



Measures of Association and Hypothesis Testing

by Deborah Rosenberg, PhD and Arden Handler, DrPH

Once best estimates are chosen, both from a statistical and epidemiologic perspective, hypotheses about the estimated association between a single mean, proportion, or rate and a fixed value, typically a standard or goal, or about the estimated association between two or more means, proportions, or rates can be tested. Hypotheses address questions about whether any observed differences in health status, risk and protective factors, health services or health systems indicators are real or spurious.

Statistics to test hypotheses take the following general form:

$$\text{Test Statistic} = \frac{\text{Observed Association} - \text{Expected Association}}{\text{Standard Error of the Association}}$$

Here, the observed association is the comparison of a single estimate and a standard or the comparison of two or more estimates, and the expected association reflects the same comparison but under what is called the *null hypothesis* or the assumption that no association exists.

Observed associations and the test statistics used for evaluating them follow known probability distributions. Critical values can be chosen to assess the probability that the observed association is different from its expected value. As with single means, proportions, or rates, measures of association are statistically unbiased estimates of the true population experience. They estimate average association just as single measures estimate average occurrence.

There are two types of error that can be made when testing hypotheses:

Type I Error: Concluding that there is an association when one does not exist

Type II Error: Concluding that there is no association when one does exist

The probability of interpreting an observed association as real when it is in fact due to chance, or the probability of committing a Type I error, is what is reported as the p-value.

By convention, associations are considered statistically significant when the p-value is < 0.05 , although other values are also used depending on the analysis. Since, as we've already seen, 95% of the probability in the standard normal distribution lies between -1.96 and 1.96, these are typically chosen as critical values. Any result from a statistical test based on the standard normal curve which is less than -1.96 or greater than 1.96 means that there is less than a 5% probability ($p < 0.05$) that observed differences are due to chance alone.

In addition to using p-values, confidence limits can also be calculated around measures of association. These confidence limits define a range of values within which the true population association is likely to lie, and tell us about the reliability of our data.

The confidence intervals for many measures of association take the general form:

$$CI = \text{Observed Association} \pm \text{Critical Value} \times \text{Standard Error of the Association}$$

MEASURES OF ASSOCIATION: DIFFERENCE MEASURES

Differences between means, proportions, and rates can theoretically range from negative infinity to positive infinity ($-\infty$ to ∞), with their expected values, assuming no association, being 0. In other words, one of the means (or proportions) being compared may theoretically be either infinitely less than or infinitely greater than the other one, and if the two means (or proportions) are equal, then the difference between them will be 0. The distribution of a difference measure is approximately normal if sample size is reasonably large. Statistical testing, therefore, can usually rely on the known values of the standard normal curve.

Two Independent Means

For two independent means, the measure of association is their observed difference:

$$\bar{X}_1 - \bar{X}_2$$

The expected association under the null hypothesis is that $\bar{X}_1 - \bar{X}_2 = 0$. Then, the formula that gives a test of statistical significance follows the t distribution, a generalization of the normal distribution, as follows:

$$t = \frac{\bar{X}_1 - \bar{X}_2 - 0}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

where S is the standard deviation of the observed values

Here, the observed value is the actual difference between the two means and the expected value is 0 under the null hypothesis. The standard error of the measure of association is the sum of the standard errors of the two means being compared. Assuming that the two variances are actually the same as shown above, a common variance can be estimated by taking the average of the two sample variances in the data, weighted by the number of observations contributing to each mean.

The formula for the confidence interval around the difference between two independent means is as follows:

$$(\bar{X}_1 - \bar{X}_2) \pm 1.96 \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Now the formula for the standard error is slightly different than that used for the statistical test. When calculating confidence limits, we are simply estimating the error around the observed association without any reference to a null hypothesis, and therefore the constraint of a common variance imposed on the statistical test does not apply. Here, the standard error is simply the unweighted sum of the variances of the two means being compared.

Suppose we want to compare the mean number of well child visits received by children in the first two years of life in two counties. The mean in County A is observed to be 6.2 with a standard deviation of 3.4 (square root of the variance) based on a sample of 200 two-year-olds. The mean in County B is observed to be 5.3 with a standard deviation of 4.5 based on a sample of 150 two-year-olds. On average, are the children in the two counties receiving the same amount of well child care? The test statistic to estimate the answer to this question is:

$$\begin{aligned} t &= \frac{(6.2 - 5.3) - 0}{\sqrt{\frac{199 \times 3.4^2 + 149 \times 4.5^2}{200 + 150 - 2} \left(\frac{1}{200} + \frac{1}{150} \right)}} \\ &= \frac{0.9 - 0}{\sqrt{15.3 \left(\frac{1}{200} + \frac{1}{150} \right)}} \\ &= \frac{0.9}{0.42} = 2.1 \end{aligned}$$

The common variance is 15.3, or a standard deviation of 3.9, a weighted average of the standard deviations of 3.4 in County A and 4.5 in County B. The weights are the proportion of the total sample contributed by each county—57% $\left(\frac{199}{348} \times 100 \right)$ for County A and 43% $\left(\frac{149}{348} \times 100 \right)$ for County B.

The result of 2.1 tells us that the observed difference in the two means is slightly more than 2 standard deviations away from the expected difference of 0. According to the values on the standard normal curve (equivalent to the t distribution with this sample size), there is only a 2% probability of seeing this test result or one even farther away from 0 if, in fact, the two-year-olds in County A and County B are receiving the same number of well child visits on average. In other words, it is very unlikely ($p \approx 0.02$) that the difference we observe is due solely to chance. Since $2.1 > 1.96$, we report that County A's mean is significantly different than County B's ($p < 0.05$).

The confidence interval around the difference between the two means is:

$$\begin{aligned}
CI &= (6.2 - 5.3) \pm 1.96 \sqrt{\frac{3.4^2}{200} + \frac{4.5^2}{150}} \\
&= 0.9 \pm 1.96 \sqrt{0.19} \\
&= 0.9 \pm 0.85 \\
&= (0.05 - 1.75)
\end{aligned}$$

Since the confidence interval of 0.05-1.75 does not include the expected value of 0, the two means are shown to be statistically different.

