Preface

Four years ago (2002), I (DGK) authored a unique educational program, ***ActivEpi*** (Springer Publishers), developed in CD-ROM format to provide a multimedia interactive "electronic textbook" on basic principles and methods of epidemiology. In 2003, the ***ActivEpi*** **Companion Text,** authored by myself (DGK), KM Sullivan and ND Barker and also published by Springer**,** was developed to provide a hard-copy of the material contained in the ***ActivEpi*** CD-ROM. The CD-ROM contains 15 chapters, with each consisting of a collection of "activities" including narrated expositions, interactive study questions, quizzes, homework questions, and web links to relevant references on the Internet.

In the nearly three years since the publication of the ***ActivEpi*** CD-ROM, we have received several suggestions from instructors of introductory epidemiology courses as well as health and medical professionals to produce an abbreviated version that narrows the discussion to the most "essential" principles and methods. Instructors expressed to us their concern that the material covered by the CD-ROM (and likewise, the Companion Text) was too comprehensive to conveniently fit the amount of time available in an introductory course. Professionals expressed their desire for a more economically time-consuming version that would conveniently fit their "after hours" availability.

To address these suggestions, we have herewith produced **A Pocket Guide to Epidemiology** which provides a much shorter, more "essential" version of the material covered by the ***ActivEpi*** CD-ROM and Companion Text. We realize that determining what is "essential" is not a simple task, especially since, from our point of view, the original CD-ROM was already restricted to "essential" topics. Nevertheless, to produce this text, we decided to remove from the original material a great many fine points of explanation and complicated topics/issues about epidemiologic principles and methods, with our primary goal a "quicker read".

A Pocket Guide to Epidemiology contains less than half as many pages as the ***ActivEpi*** Companion Text. We have continued to include in **A Pocket Guide to Epidemiology** many of the study questions and quizzes that are provided in each Lesson of the CD ROM, but we have eliminated homework exercises, computer exercises, and Internet linkages from the original CD-ROM. Nevertheless, we indicate throughout **A Pocket Guide to Epidemiology** how and where the interested reader can turn to the ***ActivEpi*** CD ROM (or the Companion Text) to pursue more detailed information.

We authors view **A Pocket Guide to Epidemiology** as a stand-alone introductory text on the basic principles and concepts of epidemiology. Our primary audience for this text is the public health student or professional, clinician, health journalist, and anyone else at any age or life experience that is interested in learning what epidemiology is all about in a convenient, easy to understand format with timely, real-world health examples. We believe that the reader of this text will also benefit from using the multi-media learner-interactive features of the ***ActivEpi*** CD ROM electronic textbook to further clarify and enhance what is covered in this more abbreviated (non-electronic) text. Nevertheless, we suggest that, on its own, **A Pocket Guide to Epidemiology** will provide the interested reader with a comfortable, time-efficient and enjoyable introduction to epidemiology.

CHAPTER 5

WHAT'S THE ANSWER? MEASURES OF EFFECT

*In epidemiologic studies, we compare disease frequencies of two or more groups using a **measure of effect**. We will describe several types of measures of effect in this chapter. The choice of measure typically depends on the study design being used.*

Ratio Versus Difference Measures of Effect

Our focus in Chapter 5 is on ratio measures of effect, which are of the form M_1/M_0, where M_1 and M_0 are two measures of disease frequency, e.g., risks, rates, or prevalences that are being compared.

We consider difference measures of effect, which are of the form M_1-M_0, in Chapter 6 on "Measures of Potential Impact". Difference measures are also called measures of attributable risk.

Ratio measures are typically used in epidemiologic studies that address the etiology of a disease/health outcome, whereas difference measures are used to quantify the public health importance of factors that are determinants of a disease/health outcome.

Smoking and Lung Cancer

Cigarette smoking became increasingly popular in America after World War I when cigarettes were handed out to soldiers as a way to boost morale. But along with the rise in smoking, came a disturbing rise in the lung cancer rate and some early warnings from a handful of doctors about possible dangers of smoking. Early studies in the 1930s and 1940s of the possible relationship between smoking and lung cancer were **case-control studies**. It became quite apparent that lung cancer patients smoked much more than controls. In one study in particular, lung cancer patients were 17 times more likely than controls to be two-pack-a-day smokers.

In the early 1950s, doctors Horn and Hammond of the American Cancer Society conducted one of the first **cohort studies** on the harmful effects of smoking. About 200,000 people were given a smoking questionnaire and then followed for four years. Death rates and cause of death for smokers and for non-smokers were compared. The preliminary study published in 1958 caused quite a sensation. It was the largest study on smoking that had been done, and it showed that smokers were ten times more likely than nonsmokers to get lung cancer.

Both the cohort and case-control studies attempted to assess the proposed relationship between smoking and lung cancer by deriving a measure of effect that quantified the extent of this relationship. The measure described in the **case-control study** is called an **odds ratio.** The measure described in the **cohort study** is called a **risk ratio**. The activities that follow discuss these two fundamental measures of effect.

Summary

- The odds ratio and the risk ratio are two fundamental measures of effect.
- These measures were used in epidemiologic studies of the relationship between smoking and lung cancer.

- ❖ The odds ratio is typically the measure of effect used in case-control studies.
- ❖ The risk ratio is typically the measure of effect used in cohort studies.

The Risk Ratio

The table below summarizes the results of a five-year follow-up study to determine whether or not smokers who have had a heart attack will reduce their risk for dying by quitting smoking. A cohort of 156 heart attack patients was studied, all of whom were regular smokers up to the time of their heart attack. Seventy-five of these patients continued to smoke after their attack. The other 81 patients quit smoking during their recovery period. Of the 75 patients that continued smoking, 27 died, so the proportion of these patients that died is 0.36. Of the 81 patients who quit smoking, 14 died, so the corresponding proportion is 0.17. These proportions estimate the five-year risks of dying for these two groups of patients. We may wonder whether those heart attack patients who continue smoking are more likely to die within 5 years after their first heart attack than those who quit.

