Proportions, Odds, and Risk[1]

Christopher L. Sistrom, MD, MPH
Cynthia W. Garvan, PhD

Index term:
Statistical analysis

Published online
10.1148/radiol.2301031028
Radiology 2004; 230:12–19

Abbreviations:
OR = odds ratio
RR = relative risk

Perhaps the most common and familiar way that the results of medical research and epidemiologic investigations are summarized is in a table of counts. Numbers of subjects with and without the outcome of interest are listed for each treatment or risk factor group. By using the study sample data thus tabulated, investigators quantify the association between treatment or risk factor and outcome. Three simple statistical calculations are used for this purpose: difference in proportions, relative risk, and odds ratio. The appropriate use of these statistics to estimate the association between treatment or risk factor and outcome in the relevant population depends on the design of the research. Herein, the enumeration of proportions, odds ratios, and risks and the relationships between them are demonstrated, along with guidelines for use and interpretation of these statistics appropriate to the type of study that gives rise to the data.

In a previous article in this series (1), the 2 × 2 contingency table was introduced as a way of organizing data from a study of diagnostic test performance. Applegate et al (2) have previously described analysis of nominal and ordinal data as counts and medians. Binary variables are a special case of nominal data where there are only two possible levels (eg, yes/no, true/false). Data in two binary variables arise from a variety of research methods that include cross-sectional, case-control, cohort, and experimental designs. In this article, we will describe three ways to quantify the strength of the relationship between two binary variables: difference of proportions, relative risk (RR), and odds ratio (OR). Appropriate use of these statistics depends on the type of data to be analyzed and the research study design.

Correct interpretation of the difference of proportions, the RR, and the OR is key to the understanding of published research results. Misuse or misinterpretation of them can lead to errors in medical decision making and may even have adverse public policy implications. An example can be found in an article by Schulman et al (3) published in the *New England Journal of Medicine* about the effects of race and sex on physician referrals for cardiac catheterization. Results of this study of Schulman et al received extensive media coverage about the findings that blacks and women were referred less often than white men for cardiac catheterization. In a follow-up article, Schwartz et al (4) showed how the magnitude of the findings of Schulman et al was overstated, chiefly because of confusion among OR, RR, and probability. The resulting controversy underscores the importance of understanding the nuances of these statistical measures.

Our purpose is to show how a 2 × 2 contingency table summarizes results of several common types of biomedical research. We will describe the four basic study designs that give rise to such data and provide an example of each one from literature related to radiology. The appropriate use of difference in proportion, RR, and OR depends on the study design used to generate the data. A key concept to be developed about using odds to estimate risk is that the relationship between OR and RR depends on outcome frequency. Both graphic and computational correction of OR to estimate RR will be shown. With rare diseases, even a corrected RR estimate may overstate the effect of a risk factor or treatment. Use of difference in proportions (expressed as attributable risk) may give a better picture of societal impact. Finally, we introduce the concept of confounding by factors extraneous to the research question that may lead to inaccurate or contradictory results.

2 × 2 CONTINGENCY TABLES

Let *X* and *Y* denote two binary variables that each have only two possible levels. Another term for binary is dichotomous. Results are most often presented as counts of observations at each level. The relationship between *X* and *Y* can be displayed in a 2 × 2 contingency table. Another name for a contingency table is a cross-classification table. A 2 × 2

[1] From the Departments of Radiology (C.L.S.) and Biostatistics (C.W.G.), University of Florida College of Medicine, PO Box 100374, Gainesville, FL 32610. Received July 2, 2003; revision requested July 30; revision received August 4; accepted August 13. **Address correspondence to** C.L.S. (e-mail: *sistrc@radiology.ufl.edu*).

contingency table consists of four cells: the cell in the first row and first column (cell 1–1), the cell in the first row and second column (cell 1–2), the cell in the second row and first column (cell 2–1), and the cell in the second row and second column (cell 2–2). Commonly used symbols for the cell contents include n with subscripts, p with subscripts, and the letters a–d. The n_{11}, n_{12}, n_{21}, and n_{22} notation refers to the number of subjects observed in the corresponding cells. In general, "n_{ij}" refers to the number of observations in the ith row (i = 1, 2) and jth column (j = 1, 2). The total number of observations will be denoted by n (ie, $n = n_{11} + n_{12} + n_{21} + n_{22}$). The p_{11}, p_{12}, p_{21}, and p_{22} notation refers to the proportion of subjects observed in each cell. In general, "p_{ij}" refers to the proportion of observations in the ith row (i = 1, 2) and jth column (j = 1, 2). Note that $p_{ij} = n_{ij}/n$. For simplicity, many authors use the letters a–d to label the four cells as follows: a = cell 1–1, b = cell 1–2, c = cell 2–1, and d = cell 2–2. We will use the a–d notation in equations that follow. Table 1 shows the general layout of a 2 × 2 contingency table with symbolic labels for each cell and common row and column assignments for data from medical studies.