Two Independent Proportions

For two proportions, there are several measures of association that can be examined. Epidemiology uses the 2×2 table as the framework for constructing these measures:

		Disease or Other Health Outcome		
		Yes	No	
Exposure or Person, Place, or Time Characteristic	Yes	a	b	a + b (n ₁)
	No	c	d	c + d (n ₂)
		a + c (m ₁)	b + d (m ₂)	a + b + c + d N

Comment

If the proportions are incidence density data, the 2×2 table is actually slightly different than above because the denominators are not counts of individuals; n_1 and n_2 are person-time data and therefore quantities for cells b and d cannot be calculated. Public health data, however, are typically cumulative incidence or prevalence data, and so the focus will be on the 2×2 table as shown in the above figure.

For the moment, we will assume that the data in the 2×2 table are either complete population data or data from a random sample. In other words, N is the population or sample total, n_1 and n_2 are the population or sample totals of those with and without the exposure, and m_1 and m_2 are the population or sample totals of those with and without the disease. These are the kind of data typically encountered by public health agencies for use in surveillance and monitoring. Later, we will discuss changes in the 2×2 table when data are not population data or data from a random sample, but are instead from more complex sampling designs used in some surveys and research studies.

Notice, that for public health purposes, the exposure variable in the 2×2 table may be a medical risk factor, a social or behavioral risk factor, a person characteristic, the place of occurrence, the time of occurrence, or a person-place-time characteristic. The health status outcome variable may be a disease, a behavior, a measure of access to care or service utilization, or any other measure that is hypothesized to be an outcome of the "exposure". In addition, the 2×2 table can be used to examine data derived from

ecologic analyses in which information is for exposures and outcomes for groups rather than individuals (e.g., percent uninsured in each area and % of children with asthma in each area). The framework of the 2×2 table accommodates data from a wide variety of sources and is a valuable tool for organizing data for a wide variety of analyses.

Examples of 2×2 Tables

1. Individual Level Data on a Risk Factor and a Health Outcome

		Health Outcome		
		Yes	No	
Smoking	Yes	a	b	n_1
	No	c	d	n_2
		m_1	m_2	N individuals

2. Individual Level Data on a Person Characteristic and a Health Outcome

		Health Outcome		
		Yes	No	
Low Income	Yes	a	b	n_1
	No	c	d	n_2
		m_1	m_2	N individuals

3. Individual Level Data on Place of Residence and a Health Outcome

		Health Outcome		
		Yes	No	
Residence	County A	a	b	n_1
	County B	c	d	n_2
		m_1	m_2	N individuals

4. Individual Level Data on Time of Occurrence and a Health Outcome

		Health Outcome		
		Yes	No	
Occurrence	Time 1	a	b	n_1
	Time 2	c	d	n_2
		m_1	m_2	N individuals

5. Individual Level Data on a Person-Place-Time Characteristic and a Health Outcome

		Health Outcome		
		Yes	No	
Residence in Low-Income Area	Yes	a	b	n_1
	No	c	d	n_2
		m_1	m_2	N individuals

6. Ecologic Analysis

Aggregate Level Data on a Risk Factor and a Health Outcome

		Health Outcome		
		%	%	
% Smokers in Area		High	Low	
		> 20 %	a	
$\leq 20\%$	c	d	n_2	
		m_1	m_2	N areas

7. Ecologic Analysis

Aggregate Level Data on a Person-Place-Time Characteristic and a Health Outcome

		Health Outcome		
		%	%	
% Low-Income in Area		High	Low	
		> 20 %	a	
$\leq 20\%$	c	d	n_2	
		m_1	m_2	N areas

Comment

An important distinction must be made between ecologic variables and ecologic analysis. Table 5 above illustrates what we have termed a person-place-time variable, or a person characteristic aggregated in units of geography and time. Person-place-time variables are examples of ecologic variables. The variables "residence in a low income community" or "percent low-income in the community" have values that can be assigned to each individual in a sample, although the information they contain pertains to the environment within which the individual lives. Risk factor and outcome variables can also be ecologic variables if they are aggregates in time and place. For example, "prevalence of smoking among women giving birth in a community", "county prevalence of unintended pregnancy", or "city infant mortality rate" are ecologic variables with values that can be assigned to individuals.

A data-set in which the observations, or units of analysis, are unique individuals may contain a mixture of individual level variables and ecologic variables. For example, values for smoking status during pregnancy and delivery of a low birthweight infant, as well as values for the smoking prevalence among women delivering in the individual's county and the low birthweight rate in the individual's county might be collected for each individual. The data for five individuals from a hypothetical MCH individual level data-set might look like:

Individual	Variables			
	Smoking during pregnancy	Infant Birthweight < 2500 grams	Smoking Prevalence in Community	Low Birthweight Prevalence in Community
1	Yes	No	15 %	8%
2	No	Yes	10 %	6%
3	No	No	20 %	10%
4	Yes	Yes	20 %	9%
5	No	No	15 %	7%

Including ecologic variables in an analysis of individuals is not equivalent to conducting an ecologic analysis. In an ecologic analysis, the observations in the data-set, or units of analysis, are either geographic areas, health facilities, schools, time periods, or other aggregates. No data for individuals are available. Instead, analysis is focused on the relationship between exposure and outcome for the groups in the data-set and not directly on the individuals they encompass.

