6

MEASURES OF ASSOCIATION

OBJECTIVES

After reading this chapter, you should be able to:

1. Calculate and interpret the following measures of association:
   - incidence risk ratio
   - odds ratio
   - incidence rate ratio
   - risk difference (attributable risk)
   - attributable fraction (exposed)
   - population attributable risk
   - attributable fraction (population).

2. Understand when to use each of the above measures of association.

3. Correctly use the concepts of strength of association and statistical significance when presenting research results.

4. Understand the basis for the common methods of computing significance tests and confidence intervals.

6.1 INTRODUCTION

Measures of association are used to assess the magnitude of the relationship between an exposure to a disease (*eg* a potential 'cause') and a disease. In contrast, measures of statistical significance cannot be used to indicate the magnitude of the effect (*ie* the strength of association) because they are heavily dependent on sample size.

In general, the material in this chapter will focus on comparing the frequency of disease in exposed subjects with the frequency of disease in subjects that are not exposed. Depending on study design, disease frequency can be expressed as:

- incidence risk (cohort study design)
- incidence rate (cohort study design)
- prevalence (cross-sectional study design)
- odds (cohort or cross-sectional study design).

Conversely, in case-control study designs, the objective is to compare the odds of exposure in 2 groups—those with the disease under investigation (the cases), and those without the disease under investigation (the controls).

If disease frequency has been measured as risk, the data for measuring the strength of association between exposure and disease are summarised in Table 6.1.

Table 6.1 Presentation of incidence risk data

	Exposure		
	Exposed	**Non-exposed**	
Diseased	a_1	a_0	m_1
Non-diseased	b_1	b_0	m_0
	n_1	n_0	n

where:
a_1 = the number of exposed animals that have the disease
a_0 = the number of non-exposed animals that have the disease
b_1 = the number of exposed animals that do not have the disease
b_0 = the number of non-exposed animals that do not have the disease.

If disease frequency has been measured as rates, the data for measuring the strength of association between exposure and disease are summarised in Table 6.2.

Table 6.2 Presentation of incidence rate data

	Exposure		
	Exposed	**Non-exposed**	
Number of cases	a_1	a_0	m_1
Animal-time at risk	t_1	t_0	t

Note For simplicity, we will refer to the frequency of disease in animals, but these could also be measured in groups of animals (*eg* number of herds affected). We will refer to associations as though we believe them to be causal. Criteria for inferring causation are reviewed in Chapter 1.

6.2 MEASURES OF ASSOCIATION

The strength of an association between an exposure and a disease usually is expressed using a 'relative' effect measure computed as the ratio of 2 estimates of disease frequency. There are 3 common ratio measures of association: the risk ratio (*RR*), the incidence rate ratio (*IR*) and the odds ratio (*OR*). The appropriate measure of association depends on the study design and its corresponding measure of disease frequency.

6.2.1 Risk ratio

RR is the ratio of the risk (*R*) of disease in the exposed group to the risk of disease in the non-exposed group.

$$RR = \mathrm{p}(D+|E+)/\mathrm{p}(D+|E-) = (a_1/n_1)/(a_0/n_0) \qquad \textit{Eq 6.1}$$

Risk ratio (also known as relative risk) can be computed in cohort studies and, in some cases, cross-sectional studies. It cannot be used in case-control studies because the p(*D*+) is an arbitrary value determined by the number of cases and controls included in the study.

RR ranges from 0 to infinity. A value of 1 means there is no association between exposure and disease:

RR < 1 exposure is protective (*eg* vaccines)
RR = 1 exposure has no effect (*ie* null value)
RR > 1 exposure is positively associated with disease.

Risk ratio says **nothing** about how much disease is occurring in the population. The actual frequency of the disease can be quite low, but the *RR* can be high. For example, in Table 6.3, which summarises the records from a large (hypothetical) herd of Hereford cattle over 5 years, the risk of 'cancer eye' in the herd is low: 40/6000=0.0067, but the risk of cancer eye in cattle with white eyelids is 3.8 times higher than that of cattle with pigmented lids.