TABLE 1
Notation for 2 × 2 Contingency Table

	Y*	
X†	Yes‡	No§
Present‖	n_{11} p_{11} a	n_{12} p_{12} b
Absent#	n_{21} p_{21} c	n_{22} p_{22} d

Note.—n = number of subjects in the cell, p = proportion of entire sample in the cell, a–d = commonly used cell labels.

*Response, outcome, or disease status variable.

† Explanatory, risk factor, or exposure variable.

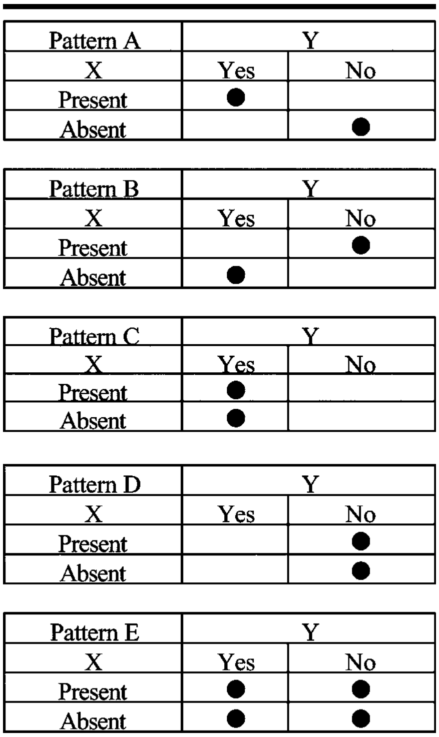
‡ Adverse outcome or disease-positive response.

§ No adverse outcome or disease-negative response.

‖ Exposed or risk-positive group.

# Unexposed or risk-negative group.

In many contingency tables, one variable is a response (outcome or dependent variable) and the other is an explanatory (independent) variable. In medical studies, the explanatory variable (X in the rows) is often a risk or a protective factor and the response (Y in the columns) is a disease state. The distribution of observed data in a 2 × 2 table indicates the strength of relationship between the explanatory and the response variables. Figure 1 illustrates possible patterns of observed data. The solid circle represents a cell containing numerous observations. Intuitively, we would expect that if the X and Y variables are associated, then pattern A or B would be observed. Patterns C, D, and E suggest that X and Y are independent of each other (ie, there is no relationship between them).



Pattern A	Y	
X	Yes	No
Present	●	
Absent		●

Pattern B	Y	
X	Yes	No
Present		●
Absent	●	

Pattern C	Y	
X	Yes	No
Present	●	
Absent	●	

Pattern D	Y	
X	Yes	No
Present		●
Absent		●

Pattern E	Y	
X	Yes	No
Present	●	●
Absent	●	●

Figure 1. Diagram shows possible patterns of observed data in a 2 × 2 table. Cells with black circles contain relatively large numbers of counts. With pattern A, n_{11} and n_{22} are large, suggesting that when X is present, Y is "yes." With pattern B, n_{12} and n_{21} are large, suggesting that when X is absent, Y is "yes." With pattern C, n_{11} and n_{21} are large, suggesting that Y is "yes" regardless of X. With pattern D, n_{12} and n_{22} are large, suggesting that Y is "no" regardless of X. With pattern E, n_{11}, n_{12}, n_{21}, and n_{22} are all about the same, suggesting that Y is "yes" and "no" in equal proportion regardless of X.

STUDY DESIGNS THAT YIELD 2 × 2 TABLES

The statistical methods used to analyze research data depend on how the study was conducted. There are four types of designs in which two-by-two tables may be used to organize study data: case-control, cohort, cross sectional, and experimental. The first three designs are often called observational to distinguish them from experimental studies; of the experimental studies, the controlled clinical trial is the most familiar. The 2 × 2 tables that result from the four designs may look similar to each other. The outcome is typically recorded in columns, and the explanatory variable is listed in the rows. The χ^2 statistic may be calculated and used to test the null hypotheses of independence between row and column variables for all four types of studies. These methods are described in a previous article in this series (2). Table 2 summarizes the features of the four types of designs in which 2 × 2 tables are used to organize study data. Each is briefly described next, with a radiology-related example provided for illustration. The ordering of the study designs in the following paragraphs reflects, in general, the strength (ie, weaker to stronger) of evidence for causation obtained from each one. The advantages and disadvantages of the different designs and situations where each is most appropriate are beyond the scope of this article, and the reader is encouraged to seek additional information in biostatistics, research design, or clinical epidemiology reference books (5–7).