Heart Attack Patients

	Smoke	Quit	Total
Death	27	14	41
Survival	48	67	115
Total	75	81	156

5-year risks of dying

continuing smokers: $27/75 = 0.36$

smokers who quit: $14/81 = 0.17$

A measure of effect gives a numerical answer to this question. Such a measure allows us to make a comparison of two or more groups, in this case, continuing smokers and smokers who quit. For follow-up studies such as described here, the typical measure of effect is a **risk ratio**. To calculate a risk ratio, we take the ratio of the two risks being compared, that is, we simply divide one risk by the other. Actually, we are getting an "estimate" of the risk ratio, which we indicate by putting a "hat" symbol over the RR notation. $\hat{\text{RR}}$ is an estimate because we are using two estimates of risk based on samples from the two groups being compared. In our example, therefore, we divide 0.36 by 0.17 to get 2.1.

$$\text{Estimated } \hat{\text{RR}} = \frac{\text{Estimated Risk for continuing smokers}}{\text{Estimated Risk for smokers who quit}} = \frac{0.36}{0.17} = 2.1$$

The estimated risk ratio of 2.1 tells us that continuing smokers are about twice as likely to die as smokers who quit. In other words, for heart attack patients the five-year risk for continuing smokers is about twice the corresponding risk for smokers who quit.

Study Questions (Q5.1) Using the five-year follow-up study comparing mortality between smokers and quitters example:

1. How would you interpret a Risk Ratio of 4.5?
2. What if the Risk Ratio was 1.1?
3. How about if the Risk Ratio was less than 1, say 0.5?
4. How would you interpret a value of 0.25?

If our estimated risk ratio had been 1.1, we would have evidence that the risk for continuing smokers was essentially equal to the risk for smokers who quit. We

call a risk ratio of 1 the **null value** of the risk ratio. This is the value that we get for the risk ratio when there is no effect, that is, the effect is null.

Summary

- The risk ratio (RR) is the ratio of the risk for one group, say group 1, to the risk for another group, say group 0.
- The value of RR can be greater than one, equal to one, or less than one.
- If the RR is greater than one, the risk for group 1 is larger than the risk for group 0.
- If the RR is below one, the risk for group 1 is less than the risk for group 0.
- And, if the RR is equal to 1, the risks for group 1 and 0 are equal, so that there is no effect of being in one group when compared to the other.

Risk Ratio Numerator and Denominator

In general, the risk ratio that compares two groups is defined to be the risk for one group divided by the risk for the other group. It is important to clearly specify which group is in the numerator and which group is in the denominator.

If, for example, the two groups are labeled group 1 and group 0, and the risk for group 1 is in the numerator, then we say that the risk ratio compares group 1 to group 0. On the other hand, if the risk for group 0 is in the numerator, then we say that the risk ratio compares group 0 to group 1.

Quiz (Q5.2) For heart attack patients, the risk ratio is defined to be the risk for continuing smokers divided by the risk for smokers who quit. For the following scenarios what would be the risk ratio?

1. Continuing smokers are twice as likely to die as smokers who quit. **???**
2. Continuing smokers are just as likely to die as smokers who quit. **???**
3. Smokers who quit are twice as likely to die as continuing smokers. **???**

Choices **0** **0.1** **0.2** **0.5** **1** **2**

Let's consider the data from a randomized clinical trial to assess whether or not taking aspirin reduces the risk for heart disease. The exposed group received aspirin every other day whereas the comparison group received a placebo. A table of the results is shown below.

		Aspirin		Placebo		Total
		n	Column %	n	Column %	
Developed Heart Disease	**Yes**	104	(1.04)	189	(2.36)	293
	No	9,896	(98.96)	7,811	(97.64)	17,707
	Total	10,000	(100.00)	8,000	(100.00)	18,000

4. The estimated risk for the aspirin group is **???**
5. The estimated risk for the placebo group is **???**
6. The estimated risk ratio that compares the aspirin group to the placebo group is given by **???**

Choices **0.0104** **0.0236** **0.44** **104/189** **2.269** **98.96/97.64**

The Odds Ratio

Epidemiologists in the Division of Bacterial Diseases at CDC, the Centers for Disease Control and Prevention in Atlanta, investigate the sources of outbreaks caused by eating contaminated foods. For example, a case-control study was carried out to determine the source of an outbreak of diarrheal disease at a Haitian Resort Club from November 30 to December 8, 1984.

The investigators wondered whether eating raw hamburger was a primary source of the outbreak. Because this is a **case-control study** rather than a follow-up study, the study design starts with **cases**, here, persons at the resort who had diarrhea during the time period of interest. The **controls** were a random sample of 33 persons who stayed at the resort but did not get diarrhea during the same time period. There were a total of 37 cases during the study period. All 37 cases and the 33 controls were interviewed by a team of investigators as to what foods they ate during their stay at the resort.

	Raw Hamburger		
	Ate	Did not eat	Total
Cases	17	20	37
Controls	7	26	33
Total	24	46	70

Proportion of cases: $17/37 = 0.46$
Proportions of controls: $7/33 = 0.21$

Of the 37 cases, 17 persons ate raw hamburger, so that the proportion of the cases that ate raw hamburger is 0.46. Of the 33 controls, 7 ate raw hamburger, so the corresponding proportion is 0.21. We may wonder, then, whether these data suggest that eating raw hamburger was the source of the outbreak.

Because this is a case-control study rather than a follow-up study, these proportions do not estimate risks for cases and controls. Therefore, we can***not*** compute a risk ratio. So, then, what measure of effect should be used in case-control studies? The answer is the **odds ratio** (**OR),** which is described in the next section.

Summary

- A case-control study was used to investigate a foodborne outbreak at a Caribbean resort.
- In a case-control study, we cannot estimate risks for cases and controls.
- Consequently, we cannot use the risk ratio (RR) as a measure of effect, but must use the odds ratio (OR) instead.

Why can't we use a risk ratio in case-control studies?

In a case-control study, we cannot estimate risk, but rather, we estimate **exposure probabilities** for cases and controls. The exposure probability for a case is the probability that a subject is exposed given that he/she is a case; this is not equivalent to the probability that a subject is a case given that he/she is exposed, which is the risk for exposed.

In other words, using conditional probability notation:

Pr(exposed | case) ≠ Pr(case | **exposed**), where "|" denotes "given".

Similarly the exposure probability for a control is not equivalent to 1 minus the risk for exposed. That is,

Pr(**exposed** | control) ≠ 1 - Pr(case | **exposed**).

The ratio of two exposure probabilities is, unfortunately, not a risk ratio. Therefore, in case-control studies we must use a different measure of effect, namely the odds ratio.