The data for five communities from a hypothetical aggregate level MCH data-set might look like:

Community	Variables	
	Smoking Prevalence in Community	Low Birthweight Prevalence in Community
1	15 %	8%
2	10 %	6%
3	20 %	10%
4	20 %	9%
5	15 %	7%

Analogous to the measure of association for two means, comparison of two proportions may also be evaluated using a difference measure:

$$\frac{a}{n_1} - \frac{c}{n_2} = p_1 - p_2 \text{ or } r_1 - r_2$$

And the test of statistical significance can either be in a form that follows the χ^2 or the z (standard normal) distribution:

		Outcome		
		Yes	No	
Risk Factor	Yes	a	b	n_1
	No	c	d	n_2
		m_1	m_2	N

$$\chi^2 = \sum_{i=a}^d \frac{(O_i - E_i)^2}{E_i}$$

where O_i = the observed value in each cell or

$$z = \frac{(p_1 - p_2) - 0}{\sqrt{p_0(1-p_0) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

and $E_i = \frac{\text{row total} \times \text{column total}}{N}$

With the Poisson assumption for rates, the formula can be rewritten:

$$z = \frac{r_1 - r_2 - 0}{\sqrt{r_0 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

For the z tests, p_1 and $r_1 = \frac{a}{n_1}$, p_2 and $r_2 = \frac{c}{n_2}$, and p_0 and $r_0 = \frac{m_1}{N}$

The formula using the chi-square distribution compares the observed and expected value of the counts in each cell of the 2×2 table, while the z test compares the observed and expected value of the difference in proportions. The two tests are equivalent; that is, for the 2×2 table, $\chi^2 = z^2$. The p-values associated with these two tests are also equivalent and therefore the conclusions drawn will be identical.

As in the t test for means, the observed value in the z test is the actual difference between the two proportions and the expected value is 0 under the null hypothesis. The standard error of the measure of association is the sum of the standard errors of the proportions being compared. Assuming that the two variances are actually the same, a common variance can be estimated by taking the average of the two sample variances in the data, weighted by the number of observations contributing to each proportion.

The formulas for calculating confidence intervals around the difference between two independent proportions are:

Binomial : $CI = p_1 - p_2 \pm 1.96 \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$

Poisson : $CI = r_1 - r_2 \pm 1.96 \sqrt{\frac{r_1}{n_1} + \frac{r_2}{n_2}}$

Once again, the formulas for the standard errors are somewhat different than for the statistical tests because the null hypothesis is not being tested.

Comment

In practice, the formula for the standard error which simply sums the variances of the means or proportions being compared is often used for calculating test statistics as well as for confidence intervals. This is a reasonable approach since the results of using this formula and those from using the one requiring estimation of a common variance are usually very similar.

Let's return to the example of County A with its infant mortality rate of 8 per 1,000 based on 1,375 live births. Suppose that County B has an infant mortality rate of 6.7 based on 900 births. On average, is the infant mortality experience in these two counties different? Below is the 2×2 table to examine these data more carefully:

		Infant Death		
		Yes	No	
County	A	11	1,364	1,375
	B	6	894	900
		17	2,258	2,275

We can use the χ^2 formula for testing the difference between these two rates. In order to do this we first need to calculate the expected values for each cell in the table:

$$E_1 = \frac{17 \times 1,375}{2,275} = 10.3, \quad E_2 = \frac{2,258 \times 1,375}{2,275} = 1,364.7$$

$$E_3 = \frac{17 \times 900}{2,275} = 6.7, \quad E_4 = \frac{2,258 \times 900}{2,275} = 893.3$$

Then the test statistic to estimate whether County A and County B have different infant mortality rates is:

$$\begin{aligned} \chi^2 &= \frac{(11 - 10.3)^2}{10.3} + \frac{(1,364 - 1,364.7)^2}{1,364.7} + \frac{(6 - 6.7)^2}{6.7} + \frac{(894 - 893.3)^2}{893.3} \\ &= 0.7^2 \left(\frac{1}{10.3} + \frac{1}{1,364.7} + \frac{1}{6.7} + \frac{1}{893.3} \right) \\ &= 0.12 \end{aligned}$$

We can also use the standard normal distribution to test the difference between the two rates. Since these are cumulative incidence rates of a fairly rare event (infant death), we will use the formula with the Poisson assumption. In order to do this we first need to calculate the estimate of the common variance (r_0) that, in this example, is the overall infant mortality rate for both counties combined:

$$\frac{m_1}{N} \times 1,000 = \frac{17}{2,275} \times 1,000 = 7.5$$

Then the test statistic to estimate whether County A and County B have different infant mortality rates is:

$$z = \frac{8 - 6.7 - 0}{\sqrt{7.5 \left(\frac{1}{1,375} + \frac{1}{900} \right)} \times 1,000}$$

$$= 0.35$$

Remember that since the two rates are in integer form (8 and 6.7) rather than decimal form (0.008 and 0.0067), we need to use the multiplier of 1,000 in the denominator.

And 0.35^2 , or z^2 , = 0.12, the same as the chi-square value we obtained above.

The result of 0.35 means that the observed difference in the two rates is much less than 1 standard deviation away from the expected difference of 0. According to the values on the standard normal curve, there is a 73% probability of seeing this test result or one even farther away from 0 if, in fact, the infant mortality experience in County A and County B is actually the same on average. In other words, it is very likely ($p \approx 0.73$) that the difference we observe is due solely to chance. Since $0.35 < 1.96$, we report that County A's infant mortality rate is not significantly different than County B's ($p > 0.05$).

The confidence limits around the difference in the two infant mortality rates is:

$$CI = (8 - 6.7) \pm 1.96 \sqrt{\left(\frac{8}{1,375} + \frac{6.7}{900} \right) \times 1,000}$$

$$= 1.3 \pm 1.96 \sqrt{13.26}$$

$$= 1.3 \pm 7.14$$

$$= (-5.8 - 8.4)$$

Since the confidence interval of $-5.8-8.4$ includes the expected value of 0, the county rates are not statistically different.

The Attributable Risk

The difference between two proportions is given a specific interpretation in epidemiology called the *attributable risk* (AR). The attributable risk quantifies the number of occurrences of a health outcome that are due to, or can be attributed to, the exposure or risk factor. In order for this interpretation to apply, the relationship between the risk factor and outcome must be hypothesized to be causal. Looking at a few examples will better illustrate this point.