The statistics used to quantify the relationship between variables (ie, the effect size) are detailed and the examples of study designs are summarized in Table 3. These will be briefly defined now. The difference in proportion is the difference in the fraction of subjects who have the outcome between the two levels of the explanatory variable. The RR is the ratio of proportion (ie, risk) of subjects who have the outcome between different levels of the explanatory variable. The OR is the ratio of the odds that a subject will have the outcome between the two levels of the explanatory variable. Each of these statistics has an associated standard error that can be calculated and used to form CIs around the estimate at any chosen level of precision. The calculation of CIs and their use for inference testing are described in a previous article in this series (8).

Cross-sectional Studies

A cross-sectional study does not involve the passage of time. A single sample is selected without regard to disease state or exposure status. Information on disease state and exposure status is determined with data collected at a single time point. Data about exposure status and

TABLE 2
Comparison of Four Study Designs

Attribute	Cross-sectional Study	Case-Control Study	Cohort Study	Experimental Study
Sample selection	One sample selected without regard to disease or exposure status	Two samples selected: one from disease-positive population, one from disease-negative population	Two samples selected: one from exposed population, one from unexposed population	One sample selected that is disease negative; sample is randomly assigned to treatment or control group
Proportions that can be estimated	Prevalence of disease in the exposed and unexposed groups	Proportion of cases and controls that have been exposed to a risk factor	Incidence of disease in exposed and unexposed groups	Incidence of disease in treated and untreated (control) groups
Time reference	Present look at time	Backward look in time	Forward look in time	Forward look in time
Effect measure	OR, difference in proportions*	OR	RR, difference in proportions*	RR, difference in proportions*

* Difference in proportions may be used as an alternate measure of effect.

TABLE 3
Quantification of Effect Size for Two Binary Variables

Definition	Difference of Proportions*	RR†	OR‡
Calculation for estimate based on sample data	$n_{11}/n_{11} + n_{12} - n_{21}/n_{21} + n_{22}$	$\frac{n_{11}/n_{11} + n_{12}}{n_{21}/n_{21} + n_{22}}$	$n_{11}n_{22}/n_{12}n_{21}$
Calculation in terms of *a*–*d*-cell labels	$[a/(a + b)] - [c/(c + d)]$	$\frac{[a/(a+b)]}{[c/(c+d)]}$	ad/bc
Cross-sectional example (Table 4)	0.22	1.69	2.52
Case-control example (Table 5)	NA§	NA§	2.22
Cohort example (Table 6)	0.046	1.07	1.25‖
Experimental example (Table 7)	0.029	1.56	1.61‖

* Proportion with outcome in exposed group minus proportion with outcome in unexposed group.
† Risk of outcome in exposed group divided by risk of outcome in unexposed group.
‡ Odds of outcome in exposed group divided by odds of outcome in unexposed group.
§ NA = not typically calculated or reported, since the statistic is not meaningful for this study design.
‖ ORs may be calculated for cohort and experimental studies, but RR is preferred.

TABLE 4
Example of Cross-sectional Study Data

Major Risk Factor	DVT Positive*	DVT Negative*	Total
Present	81	67	148
Absent	90	188	278

Source.—Reference 9.
* DVT = deep venous thrombosis.

disease state can be organized into a 2 × 2 contingency table, and the prevalence (ie, the proportion of a group that currently has a disease) can be compared for the exposed and unexposed groups. Effect size from cross-sectional studies may be assessed with difference in proportions, RR, or OR. For example, Cogo et al (9) studied the association between having a major risk factor (eg, immobilization, trauma, and/or recent surgery) and deep vein thrombosis. This study was performed in an outpatient setting by using a cross-sectional design. A total of 426 subjects who were referred by general practitioners underwent contrast material–enhanced venography to determine deep vein thrombosis status (positive or negative). Concurrently, information on major risk factors was recorded as being present or absent. They found that deep vein thrombosis was more likely to occur when a major risk factor was present (81 [55%] of 148) than when none was present (90 [32%] of 278). The data are shown in Table 4, and effect measures are included in Table 3.

Case-Control Studies

In a case-control study, the investigator compares instances of a certain disease or condition (ie, the cases) with individuals who do not have the disease or condition (ie, the "controls" or control subjects). Control subjects are usually selected to match the patients with cases of disease in characteristics that might be related to the disease or condition of interest. Matching by age and sex is commonly used. Investigators look backward in time (ie, retrospectively) to collect information about risk or protective factors for both cases and controls. This is achieved by examining past records, interviewing the subject, or in some other way. The only correct measure of effect size for a case-control study is the OR. However, the calculated OR may be used to estimate RR after appropriate correction for disease frequency in the population of interest, which will be explained later. For example, Vachon et al (10) studied the association between type of hormone replacement therapy and increased mammographic breast density by using a case-control study design. They identified 172 women who were undergoing hormone replacement therapy who had increased breast density (cases) and 172 women who were undergoing hormone replacement therapy who did not have increased breast density (controls). The type of hormone replacement therapy used by all subjects was then determined. They found that combined hormone replacement therapy was associated with increased breast density more often than was therapy with estrogen alone (OR = 2.22). The data are presented in Table 5.