The Odds Ratio (continued)

To understand odds ratios, we must start with the concept of an **odds**. The term **odds** is commonly used in sporting events. We may read that the odds are 3 to 1 against a particular horse winning a race, or that the odds are 20 to 1 against Spain winning the next World Cup, or that the odds are 1 to 2 that the New York Yankees will reach the World Series this year. When we say that the odds against a given horse are 3 to 1, what we mean is that the horse is 3 times more likely to lose than to win.

The odds of an event are easily calculated from its probability of occurrence. The odds can be expressed as **P**, the probability that the event will occur, divided by 1 - P, the probability that the event will not occur.

$$\text{Odds} = \frac{\mathbf{P}}{\mathbf{1 - P}} = \frac{\text{P(Event will occur)}}{\text{P(Event will not occur)}}$$

In our horse race example, if P denotes the probability that the horse will lose, then 1 - P denotes the opposite probability that the horse will win. So, if the probability that the horse will lose is 0.75, then the probability that the horse will win is 0.25, and the odds are 3, or 3 to 1.

$$\text{Odds} = \frac{\text{P}}{\text{1 - P}} = \frac{\text{P(horse will lose)}}{\text{P(horse will win)}} = \frac{0.75}{0.25} = 3 \text{ or } \frac{3}{1}$$

In the Haitian resort case-control study, recall that the event of interest occurs if a study subject ate raw hamburger, and, if so, we say this subject is **exposed**. The estimated probability of exposure for the cases was 0.46, so the estimated **odds of being exposed for cases** is 0.46 divided by 1 - 0.46:

$$\hat{\text{Odds}}_{\text{Cases}} = \frac{0.46}{1 - 0.46} = .85$$

Similarly, the estimated probability of exposure for controls was 0.21, so the estimated odds for controls is 0.21 divided by 1 - 0.21:

$$\hat{\text{Odds}}_{\text{Controls}} = \frac{0.21}{1 - 0.21} = .27$$

The **estimated odds ratio** for these data is the ratio of the odds for cases divided by the odds for controls, which equals 3.2.

$$\text{Odds Ratio } (\hat{\text{OR}}) = \frac{\hat{\text{Odds}}_{\text{Cases}}}{\hat{\text{Odds}}_{\text{Controls}}} = \frac{.85}{.27} = 3.2$$

How do we interpret this odds ratio estimate? One interpretation is that the **exposure odds for cases** is about 3.2 times the **exposure odds for controls**. Since those who ate raw hamburger are the exposed subjects, the odds that a case ate raw hamburger appear to be about 3.2 times the odds that a control subject ate raw hamburger.

Study Questions (Q5.3)

Using the Haiti case-control study example:

1. How would you interpret an odds ratio of 2.5?
2. What if the odds ratio was 1.1?

3. How about if the odds ratio less than 1, say 0.5?
4. How would you interpret a value of 0.25?

Odds ratios, like risk ratios, can be greater than one, equal to one, or less than one. An odds ratio greater than one says that the exposure odds for cases is **larger** than the exposure odds for controls. An odds ratio below one says that the exposure odds for cases is **less** than the exposure odds for controls. An odds ratio equal to 1 says that the exposure odds for cases and controls are equal.

Summary

- The odds of an event can be calculated as P/(1-P) where P is the probability of the event.
- The odds ratio (OR) is the ratio of two odds.
- In case-control studies, the OR is given by the exposure odds for the cases divided by the exposure odds for controls.
- Odds ratios, like risk ratios, can be greater than 1, equal to 1, or less than 1, where 1 is the null value.

Quiz (Q5.4) A causal relationship between cigarette smoking and lung cancer was first suspected in the 1920s on the basis of clinical observations. To test this apparent association, numerous studies were conducted between 1930 and 1960. A classic **case-control study** was done in 1947 to compare the smoking habits of lung cancer patients with the smoking habits of other patients.

1. In this case-control study, it is **???** to calculate the risk of lung cancer among smokers, and thus, the appropriate measure of association is the **???**.

Choices **Not possible** **odds ratio** **possible** **risk ratio**

Let's consider the data below from this classic case-control study to assess the relationship between smoking and lung cancer. Cases were hospitalized patients newly diagnosed with lung cancer. Controls were patients with other disorders. This 2 x 2 table compares smoking habits for the male cases and controls.

2. The probability of being a smoker among cases is **???**
3. The probability of being a smoker among controls is **???**
4. The odds of smoking among cases is **???**
5. The odds of smoking among controls is **???**
6. The odds ratio is **???**

Choices **0.11** **1.04** **10.50** **1296/1357** **1350/1357** **1350/2646** **192.86** **21.25** **7/68** **9.08**

	Cigarette Smoker	Non-Smoker	Total
Cases	1350	7	1357
Controls	1296	61	1357
Total	2646	68	2714

In a case-control study to find the source of an outbreak, the odds ratio for eating coleslaw is defined to be the odds for cases divided by the odds for controls. For

the following scenarios what would be the odds ratio?

7. Cases have an odds for eating coleslaw three times higher than controls **???**
8. Cases have the same odds for eating coleslaw as controls **???**
9. Controls have three times the odds for eating coleslaw as cases **???**

Choices **0** **0.25** **0.333** **1** **3** **4**

Calculating the Odds Ratio

This layout for a two by two table provides a more convenient way to calculate the odds ratio. The formula is ***a*** times ***d*** over ***b*** times ***c***. It is called the **cross product ratio** formula because it is the ratio of one product that crosses the table divided by the other product that crosses the table.

General Form of the 2 by 2 Table

	Exposure Status: Yes	No	Total
Cases	a	b	m_1
Controls	c	d	m_0
Total	n_1	n_0	m

Cross Product Ratio

$$\hat{OR} = \frac{a \times d}{b \times c}$$

To illustrate this formula consider the data from the Haitian resort outbreak. The cross product formula gives us the same result, 3.2, as we obtained originally from the ratio of exposure odds for cases and controls.