Table 6.3 Data on ocular carcinoma and eyelid pigmentation from a hypothetical longitudinal study of a large herd of Hereford cattle

		Eyelids		
		Non-pigmented	Pigmented	
Ocular carcinoma	Present	38	2	40
	Absent	4962	998	5960
		5000	1000	6000

$$RR = (38/5000)/(2/1000) = 3.8$$

As noted, *RR* can be computed from cross-sectional studies. Cross-sectional studies normally measure the prevalence of disease, but in certain situations (*eg* a short period of risk of disease that has been completed for all animals) the prevalence might be a valid estimate of the incidence risk. In this situation, *RR* can be used. In other situations, the term prevalence ratio (*PR*) would be preferred. It is computed in the same way as *RR* (and the term *RR* is sometimes used instead of *PR*).

6.2.2 Incidence rate ratio

The incidence rate ratio (*IR*) is the ratio of the disease frequency (measured as incidence rate) in an exposed group to the incidence rate (*I*) in a non-exposed group.

$$IR=(a_1/t_1)/(a_0/t_0) \qquad \textit{Eq 6.2}$$

IR can only be computed from studies in which an incidence rate can be calculated (*ie* cohort studies). It is sometimes referred to as the incidence density ratio. *IR* ranges from 0 to infinity. A value of 1 means there is no association between the exposure and disease, with values <1 indicating protection and values >1 indicating an increased rate of disease in the exposed group.

Table 6.4 presents some hypothetical data on teat pre-dipping and cases of clinical mastitis in dairy herds.

Table 6.4 Data on cases of mastitis and pre-dipping in a hypothetical dairy herd

	Not pre-dipped	Pre-dipped	
# of cases of mastitis	18	8	26
# of cow months	250	236	486

$$IR=(18/250)/(8/236)=2.12$$

In this example, the rate of mastitis is 2.1 times higher in cows whose teats are not pre-dipped than in cows whose teats are pre-dipped prior to milking.

6.2.3 Odds ratio

The *OR* is the odds of the disease (*O*) in the exposed group divided by the disease odds in the non-exposed group.

$$\begin{aligned} OR&=\text{odds}(D+|E+)/\text{odds}(D+|E-)\\ &=(a_1/b_1)/(a_0/b_0)\\ &=(a_1*b_0)/(a_0*b_1) \end{aligned} \qquad \textit{Eq 6.3}$$

Alternatively, it can be calculated as the odds of exposure in the diseased group divided by the odds of exposure in the non-diseased group.

$$\begin{aligned} OR&=\text{odds}(E+|D+)/\text{odds}(E+|D-)\\ &=(a_1/a_0)/(b_1/b_0)\\ &=(a_1*b_0)/(a_0*b_1) \end{aligned} \qquad \textit{Eq 6.4}$$

Based on the data in Table 6.3, the *OR*=(38/2)/(4962/998)=3.82.

Note The odds ratio is the only measure of association that exhibits this 'symmetry' which enables you to switch the exposure and the disease (outcome). Consequently, *OR* is the only measure of strength of association applicable to case-control studies. (Because disease frequency in the sample is artificially established in case-control studies, the relative risk is not an appropriate measure of strength of association.)

The interpretation of *OR* is the same as *RR* and *IR*. An *OR*=1 indicates no effect while values <1 and >1 are indicative of reduced risk (protection) and increased risk, respectively.

6.2.4 Relationships among *RR*, *IR* and *OR*

In general, the relationships among *RR*, *IR* and *OR* is such that *IR*s are further from the null value (1) than *RR*s, and the *OR*s are even further away as can be seen in Fig. 6.1.

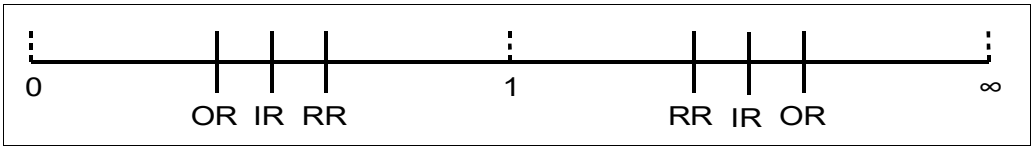

Fig. 6.1 General relationship among RR, IR and OR

RR* and *OR

If the disease occurs infrequently in the underlying population (prevalence or incidence risk <5%), *OR* is approximately equal to *RR*. In this situation,

$$RR = \frac{\frac{a_1}{a_1+b_1}}{\frac{a_0}{a_0+b_0}} \approx \frac{\frac{a_1}{b_1}}{\frac{a_0}{b_0}} = OR$$

because if the disease is rare, a_1 is very small and a_1+b_1 approaches b_1 and a_0 is very small so a_0+b_0 approaches b_0.