Cohort Studies

A cohort is simply a group of individuals. The term is derived from Roman military tradition; according to this tradition, legions of the army were divided into 10 cohorts. This term now means any specified subdivision or group of people marching together through time. In other words, cohort studies are about the life histories of sections of populations and the individuals who are in-

TABLE 5
Example of Case-Control Study Data

Condition	Cases*	Controls†
Exposed‡	111	79
Nonexposed§	50	79

Source.—Reference 10.
* Increased breast density.
† No increased breast density.
‡ Combined therapy.
§ Estrogen alone.

TABLE 6
Example of Cohort Study Data

	No. Undergoing Mammography within 2 Years*		
Result at Last Mammography	Returned	Did Not Return	Total
False-positive	602	211	813
True-negative	3,098	1,359	4,457

Source.—Reference 11.
* Within 2 years after last mammography.

TABLE 7
Example of Experimental Study Data

	Underwent Mammography within 2 Years		
Treatment	No. Who Did	No. Who Did Not	Total
Intervention*	100	1,129	1,229
Control†	64	1,165	1,229

Source.—Reference 12.
* Intervention indicates that subjects received a mailed reminder.
† Control indicates that subjects did not receive a mailed reminder.

cluded in them. In a prospective study, investigators follow up subjects after study inception to collect information about development of disease. In a retrospective study, disease status is determined from medical records produced prior to the beginning of the study but after articulation of the cohorts. In both types of studies, initially disease-free subjects are classified into groups (ie, cohorts) on the basis of exposure status with respect to risk factors. Cumulative incidence (ie, the proportion of subjects who develop disease in a specified length of time) can be computed and compared for the exposed and unexposed cohorts. The main difference between prospective and retrospective cohort studies is whether the time period in question is before (retrospective) or after (prospective) the study begins. Effect size from a cohort study is typically quantified with RR and/or difference in proportions. For example, Burman et al (11) studied the association between false-positive mammograms and interval breast cancer screening by using a prospective cohort study design. Women in whom a false-positive mammogram was obtained at the most recent screening formed one cohort, and women in whom a previously negative mammogram was obtained formed the other cohort. All of the women were followed up to determine if they obtained a subsequent screening mammogram within the recommended interval (up to 2 years, depending on age). Burman et al found no significant difference in the likelihood that a woman would obtain a mammogram between the two cohorts (RR = 1.07). The data are included in Table 6.

Experimental Studies

The characteristic that distinguishes any experiment is that the investigator directly manipulates one or more variables (not the outcome!). A clinical trial is the most common type of experimental study used in medical research. Here, the investigator selects a sample of subjects and assigns each to a treatment. In many cases, one treatment may be standard therapy or an inactive (ie, placebo) treatment. These subjects are the controls, and they are compared with the subjects who are receiving a new or alternative treatment. Treatment assignment is almost always achieved randomly so that subjects have an equal chance of receiving one of the treatments. Subjects are followed up in time, and the cumulative incidence of the outcome or disease is compared between the treatment groups. RR and/or difference in proportions is typically used to quantify treatment effect on the outcome. An example of such a study can be found in an article by Harrison et al (12) in which they describe their trial of direct mailings to encourage attendance for mammographic screening. At the start of the study, half of the women were randomly assigned to receive a personally addressed informational letter encouraging them to attend screening. The other half (ie, controls) received no intervention. The number of women in each group who underwent mammography during the subsequent 2 years was enumerated. The women to whom a letter was sent were more likely to obtain a mammogram (RR = 1.56). The data are listed in Table 7.

CALCULATION OF PROPORTIONS FROM A 2 × 2 TABLE

Various proportions can be calculated from the data represented in a 2 × 2 contingency table. The cell, row, and column proportions each give different information about the data. Proportions may be represented by percentages or fractions, with the former having the advantage of being familiar to most people. Cell proportions are simply the observed number in each of the four cells divided by the total sample size. Each cell also has a row proportion. This is the number in the cell divided by the total in the row containing it. Likewise, there are four column proportions that are calculated by dividing the number in each cell by the total in the column that contains it.

In a 2 × 2 table organized with outcome in the columns and exposure in the rows, the various proportions have commonly understood meanings. Cell proportions are the fractions of the whole sample found in each of the four combinations of exposure and outcome status. Row proportions are the fractions with and without the outcome. It may seem counterintuitive that row proportions give information about the outcome. However, remembering that cells in a given row have the same exposure status helps one to clarify the issue. Similarly, column proportions are simply the fraction of exposed and unexposed subjects.