Consider the association between smoking and low birthweight shown in the 2×2 table below:

		Low Birthweight		
		Yes	No	
Smoker	Yes	280	2,720	3,000
	No	720	11,280	12,000
		1,000	14,000	15,000

The proportion of smokers who deliver low birthweight infants is $280/3,000$ or 9.33% and the proportion of non-smokers who deliver low birthweight infants is $720/12,000$ or 6.0%. The difference between these two proportions, or the attributable risk is:

$$\text{Attributable Risk} = p_1 - p_2 = 9.33 - 6.0 = 3.33$$

Comment

The test of statistical significance for the difference between these 2 binomial proportions is:

$$z = \frac{9.33 - 6.0}{\sqrt{6.67 \times 93.33 \left(\frac{1}{3,000} + \frac{1}{12,000} \right)}} = 6.5$$

which is highly significant ($p < 0.0001$) since 6.5 is much greater than the critical value of 1.96.

The attributable risk defines the excess risk of delivering a low birthweight infant among the *group with the risk factor* of interest, in this case the women who smoke. For this example, 3.33 out of every 100 live births to smokers are infants whose low birthweight can be attributed to smoking.

From a public health perspective, it is often more useful to re-define the attributable risk in terms of the *whole population*, in this case all live births regardless of whether their mothers smoke. This modified measure is called the population attributable risk (PAR).

Instead of comparing the outcomes for those with and without a risk factor, as is done for statistical testing (and the AR), the PAR compares the outcomes for the whole population to those without the risk factor. In this example, the proportion of all women delivering who had a low birthweight infant is $1,000/15,000$ or 6.67%, and, as we've already seen, the proportion of non-smokers delivering a low birthweight infant is 6.0%. The PAR is:

$$\text{Population Attributable Risk} = p_0 - p_2 = 6.67 - 6.0 = 0.67$$

Here the excess in low birthweight is expressed in terms of the experience of all live births. For this example, 0.67 out of every 100 live births are infants whose low birthweight can be attributed to smoking.

For both the attributable risk (AR) and the population attributable risk (PAR), the low birthweight percent among the non-smokers, or 6.0%, is considered the low birthweight percent expected among all live

births if no women smoked during pregnancy. If the prevalence of smoking were reduced to 0, the 2×2 table would become a 1×2 table since there would no longer be any exposed group:

		Low Birthweight		
		Yes	No	
Smoking	No	900	14,100	15,000
		900	14,100	15,000

The 900 expected low birthweight infants in the table is arrived at by multiplying the expected population proportion of 0.06 by the 15,000 total live births:

$$0.06 \times 15,000 = 900$$

which is 100 fewer than the 1,000 low birthweight infants seen in the original 2×2 table. The 100 preventable low birthweight births can also be found directly by multiplying the AR by the number of smokers ($0.0333 \times 3,000=100$), or by multiplying the PAR by the total number of live births ($0.0067 \times 15,000=100$).

Thus far, we have used the attributable risk or the population attributable risk to assess the impact of completely eliminating a risk factor. The 2×2 table can also be used as a tool to examine the impact of reducing the prevalence of a risk factor to something more than 0. This fits a more realistic scenario in which a public health intervention might be able to reduce partially, though not completely, the prevalence of a risk factor.

In the smoking and low birthweight example, the observed prevalence of smoking is 20% ($3,000/15,000$), and from a public health perspective it would be important to estimate the impact of reducing this to, say, 15%. The numbers in the 2×2 table can be rearranged to conform to this hypothetical situation:

		Low Birthweight		
		Yes	No	
Smoker	Yes	210	2,040	2,250
	No	765	11,985	12,750
		975	14,025	15,000

The percent low birthweight among smokers and non-smokers has not been changed and neither, therefore, has the attributable risk which is still 3.33 per 100 births among the smokers. The number of low birthweight births, however, has dropped from 1,000 to 975, resulting in an overall low birthweight percent of 6.5. Only the prevalence of smoking has changed, from 20% ($3,000/15,000$) to 15% ($2,250/15,000$), and this 5% decrease means 25 fewer low birthweight births.

By having an impact on the prevalence of smoking among pregnant women, the population attributable risk has changed to:

$$\text{Population Attributable Risk} = p_0 - p_2 = 6.5 - 6.0 = 0.5$$

Now, only 0.5 out of every 100 births are infants whose low birthweight is attributed to smoking during pregnancy. The population attributable risk, then, is a function of the attributable risk and the prevalence

of the risk factor in the population. It makes sense that as the prevalence of a risk factor decreases, the less of a contribution it makes to the overall occurrence (p_0) of an adverse outcome. The PAR can be rewritten as:

$$\text{Population Attributable Risk} = \text{Attributable Risk} \times \text{Prevalence of Risk}$$

$$p_0 - p_2 = (p_1 - p_2) \times \frac{n_1}{N}$$

For example :

$$6.67 - 6.0 = 3.33 \times 0.20 = 0.67$$

and following the reduction in smoking prevalence :

$$6.5 - 6.0 = 3.33 \times 0.15 = 0.50$$

Two other measures based on the attributable risk and the population attributable risk, are perhaps the most useful in describing the importance of a risk factor's contribution to an adverse outcome. These two measures are the attributable risk percent (AR%) and the population attributable risk percent (PAR%). The AR% simply converts the attributable risk into the percent of adverse outcomes preventable in the group with the risk factor, while the PAR% simply converts the population attributable risk into the percent of adverse outcomes preventable in the entire population at risk. The conversion into percents makes these measures more readily understandable and therefore more meaningful to report than the AR and PAR themselves.

Continuing with the smoking and low birthweight data, the formulas for the AR% and PAR% are:

$$\text{Attributable Risk \%} = \frac{p_1 - p_2}{p_1} \times 100 = \frac{3.33}{9.33} \times 100 = 36\%$$

$$\text{Population Attributable Risk \%} = \frac{p_0 - p_2}{p_0} \times 100 = \frac{0.67}{6.67} \times 100 = 10\%$$

Thirty-six percent of the low birthweight infants born to the 3,000 smokers may be attributable to smoking; 10% of the low birthweight infants born to the 15,000 pregnant women may be attributable to smoking.

