doing a cohort study?
British Medical Journal 1950


Because of weaknesses inherent to case-control studies, they would have done their cohort study no matter what.
What are those weaknesses?
The best way to answer your question is to compare cohort with case-control studies. Remind me how we do a cohort study.
Case-Control
Cohort

1


First, we identify an exposure that we believe is associated with a disease. Next, we identify a population that is healthy but is at risk for developing our disease of interest. Then, we classify members of that population as exposed or unexposed. Finally, we wait for new cases of the disease to develop and compare the rates of disease occurrence in the exposed and unexposed groups.
Cohort Study Design
Study Population
Exposed
Not exposed
Disease +
Disease -
Disease +
Disease -
Time

Correct. You have described a prospective cohort trial. It is possible to perform a retrospective cohort, but for now we will keep things simple and stick with the prospective cohort. What are some advantages of this study design?

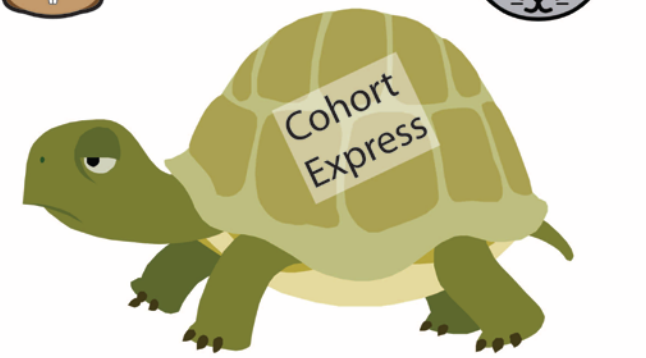
It insures that the exposure precedes the disease that the exposure supposedly causes.

True. Cohort studies are also used to study rare exposures, track multiple effects of a single exposure and are used to measure disease incidence.

There must be some disadvantages.

Yes. If follow up takes many years, results won't be readily available.

Why is that a problem?



In 2006, the CDC investigated an outbreak of disease that killed three people and caused hemolytic-uremic syndrome in multiple victims.

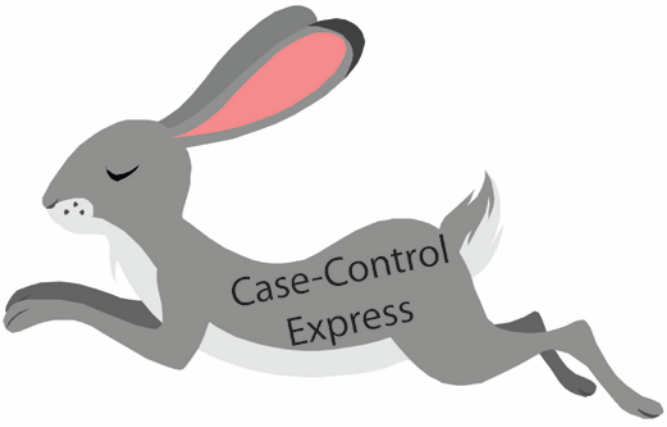
Hemolytic-uremic syndrome; sounds awful!

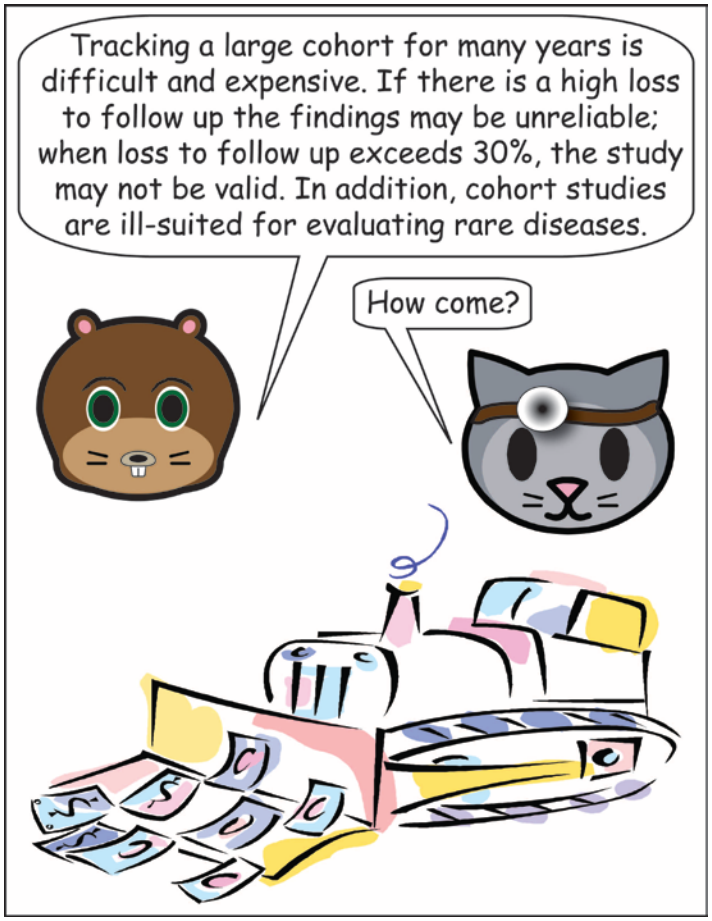
Yes, it occurs when an ingested toxin destroys red blood cells, causing kidney injury. If the CDC had done a cohort study to find the cause of this outbreak, they would have had to enroll thousands of subjects and wait for people to become diseased.

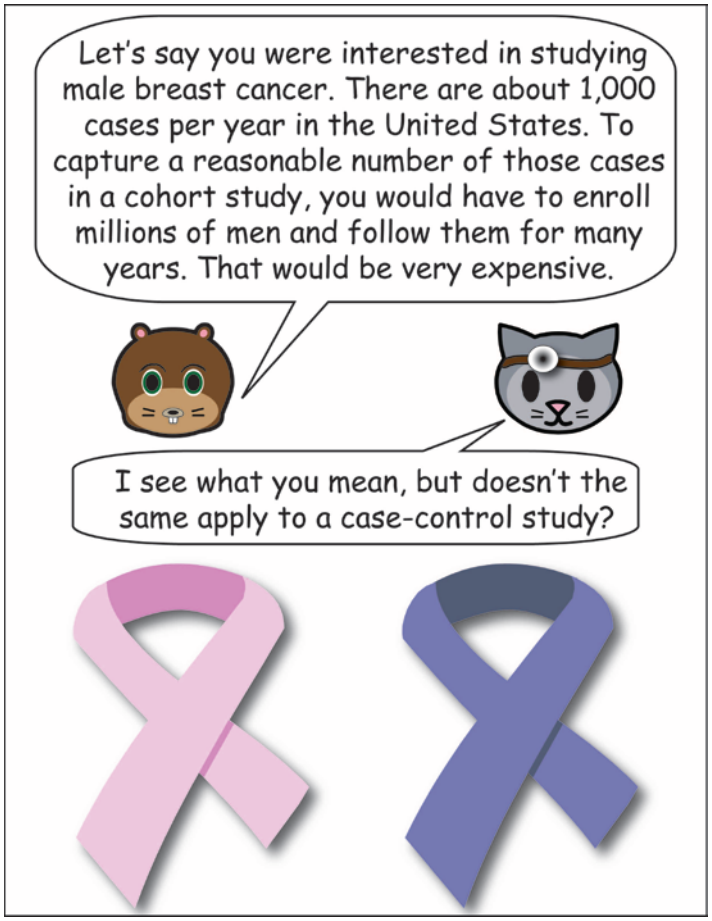
But meanwhile people were dying!

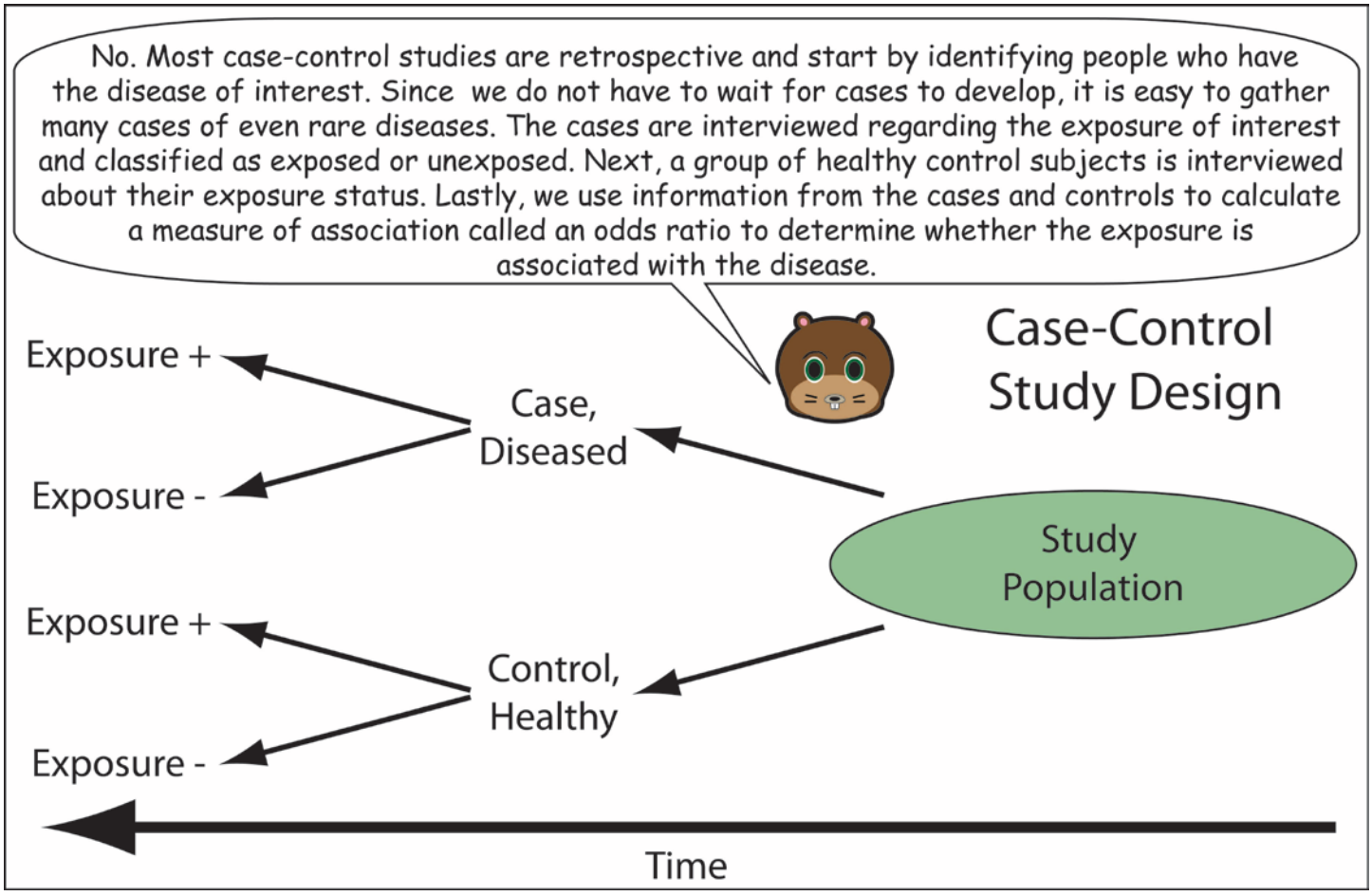
That's why the CDC did a case-control study and quickly established that the cause of the outbreak was fresh spinach contaminated with E. coli. The public was then warned not to consume tainted spinach.