Case-control Study: Outbreak of Diarrheal Disease at a Haitian Resort Club

Raw Hamburger	Yes	No	Total
Cases	a=17	b=20	m_1=37
Controls	c=7	d=26	m_0=33
Total	n_1=24	n_0=46	m=70

Cross Product Ratio

$$\hat{OR} = \frac{a \times d}{b \times c} = \frac{(17)(26)}{(20)(7)} = 3.2$$

$$= \frac{\hat{Odds}_{Cases}}{\hat{Odds}_{Controls}}$$

Study Question (Q5.5)

1. Should we calculate the OR for other foods eaten during the outbreak before we blame raw hamburger as the source?

Although the odds ratio must be computed in case-control studies for which the risk ratio cannot be estimated, the odds ratio can also be computed in follow-up studies. (Note that the OR and RR can also be calculated in randomized clinical trials that have cumulative incidence measures.)

	OR	RR
Case-Control Studies	☑	☒
Follow-up (Cohort)	☑	☑

For example, let us consider the "quit smoking" study for heart attack patients. The study design here is a follow-up study. We previously estimated that the risk for patients who continued to smoke was 2.1 times greater than the risk for those who quit.

Using the cross product formula on these follow-up data

Follow-up (Cohort) Study

	Smoke	Quit	Total
Death	27	14	41
Survival	48	67	115
Total	75	81	156

$$\hat{RR} = \frac{\hat{\text{Risk for continuing smokers}}}{\hat{\text{Risk for smokers who quit}}} = \frac{27/75}{14/81} = \frac{0.36}{0.17} = 2.1$$

$$\hat{OR} = \frac{a \times d}{b \times c} = \frac{(27)(67)}{(14)(48)} = 2.7$$

yields 2.7. The fact that these two numbers (the risk ratio and odds ratio) are not equal should not be surprising, since the risk ratio and odds ratio are two different measures. But the values in this example are not very different. In fact, these two estimates have similar interpretations since they both suggest that there is a moderate relationship between quit smoking status and survival status.

Summary

- A convenient formula for the OR is the cross product ratio: (ad)/(bc)
- The OR can be estimated in both case-control and follow-up studies using the cross-product formula.

(See below for discussion of how the risk ratio can be approximated by the odds ratio.)

Quiz (Q5.6) To study the relationship between oral contraceptive use and ovarian cancer, CDC initiated the Cancer and Steroid Hormone Study in 1980 (see table below). It was a case-control study.

1. Using the cross product ratio formula, the OR comparing the exposure status of cases versus controls is (93) * (**???**) / (**???**) * (959) which equals **???**.
2. This means that the **???** of **???** among the cases was **???** the **???** of exposure among the **???**.

Choices **0.23** **0.77** **1.3** **683** **86** **cases** **controls** **disease** **exposed** **exposure** **greater than** **less than** **non-exposed** **odds** **risk**

	Ever Used OCs	Never Used OCs	Total
Cases	93	86	179
Controls	959	683	1642
Total	1052	769	1821

The Odds Ratio in Different Study Designs

The odds ratio can be computed for both case-control and follow-up (cohort) studies. Because a case-control study requires us to estimate exposure probabilities rather than risks, we often call the odds ratio computed in case-control studies the **exposure odds ratio** (**EOR**). In contrast, because a follow-up study allows us to estimate risks, we often call the odds ratio computed from follow-up studies the **risk odds ratio** (**ROR**).

Odds Ratio

Case-control studies (exposure probabilities):

Exposure odds ratio (EOR)

Follow-up (Cohort) studies (risks):

Risk odds ratio (ROR)

Cross-sectional studies (prevalences):

Prevalence odds ratio (POR)

The odds ratio can also be computed for cross-sectional studies. Since a cross-sectional study measures prevalence or existing conditions at a point in time, we usually call an odds ratio computed from a cross-sectional study a **prevalence odds ratio** (**POR**).

As an example of the computation of a prevalence odds ratio for cross-sectional data, consider these data that were collected from a cross-sectional

survey designed to assess the relationship between coronary heart disease and various risk factors, one of which was personality type. For these cross-sectional data, we can use the general cross product ratio formula to compute a prevalence odds ratio. The odds of having a type A personality among those with coronary heart disease is 5 times the odds of those without the disease.

Cross-sectional data (POR):

		Personality Type A	Personality Type B	Total
CHD	Yes	93	36	129
	No	46	89	135
	Total	139	125	264

Prevalence odds ratio (POR) = ad/bc

$$\hat{POR} = \frac{(93)(89)}{(46)(36)} = 5.0$$

Odds of Type A_{CHD} = 5x Odds of Type $A_{NO\text{-}CHD}$

In general we can use the cross product ratio formula to compute an exposure odds ratio, a risk odds ratio, or a prevalence odds ratio depending on the study design used.

Summary

- The OR computed from a case-control study is called the exposure odds ratio (EOR).
- The OR computed from a follow-up study is called the risk odds ratio (ROR)
- The OR computed from a cross-sectional study is called the prevalence odds ratio (POR)
- We can use the general cross-product ratio formula to calculate the EOR, ROR, or POR depending on the study design used.

Does ROR = EOR = POR?

Not necessarily. Although the calculation formula (i.e., ad/bc) is the same regardless of the study design, different values of the estimated odds ratio from a 2 x 2 table might be obtained for different study designs. This is because of the possibility of selection bias (described in Chapter 8). For example, a case-control study that uses prevalent cases could yield a different odds ratio estimate than a follow-up study involving only incident cases.

Quiz (Q5.7) Data is shown below for a cross-sectional study to assess whether maternal cigarette smoking is a risk factor for low birth weight.

1. Calculate the odds ratio that measures whether smokers are more likely than non-smokers to deliver low birth weight babies. OR=**???**
2. This odds ratio estimate suggests that smokers are **???** than non-smokers to have low birth weight babies.
3. This odds ratio is an example of a(n) **???** odds ratio.

Choices **0.48** **2.04** **2.18** **exposure** **less likely** **more likely** **prevalence** **risk**

	Smokers	Non-Smokers	Total
Low Birth weight	1,556	14,974	16,530
High Birth weight	694	14,532	15,226
Total	2,250	29,506	31,756

Comparing the Risk Ratio and the Odds Ratio in Follow-up Studies

We have described two widely used measures of effect, the risk ratio and the odds ratio. Risk ratios are often preferred because they are easier to interpret. But, as we have seen, in case-control studies, we cannot estimate risks and must work instead with an exposure odds ratio (EOR). In follow-up studies, however, we have the option of computing both a risk ratio and a risk odds ratio (ROR). Which should we prefer?