DIFFERENCE IN PROPORTIONS

The difference in proportions is used to compare the response *Y* (eg, disease: yes or no) according to the value of the explanatory variable *X* (eg, risk factor: exposed or unexposed). The difference is defined as the proportion with the outcome in the exposed group minus the proportion with the outcome in the unexposed group. By using the *a–d* letters for cell labels (Table 1), the calculation is as follows:

$$[a/(a + b)] - [c/(c + d)].$$

For the cross-sectional study data (Table 4) we would calculate the following equation:

$$[81/(81 + 67)] - [90/(90 + 188)] = 0.547 - 0.324 = 0.223.$$

The difference in proportions always is between −1.0 and 1.0. It equals zero when the response *Y* is statistically independent of the explanatory variable *X*. When *X* and *Y* are independent, then there is no association between them. It is appropriate to calculate the difference in proportions for the cohort, cross-sectional, and experimental study designs. In a case-control study, there is no information about the proportions that are outcome (or disease) positive in the population. This is because the investigator actually selects subjects to get a fixed number at each level of the outcome (ie, cases and controls). Therefore, the difference in proportions statistic is inappropriate for estimating the association between exposure and outcome in a case-control study. Table 3 lists the difference in proportions estimate for cross-sectional, cohort, and experimental study examples (9,11,12).

RISK AND RR

Risk is a term often used in medicine for the probability that an adverse outcome, such as a side effect, development of a disease, or death, will occur during a specific period of time. Risk is a parameter that is completely known only in the rare situation when data are available for an entire population. Most often, an investigator estimates risk in a particular population by taking a representative random sample, counting those that experience the adverse outcome during a specified time interval, and forming a proportion by dividing the number of adverse outcomes by the sample size. For example, the estimate of risk is equal to the number of subjects who experience an event or outcome divided by the sample size.

The epidemiologic term for the resulting rate is the cumulative incidence of the outcome. The incidence of an event or outcome must be distinguished from the prevalence of a disease or condition. Incidence refers to events (eg, the acquisition of a disease), while prevalence refers to states (eg, the state of having a disease). RR is a measure of association between exposure to a particular factor and risk of a certain outcome. The RR is defined to be the ratio of risk in the exposed and unexposed groups. An equivalent term for RR that is sometimes used in epidemiology is the cumulative incidence ratio, which may be calculated as follows: RR is equal to the risk among exposed subjects divided by the risk among unexposed subjects.

In terms of the letter labels for cells (Table 1), RR is calculated as follows:

$$\mathrm{RR} = [a/(a + b)]/[c/(c + d)].$$

The RR for our cohort study example (Table 6) would be calculated by dividing the fraction with a mammogram in the false-positive cohort (602/813 = 0.74) by the fraction with a mammogram in the true-negative cohort (3,098/4,457 = 0.695). This yields 1.065. The value of RR can be any nonnegative number. An RR of 1.0 corresponds to independence of (or no association between) exposure status and adverse outcome. When RR is greater than 1.0, the risk of disease is increased when the risk factor is present. When RR is less than 1.0, the risk of disease is decreased when the risk factor is present. In the latter case, the factor is more properly described as a protective factor. The interpretation of RR is quite natural. For example, an RR equal to 2.0 means that an exposed person is twice as likely to have an adverse outcome as one who is not exposed, and an RR of 0.5 means that an exposed person is half as likely to have the outcome. Table 3 illustrates the calculation of the RR estimates from the various examples.

Any estimate of relative risk must be considered in the context of the absolute risk. Motulsky (5) gives an example to show how looking only at RR can be misleading. Consider a vaccine that halves the risk of a particular infection. In other words, the vaccinated subjects have an RR of 0.5 of getting infected compared with their unvaccinated peers. The public health impact depends not only on the RR but also on the absolute risk of infection. If the risk of infection is two per million unvaccinated people in a year, then halving the risk to one per million is not so important. However, if the risk of infection is two in 10 unvaccinated people in a year, then halving the risk is of immense consequence by preventing 100,000 cases per million. Therefore, it is more informative to compare vaccinated to unvaccinated cohorts by using the difference in proportions. With the rare disease, the difference is 0.0000001, while for the common disease it is 0.1. This logic underlies the concept of number needed to treat and number needed to harm. These popular measures developed for evidence-based medicine allow direct comparison of effects of interventions (ie, number needed to treat) or risk factors (ie, number needed to harm). In our vaccination scenario, the number needed to harm (ie, to prevent one infection) for the rare disease is 1 million and for the common disease is 10.

ODDS AND THE OR

The OR provides a third way of comparing proportions in a 2 × 2 contingency table. An OR is computed from odds. Odds and probabilities are different ways of expressing the chance that an outcome may occur. They are defined as follows: The probability of outcome is equal to the number of times the outcome is observed divided by the total observations. The odds of outcome is equal to the probability that the outcome does occur divided by the probability that the outcome does not occur.