Alternatively, from the 1 x 2 table that resulted when the prevalence of smoking was hypothetically reduced to 0, we saw that the number of low birthweight births was reduced by 100. The AR% is this number of potentially preventable adverse outcomes divided by the 280 smokers who delivered low birthweight infants ($100/280=0.36$) and the PAR% is this number of potentially preventable adverse outcomes divided by the total of 1,000 women who delivered low birthweight infants ($100/1,000=10\%$).

The meaning of the attributable risk and its associated measures shifts when the exposure is not a risk factor. Suppose, instead of smoking, the rows of the 2 x 2 table with low birthweight are County A and County B. While it would be inappropriate to consider a county to be a cause of low birthweight, and therefore to attribute 3.3 low birthweight infants per 100 births directly to some feature of residence in County A, this value can be used to provide a practical interpretation to the significant statistical test results that were previously obtained. It might be fair to state that, on average, County A has an excess of 36% low birthweight births compared to County B.

Test Yourself

Question:

Using the data below for the same population of 15,000 live births described above, calculate the attributable risk, the population attributable risk, the attributable risk % and the population attributable risk % for the association of cocaine use during pregnancy and low birthweight.

		Low Birthweight		
		Yes	No	
Cocaine Use	Yes	130	370	500
	No	870	13,630	14,500
		1,000	14,000	15,000

Answer:

LBW among cocaine users is $130/500 \times 100$ or 26% (p_1)

LBW among non-users is $870/14,500 \times 100$ or 6% (p_2)

Overall LBW is $1,000/15,000 \times 100$ or 6.67 % (p_0)

$$AR = 26.0 - 6.0 = 20.0$$

$$PAR = 6.67 - 6.0 = 0.67$$

$$AR\% = \frac{20.0}{26.0} \times 100 = 77\%$$

$$PAR\% = \frac{0.67}{6.67} \times 100 = 10\%$$

Note that the PAR and PAR% in this example are exactly the same as those calculated for the association between smoking and low birthweight in the same population. This is because, while cocaine use is a much more potent risk factor for low birthweight than smoking and therefore the cocaine AR and AR% are much higher, the prevalence of cocaine use is much lower than the prevalence of smoking. These two forces—the magnitude of the risk difference and the magnitude of the difference in prevalence—offset each other, leading to the same PAR and PAR%.

Hypotheses about a Mean, Proportion, or Rate in Comparison to a Standard

Thus far, excess risk, causal or otherwise, has been defined in terms of the observed proportion for the lower risk group. In the low birthweight example on page 57, the low birthweight percent in the non-smokers was, in effect, the expected value; in the infant mortality example on pages 55-56, the infant mortality rate in County B was, in effect, the expected value. Often, instead of defining the expected value in these terms, it is useful to use a standard. The standard might be the current level of an outcome in the nation or the state, or a national or state goal.

The test of significance for the difference between a proportion and a standard is:

$$z = \frac{p_1 - \text{Standard}}{\sqrt{\frac{\text{Standard}(1 - \text{Standard})}{n_1}}} \quad \text{or} \quad z = \frac{r_1 - \text{Standard}}{\sqrt{\frac{\text{Standard}}{n_1}}}$$

This formula is exactly the same as that for the difference between two proportions if we treat the standard as though it was a proportion from an infinite (or very large) population. With this assumption of extremely large sample size, the value of the standard dominates the overall proportion or rate, and its standard error equals (or approaches) 0. The formula for comparing two proportions, then, reduces to the one above. For example, comparing County A's infant mortality rate to the Year 2000 Objective of 7 per 1,000 live births, assuming the standard is based on a population of 10,000,000 live births, the data can be arranged in a 2×2 table as follows:

		Infant Death		
		Yes	No	
County A	11	1,364	1,375	
Year 2000 Objective	70,000	9,930,000	10,000,000	
	70,011	9,931,364	10,001,375	

The equivalence of the test for two proportions (or two rates) with the test for a proportion (or rate) and a standard can now be shown. The following example is for the case of a Poisson variable (rate):

The formula for testing the difference between two rates is (see page 54)

$$\frac{8-7}{\sqrt{7\left(\frac{1}{1,375} + \frac{1}{10,000,000}\right)} \times 1,000}$$

and it reduces to the formula for testing the difference between one rate and a standard

$$\begin{aligned} &\equiv \frac{8-7}{\sqrt{\frac{7}{1,375}} \times 1,000} \\ &= 0.44 \end{aligned}$$

The overall proportion (70,011/10,001,375) is essentially equal to the standard, and the standard error of the standard (7/10,000,000) essentially equals 0. Similar to the results when comparing County A to County B (pages 55-56), the conclusion from this statistical test is that County A's rate is not different from the Year 2000 Objective; since $0.44 < 1.96$, the critical value from the standard normal distribution.

Comment

If the standard to be used is not simply a goal, but actual data from the nation or the state, the comparison is technically not between two independent proportions since the data for the geographic area of interest is also data included in the standard. This violates one of the assumptions underlying the statistical test. When the estimates are not independent, the standard error of their difference will be underestimated by the usual calculation. On the other hand, the difference itself will be underestimated since the inclusion of the local data in the standard makes the standard and the local area more similar. These two errors—the underestimate of the standard error and the underestimate of the difference itself—may counteract each other, but the analyst must be aware of the potential for spurious results. If the population size of the geographic area of interest is much smaller than that of the standard, it is fairly safe to ignore the assumption of independence and proceed with statistical testing as usual.

Example

A statistical test of the difference between the 6.67% low birthweight seen previously in the hypothetical population of 15,000 live births and the Year 2000 Objective of 5 % is calculated as follows [for low birthweight, we use the formula for a binomial variable (percent)]:

$$z = \frac{6.67 - 5}{\sqrt{\frac{5.0(100 - 5.0)}{15,000}}} = 9.6$$

Remember that since the proportion and the standard are in integer form (6.67 and 5) rather than decimal form (0.067 and 0.05), we need to use the multiplier of 100 in the denominator.