It can be shown mathematically that if a risk ratio estimate is equal to or greater than one, then the corresponding risk odds ratio is at least as large as the risk ratio. For example, using the follow-up data for the quit smoking study of heart attack patients, we saw that the estimated risk ratio was 2.1, which is greater than one; the corresponding odds ratio was 2.7, which is larger than 2.1.

Follow-up (Cohort) Studies: RR vs. ROR

If $\hat{RR} \geq 1$, then $\hat{ROR} \geq \hat{RR}$

Quit Smoking Data for Heart Attack Patients

	Smoke	Quit	Total
Death	27	14	41
Survival	48	67	115
Total	75	81	156

$\hat{RR} = 2.1$

$\hat{ROR} = 2.7$

$$\text{If } \hat{RR} \geq 1, \text{ then } \hat{ROR} \geq \hat{RR}$$

Similarly if the risk ratio is less than one, the corresponding odds ratio is as small or smaller than the risk ratio. For example, if we switch the columns of the quit smoking table, then the risk ratio is 0.48, which is less than one, and the corresponding odds ratio is 0.37, which is less than 0.48.

If $\hat{RR} \leq 1$, then $\hat{ROR} \leq \hat{RR}$

Quit Smoking Data for Heart Attack Patients

	Quit	Smoke	Total
Death	14	27	41
Survival	67	48	115
Total	81	75	156

$\hat{RR} = 0.48$

$\hat{ROR} = 0.37$

$$\text{If } \hat{RR} \leq 1, \text{ then } \hat{ROR} \leq \hat{RR}$$

It can also be shown that if a disease is "rare", then the risk odds ratio will closely approximate the risk ratio. For follow-up studies, this *rare disease assumption* means that the risk that any study subject will develop the disease is small enough so that the corresponding odds ratio and risk ratio estimates give essentially the same interpretation of the effect of exposure on the disease.

Typically a rare "disease", is considered to be a disease that occurs so infrequently in the population of interest that the risk for any study subject is approximately zero. For example, if one out of every 100,000 persons develops the disease, the risk for this population is zero to 4 decimal places. Now that's really rare!

Study Questions (Q5.8)

1. Is a risk of .01 rare?
2. Suppose that for a given follow-up study, the true risk is not considered to be rare. Is it possible for the ROR and RR to be approximately the same?

We can write a formula that expresses the risk odds ratio in terms of the risk ratio:

$$\text{ROR} = \text{RR} \times f \quad \text{where} \quad f = \frac{(1 - R_0)}{(1 - R_1)}$$

and $\mathbf{R_0}$ is the risk for the unexposed, $\mathbf{R_1}$ is the risk for the exposed, and $\mathbf{RR=R_1/R_0}$

This formula says that the risk odds ratio is equal to the risk ratio multiplied by the factor ***f***, where ***f*** is defined as 1 minus the risk for the unexposed group (R_0) divided by 1 minus the risk for the exposed group (R_1). You can see from this equation that if both $\mathbf{R_1}$ and $\mathbf{R_0}$ are approximately 0, then ***f*** is approximately equal to one, and the risk odds ratio is approximately equal to the risk ratio.

Study Questions (Q5.9)

1. In the quit smoking example, where R_0 is 0.17 and R_1 equals 0.36, what is ***f***?
2. For this value of ***f***, is the ROR close to the RR?
3. What happens to ***f*** if the risks are halved, i.e., $R_0 = 0.17/2 = 0.085$ and $R_1 = 0.36/2 = 0.180$?
4. Are the ROR and RR estimates close for this ***f***?
5. What happens to f if we again halve the risks, so that $R_0=0.0425$ and $R_1=0.09$?
6. Is the approximation better?
7. Based on your answers to the above questions, how "rare" do the risks have to be for the odds and risk ratios to be approximately equal?

Summary

- If an estimate of RR $\geq$ 1, then the corresponding estimate of ROR is at least as large as the estimate of the RR.
- If an estimate of RR $\leq$ 1, then the corresponding estimate of ROR is as small or smaller than the estimate of RR.
- In follow-up (cohort) studies, the "rare disease assumption" says that the risk for any study subject is approximately zero.
- Under the rare disease assumption, the risk odds ratio (ROR) computed in a follow-up study approximates the risk ratio (RR) computed from the same study.

Comparing the RR and the OR in the Rotterdam Study

Osteoporosis is a common disease in the elderly, and leads to an increased risk of bone fractures. To study this disease, a cohort consisting of nearly 1800 postmenopausal women living in Rotterdam, the Netherlands, was followed for four years. The Rotterdam Study investigators wanted to know which genetic factors determine the risk of fractures from osteoporosis. They focused on a gene coding for one of the collagens that are involved in bone formation. Each person's genetic make-up consists of two alleles of this gene, and each allele can have one of two alternative forms, called allele A or allele B. The investigators showed that

women with two A alleles had a higher bone mass than women with at least one B allele. They therefore hypothesized that the risk of fractures would be higher in women with allele B.

Of the 1194 women with two A alleles, 64, or 5.36%, had a fracture during follow-up. Of the 584 women with at least one B allele, 47, or 8.05%, had a fracture.

Study Questions (Q5.10)

1. Calculate the risk ratio for the occurrence of fractures in women with at least one B allele compared to women with two A alleles.

Because the risk ratio estimate is greater than 1, we expect the risk odds ratio to be at least as large as the risk ratio.

Study Questions (Q5.10) continued

2. Calculate the risk odds ratio for the occurrence of fractures in women with at least one B allele compared to women with two A alleles.

Note that the risk of fractures is relatively rare in this population; therefore the risk odds ratio is approximately equal to the risk ratio. Recall the formula ROR = RR * f. Here, f is defined as 1 minus the risk in women with two A alleles divided by 1 minus the risk in women with at least one B allele.

$$\text{ROR} = \text{RR} \times \frac{1 - \text{R(2 A alleles)}}{1 - \text{R(1 B allele)}}$$

Study Questions (Q5.10) continued

3. Using the formula ROR = RR x f, can you show that we computed the correct risk odds ratio?

In this study, both the risk ratio and the risk odds ratio lead to the same conclusion: women with at least one B allele have a 50% higher chance of fractures than women with two A alleles. The Rotterdam Study investigators concluded that genetic make-up can predispose women to osteoporotic fractures.