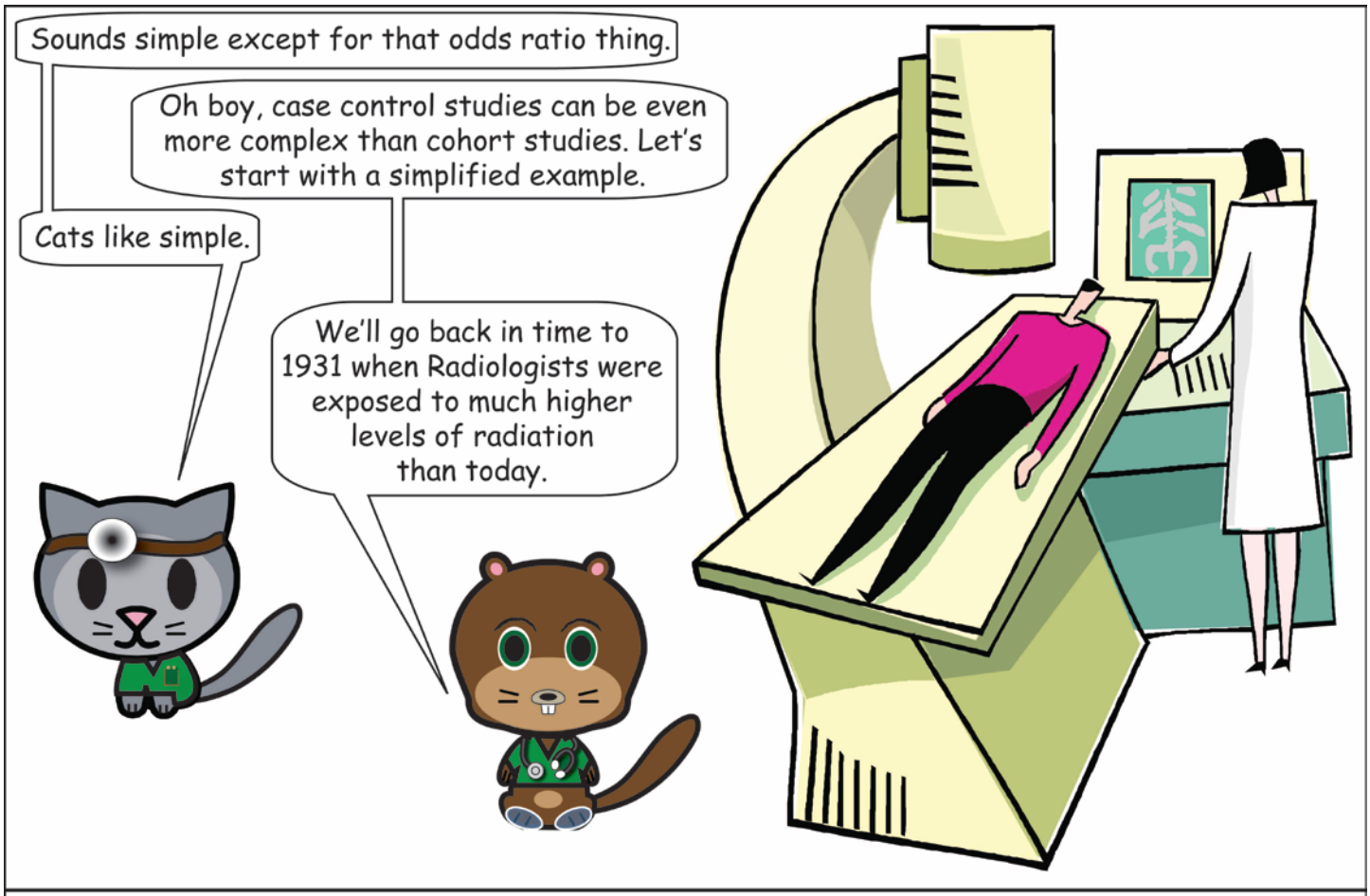
Cats don't like spinach anyway. Are there other disadvantages of cohort studies?




Tracking a large cohort for many years is difficult and expensive. If there is a high loss to follow up the findings may be unreliable; when loss to follow up exceeds 30%, the study may not be valid. In addition, cohort studies are ill-suited for evaluating rare diseases.
How come?


Let's say you were interested in studying male breast cancer. There are about 1,000 cases per year in the United States. To capture a reasonable number of those cases in a cohort study, you would have to enroll millions of men and follow them for many years. That would be very expensive.
I see what you mean, but doesn't the same apply to a case-control study?


No. Most case-control studies are retrospective and start by identifying people who have the disease of interest. Since we do not have to wait for cases to develop, it is easy to gather many cases of even rare diseases. The cases are interviewed regarding the exposure of interest and classified as exposed or unexposed. Next, a group of healthy control subjects is interviewed about their exposure status. Lastly, we use information from the cases and controls to calculate a measure of association called an odds ratio to determine whether the exposure is associated with the disease.
Case-Control Study Design
Exposure +
Exposure -
Case, Diseased
Study Population
Exposure +
Exposure -
Control, Healthy
Time



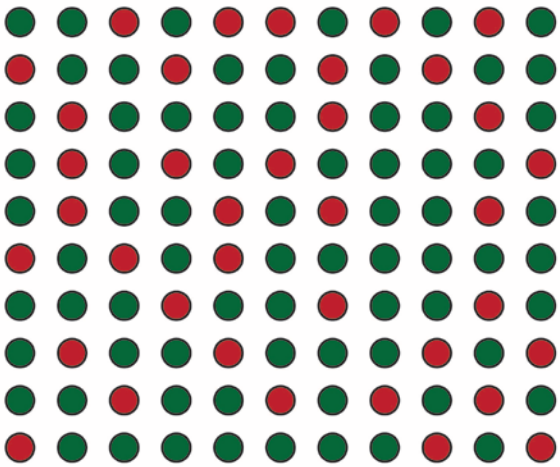
A 1930 case series suggested that Radiologists were more likely to be diagnosed with leukemia than other physicians. You decide to do a cohort study and identify a 110 member group practice and classify Docs as exposed (Radiologists) or unexposed (all other Docs). You wait one year for new cases of leukemia to develop. Results of your (fake) study are below.

● = Healthy + Unexposed

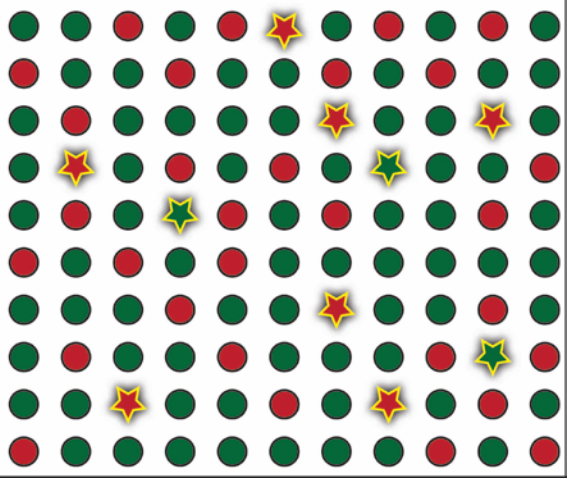
● = Healthy + Exposed

★ = Leukemia + Unexposed

★ = Leukemia + Exposed



One Year Later



Boys, there is a better way to present this data using a 2x2 table. First check out the anatomy and physiology of a 2x2 table, then use the table from our leukemia study to calculate the cumulative incidence in each group and the relative risk.

Disease

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

Leukemia

Radiologist	Yes	No	
Yes	7	30	37
No	3	70	73
	10	100	110

a=exposed+diseased
b=exposed+well
c=unexposed+diseased
d=unexposed+well

We have 37 exposed and 73 non-exposed. There were 7 cases of leukemia among the 37 exposed and 3 cases among the 73 non-exposed. The incidence among the exposed is .189 and the incidence among the unexposed is .041. The relative risk is .189/.041= 4.6. That means that Radiologists have 4.6 times the risk of other doctors for developing leukemia.

● = Healthy + Unexposed

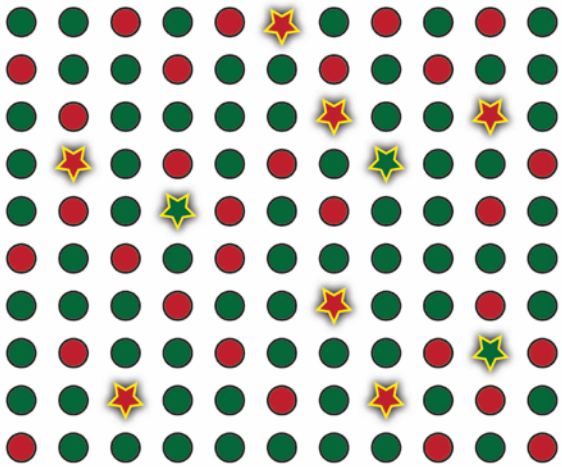
● = Healthy + Exposed

★ = Leukemia + Unexposed

★ = Leukemia + Exposed

Leukemia

Radiologist	Yes	No	
Yes	7	30	37
No	3	70	73
	10	100	110



Good, now let's look at the case-control version of this study. What would we do first?

Identify the cases.

Next, you sample from your population to get the controls. Then what?