Similarly, if *RR* in a population is close to the null (*ie* $RR \approx 1$) then *RR* and *OR* will be very close. (If $RR=1$, then $RR=OR$). *OR*s are commonly used because they can be derived easily from logistic regression analyses (Chapter 16).

RR* and *IR

RR and *IR* will be close to each other if the exposure has a negligible impact on the total time at risk in the study population. This occurs if the disease is rare or if *IR* is close to the null value ($IR=1$). (See Chapter 4 for details on the role of time at risk in computation of incidence rates.)

OR* and *IR

OR is a good estimator of *IR* under 2 conditions. If controls are selected in a case-control study using 'cumulative' or risk-based sampling (*ie* controls selected from all non-cases once all cases have occurred—see Chapter 9), then *OR* will be a good estimate of *IR* only if the disease is rare. However, if controls are selected using 'density' sampling (*ie* a control selected from the non-cases each time a case occurs), then *OR* is a direct estimate of *IR*, regardless of whether or not the disease is rare.

6.3 MEASURES OF EFFECT

The effect (or impact) of a risk factor on a disease usually is expressed using an 'absolute' effect measure which is computed as the difference between 2 measures of disease frequency. Some prefer to use 'effect' measures rather than 'strength' measures because effect measures more closely relate to the number of cases an exposure causes (or prevents) than measures of association based on strength. The effect can be computed just for the exposed group (Section 6.3.1) or for the population (Section 6.3.2). Although we use the term 'effect', it is well to remember that we are measuring associations. Thus, the 'effect' will only be the result of exposure if the association is causal.

6.3.1 Measures of effect in the exposed group

Even when an exposure is very strongly associated with disease occurrence (*eg* smoking and lung cancer in humans), typically some disease cases occur in the non-exposed population (lung cancer does occur rarely in non-smokers). The incidence in the non-exposed population can be viewed as the 'baseline' level of risk for individuals if the exposure were completely absent from the population. To evaluate the effect of an exposure on disease frequency in exposed subjects, we can consider both the absolute difference in risk between the exposed and non-exposed groups (risk difference (*RD*)) and the proportion of disease in the exposed group that is attributable to the exposure (attributable fraction (AF_e)). Both these measures incorporate the baseline risk in the non-exposed population, and assume that all other risk factors are common to both the exposed and non-exposed groups (*ie* absence of confounding, see Chapter 13).

Risk difference, incidence rate difference

RD is the risk of disease in the exposed group minus the risk of disease in the non-exposed group. It is also referred to as the **attributable risk**.

$$\begin{aligned} RD &= \mathrm{p}(D+|E+) - \mathrm{p}(D+|E-) \\ &= (a_1/n_1) - (a_0/n_0) \end{aligned} \qquad \textit{Eq 6.5}$$

RD indicates the increase in the probability of disease in an exposed group, beyond the baseline risk, that results from the exposure.

The incidence rate difference (*ID*) can similarly be calculated as the difference between 2 incidence rates:

$$ID = (a_1/t_1) - (a_0/t_0) \qquad \textit{Eq 6.6}$$

Difference measures are interpreted as follows:

RD or *ID* < 0 exposure is protective
RD or *ID* = 0 exposure has no effect
RD or *ID* > 1 exposure is positively associated with disease.