Since 9.6 is greater than the critical value of 1.96, the extent of low birthweight births in this population is significantly higher than the Year 2000 Objective. Because the sample size is large (15,000 live births), the estimate of 6.7% is very reliable and so we are quite confident that it is truly different from the goal of 5%. If this same test were carried out for a much smaller population of, say, 500 live births with the same estimate of 6.7 % low birthweight, the results would be:

$$z = \frac{6.67 - 5}{\sqrt{\frac{5.0(100 - 5.0)}{500}}} = 1.7$$

Since 1.7 is less than the critical value of 1.96, the extent of low birthweight births in this population is not shown to be significantly different than the Year 2000 Objective. Because the sample size is small (500 live births) and much less reliable than for the larger population, we are not as confident that the estimate of 6.67% is truly different from the goal of 5%.

MEASURES OF ASSOCIATION: RATIO MEASURES

The Relative Risk, The Relative Prevalence, and The Odds Ratio

In addition to the measures of association based on differences between means, proportions, or rates, epidemiology has contributed two measures of association based on ratios. The relative risk (RR) is the cumulative incidence in the exposed population divided by the cumulative incidence in the unexposed. When the data are based on prevalence rather than incidence, this measure is termed relative prevalence (RP). The relative risk and relative prevalence are calculated as follows:

$$\text{RR and RP} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}} = \frac{\frac{a}{n_1}}{\frac{c}{n_2}} = \frac{r_1}{r_2} \text{ or } \frac{p_1}{p_2}$$

The other measure contributed by epidemiology, the *odds ratio*, is the odds of having the outcome in the exposed divided by the odds of having the outcome in the unexposed. The odds ratio does not use the population denominators of n_1 and n_2 , but just the numbers in the cells of the 2×2 table. It approximates the relative risk when the health outcome is rare:

$$\text{OR} = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{ad}{bc}$$

To see why the odds ratio approximates the relative risk when the health outcome is rare, consider an example with child injury mortality. Suppose the child injury mortality rate in one state is 34 per 100,000 children, and 26 per 100,000 children in another. For illustration, assume that each state has a child population of 100,000. Then the 2×2 table to compare these two mortality rates would look as follows:

		Child Injury Mortality		
		Yes	No	
State I	Yes	34	99,966	100,000
State II	No	26	99,974	100,000
		60	199,940	200,000

$$RR = \frac{\frac{34}{100,000}}{\frac{26}{100,000}} = \frac{34}{26} = 1.3$$

$$OR = \frac{\frac{34}{99,966}}{\frac{26}{99,974}} = \frac{34.01}{26.01} = 1.3$$

When the incidence or prevalence rate of an outcome in a population is very small, the difference between using a + b and c + d, or only b and d in the denominators is negligible. As the incidence or prevalence becomes higher, however, the difference between the two measures becomes greater with the odds ratio always being greater than the relative risk.

Suppose the data in the above example are for the number of two-year-old children receiving fewer than five well child visits so that the prevalence rates are 34 per 100 two-year-olds (34%) and 26 per 100 two-year-olds (26%) instead of 34 and 26 per 100,000:

		< 5 Well Child Visits in the First Two Years of Life		
		Yes	No	
State I	Yes	34,000	66,000	100,000
State II	No	26,000	74,000	100,000
		60,000	140,000	200,000

$$RR = \frac{\frac{34,000}{100,000}}{\frac{26,000}{100,000}} = \frac{34}{26} = 1.3$$

$$OR = \frac{\frac{34,000}{66,000}}{\frac{26,000}{74,000}} = \frac{51.5}{35.1} = 1.5$$

This more commonly occurring outcome yields a relative risk and odds ratio that are noticeably different. With a slightly less common occurrence such as low birthweight which, although measured per 100 births, typically ranges from 5% to 15 %, the relative risk and odds ratio will be quite close. If we were to compare low birthweight rates of 13 % and 10 %, for example, the odds ratio would be 1.34, barely larger than the relative risk of 1.3.

Because the relative risk is a comparison of two cumulative incidence rates, and the odds ratio is a comparison of two odds, the interpretation of the results is somewhat different. The relative risk is a direct comparison of probabilities so that we can say from our example with childhood injury that the children in State I have 1.3 times the risk (probability) of dying from an injury than the children in State II. On the other hand, the odds ratio is not a direct comparison of probabilities so we say that the children in State I are 1.5 times as likely to die from an injury as the children in State II. When relative prevalence is used, the comparison is between two probabilities of "having" a health status outcome.

The relative risk and relative prevalence are generally considered to be preferable to the odds ratio because they are directly related to the probability of developing or having a health outcome. The odds ratio, however, can be used when sampling strategies aimed at studying rare events yield artificial estimates of n_1 and n_2 making the RR and RP impossible to calculate. In addition, the odds ratio has become accepted as a robust measure in and of itself and is sometimes used even when it would be possible to calculate the RR and RP.

Test Yourself

Question:

Calculate the relative risk (RR) and odds ratio (OR) using the 2×2 table for cocaine use during pregnancy and low birthweight:

		Low Birthweight		
		Yes	No	
Cocaine Use	Yes	130	370	500
	No	870	13,630	14,500
		1,000	14,000	15,000

Answer:

LBW among cocaine users is $130/500 \times 100$ or 26% (p_1)

LBW among non-users is $870/14,500 \times 100$ or 6% (p_2)

Overall LBW is $1,000/15,000 \times 100$ or 6.7% (p_0)

$$RR = \frac{\frac{130}{500}}{\frac{870}{14,500}} = \frac{0.26}{0.06} = 4.3$$

$$OR = \frac{\frac{130}{370}}{\frac{870}{13,630}} = \frac{0.35}{0.064} = 5.5$$

The relative risk, relative prevalence, and odds ratio can theoretically range from 0 to positive infinity, with their expected values, assuming no association, being 1. In other words, the ratio is 0 when the proportion in the numerator is 0 regardless of the value of the proportion in the denominator. The ratio can be infinitely large when the proportion in the numerator is infinitely larger than the one in the denominator. If the two proportions are equal, then the ratio of one to the other is 1.