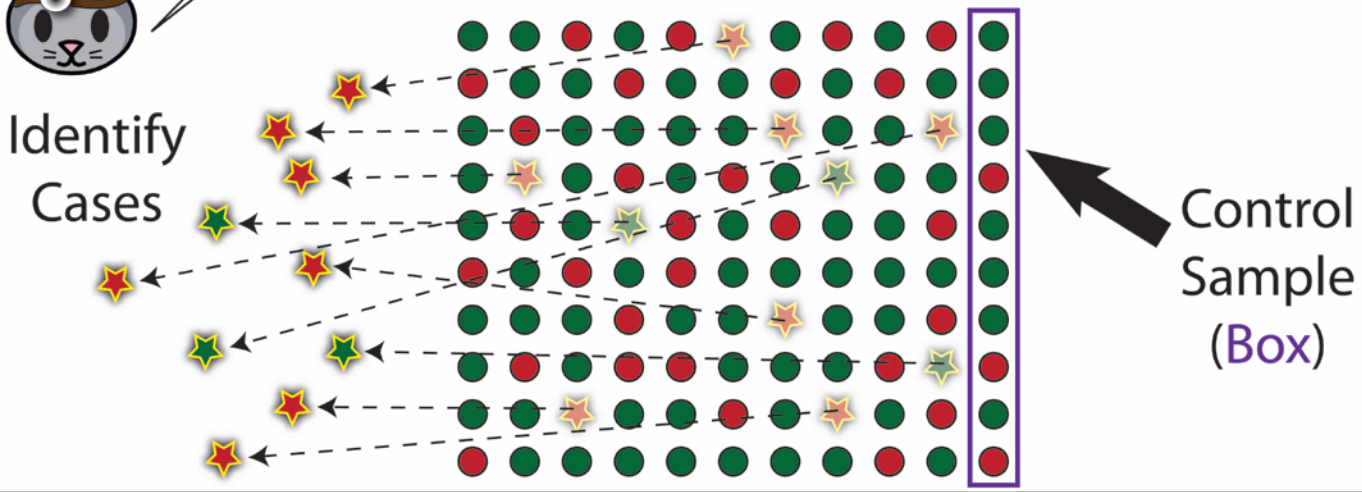
We classify the cases and the controls as exposed or unexposed.

● = Healthy + Unexposed

● = Healthy + Exposed

★ = Leukemia + Unexposed

★ = Leukemia + Exposed

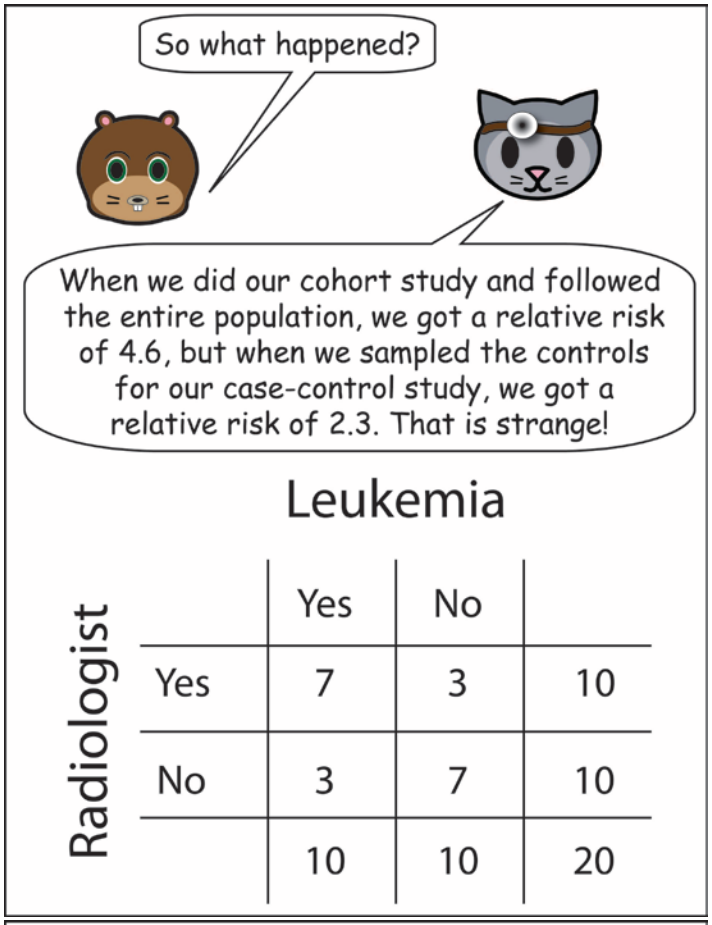


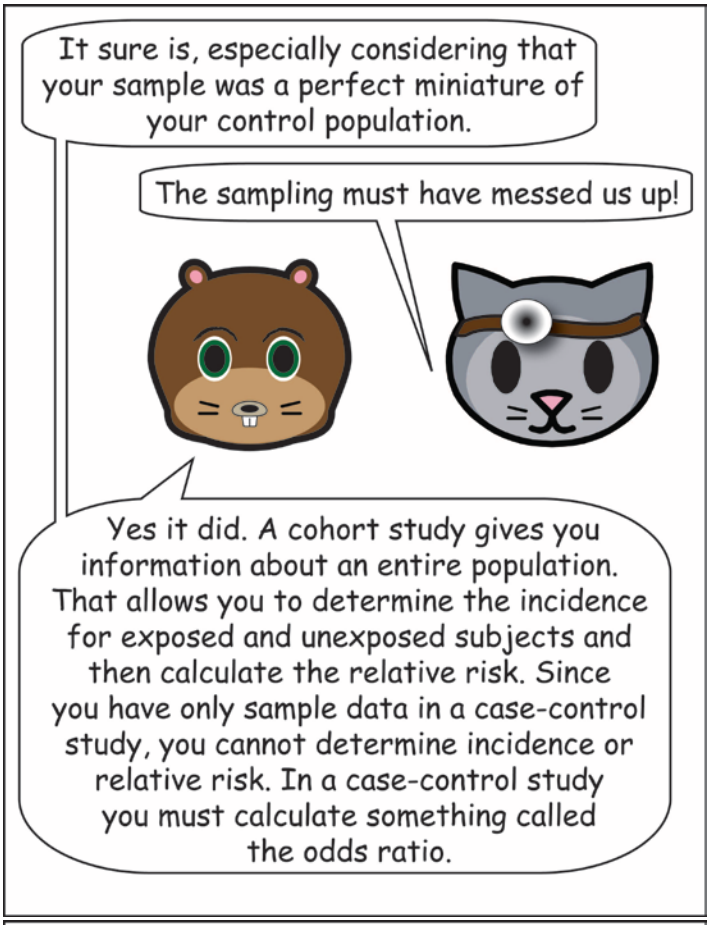
The results are shown in table form below. Note that our control sample is a perfect representation of the unexposed portion from our cohort study; in the cohort study, 70 out of 100 healthy subjects were unexposed, while in the case-control study, 7 of 10 healthy controls are unexposed. Recalculate the incidence rates and the relative risk.

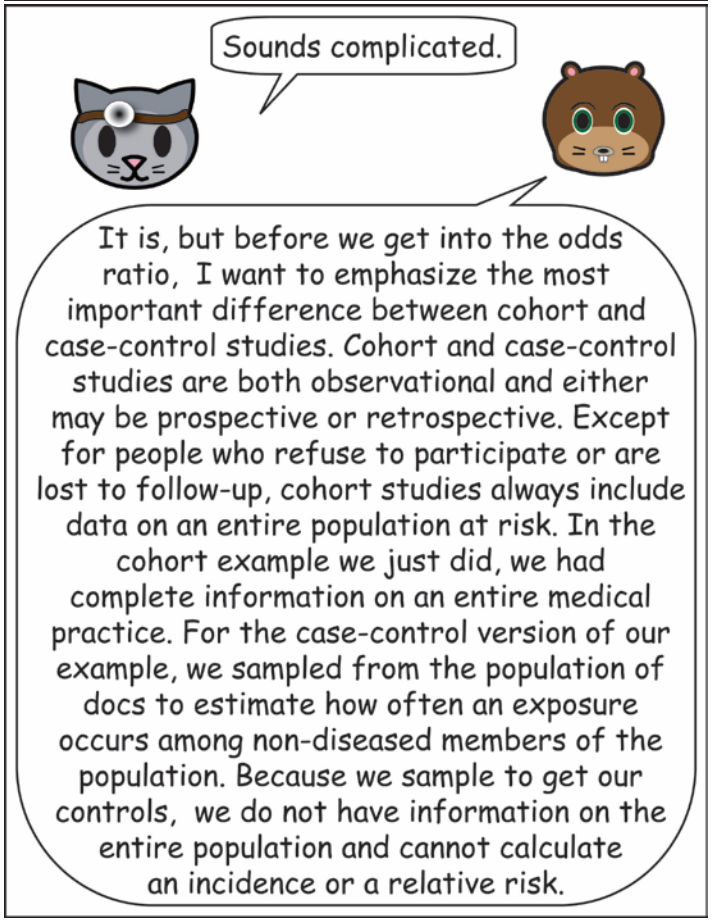
Now we have 10 exposed and 10 unexposed. Among the 10 exposed, there are 7 cases and among the 10 unexposed, we have 3 cases. The incidence among exposed is 7/10 or .7, while the incidence among the unexposed is 3/10 or .3. The relative risk is .7/.3 or 2.3.

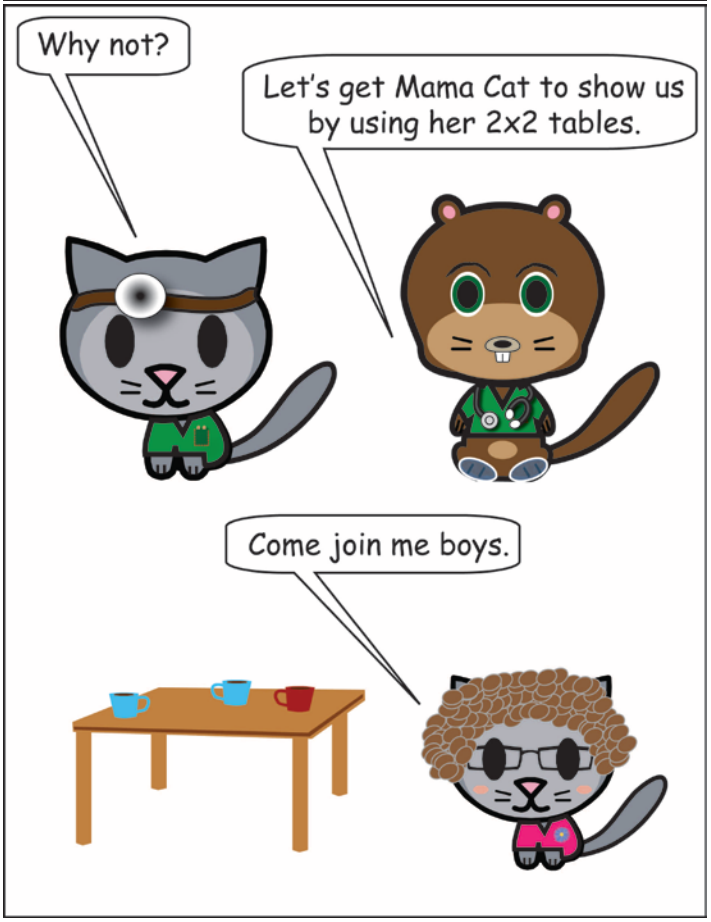
Leukemia

Radiologist	Yes	No	
Yes	7	3	10
No	3	7	10
	10	10	20


So what happened?
When we did our cohort study and followed the entire population, we got a relative risk of 4.6, but when we sampled the controls for our case-control study, we got a relative risk of 2.3. That is strange!
Leukemia
Radiologist
Yes
No
Yes 7 3 10
No 3 7 10
10 10 20


It sure is, especially considering that your sample was a perfect miniature of your control population.
The sampling must have messed us up!
Yes it did. A cohort study gives you information about an entire population. That allows you to determine the incidence for exposed and unexposed subjects and then calculate the relative risk. Since you have only sample data in a case-control study, you cannot determine incidence or relative risk. In a case-control study you must calculate something called the odds ratio.