Quiz (Q5.11): RR versus OR in follow-up studies A questionnaire was administered to persons attending a social event in which 39 of the 87 participants became ill with a condition diagnosed as salmonellosis. The 2 x 2 table below summarizes the relationship between consumption of potato salad and illness.

1. The risk ratio comparing the exposed to the non-exposed is **???**
2. The odds ratio is **???**
3. Does the odds ratio closely approximate the risk ratio? **???**
4. Do you consider this illness to be "rare"? **???**

Choices **0.25** **1.7** **3.7** **36.0** **9.8** **no** **yes**

	Exposed	**Non-Exposed**	**Total**
Ill	36	3	39
Well	12	36	48
Total	48	39	87

Let's consider data from a classic study of pellagra. Pellagra is a disease caused by dietary deficiency of niacin and characterized by dermatitis, diarrhea, and dementia. Data comparing cases by gender are shown below.

5. The risk ratio of pellagra for females versus males is (1 decimal place) **???**
6. The odds ratio is (to one decimal place) **???**
7. Does the odds ratio closely approximate the risk ratio? **???**
8. Do you consider this illness to be "rare"? **???**

Choices **1.4** **2.4** **2.5** **24.2** **no** **yes**

	Females	Males	Total
Ill	46	18	64
Well	1438	1401	2839
Total	1484	1419	2903

Comparing the RR and OR in Case-Control Studies

We have already seen that, for follow-up studies, if the disease is "rare", then the risk odds ratio will be a close approximation to the risk ratio computed from the same follow-up data. However, in case-control studies, a risk ratio estimate cannot be computed, and an exposure odds ratio must be used instead. So, for case-control data, if the disease is "rare", does the exposure odds ratio approximate the risk ratio that would have resulted from a comparable follow-up study? The answer is yes, depending on certain conditions that must be satisfied, as we will now describe.

This two-way table categorizes lung cancer and smoking status for a cohort of physicians in a large metropolitan city that are followed for 7 years. Forty smokers and twenty non-smokers developed lung cancer. The risk ratio is 2. Also, for this population, the risk odds ratio is equal to 2.02, essentially the same as the risk ratio. Since these are measures of effect for a population, we have not put the hat symbol over the risk ratio and risk odds ratio terms.

Cohort of Physicians (7-year follow-up)

	Smoker ? Yes	No	Total
LC	40	20	60
No LC	1960	1980	3940
Total	2000	2000	4000

population measures

$$RR = \frac{40/2000}{20/2000} = 2$$

$$ROR = \frac{40 \times 1980}{1960 \times 20} = 2.02$$

We now consider the results that we would expect to obtain if we carried out a case-control study using this cohort as our source population. We will assume that the 7-year follow-up has occurred. We also assume that there exists a comprehensive cancer registry, so that we were able to find al1 60 incident cases that developed over the 7year period. These would be our cases in our case-control study. Now suppose we randomly select 60 controls from the source population as our comparison group. Since half of the entire cohort of 4000 physicians was exposed and half was unexposed, we would

Case-Control Study

		E	not E	Total
(incident)	Cases	40	20	60
	Controls	30	30	60
	Total	70	50	120

expect 30 exposed and 30 unexposed out of the 60 controls.

$\hat{EOR} = 2.0 \approx \hat{RR} = 2$

We can use the cross product ratio formula to compute the expected exposure odds ratio, which turns out to be 2. This value for the exposure odds ratio obtained from case-control data is the same that we would have obtained from the risk ratio and the risk odds ratio if we had carried out the follow-up study on this population cohort. In other words, the expected EOR from this case-control study would closely approximate the RR from a corresponding follow-up study, even if the follow-up study was never done!

$\hat{EOR} = 1.0 \not\approx \hat{RR} = 2$

Case-Control Study

		E	not E	Total
(incident)	Cases	40	20	60
	Controls	40	20	60
	Total	80	40	120

$\hat{EOR} = 1.0$

We may wonder whether the EOR would approximate the RR even if the 60 controls did not split equally into exposed and unexposed groups as expected. This can occur by chance from random selection or if we do a poor job of picking controls. For example, suppose there were 40 exposed and 20 unexposed among the controls. Then the estimated exposure odds ratio would equal 1 instead of 2, so in this situation, the EOR would be quite different from the RR obtained from a comparable follow-up study.

What we have shown by example actually reflects an important caveat when applying the rare disease assumption to case-control data. The choice of controls in a case-control study must be representative of the source population from which the cases developed. If not, either by chance or a poor choice of controls, then the exposure odds ratio will not necessarily approximate the risk ratio even if the disease is rare. There is another important caveat for applying the rare disease assumption in a case-control study. The cases must be incident cases, that is, the cases need to include all new cases that developed over the time-period considered for determining exposure status. If the cases consisted only of prevalent cases at the time of case-ascertainment, then a biased estimate may result because the measure of effect would be estimating prevalence rather than incidence.

Summary

- In case-control studies, the EOR approximates an RR when the following 3 conditions are satisfied:
  1) The rare disease assumption holds
  2) The choice of controls in the case-control study must be representative of the source population from which the cases developed.
  3) The cases must be incident cases.

Quiz (Q5.12): Understanding Risk Ratio In a case-control study, if the rare disease assumption is satisfied, then:

1. The **???** approximates the **???** provided that there is no **???** in the selection of **???**, and the cases are **???** rather than **???** cases.

Choices **EOR** **ROR** **RR** **bias** **cases** **controls** **incidence** **prevalent** **randomness**

In a community of 1 million persons, 100 cases of a disease were reported, distributed by exposure according to the table below.

2. Calculate the RR. **???**
3. Calculate the ROR **???**
4. Is this a rare disease? **???**

	Exposed	**Non-Exposed**	**Total**
Ill	90	10	100
Well	499,910	499,990	999,900
Total	500,000	500,000	1,000,000

If the exposure status of all one million persons in the study population had not been available, the investigator may have conducted a case-control study. Suppose a random sample of 100 controls were selected.

5. Approximately what percentage of these controls would you expect to be exposed? **???**
6. What is the expected EOR in the case-control study? **???**

Choices **0.11** **10** **50** **9.00** **90** **no** **yes**

Note: On Lesson Page 5-3 of the ActivEpi CD-ROM, there is an activity (and corresponding asterisk) that provides a mathematical proof of the odds ratio approximation to the risk ratio in case control studies. This proof makes use of conditional probability statements and Bayes Theorem.