We are familiar with the concept of odds through gambling. Suppose that the odds a horse named Lulu will win a race are 3:2 (ie, read as "three to two"). The 3:2 is equivalent to 3/2 or 1.5. The probability that Lulu will be the first to cross the finish line can be calculated from the odds, since there is a deterministic relationship between odds and probability (ie, if you know the value of one, then you can find the value of the other). We know that:

$$Pr = \mathrm{Odds}/(1 + \mathrm{Odds})$$

and

$$\mathrm{Odds} = Pr/(1 - Pr),$$

where *Pr* is probability.

The probability that Lulu will win the race is (1.5)/1 + (1.5) = 0.60, or 60%.

Probabilities always range from 0 to 1.0, while odds can be any nonnegative number. The odds of a medical outcome in exposed and unexposed groups are defined as follows: Odds of disease in the exposed group is equal to the probability that the disease occurs in the exposed group divided by the probability that the disease does not occur in the exposed group. Odds of disease in the unexposed group is equal to the probability that the disease occurs in the unexposed group divided by the probability that the disease does not occur in the unexposed group.

It is helpful to define odds in terms of the *a*–*d* notation shown in Table 1. Useful mathematical simplifications (where Odds_{exp} is the odds of outcome in the exposed group and $\text{Odds}_{\text{unex}}$ is the odds of outcome in the unexposed group) that arise from this definition are as follows:

$$\text{Odds}_{\text{exp}} = [a/(a + b)]/[b/(a + b)] = a/b$$

and

$$\text{Odds}_{\text{unex}} = [c/(c + d)]/[d/(c + d)] = c/d.$$

Note that these simplifications mean that the outcome probabilities do not have to be known in order to calculate odds. This is especially relevant in the analysis of case-control studies, as will be illustrated. The OR is defined as the ratio of the odds and may be calculated as follows: OR is equal to the odds of disease in the exposed group divided by the odds of disease in the unexposed group. By using the *a*–*d* labeling,

$$\text{OR} = (a/b)/(c/d) = ad/bc.$$

The OR for our case-control example (Table 5) would be calculated as follows:

$$\text{OR} = [(111)(79)]/[(79)(50)] = 2.22.$$

The OR has another property that is particularly useful for analyzing case-control studies. The OR we calculate from a case-control study is actually the ratio of odds of exposure, not outcome. This is because the numbers of subjects with and without the outcome are always fixed in a case-control study. However, the calculation for exposure OR and that for outcome OR are mathematically equivalent, as shown here:

$$(a/c)/(b/d) = ad/bc = (a/b)/(c/d).$$

Therefore, in our example, we can correctly state that the odds of increased breast density is 2.2 times greater in those receiving combined hormone replacement therapy than it is in those receiving estrogen alone. Authors, and readers, must be very circumspect about continuing with the chain of inference to state that the risk of increased breast density in those receiving combined hormone replacement therapy is 2.2 times higher than is the risk of increased breast density in those receiving estrogen alone. In doing this, one commits the logical error of assuming causation from association. Furthermore, the OR is not a good estimator of RR when the outcome is common in the population being studied.

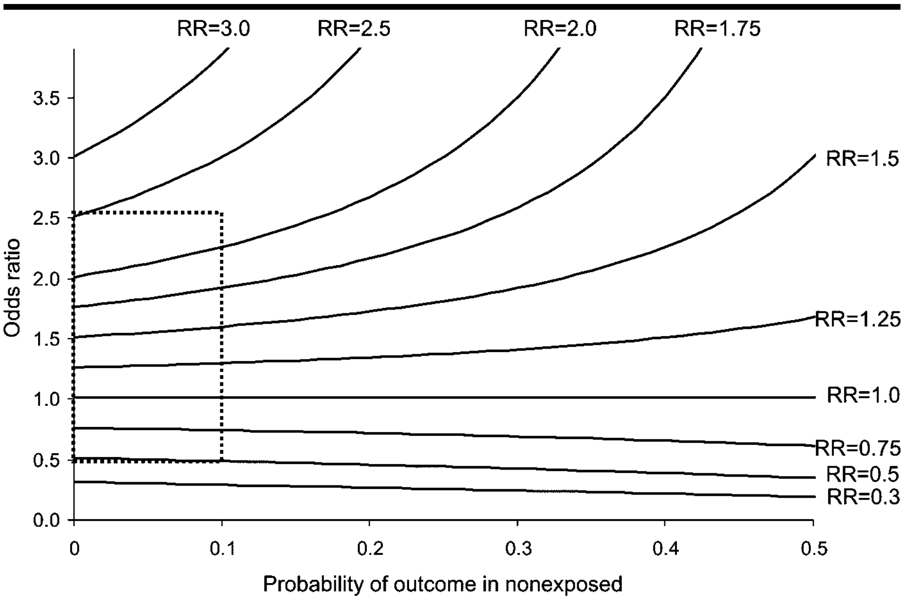


Figure 2. Graph shows relationship between OR and probability of outcome in unexposed group. Curves represent values of underlying RR as labeled. Rectangle formed by dotted lines represents suggested bounds on OR and probability of outcome in the unexposed group within which no correction from OR to RR is needed. When OR is more than 2.5 or less than 0.5 or probability of outcome in the unexposed group is more than 0.1, a correction (calculation in text) should be applied.