Sounds complicated.
It is, but before we get into the odds ratio, I want to emphasize the most important difference between cohort and case-control studies. Cohort and case-control studies are both observational and either may be prospective or retrospective. Except for people who refuse to participate or are lost to follow-up, cohort studies always include data on an entire population at risk. In the cohort example we just did, we had complete information on an entire medical practice. For the case-control version of our example, we sampled from the population of docs to estimate how often an exposure occurs among non-diseased members of the population. Because we sample to get our controls, we do not have information on the entire population and cannot calculate an incidence or a relative risk.


Why not?
Let's get Mama Cat to show us by using her 2x2 tables.
Come join me boys.

Compare the 2x2 tables for the cohort (top) and case-control (bottom) versions of our leukemia study and look at how the differences in cells b and d (red) alter our relative risk calculations. If we do a case-control study, we do not have complete data for either the exposed or unexposed people (purple), so our incidence calculations for both the exposed and the unexposed groups are incorrect. Compared to a cohort study, in a case-control study diseased people are overrepresented because our healthy controls are only a sample of the entire poplation. In the cohort study, we had complete information with respect to exposure and disease status for an entire population, but in the case-control study, we are missing data on 90 healthy people because we only sampled 10 controls.

Remember, in real life we would never calculate incidence or a risk ratio from a case-control study. The example at bottom is done only to illustrate why doing so is wrong.

Cohort Study Table and Calculations

Leukemia

Radiologist	Yes	No	
Yes	7	30	37
No	3	70	73
	10	100	110

Incidence Exposed=7/37=.189

Incidence Unexposed=3/73=.041

Relative Risk=.189/.041=4.6

Case-Control Study Table and Calculations

Leukemia

Radiologist	Yes	No	
Yes	7	3	10
No	3	7	10
	10	10	20

Incidence Exposed=7/10=.7

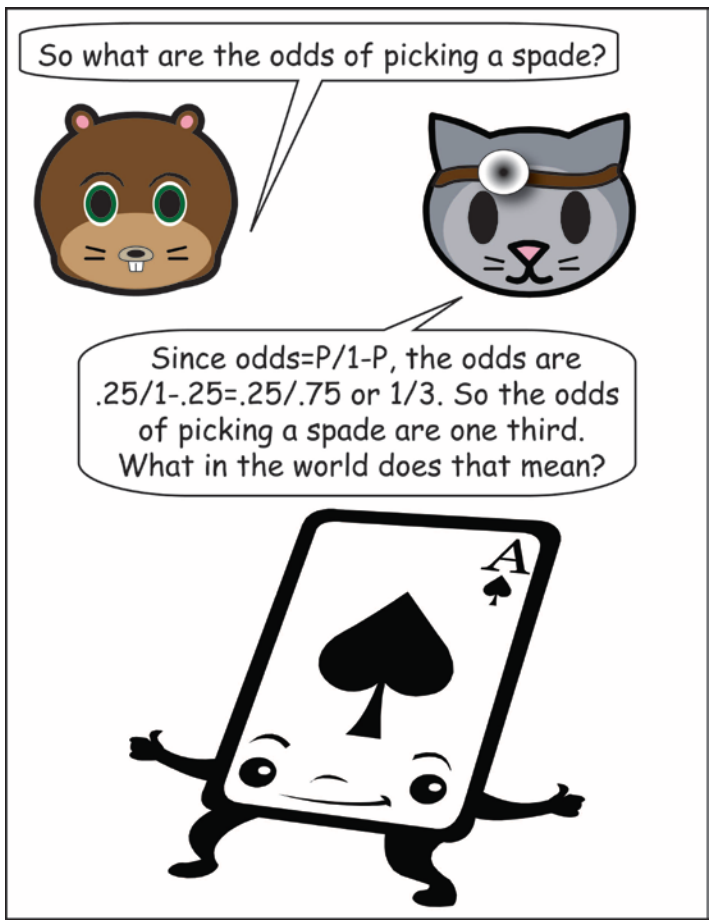
Incidence Unexposed=3/10=.3

Relative Risk=.7/.3=2.3


My brain hurts.
Trust me, it gets worse.
So what is an odds ratio?
We'll start by figuring out what odds are.
Odds, sounds sleazy in a Las Vegas sort of way. Cats don't like Vegas!
Don't worry, odds are perfectly respectable.


Odds are defined as the probability of an event happening divided by the probability of that event not happening. We can write that definition as an equation: Odds=P/1-P.
Example please.


Let's try an easy one: What are the odds of randomly picking a spade out of a deck of cards?
The probability of picking a spade is .25. Since the total probability has to equal 1, the probability of not picking a spade is 1-.25 or .75. That makes sense; there are 3 other suits in a deck.


So what are the odds of picking a spade?
Since odds=P/1-P, the odds are .25/1-.25=.25/.75 or 1/3. So the odds of picking a spade are one third. What in the world does that mean?

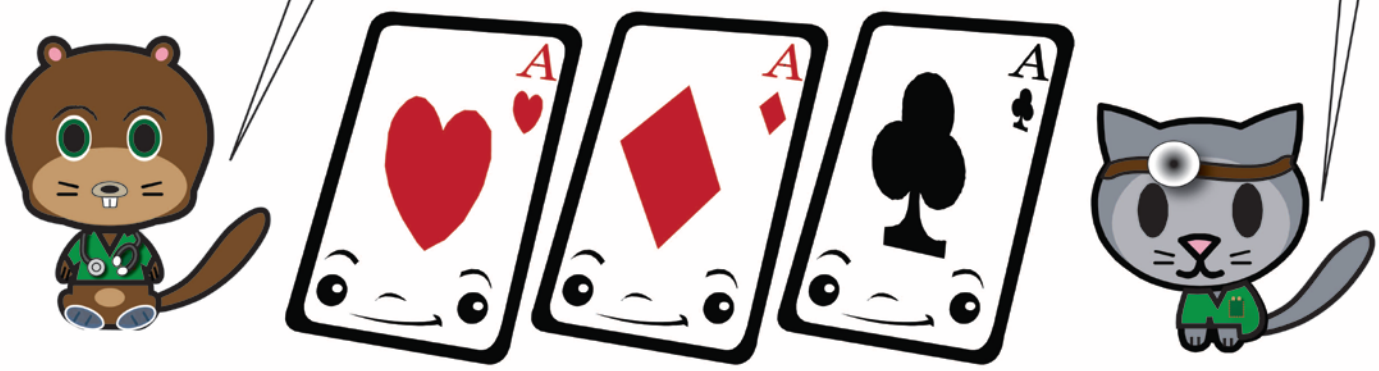
Odds are difficult! In this case, an odds of one third means that the chance of randomly drawing a spade are one third the probability of not drawing a spade. Or, the chances of drawing a card other than a spade are 3 times the probability of drawing a spade.

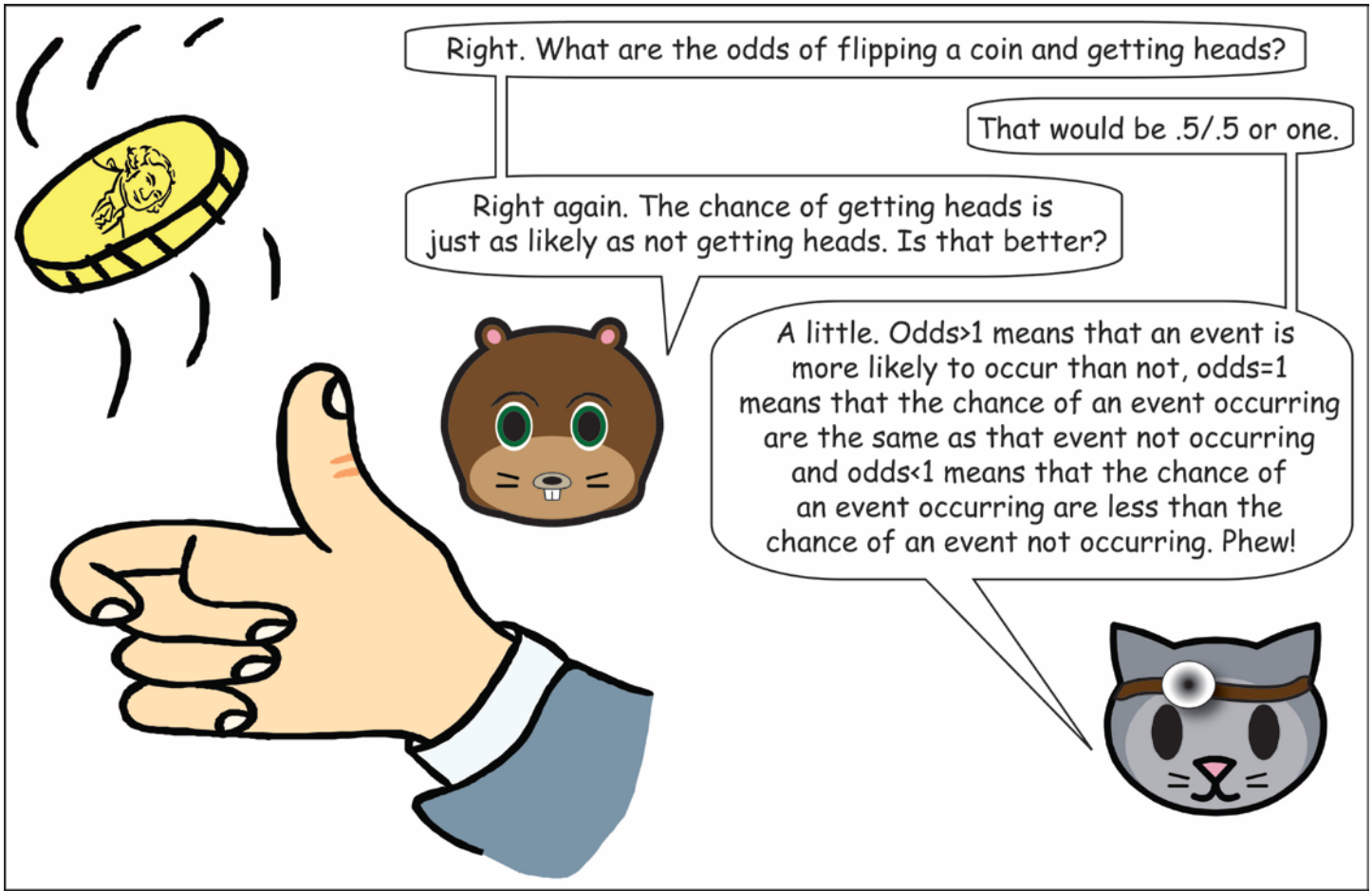
Let's try again, this time in reverse. What are the odds of drawing a heart, a club or a diamond?