The Rate Ratio

A **rate ratio** is a ratio of two average rates. It is sometimes called an **incidence density ratio** or a **hazard ratio**. Recall the general formula for an average rate: **I** denotes the number of new cases of the health outcome, and **PT** denotes the accumulation of person-time over the follow-up.

The general data layout for computing a rate ratio is shown below. $\mathbf{I_1}$ and $\mathbf{I_0}$ denote the number of new cases in the exposed and unexposed groups, and $\mathbf{PT_1}$ and $\mathbf{PT_0}$ denote the corresponding person time accumulation for these two groups. The formula for the **rate ratio** or the **incidence density ratio** (**IDR**) is also provided. We have used the notation IDR instead of RR to denote the rate ratio in order to avoid confusion with our previous use of RR to denote the risk ratio.

Average Rate: $\frac{I}{PT}$

Layout for computing a Rate Ratio (i.e., IDR)

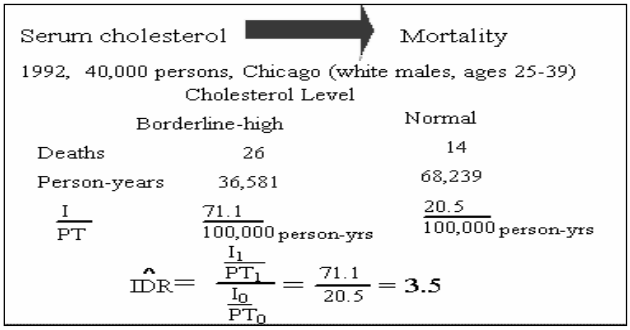
	Exposed	Unexposed	Total
New Cases	I_1	I_0	I
Person Time	PT_1	PT_0	PT

Rate Ratio:

$$IDR = \frac{\frac{I_1}{PT_1}}{\frac{I_0}{PT_0}}$$

As with both the risk ratio and odds ratio measures, the rate ratio can be >1, <1, or =1. If the rate ratio is equal to 1, it means that there is no relationship

between the exposure and disease using this measure of effect.



To illustrate the calculation of a rate ratio, we consider data on the relationship between serum cholesterol level and mortality from a 1992 study of almost 40,000 persons from the Chicago area. The data shown compares white males with borderline-high cholesterol levels and white males with normal cholesterol levels. Subjects, including persons from other race and sex categories, were enrolled into the study between 1967 and 1973, screened for cardiovascular disease (CVD) risk factors, and then followed for an average of 14 to 15 years. There were a total of 26 CHD-related deaths based on 36,581 person-years of follow-up among white males aged 25 to 39 with borderline-high cholesterol at entry into the study. This yields a rate of 71.1 deaths per 100,000 person-years. Among the comparison group there were 14 CHD-related deaths based on 68,239 person-years of follow-up, this yields a rate of 20.5 deaths per 100,000 person-years. Thus, white males aged 25-39 with borderline high cholesterol have 3.5 times the mortality rate as those with normal cholesterol, indicating that persons with even moderately high cholesterol carry an increased risk for CHD mortality.

Summary: Rate Ratio

- A ratio of two average rates is called a rate ratio (i.e., an incidence density ratio, hazard ratio)
- The formula for the rate ratio (IDR) is given by:

$$IDR = \frac{I_1/PT_1}{I_0/PT_0}$$

where I_1 and I_0 are the number of new cases and PT_1 and PT_0 are the accumulated person-time for groups 1 and 0, respectively.

- As with the RR and the OR, the IDR can be >1, <1, or =1.

Quiz (Q5.13) Data is shown on the next page for a follow-up study to compare mortality rates among diabetics and non-diabetics.

1. The mortality rate for diabetics is **???**
2. The mortality rate for non-diabetics is **???**
3. The rate ratio is **???**

Choices **13.9** **13.9 per 1000 person-years** **2.8**
2.8 per 1000 person-years **38.7** **38.7 per 1000 person-years**

	Diabetic	Non-diabetic	Total
Dead	72	511	583
Alive	146	3,312	3,458
Person-Years	1,862.4	36,532.2	38,394.6

4. The rate ratio comparing the mortality rates of diabetics with non-diabetics is 2.8. Which of the following is the correct interpretation of this measure?
   A. Those with diabetes are 2.8 times more likely to die than those without.
   B. People are 2.8 times more likely to die of diabetes than any other illness
   C. Death among diabetics is occurring at a rate 2.8 times that of non-diabetics

Nomenclature

Table setup for cohort, case-control, and prevalence studies:

	Exposed	Not Exposed	Total
Disease/cases	a	b	m_1
No Disease/controls	c	d	m_0
Total	n_1	n_0	n

Table setup for cohort data with person-time:

	Exposed	Not Exposed	Total
Disease (New cases)	I_1	I_0	I
No Disease	-	-	-
Total disease-free person-time	PT_1	PT_0	PT

Δt	Change in time
EOR	Exposure odds ratio; odds of exposure in diseased divided by the odds of exposure in nondiseased
I	Average incidence or total number of new cases
I_0	Number of new cases in nonexposed
I_1	Number of new cases in exposed
IDR	Incidence density ratio; rate in exposed/rate in nonexposed (also called the rate ratio)
N	Size of population under study
N_0	Size of population under study in nonexposed at time zero
N_1	Size of population under study in exposed at time zero
OR	Odds ratio: ad/bc
P	Probability of an event
P(D | E)	Probability of disease given exposed
P(D | not E)	Probability of disease given not exposed
P(E | D)	Probability of exposure given diseased
P(E | not D)	Probability of exposure given not diseased
POR	Prevalence odds ratio; an odds ratio calculated with prevalence data
PT	Disease-free person-time
PT_0	Disease-free person-time in nonexposed
PT_1	Disease-free person-time in exposed
R_0	Risk in unexposed
R_1	Risk in exposed

ROR	Risk odds ratio; an odds ratio calculated from cohort risk data
RR	Risk ratio: risk in exposed divided risk in unexposed
T or **t**	Time

Formulae

$RR = R_1 / R_0$
Odds = P / (1-P)
Odds ratio = ad/bc
$ROR = RR * f$ where $f=(1-R_0)/(1-R_1)$
$IDR = (I_1/PT_1) / (I_0/PT_0)$

References

Doll R, and Hill AB. Smoking and carcinoma of lungs: preliminary report. Br Med J 1950;2:739-48.