RELATIONSHIP BETWEEN RR AND OR

There is a strict and invariant mathematic relationship between RR and OR when they both are from the same population, as may be observed with the following equation:

$$\text{RR} = \text{OR}/[(1 - Pr_{\text{o}}) + (Pr_{\text{o}})(\text{OR})],$$

where Pr_{o} is the probability of the outcome in the unexposed group.

The relationship implies that the magnitude of OR and that of RR are similar only when Pr_{o} is low (13). In epidemiology, Pr_{o} is referred to as the incidence of outcome in the unexposed group. Thus, the OR obtained in a case-control study accurately estimates RR only when the outcome in the population being studied is rare. Figure 2 shows the relationship between RR and OR for various values of Pr_{o}. Note that the values of RR and OR are similar only when the Pr_{o} is small (eg, 10/100 = 10% or less). At increasing Pr_{o}, ORs that are less than 1.0 underestimate the RR, and ORs that are greater than 1.0 overestimate the RR. A rule of thumb is that the OR should be corrected when incidence of the outcome being studied is greater than 10% if the OR is greater than 2.5 or the OR is less than 0.5 (4). Note that with a case-control study, the probability of outcome in the unexposed must be obtained separately because it cannot be estimated from the sample.

This distinction was at the heart of the critique of Schwartz et al (4) regarding the article by Schulman et al (3). Schulman et al had reported an OR of 0.6 for referral to cardiac catheterization (outcome) between blacks and whites (explanatory). However, referral for catheterization occurred up to 90% of the time, so the corresponding RR should have been 0.93. This information would have created a rather unspectacular news story (ie, blacks referred 7% less often than whites) compared with the initial, and incorrect, headlines stating that blacks were referred 40% less often than whites.

In our examples, this relationship is also apparent. In the studies where both RR and OR were calculated, the OR is

TABLE 8
Death Penalty Verdicts Following Murder Convictions in Florida, 1976–1987

		Death Penalty*		
Victim's Race	Defendant's Race	Yes	No	RR, White/Black
White	White	53 (11)	414 (89)	0.495
White	Black	11 (23)	37 (77)	. . .
Black	White	0 (0)	16 (100)	0
Black	Black	4 (3)	139 (97)	. . .
Combined	White	53 (11)	430 (89)	UNC, 1.40[†] MH, 0.48[‡]
Combined	Black	15 (8)	176 (92)	. . .

Source.—Reference 15.

* Data are numbers of verdicts. Data in parentheses are percentages of the totals.

[†] UNC = uncorrected.

[‡] MH = Mantel-Haenszel estimate.

larger than the RR, as we now expect. In the experimental study example, the OR of 1.61 is only slightly larger than the RR of 1.56. This is because the probability of mammography (the outcome) is rare at 6.7 per 100. In contrast, the cohort study yields an OR of 1.25, which is considerably larger than the RR of 1.07, with the high overall probability of mammography of 73 per 100 explaining the larger difference.

CONFOUNDING VARIABLES AND THE SIMPSON PARADOX

An important consideration in any study design is the possibility of confounding by additional variables that mask or alter the nature of the relationship between an explanatory variable and the outcome. Consider data from three binary variables, the now familiar X and Y, as well as a new variable Z. These can be represented in two separate 2 × 2 contingency tables: one for X and Y at level 1 of variable Z and one for X and Y at level 2 of variable Z. Alternatively, the X and Y data can be represented in a single table that ignores the values of Z (ie, pools the data across Z).

The problem occurs when the magnitude of association between X (explanatory) and Y (outcome) is different at each level of Z (confounder). Estimation of the association between X and Y from the pooled 2 × 2 table (ignoring Z) is inappropriate and often incorrect. In such cases, it is misleading to even list the data in aggregate form. There are statistical methods to correct for confounding variables, with the assumption that they are known and measurable. A family of techniques attributed to Cochran (14) and Mantel and Haenszel (15) are commonly used to produce summary estimates of association between X and Y corrected for the third variable Z.

The entire rationale for the use of randomized clinical trials is to eliminate the problem of unknown and/or immeasurable confounding variables. In a clinical trial, subjects are randomly assigned to levels of X (the treatment or independent variable). The outcome (Y) is then measured during the course of the study. The beauty of this scheme is that all potentially confounding variables (Z) are equally distributed among the treatment groups. Therefore, they cannot affect the estimate of association between treatment and outcome. For randomization to be effective in elimination of the potential for confounding, it must be successful. This requires scrupulous attention to detail by investigators in conducting, verifying, and documenting the randomization used in any study.