Attributable fraction (exposed)

The AF_e expresses the proportion of disease in exposed individuals that is due to the exposure, assuming that the relationship is causal. Alternatively, it can be viewed as the proportion of disease in the exposed group that would be avoided if the exposure were removed. AF_e can be calculated from either incidence data in both exposed and non-exposed groups, or directly from the *RR*.

$$\begin{aligned} AF_e &= RD/\mathrm{p}(D+|E+) \\ &= \{(a_1/n_1) - (a_0/n_0)\}/(a_1/n_1) \\ &= (RR-1)/RR \\ &\simeq (OR-1)/OR \qquad (\text{approximate } AF_e) \end{aligned} \qquad \textit{Eq 6.7}$$

These calculations assume that exposure is positively associated with disease. Values for attributable fraction range theoretically from 0 (where risk is equal regardless of exposure; *RR*=1) to 1 (where there is no disease in the non-exposed group and all disease is due to the exposure; *RR*=∞). If exposures are negatively associated with disease, attributable fraction can be calculated in the same manner by regarding 'lack of exposure' to the protective factor as the factor that enhances risk. One example of this approach is estimation of vaccine efficacy. In

case-control studies when actual disease frequencies in the exposed and non-exposed groups are unknown, attributable fraction can be approximated by substituting the *OR* for *RR* (as shown in Eq 6.7).

Vaccine efficacy is one form of AF_e with 'not vaccinated' equivalent to being 'factor positive' ($E+$). For example, if 20% of non-vaccinated animals develop disease [$p(D+|E+)=0.20$] and 5% of vaccinated animals develop disease [$p(D+|E-)=0.05$], the following can be calculated:

$$RD = 0.20 - 0.05 = 0.15$$
$$AF_e = 0.15/0.20 = 0.75 = 75\%$$

The vaccine has prevented 75% of the cases of disease that would have occurred in the vaccinated group if the vaccine had not been used. This is known as vaccine efficacy.

Note When based on incidence rates, the measures of effect in the exposed group (AF_e) or in the population (AF_p—see below) relate to proportional or absolute changes in the rates, but not necessarily to the proportion or number of cases. This technical difference arises because the exposure might affect the timing (*ie* when) of disease occurrence but not the actual number of cases. Thus, the actual number of cases could be constant but the time at risk, and hence the rate, would differ.

Etiologic fraction

A distinction can be made between an attributable fraction (as calculated above—also called **excess fraction**) and **etiologic fraction** (Greenland & Robins, 1988; Rothman *et al*, 2008). While the former represents the excess cases observed in the exposed group, the etiologic fraction is the proportion of cases (in the exposed group) for which exposure was a component of the sufficient cause (see Chapter 1). Unfortunately, they are not equal as the following hypothetical example will show. Assume that in a population of animals a disease is inevitable (all animals get it) by 2 years of age. Animals without the exposure of interest will develop the disease between 1 and 2 years of age. However, exposure contributes to another sufficient cause and exposed animals all develop the disease by 1 year of age. If a population is followed for 2 years, the risks in the exposed and non-exposed groups are both 1, so the AF_e is zero. However, in all exposed animals, development of the disease was associated with exposure so the etiologic fraction is 1.

Unfortunately, the etiologic fraction cannot be estimated from epidemiological data because we never know what sufficient cause resulted in an observed case. Under certain specific conditions, the AF_e will equal the etiologic fraction (see Rothman *et al* (2008)), but in general all we can say is that the AF_e provides a lower bound for the etiologic fraction (*ie* the minimum value it can take).

6.3.2 Measures of effect in the population

Attributable risk and attributable fraction are useful for quantifying the effect of an exposure in the exposed group, but do not reflect the effect of the exposure in the whole population. For example, there might be a strong association between neonatal beef-calf loss and the use of prophylactic neomycin boluses at calving (RR=5, AF_e=0.8), but if the practice of giving neonatal calves a neomycin bolus is infrequent, it will not contribute much to neonatal mortality in beef calves. On the other hand, a relatively weak risk factor that is common might be a more important determinant of neonatal mortality in the population as a whole. In terms of national or regional disease-control programmes, information about the effect of a factor in the total

population is useful in allocating resources for health-promotion and disease-control programmes.

Population attributable risk

PAR is analogous to *RD*, in that it indicates a simple difference in risk between 2 groups. However, the focus of *PAR* is the increase in risk of disease in the entire population that is attributable to the exposure. Therefore it is calculated as the overall observed risk (combining exposed and non-exposed groups) in the population minus the baseline risk (risk in the non-exposed). Clearly, *PAR* is determined by both the strength of the association and the frequency of exposure to the risk factor.

$$\begin{aligned} PAR &= \mathrm{p}(D+) - \mathrm{p}(D+|E-) \\ &= (m_1/n) - (a_0/n_0) \\ &= RD * \mathrm{p}(E+) \end{aligned} \qquad \textit{Eq 6.8}$$

Note Logically, *PAR* might be called the risk difference (population), but generally isn't.