Dyer AR, Stamler J, Shekelle RB. Serum cholesterol and mortality from coronary heart disease in young, middle-aged, and older men and women in three Chicago epidemiologic studies. Ann of Epidemiol 1992;2(1-2): 51-7.

Greenberg RS, Daniels SR, Flanders WD, Eley JW, Boring JR. Medical Epidemiology (3rd Ed). Lange Medical Books, New York, 2001.

Hammond EC, Horn D. The relationship between human smoking habits and death rates. JAMA 1958;155:1316-28.

Johansson S, Bergstrand R, Pennert K, Ulvenstam G, Vedin A, Wedel H, Wilhelmsson C, Wilhemsen L, Aberg A. Cessation of smoking after myocardial infarction in women. Effects on mortality and reinfarctions. Am J Epidemiol 1985;121(6): 823-31.

Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic Research: Principles and Quantitative Methods. John Wiley and Sons Publishers, New York, 1982.

Spika JS, Dabis F, Hargett-Bean N, Salcedo J, Veillard S, Blake PA. Shigellosis at a Caribbean Resort. Hamburger and North American origin as risk factors. Am J Epidemiol 1987;126 (6): 1173-80.

Steenland K (ed.). Case studies in occupational epidemiology. Oxford University Press, New York, NY: 1993.

Uitterlinden AG, Burger H, Huang Q, Yue F, McGuigan FE, Grant SF, Hofman A, van Leeuwen JP, Pols HA, Ralson SH. Relation of alleles of the collagen type I alpha1 gene to bone density and the risk of osteoporotic fractures in postmenopausal women. N Engl J Med 1998;338(15):1016-21.

Wynder EL, Graham EA. Tobacco smoking as a possible etiologic factor in bronchogenic carcinoma: a study of six hundred and eighty-four proved cases. JAMA 1950;143:329-36.

Answers to Study Questions and Quizzes

Q5.1

1. The five-year risk for continuing smokers is 4½ times greater than the risk for smokers who quit.
2. The risk ratio is very close to 1.0, which indicates no meaningful difference between the risks for the two groups.
3. Think of an inverse situation.
4. You should have the hang of this by now.

Q5.2

1. 2 2. 1 3. 0.5 4. 0.0104
5. . 0.0236
6. 0.44 – In general, the risk ratio that compares two groups is defined to be the risk for one group divided by the risk for the other group. It is important to clearly specify which group is in the numerator and which group is in the denominator. If, for example, the two groups are labeled *group 1* and *group 0*, and the risk for group 1 is in the numerator, then we say the risk ratio compares group 1 to group 0.

Q5.3

1. The odds that a case ate raw hamburger is about two ½ times the odds that a control subject ate raw hamburger.
2. Because the odds ratio is so close to being equal to one, this would be considered a null case, meaning that the odds that a case ate raw hamburger is about the same as the odds that a control subject age raw hamburger.
3. An odds ratio less than one means that the odds that a case subject ate raw hamburger is less than the odds that a control subject ate raw hamburger.
4. You should have the hang of this by now.

Q5.4

1. Not possible, odds ratio – The risk of disease is defined as the proportion of initially disease-free population who develop the disease during a specified period of time. In a case-control study, the risk cannot be determined.
2. 1350/1357
3. 1296/1357
4. 192.86
5. 21.25
6. 9.08 – In general, the odds ratio that compares two groups is defined to be the odds for the cases divided by the odds for the controls. The odds for each group can be calculated by the formula P/(1-P), where P is the probability of exposure.
7. 3
8. 1
9. 0.333

Q5.5

1. Of course! It is possible, for example, that mayonnaise actually contained the outbreak-causing bacteria and maybe most of the cases that ate raw hamburger used mayonnaise.

Q5.6

1. 683, 86, 0.77
2. odds, exposure, less than, odds, controls – If the estimated odds ratio is less than 1, then the odds of exposure for cases is less than the odds of exposure for controls. If the estimated odds ratio is greater than 1, then the odds of exposure for cases is greater than the odds of exposure for controls.

Q5.7

1. 2.18
2. more likely
3. prevalence

Q5.8

1. That depends on the disease being considered and on the time-period of follow-up over which the risk is computed. However, for most chronic diseases and short time periods, a risk of .01 is not rare.
2. Yes, because even though the risk may not be rare, it may be small enough so that the ROR and the RR are approximately the same.

Q5.9

1. f = (1 – 0.17) / (1 – 0.36) = 1.30
2. No, since for these data, the estimated RR equals 2.1 whereas the estimate ROR equals 2.7.
3. f = (1 – 0.085) / (1 – 0.180) = 1.12
4. Yes, since the estimated RR is again 2.1, (0.180/0.085), but the estimated ROR is 2.4.
5. f=1.05
6. Yes, since the estimated ROR is now 2.2.
7. In the context of the quit smoking example, risks below 0.10 for both groups indicate a “rare” disease.

Q5.10

1. The risk ratio in this study is 0.0805 divided by 0.0536, which equals 1.50.
2. The risk odds ratio is 47/537 divided by 64/1130 equals 1.54.
3. f=(1-0.0536) / (1-0.0805) = 1.03. The ROR = 1.03*RR = 1.03*1.50=1.54.

Q5.11

1. 9.8
2. 36.0
3. No
4. No – The risk ratio that compares two groups is defined to be the risk for one group divided by the risk for the other group. The odds ratio can be calculated by the cross product formula ad/bc. In general, a disease is considered “rare” when the OR closely approximates the RR.
5. 2.44
6. 2.49

7. Yes
8. Yes

Q5.12

1. EOR, RR, bias, controls, incident, prevalent
2. 9
3. 9
4. Yes – A disease is considered rare when the ROR closely approximates the RR.
5. 50
6. 9.00

Q5.13

1. 38.7 per 1000 person-years – The mortality rate for diabetics equals 72/1,862.4 person-years = 38.7 per 1000 person-years.
2. 13.9 per 1000 person-years – The mortality rate for non-diabetics equals 511/36,653.2 person-years = 13.9 per 1000 person-years.
3. 2.8 – The rate ratio is 38.7 per 1000 person-years/13.9 per 1000 person-years = 2.8.
4. C