It is possible for the relationship between X and Y to actually change direction if the Z data are ignored. For instance, OR calculated from pooled data may be less than 1.0, while OR calculated with the separate (so-called stratified) tables is greater than 1.0. This phenomenon is referred to as the Simpson paradox, as Simpson is credited with an article in which he explains mathematically how this contradiction can occur (16). Such paradoxic results are not only numerically possible but they actually arise in real-world situations and can have profound social implications.

Agresti (13) illustrated the Simpson paradox by using death penalty data for black and white defendants charged with murder in Florida between 1976 and 1987 (17). He showed that when the victim's race is ignored, the percentage of death penalty sentences was higher for white defendants. However, after controlling for the victim's race, the percentage of death penalty sentences was higher for black defendants. The paradox arose because juries applied the death penalty more frequently when the victim was white, and defendants in such cases were mostly white. The victim's race (Z) dominated the relationship between the defendant's race (X) and the death penalty verdict (Y). Table 8 lists the data stratified by the victim's race and combined across the victim's race. As indicated in the table, unadjusted RR of receiving the death penalty (white/black) with white victims is 0.495; with black victims, 0.0; and with combined victims, 1.40. The Mantel-Haenszel estimate of the common RR is 0.48, thus solving the paradox. The method for calculating the Mantel-Haenszel summary estimates is beyond the scope of this article. However, confounding variables, the Simpson paradox, and how to handle them are discussed in any comprehensive text about biostatistics or medical research design (5,6).

CONCLUSION

This article has focused on statistical analysis of count data that arise from the four basic designs used in medical research (ie, cross-sectional, case-control, cohort, and experimental study designs). Each of these designs often yields data that are best summarized in a 2 × 2 contingency table. Statistics calculated from such tables include cell, row, and column proportions; differences in proportion; RRs; and ORs. These statistics are used to estimate associated population parameters and are selected to suit the specific aims and design of the study. For inference concerning association between study variables, one must use the correct statistic, allow for variability, and account for any confounding variables.

References

1. Langlotz CP. Fundamental measures of diagnostic examination performance: usefulness for clinical decision making and research. Radiology 2003; 228:3–9.
2. Applegate KE, Tello R, Ying J. Hypothesis testing III: counts and medians. Radiology 2003; 228:603–608.
3. Schulman KA, Berlin JA, Harless W, et al. The effect of race and sex on physicians' recommendations for cardiac cath-

eterization. N Engl J Med 1999; 340:618–626.

4. Schwartz LM, Woloshin S, Welch HG. Misunderstandings about the effects of race and sex on physicians' referrals for cardiac catheterization. N Engl J Med 1999; 341:279–283.
5. Motulsky H. Intuitive biostatistics. Oxford, England: Oxford University Press, 1995.
6. Riegelman RK. Studying a study and testing a test. 4th ed. Philadelphia, Pa: Lippincott Williams & Wilkins, 2000.
7. Sackett DL. Clinical epidemiology: a basic science for clinical medicine. 2nd ed. Boston, Mass: Little, Brown, 1991.
8. Medina LS, Zurakowski D. Measurement variability and confidence intervals in medicine: why should radiologists care? Radiology 2003; 226:297–301.
9. Cogo A, Bernardi E, Prandoni P, et al. Acquired risk factors for deep-vein thrombosis in symptomatic outpatients. Arch Intern Med 1994; 154:164–168.
10. Vachon CM, Sellers TA, Vierkant RA, Wu FF, Brandt KR. Case-control study of increased mammographic breast density response to hormone replacement therapy. Cancer Epidemiol Biomarkers Prev 2002; 11:1382–1388.
11. Burman ML, Taplin SH, Herta DF, Elmore JG. Effect of false-positive mammograms on interval breast cancer screening in a health maintenance organization. Ann Intern Med 1999; 131:1–6.
12. Harrison RV, Janz NK, Wolfe RA, et al. Personalized targeted mailing increases mammography among long-term noncompliant medicare beneficiaries: a randomized trial. Med Care 2003; 41:375–385.
13. Agresti A. Categorical data analysis. Hoboken, NJ: Wiley, 2003.
14. Cochran WG. Some methods for strengthening the common chi-squared tests. Biometrics 1954; 10:417–451.
15. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 1959; 22:719–748.
16. Simpson EH. The interpretation of interaction in contingency tables. J R Stat Soc B 1951; 13:238–241.
17. Radelet ML, Pierce GL. Choosing those who will die: race and the death penalty in Florida. Florida Law Rev 1991; 43:1–34.