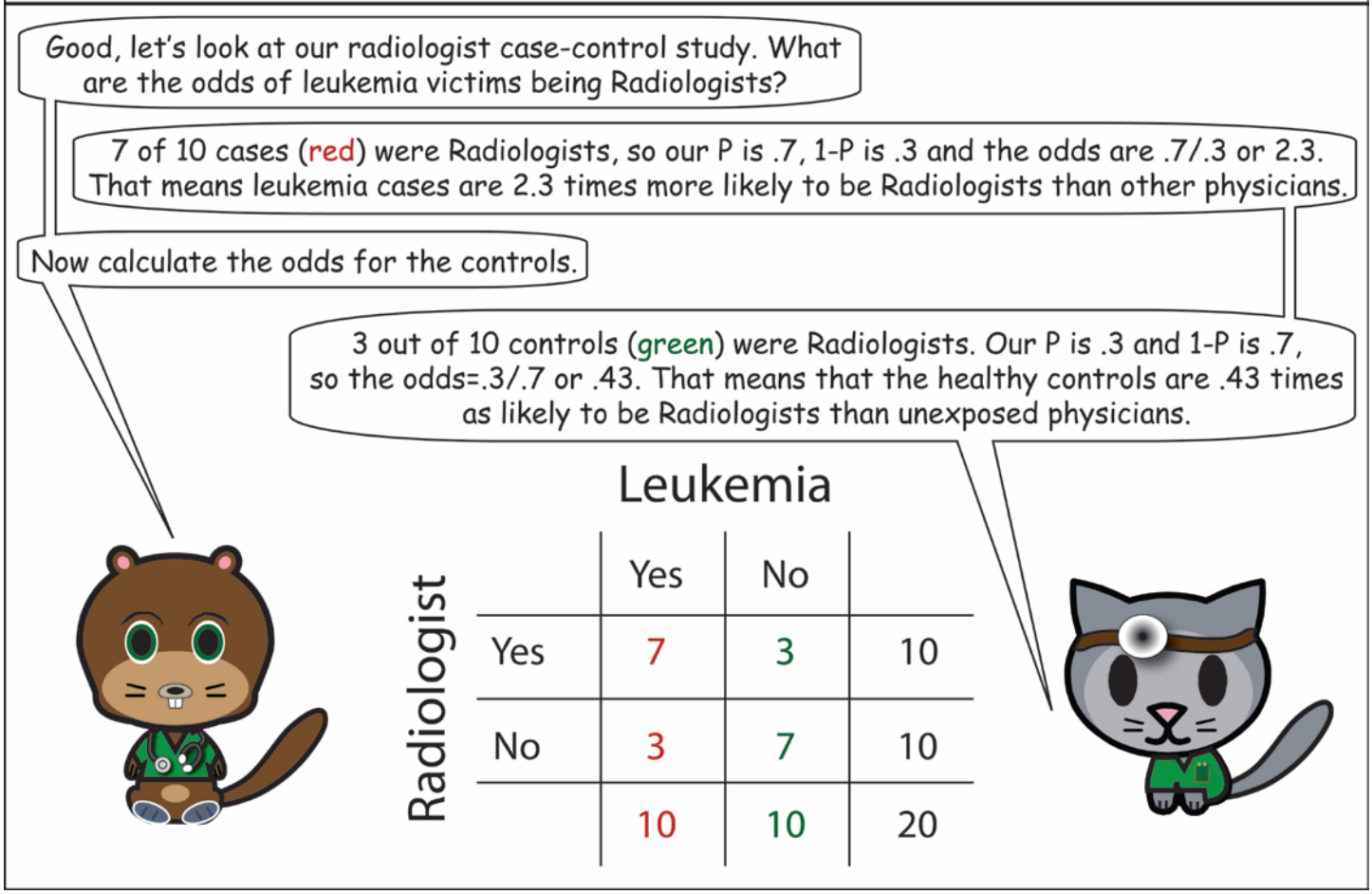
Since those 3 suits make up 75% of a deck, the probability of randomly drawing one of these suits is .75. The chance of not drawing a heart, club or diamond is .25, so our odds are .75/.25 or 3. Help! What does that mean?

An odds of 3 means that we are 3 times more likely to have our event occur than not occur. In this case we are 3 times more likely to randomly draw a heart, club or diamond as not. What would odds of one mean?

That our event is equally likely to occur as not occur.





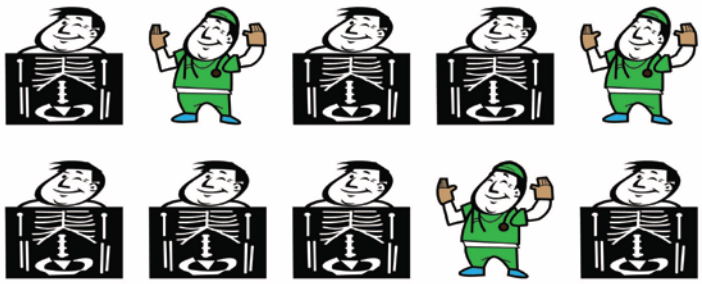


Leukemia

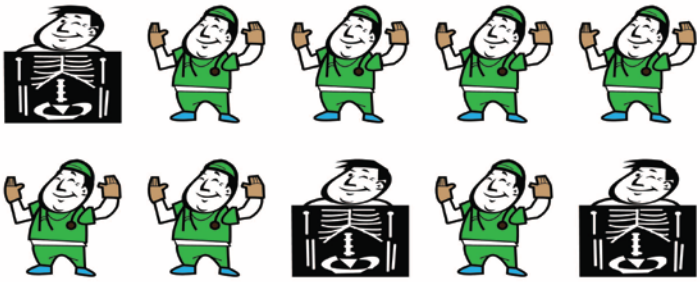
Radiologist	Yes	No	
Yes	7	3	10
No	3	7	10
	10	10	20

Our cases are more likely to be exposed Radiologists than unexposed docs. Healthy controls are less likely to be Radiologists than other physicians. We will use the control group to compare the distribution of exposure in the underlying population at risk to that of the cases.

Leukemia Cases



Healthy Controls



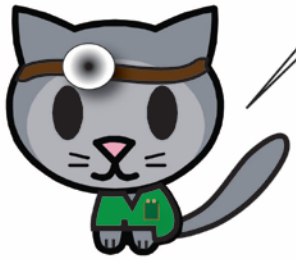
Leukemia

Radiologist	Yes	No	
Yes	7	3	10
No	3	7	10
	10	10	20

I think I get odds, what about the odds ratio?

Just divide the odds for the cases by the odds for the controls.

So odds cases/odds controls=2.3/.43=5.4. Now what does that mean?



An odds ratio is similar to a relative risk. If the odds ratio is greater than one, then there is an association between the disease and the exposure. If the odds ratio is equal to one, then there is no association. But if the odds ratio is less than one, then the exposure is protective. In this case, an OR of 5.4 means that if you have leukemia, your odds of being exposed (Radiologist) are 5.4 times greater than your odds of being unexposed. Look at the comparison of 2x2 tables for cohort and case-control studies on the next panel and the one sentence summaries of a risk ratio and an odds ratio. The structure of the summary sentences reflects the trial design; cohort studies first establish exposure status, then disease status, while case-control studies start with disease and then establish exposure status.

Cohort Study Table and Calculations

Radiologist	Leukemia: Yes	Leukemia: No	
Yes	7	30	37
No	3	70	73
	10	100	110

Incidence Exposed=7/37=.189

Incidence Unexposed=3/73=.041

Relative Risk=.189/.041=4.6

A relative risk of 4.6 means that if you are exposed, you are 4.6 times more likely to develop leukemia than if you were unexposed.