This non-symmetric distribution, with half of the possible values lying between 0 and 1 and the other half lying between one and infinity, is not easy to evaluate using the conventional statistical tests. These ratio measures are usually transformed, therefore, using the natural logarithm to yield distributions symmetric around an expected value of 0 and approximately normal in shape, analogous to the distributions of the difference measures. For ease of interpretation and reporting, the measures and their confidence limits are transformed back to their original form after performing the desired statistical tests.

The statistical tests for the relative risk and odds ratio are as follows:

Relative Risk and Relative Prevalence

$$z = \frac{\ln\left(\frac{r_1}{r_2}\right) - 0}{\sqrt{\left(\frac{1}{a} \times \frac{b}{n_1}\right) + \left(\frac{1}{c} \times \frac{d}{n_2}\right)}}$$

Odds Ratio

$$z = \frac{\ln\left(\frac{a \times d}{b \times c}\right) - 0}{\sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}}$$

The observed value in the z test is now either the natural logarithm of the relative risk or odds ratio and the expected value is again 0 (the natural logarithm of 1) under the null hypothesis. Standard errors have been determined for the natural logarithm of these ratio measures of association that, as usual, take account of the number of observations that contribute to the two rates, or odds. Note that for a given 2×2 table, the z tests for ratio measures and difference measures will be very close.

The formulas for the confidence limits around the relative risk, relative prevalence, and odds ratio are as follows:

$$CI_{RR \text{ and } RP} = e^{\left(\ln\left(\frac{r_1}{r_2}\right) \pm 1.96 \sqrt{\left(\frac{1}{a} \times \frac{b}{n_1}\right) + \left(\frac{1}{c} \times \frac{d}{n_2}\right)}\right)}$$

$$CI_{OR} = e^{\left(\ln\left(\frac{a \times d}{b \times c}\right) \pm 1.96 \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}\right)}$$

After using critical values from the standard normal curve to calculate confidence limits around the natural logarithm of the relative risk, relative prevalence, or odds ratio, the limits are exponentiated in order to transform them back onto the original scale. Remember that on their original scale, the relative risk, relative prevalence and odds ratio have expected values of 1.

Example

Remember that when we tested the difference between the hypothetical infant mortality rates of 8.0 and 6.7 per 1,000 live births in County A and County B (pages 55-56) the z statistic was 0.35 and the confidence interval around the rate difference was (-5.8-8.4). Now, let's calculate the z test and confidence interval using the relative risk for the same data:

$$z = \frac{\ln\left(\frac{8.0}{6.7}\right) - 0}{\sqrt{\left(\frac{1}{11} \times \frac{1,364}{1,375}\right) + \left(\frac{1}{6} \times \frac{894}{900}\right)}}$$

$$= \frac{0.18}{0.51}$$

$$= 0.35$$

$$CI_{RR} = e^{\left(\ln\left(\frac{8.0}{6.7}\right) \pm 1.96 \sqrt{\left(\frac{1}{11} \times \frac{1,364}{1,375}\right) + \left(\frac{1}{6} \times \frac{894}{900}\right)}\right)}$$

$$= e^{0.18 \pm 1.96 \times 0.51}$$

$$= 0.44 - 3.3$$

As before, $z=0.35$ and since the confidence interval of 0.44-3.3 includes the expected value of 1, the county rates are not statistically different.

When we discussed measures of association based on differences, there was no good way to determine how strong the association was based on the difference alone. With ratio measures, however, there is a direct connection between the value of the ratio itself and the strength of the association. For example, two associations may be equivalently strong, with relative risks of 1.5. One of these associations may be based on proportions of 30% and 20%, a difference of 10%. The other may be based on proportions of 9% and 6%, a difference of 3%. You can see that the difference measures do not directly show the equivalence in the strength of the two associations.

The following is a guide to assessing the strength of a ratio measure of association:

Relative Risk, Relative Prevalence, or Odds Ratio		Strength of the Association
0.83-1.00	1.0- 1.2	None
0.67-0.83	1.2- 1.5	Weak
0.33-0.67	1.5- 3.0	Moderate
0.10-0.33	3.0-10.0	Strong
< 0.01	>10.0	Approaching Infinite

Of course, appropriate confidence intervals (and p-values) will influence the interpretation of the strength of an association. Here are a few examples:

RR, RP, OR	Confidence Interval	Possible Interpretation
4.3 or 0.2	0.8-8.2 or 0.1-1.3	The estimate of association is strong, but is also very unstable since the confidence interval includes 1 and is very wide. Given the high value, though, these data provide some indication that this association may be real.
3.8 or 0.3	3.0-4.4 or 0.22-0.34	The estimate of association is strong, and is also very stable since the confidence interval does not include 1 and is quite narrow. These data provide evidence of a real association.
1.6 or 0.6	1.2-2.8 or 0.4-0.8	The estimate of association is moderate and quite stable since the confidence interval does not include 1 and is fairly narrow. These data provide evidence of a real association.
1.3 or 0.8	0.6-4.6 or 0.2-1.7	The estimate of association is weak and also very unstable since the confidence interval includes 1 and is very wide. Given the low value, these data provide no evidence that this association is real.

Test Yourself

Question:

How would you characterize the associations between smoking and low birthweight and cocaine use and low birthweight from the hypothetical population of 15,000 live births?

RR smoking and low birthweight = $9.33 / 6.0 = 1.6$ (1.4-1.8)

RR cocaine and low birthweight = $26.0 / 6.0 = 4.3$ (3.7-5.1)

Answer:

The association between smoking and low birthweight is moderate; that between cocaine and low birthweight is strong. The confidence intervals are quite narrow in each case, providing evidence that the associations are real.

The Preventive Fraction

The 2×2 table does not always have to be organized in terms of a risk factor and adverse outcome. It can also be organized to focus on a protective effect. This may be useful in a wide variety of analyses; in particular, this configuration is useful in the context of program evaluation, where a program has been designed to have a positive impact and measuring the magnitude of this impact is desired.