Population attributable fraction

Population attributable fraction (AF_p) is analogous to AF_e, but reflects the effect of the disease in the entire population rather than the exposed group. Assuming a causal relationship, AF_p indicates the proportion of disease in the population that is attributable to the exposure, and which would be avoided if the exposure were removed from the population (and nothing else changed). There are a number of ways of computing the AF_p (Rockhill *et al*, 1998). Most commonly, it is calculated as the ratio of *PAR* to overall risk p(*D*+) in the population (first line in Eq 6.9), and again is a function of the strength of the association and the prevalence of exposure. An equivalent formula is based on an estimate of the *RR* and the proportion of the population exposed (second line in Eq 6.9).

$$\begin{aligned} AF_p &= PAR / \mathrm{p}(D+) \\ &= \frac{\mathrm{p}(E+)(RR-1)}{\mathrm{p}(E+)(RR-1)+1} \end{aligned} \qquad \textit{Eq 6.9}$$

These formulae are appropriate for data derived from cross-sectional and longitudinal (single cohort) studies (see Chapter 7 and 8) from which the risks of disease and the prevalence of exposure are both known, and there is no confounding.

If confounding is present and adjusted estimates of the *RR* are available, the AF_p can be estimated using:

$$AF_p = pd\left(\frac{aRR-1}{aRR}\right) \qquad \textit{Eq 6.10}$$

where *pd* is the proportion of cases exposed to the risk factor and *aRR* is the adjusted *RR*. (**Note** See Chapter 13 for a discussion of confounding and computing adjusted risk ratios to remove confounding effects). This approach often is extended to the analysis of case-control studies and *OR* is used instead of *RR* (also in Eq 6.11).

If exposure has multiple (k>2) categories or if multiple exposure factors are evaluated simultaneously, an estimate of the overall AF_p can be computed using

$$AF_p = 1 - \sum_{i=0}^{k} \frac{pd_i}{aRR_i} \qquad \textit{Eq 6.11}$$

where pd_i is the proportion of cases in the i^{th} exposure level and aRR_i is the adjusted risk ratio comparing the i^{th} exposure level to the unexposed group.

6.4 STUDY DESIGN AND MEASURES OF ASSOCIATION

Table 6.5 presents a summary of the measures of association that can be computed from various study designs. Example 6.1 shows sample calculations of all these parameters.

Table 6.5 Summary of calculation of various measures of association by study type

	Cross-sectional	Cohort study	Case-control
RR	X	X	
IR		X	
OR	X	X	X
RD	X	X	
AF_e	X	X	X[b]
PAR	X	X[a]	
AF_p	X	X[a]	X[c]

[a] The *PAR* and AF_p can be estimated from a cohort study provided that an independent estimate of the p(*D*+) or the p(*E*+) in the source population is available. These are available directly from a single cohort (longitudinal) study.
[b] Estimated using *OR* as an approximation of *RR* .
[c] Estimated using *OR* as an approximation of *RR* and an independent estimate of p(*E*+|*D*+).

6.5 HYPOTHESIS TESTING AND CONFIDENCE INTERVALS

The material presented in previous sections has focused on the computation of point estimates of parameters. Investigators usually want to evaluate the statistical significance of parameters and examine the variability of their point estimates as well. There are 3 general approaches:

1. A standard error (SE) of the parameter can be computed to provide a measure of the precision of the point estimate (*ie* how much uncertainty there is in the estimate).
2. A significance (hypothesis) test can be carried out to determine if the point estimate is significantly different from some value specified by the null hypothesis test.
3. A confidence interval (CI) for the estimate can be computed.

What follows is a non-technical introduction to hypothesis-testing and confidence intervals in the context of unconditional (*ie* one exposure and one outcome) associations. These procedures are based on a classical (sometimes denoted 'frequentist') approach to statistics. An alternative approach, one based on Bayesian statistics, is less commonly used (see Chapter 23).

Note Throughout this section, all references to parameters in the text and in the formulae will refer to estimates derived from the data unless otherwise stated. 'Population parameters' (*ie* true, unknown values) will be referred to as such in the text.