Case-Control Study Table and Calculations

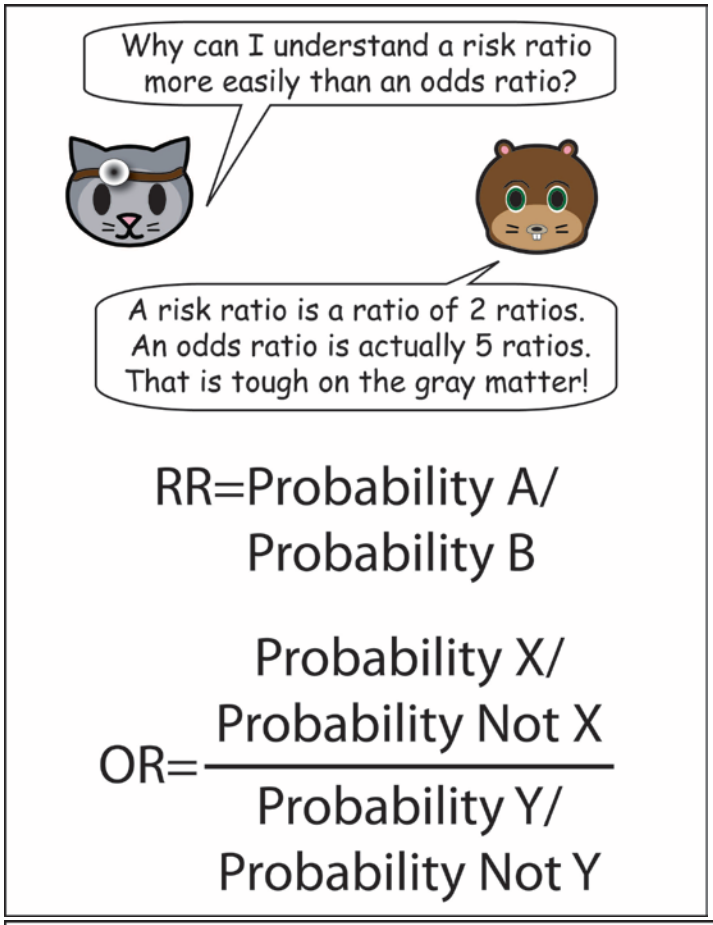
Radiologist	Leukemia: Yes	Leukemia: No	
Yes	7	3	10
No	3	7	10
	10	10	20

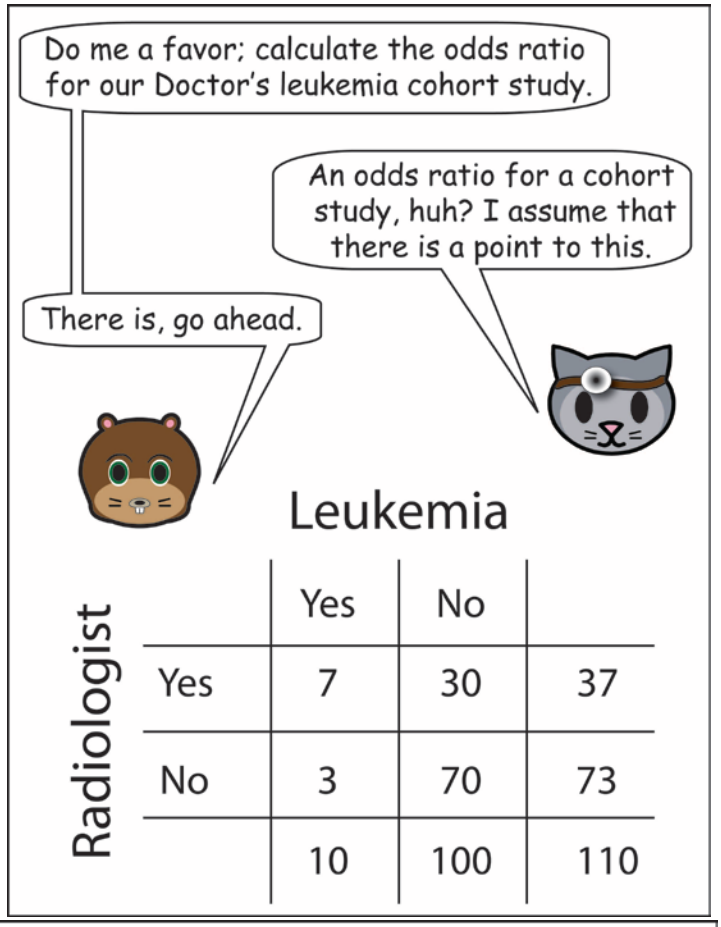
Odds Cases=.7/.3=2.3

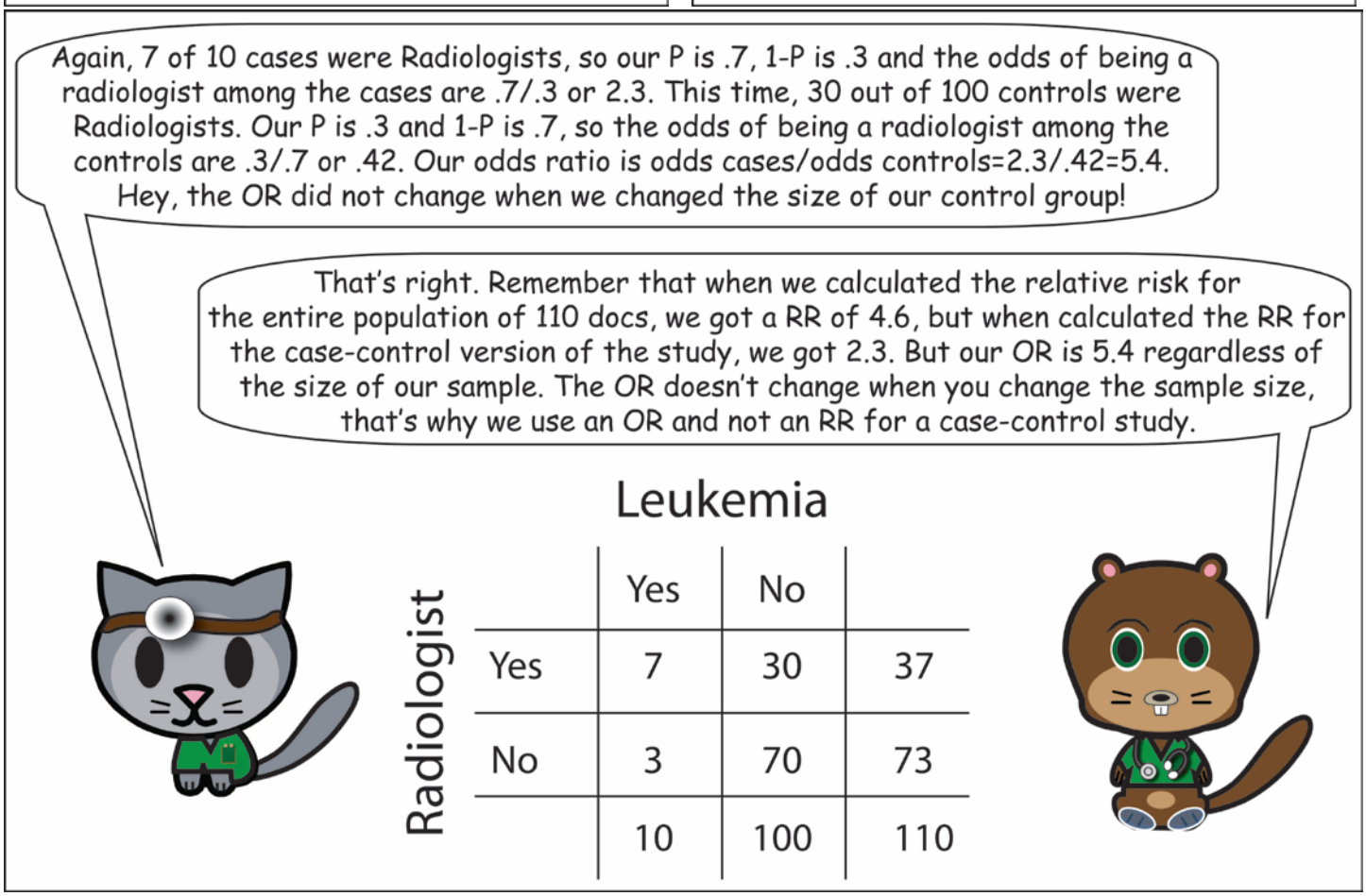
Odds Controls=.3/.7=.43

Odds Ratio=2.3/.43=5.4

An odds ratio of 5.4 means that if you have leukemia, your odds of exposure are 5.4 times greater than your odds of being unexposed.


Why can I understand a risk ratio more easily than an odds ratio?
A risk ratio is a ratio of 2 ratios. An odds ratio is actually 5 ratios. That is tough on the gray matter!
RR=Probability A/ Probability B
OR= Probability X/ Probability Not X / Probability Y/ Probability Not Y

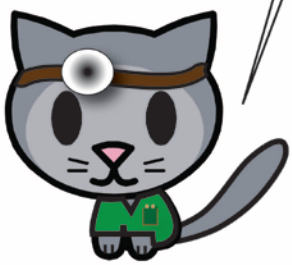

Do me a favor; calculate the odds ratio for our Doctor's leukemia cohort study.
An odds ratio for a cohort study, huh? I assume that there is a point to this.
There is, go ahead.
Leukemia
Radiologist
Yes
No
Yes 7 30 37
No 3 70 73
10 100 110


Again, 7 of 10 cases were Radiologists, so our P is .7, 1-P is .3 and the odds of being a radiologist among the cases are .7/.3 or 2.3. This time, 30 out of 100 controls were Radiologists. Our P is .3 and 1-P is .7, so the odds of being a radiologist among the controls are .3/.7 or .42. Our odds ratio is odds cases/odds controls=2.3/.42=5.4. Hey, the OR did not change when we changed the size of our control group!
That's right. Remember that when we calculated the relative risk for the entire population of 110 docs, we got a RR of 4.6, but when calculated the RR for the case-control version of the study, we got 2.3. But our OR is 5.4 regardless of the size of our sample. The OR doesn't change when you change the sample size, that's why we use an OR and not an RR for a case-control study.
Leukemia
Radiologist
Yes
No
Yes 7 30 37
No 3 70 73
10 100 110

The calculations are clumsy, there wouldn't happen to be a shortcut?

There is. If you arrange your 2x2 table as shown, the OR=ad/bc.

How about another example?



Disease

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

We'll do 2 examples, a short one and a long one. For the first one we will convert John Snow's cohort study into a case-control study. The 2x2 tables for both the cohort (on the left) and the made up case-control (on the right) studies are shown below. Go ahead and calculate the RR for the cohort study and the OR for the case-control study.

The RR for the cohort study is Risk (per 10,000) for the S&V group divided by the Risk for the Lambeth group which is 15.4/2.7=5.7. For the case-control study, the OR=ad/bc=(4093)(3946)/(461)(6054)=5.8.

Cholera Death

Water Co.	Yes	No
S & V	4,093	262,423
Lamb.	461	173,287

Cholera Death

Water Co.	Yes	No
S & V	4,093	6054
Lamb.	461	3946

The RR and the OR are nearly the same. As you can see, if the disease you are investigating is relatively uncommon, the OR provides a good estimate of the RR. The table and equations below show why the "rare disease assumption" is valid.

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

a=exposed+diseased
b=exposed+well
c=unexposed+diseased
d=unexposed+well

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

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

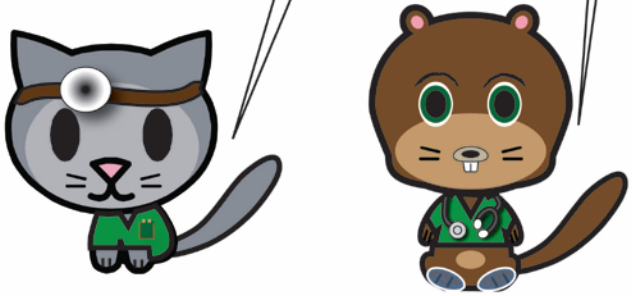
If a<<b and c<<d, then the RR and the OR are nearly equal:

$$(a/a+b)/(c/c+d)\approx(a/b)/(c/d)$$

What if the disease is common?