The 2×2 table below shows the data layout for enrollment in a hypothetical program designed to reduce a hypothetical adverse outcome:

		Adverse Outcome		
		Yes	No	
Program	Yes	80	920	1,000
	No	990	8,010	9,000
		1,070	8,930	10,000

Here, the relative risk is $0.08/0.11=0.73$ which is protective, that is, those enrolled in the program had less risk of developing the adverse outcome than those not in the program. The rows in the 2×2 table have been reversed from before, with the presumably low risk group now in the first row and the higher risk group now in the second. To make sense in this situation, the attributable risk percent must be calculated as though the table were still organized in the usual form; the high-risk group as always anchors the formula, now p_2 instead of p_1 . The result is the preventive fraction and can be interpreted as a measure of program effectiveness.

For the above table:

$$\text{Preventive Fraction} = \frac{p_2 - p_1}{p_2} = 1 - \text{Relative Risk} = \frac{0.11 - 0.08}{0.11} = 0.27$$

It can be said that the program is 27 % effective in reducing the adverse outcome.

Test Yourself

Question:

Calculate the preventive fraction for the following data on infant car seat use and motor vehicle injury

		Injury		
		Yes	No	
Car Seat Use	Yes	20	4,980	5,000
	No	50	4,950	5,000
		70	9,930	10,000

Answer:

Injury rate among infants in car seats is $20/5,000 = 0.004$

Injury rate among infants not in car seats is $50/5,000 = 0.01$

$RR = 0.004/0.01 = 0.4$

$1-RR$, or the preventive fraction is $1-0.4 = 0.60$

Using car seats for infants is 60 % effective in preventing motor vehicle injury.

ECOLOGIC ANALYSIS

The examples in the last section used the 2×2 table to examine the association between a risk factor and an outcome and between geographic areas and an outcome, in each case using individual level data. In other words, the smoking status of each pregnant woman, the county of residence for each pregnant woman, as well as the birthweight for each infant was known, and the state of residence and the number of well child visits for each child was known. The sample size (N) in each of the 2×2 tables for these examples equaled the total number of individuals for whom there was data.

Sometimes, however, an association is to be evaluated at the aggregate level and then the N in the 2×2 table equals the total number of *groups* observed. The groups, or units of analysis, may be geographic areas or health facilities. For instance, the smoking and low birthweight data might look like the following for 50 hypothetical counties:

	Low Birthweight		
	> 5 %*	≤ 5 %	
≥ 20 % Smokers	25	5	30
< 20% Smokers	10	10	20
	35	15	50 counties

*5% is the Year 2000 Objective

Now, the prevalence of smoking among pregnant women in a county is the risk factor and the outcome is prevalence of low birthweight in a county. Consider some of the familiar statistics and measures of association, and the possible interpretation in this context:

$$p_1 = \frac{25}{30} = 0.83$$

$$p_2 = \frac{10}{20} = 0.50$$

$$\text{Outcome Prevalence} : p_0 = \frac{35}{50} = 0.70$$

$$\text{"Exposure" Prevalence} : \frac{30}{50} = 0.60$$

$$\text{Relative Risk} : \frac{0.83}{0.50} = 1.7$$

$$\text{Population Attributable Risk\%} : \frac{0.70 - 0.50}{0.70} \times 100 = 29\%$$

Counties with a high prevalence of smoking among pregnant women have 1.7 times the risk of not meeting the Year 2000 Objective for low birthweight. The PAR% of 29%, means that the number of counties that might meet the Year 2000 Objective if smoking prevalence could be decreased to < 20% in all of the counties would be $0.29 \times 35 \cong 10$. In other words, only 25 rather than 35 of the 50 counties would fall short of the Objective.

Historically, epidemiologists have questioned the value of ecologic analysis, since the connection between the exposure and outcome is not linked at the individual level. In other words, in the above example, it is unknown whether the women who smoked during their pregnancies are the same ones who delivered low birthweight infants.

Example

The statistical test with county as the unit of analysis is:

$$z = \frac{0.83 - 0.50}{\sqrt{0.70(1 - 0.70)\left(\frac{1}{30} + \frac{1}{20}\right)}} = 2.5, \quad z = \frac{\ln\left(\frac{0.83}{0.50}\right)}{\sqrt{\left(\frac{1}{25} \times \frac{5}{30}\right) + \left(\frac{1}{10} \times \frac{10}{20}\right)}} = 2.1, \quad z = \frac{\ln\left(\frac{25 \times 10}{5 \times 10}\right)}{\sqrt{\frac{1}{25} + \frac{1}{5} + \frac{1}{10} + \frac{1}{10}}} = 2.4$$

Difference
Relative Risk
Odds Ratio

Whichever test is used, the z value is greater than the usual critical value of 1.96, so the p-value is less than 0.05. The tests indicate that higher prevalence of smoking among pregnant women in a county is associated with higher prevalence of low birthweight.

Categorizing the county prevalence rates of smoking at < 20 % and $\geq 20\%$ and the rates of low birthweight at $\leq 5\%$ and $> 5\%$ illustrates the importance of choosing appropriate breakpoints for continuous variables that are transformed to fit into the framework of the 2×2 table. This is true for individual level data as well, when the exposure is not simply present or absent but has many values that are converted to a dichotomous form. If the categories are not chosen appropriately, misclassification has in effect occurred, and the results of the risk measures may be biased.

FINAL NOTES ON INTERPRETING 2 X 2 TABLES

We have seen that ratio measures and difference measures derived from a 2×2 table provide different types of information. The relative risk and odds ratio are used to measure the strength or the magnitude of an association between an exposure or risk factor, and a health outcome; the attributable risk measures are used to gauge the potential public health impact of an exposure.

In addition, the interpretation of the epidemiologic measures of risk must take into account the structure and type of data being used. Just as the assumption of causality did not hold when county was the exposure variable, so too this assumption may be questioned when data are at the aggregate level or when the data are for service or systems variables. Since the traditional epidemiologic risk measures are used

with many kinds of data, however, terminology and assumptions may have to be modified to accommodate a broader public health interpretation.

In any case, the 2×2 table remains a very useful tool for organizing data whether or not particular risk measures will be reported. It helps the analyst clarify the questions of interest, permits exploration of the impact of different categorization schemes, and provides a format for exploration of hypothetical changes in health status, health services, or health systems.