In that case, the OR still tells us about the strength of the association between exposure and disease. Depending on the case-control study design, the OR is a valid substitute for the risk ratio even if the disease is not rare, but we'll leave that topic for the really thick epidemiology texts.

So what is our long example?



We will look at Doll and Hill's 1950 case-control study on smoking and lung cancer. They and others had noted a nearly parallel increase in tobacco consumption and lung cancer deaths in the UK.

That ecological study suggests a link between smoking and cancer.

Yes, but some researchers claimed that the increase in lung cancer deaths was more apparent than real, with more accurate diagnosis spuriously increasing the number of lung cancer cases. Others argued that the increase was real, but due to atmospheric pollution.

Even if we did not have these other explanations for the increase in lung cancer, we cannot use ecological data to prove an association between smoking and lung cancer. After all, an ecological study cannot tell us if the smokers are the ones that developed lung cancer.

Good. That's why we must do a study that looks at the association at the individual level, to make that connection between exposure and disease. Doll and Hill chose to do a case-control study; can you guess why?

A case-control study is faster and cheaper than a cohort study.

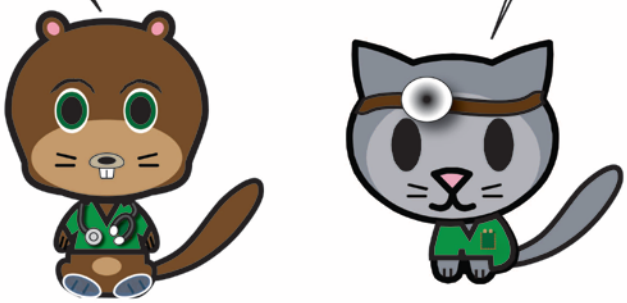
Correct. What was their first step?

Finding lung cancer cases.

Where did they get their cases?

From local hospitals.

Right again.



Twenty London hospitals were asked to notify the investigators of lung cancer diagnoses. What advantages are there in collecting your cases this way?

It's easy.

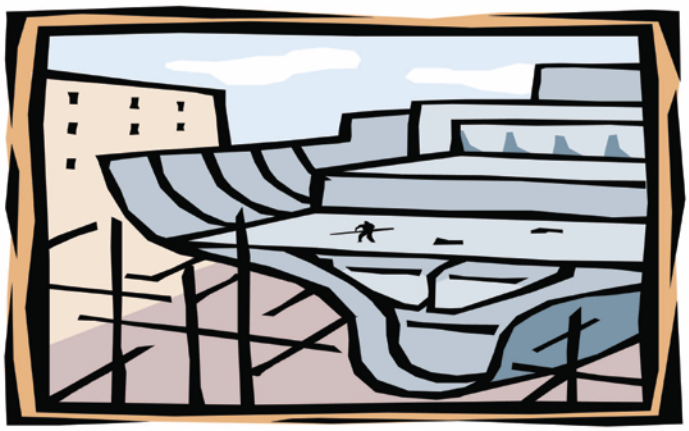
Definitely. This is called a hospital based case-control study. Are there any disadvantages of this approach?

Sure, collecting cases this way will reflect whatever biases there are in referral patterns to these hospitals.

For example...

Maybe patients referred to these hospitals are the most advanced, complex cases or are unusual in other ways. Maybe these patients tend to be the heaviest smokers. Or they have multiple other risk factors for lung cancer in addition to smoking that were unknown in 1950 like asbestos or radon exposure. These biases would result in overestimating the association between smoking and lung cancer.

Exactly. Our cases derived from high-powered London hospitals may not be representative. The alternative is to do a population based case-control study where all cases in a random sample of the population are identified. Since a population based approach is expensive, hospital based case-controls are much more common, but may be biased.

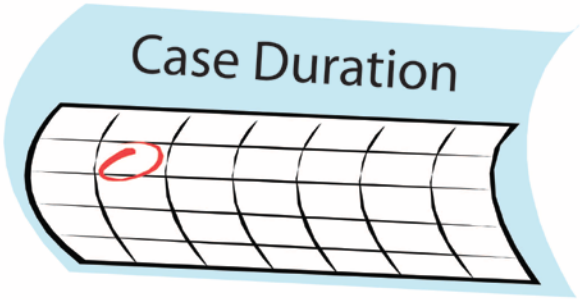


Did Doll and Hill want new (incident) cases or both old and new (prevalence) cases?

I can't tell whether their cases were new or whether they were a combination of old and new cases. What difference does it make?

Good question. When you are trying to establish a causal link between a disease and an exposure, you want your cases to be new. If your study includes prevalent cases, then you are studying factors that relate to both development and duration of disease. Of course, it is harder to prove that exposure precedes disease when you include prevalence cases. I also can't tell whether Doll and Hill studied only incident cases, but given how rapidly fatal lung cancer was in 1950, it seems reasonable to assume that most if not all of their cases were newly diagnosed. What was their next step?

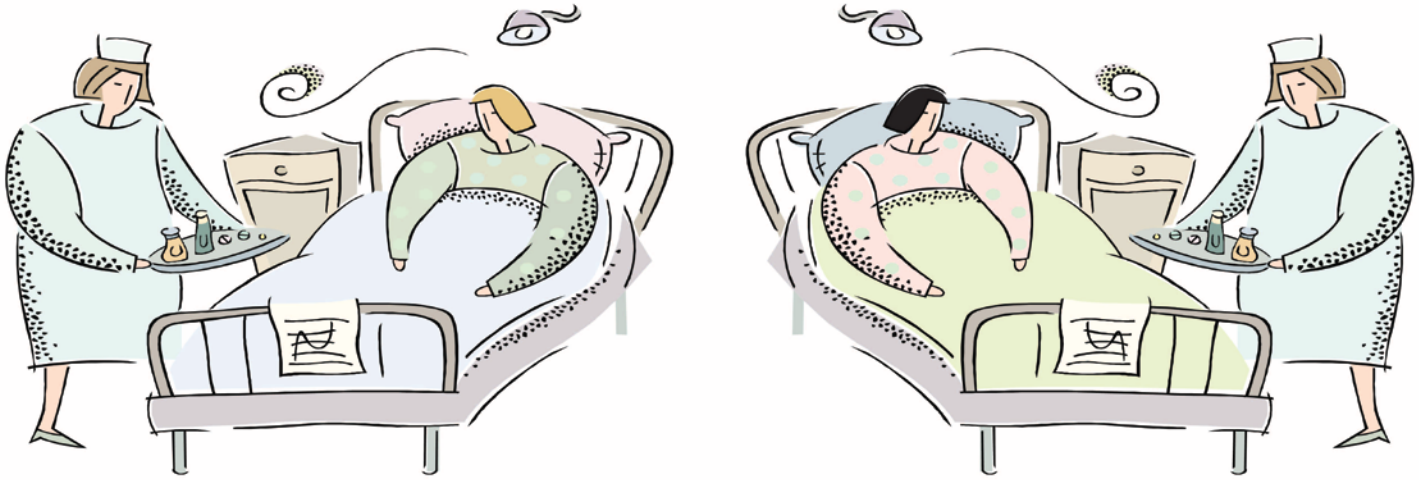
Prevalence= 

+ 

Identifying controls.

How did they do that?

According to the paper, "for each lung-carcinoma patient visited at a hospital the almoners [social workers] were instructed to interview a patient of the same sex, within the same five year age group, and in the same hospital at or about the same time." So both the cases and the controls are hospital based.

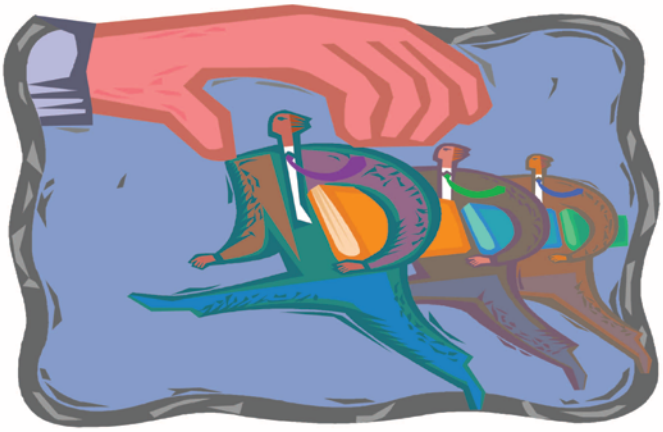


Right. Selecting controls is tricky; as a rule, you want the cases and controls drawn from the same source population. You can think of controls as people who would have been classified as cases if they had developed the disease being studied. Where else could you get your controls besides hospitals?

You could randomly select healthy people from the population.

True, but you need to keep a couple of things in mind. For example, the general population and the source population are not the same.

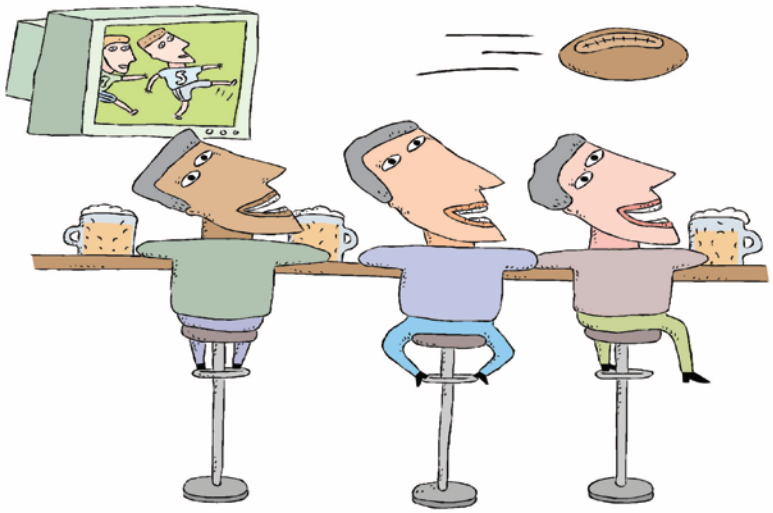
How so?



If you were studying prostate cancer, your source population consists of men over the age of 50. Including women, children and young men as controls is inappropriate since they are not at risk for prostate cancer and their exposure status tells you nothing about the relationship of the exposure and the disease. The source population for a case-control population is similar to the population at risk that you evaluate when doing a cohort study.

General Population

Prostate Population



I'll bet putting together a population based control group is expensive.

It is; that's why many case- control studies use hospital based controls like Doll and Hill did. In the 1950 smoking case-control study, there were plenty of willing control subjects in the next ward.

Of course, there must be a down side.

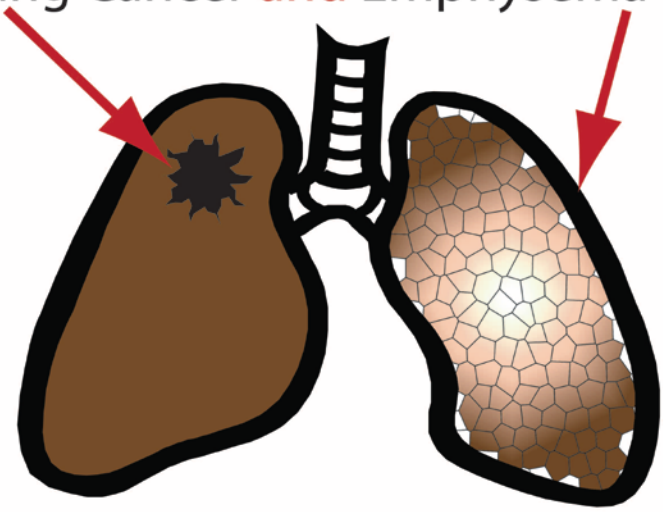
Yes. If the exposure of interest played a role in the admission of controls to the hospital, you may underestimate the association between the exposure and the disease.

How about an example?

Sure. We'll pretend that the controls are drawn from the emphysema ward down the hall and that we end up with the 2x2 table shown. Calculate the OR.

The OR=ad/bc=(9)(1)/(9)(1)=1. If the controls are admitted for another smoking related illness, then we may miss an association between smoking and lung cancer.

Smoking → Causes Lung Cancer and Emphysema



Lung Cancer

Smoker	Yes	No	
Yes	9	9	10
No	1	1	10
	10	10	20

Not surprisingly, hospital patients are sick and have higher rates of smoking and alcohol use which may contribute to their illness. As we have seen, this may bias an OR toward the null hypothesis. How comparable were Doll and Hills' cases and controls?

They provided a table for comparison. The 2 groups were similar in age distribution, gender and social class but differed slightly with respect to place of residence; although most cases and controls were from the greater London area, more cases (176/709) came from outside London compared to controls (101/709). Doll and Hill believed that "[t]he difference can be explained on the grounds that people with cancer came to London from other parts of the country for treatment at special centers."

This difference was probably trivial, but there is a technique called matching that can be used to insure that the 2 groups are similar with respect to observable characteristics.

How does matching work?

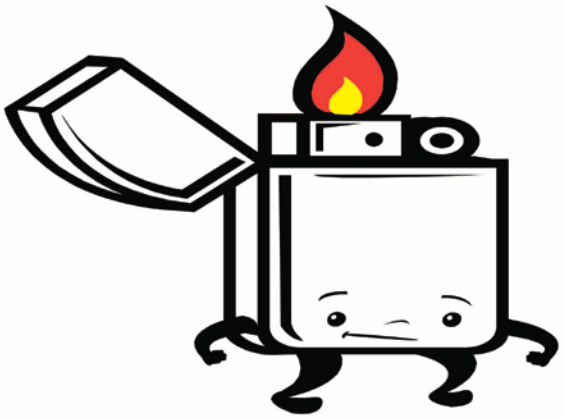
Basically, you select those characteristics that you want to control for. Each case is then matched with one or more controls that have those characteristics.

Sounds simple.

It is, but it can still be problematic. Since it involves extra work on the part of the investigators, matching is expensive. You also have to be careful not to overmatch.

What is overmatching?

The best way to explain is to hop right into an example. What if Doll and Hill had noticed that lung cancer patients tended to be older men who owned lighters? Should they have found matching controls who also owned lighters?

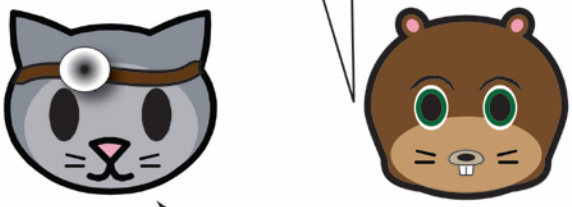


Matching cases and controls on the basis of lighter ownership is almost the same as matching for smoking exposure since lighters are associated with smoking.

Right. Remember, the controls are used to estimate the proportion of the source population that is exposed. If you find matching controls using a characteristic that is associated with the exposure, you will overestimate the amount of exposure in the underlying population and your study will be biased toward the null hypothesis.

I can see how this can get very complex.

Yes, it can. Even defining disease and exposure is tricky. You would think that lung cancer would be easy to define, but in 220/709 of Doll and Hill's lung cancer cases, there was no pathologic proof. What if most of these cases were not cancer but some inflammatory process?



You could end up showing an association between smoking and cancer that really isn't there. How did they define exposure?

I will let the authors speak for themselves: "Did the term [smoker], for example, include the woman who took one cigarette annually after her Christmas dinner, or the man of 50 who as a youth smoked a couple of cigarettes to see whether he liked it and decided he did not? If so, it is doubtful whether anyone at all could be described as a nonsmoker. A smoker was therefore defined in this inquiry as a person who had smoked as much as one cigarette a day for as long as one year, and any less consistent amount was ignored."

Charming and straightforward.

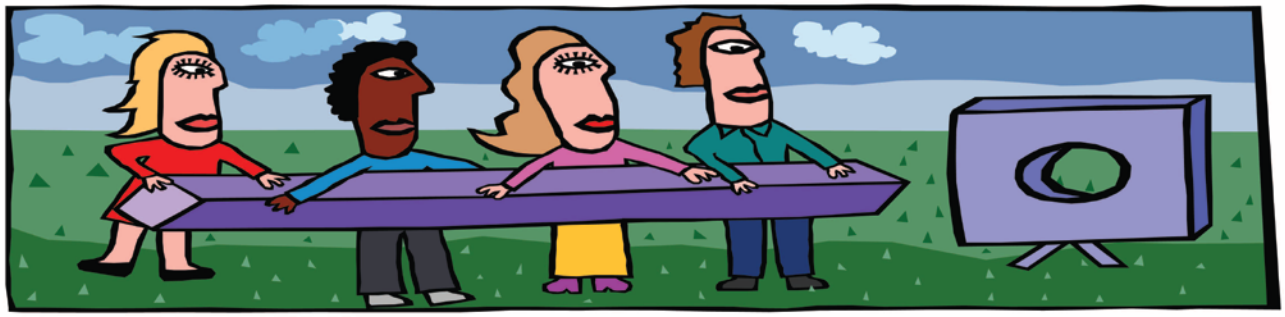
Quite, but there is still the possibility that Doll and Hill overestimated the amount of smoking among the cases because of recall bias. Perhaps cases were more likely to report a smoking history because they were looking for an explanation for their cancer, but "[m]ost of these patients cannot have known that they were suffering from cancer".

Wow, medicine must have been much more paternal in 1950; I cannot imagine withholding a cancer diagnosis.

Even though the patients were not told that they had cancer, in the vast majority of cases, the interviewers knew the diagnosis. Doll and Hill were aware of the potential for bias "by the interviewers tending to scale up the smoking habits of the lung-carcinoma cases".

How did they exclude that possibility?

There were some patients interviewed who were incorrectly thought to have lung cancer but were later shown to be cancer free. "[T]he smoking habits of the patients who were incorrectly thought to have carcinoma of the lung at the time of interview are sharply distinguished from the habits of those patients who did in fact have carcinoma of the lung, but they do not differ significantly from the habits of the other patients interviewed. "So recall bias among cases and misclassification by the interviewers is unlikely.



What was the odds ratio that Doll and Hill got?

They never calculated one! We will have to do it for them; in fact we will have to create their 2x2 table for them since they presented their findings stratified by sex. What do you think?

Virtually every single one of their cases smoked. The vast majority of their controls also smoked. The OR is 2.97; if you are a case, your odds of smoking are almost 3 times your odds of not smoking.

From our perspective, the proportion of smokers in the mid 20th century is astonishing. What would you say about this data if you were a tobacco company scientist?

Lung Cancer

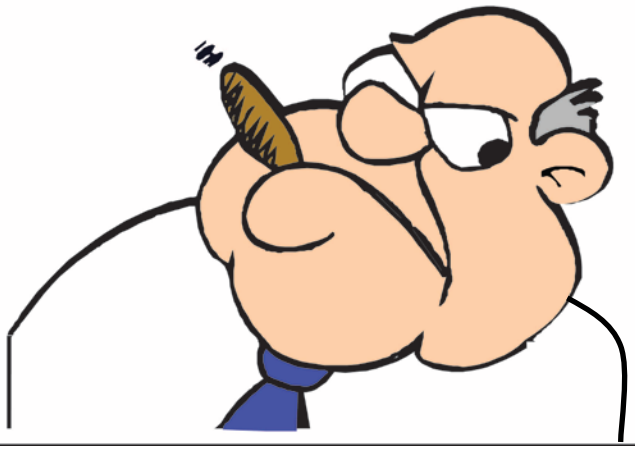
Smoker	Yes	No	
Yes	688	650	1338
No	21	59	80
	709	709	1418

I would argue that the small difference in smoking among the male cases and controls could easily be due to biases like misclassification of the exposure. In addition, most of the control non-smokers are women, who only make up 8.5% of the study subjects. I would point out that among the women 32% of lung cancer cases occurred in non-smokers. In men, only 4.2% of cancers occurred in non-smokers. Finally, I would argue that you can never use a case-control design to prove that the exposure preceded the disease.

The tobacco companies spent decades trying to cast doubt on the link between smoking and lung cancer. Lung cancer was bad for business. What might explain the relatively high rate of lung cancer among non-smoking women?

Since virtually every man smoked, these women were exposed to prodigious amounts of second hand smoke.

That's what I suspect, but I can't prove it. Now that you are an expert in observational studies, we will learn about clinical trials.



References, Acknowledgements etc.

Many of the illustrations are modified clipart from Microsoft (Redmond, Washington) Office except "Doc" Squirrel is an original creation. All artwork was created or modified using Adobe Illustrator CS4 and/or Photoshop CS4 (San Jose, California).

Look for a chapter on randomized trials coming soon!