Example 6.1 Measures of association

Assume that you want to determine if being over-conditioned (*ie* fat) at the time of calving affects a cow's risk of developing ketosis. A body condition score (BCS) of 4.0 or above would be considered over-conditioned (*ie* not desirable). You carry out a cohort study in a single large dairy herd (your population of interest) and all cows are observed from the time of calving through the first 4 months of lactation (the period at which they are at risk of developing ketosis). In addition to recording the number of cows in each BCS group that developed and did not develop ketosis, you record the number of cow-months at risk. Once a cow had a case of ketosis, she stopped contributing to the number of cow-months at risk. This occurred, on average, at 2 months' post-calving.

	BCS		
	≥ 4	< 4	
Ketosis +	60	157	217
Ketosis -	41	359	400
cows	101	516	617
cow-months	284	1750	2034

101 'fat' cows contributed 284 cow-months at risk and had 60 cases of ketosis.

516 'normal' cows contributed 1,750 cow-months at risk and had 157 cases of ketosis.

Measures of disease frequency	**Practical interpretation**
$R = p(D+) = 217/617 = 0.352$	35% of all cows had ketosis
$R_{E-} = p(D+\|E-) = 157/516 = 0.304$	30% of normal cows had ketosis
$R_{E+} = p(D+\|E+) = 60/101 = 0.594$	59% of fat cows had ketosis
$I = 217/2034 = 0.11$	0.11 cases of ketosis per cow-month in population
$I_{E-} = 157/1750 = 0.09$	0.09 cases of ketosis per cow-month in normal cows
$I_{E+} = 60/284 = 0.21$	0.21 cases of ketosis per cow-month in fat cows
Measures of association	
$RR = 0.594/0.304 = 1.95$	Fat cows were 1.9 (or 2) times as likely to develop ketosis as normal cows
$IR = (60/284)/(157/1750) = 2.34$	The rate of ketosis in fat cows was 2.3 times higher than the rate in normal cows
$OR = (359*60)/(157*41) = 3.35$	The odds of ketosis in fat cows was 3.4 times higher than the odds in normal cows
Measures of effect	
$RD = 0.594-0.304 = 0.290$	For every 100 fat cows, 29 had ketosis due to the being fat (assuming a causal relationship)
$AF_e = 0.290/0.594 = 0.488$	49% of the ketosis occurring in fat cows was attributable to them being fat
$PAR = 0.352-0.304 = 0.048$	For any 100 cows in this population, 5 had ketosis that was attributable to them being fat
$AF_p = 0.048/0.352 = 0.136$	14% of the ketosis in the population was attributable to fat cows

6.5.1 Standard error

For some of the parameters described in previous sections, estimates of the variance of the parameter can be computed directly and the square root of this variance is the estimated SE of the parameter. For example, based on the incidence rate data presented in Table 6.2, the variance of the *ID* is:

$$\text{var}(ID) = \frac{a_1}{t_1^2} + \frac{a_0}{t_0^2} \qquad \textit{Eq 6.12}$$

The variance of the *RD* can be computed directly as:

$$\text{var}(RD) = \frac{\frac{a_1}{n_1}\left(1 - \frac{a_1}{n_1}\right)}{n_1} + \frac{\frac{a_0}{n_0}\left(1 - \frac{a_0}{n_0}\right)}{n_0} \qquad \textit{Eq 6.13}$$

For other population parameters, it is not possible to directly compute their variance although methods for estimating the variance based on large sample approximations are available. This approximation is commonly done using a Taylor series approximation. Alternatively, a test-based method (sometimes referred to as the delta method) can be used (Kleinbaum *et al*, 1982) but it generally results in estimates of the SE which are too small, so this approach will not be discussed further.

For ratio measures (*eg IR*), the variance is computed on the log scale. However, there is no simple expression for the var(lnθ), so it is usually estimated using a first-order Taylor series approximation. The formulae for Taylor series approximation estimates of the var(ln *RR*) and var(ln *OR*) are:

$$\text{var}(\ln RR) = \frac{1}{a_1} - \frac{1}{n_1} + \frac{1}{a_0} - \frac{1}{n_0} \qquad \textit{Eq 6.14}$$

$$\text{var}(\ln OR) = \frac{1}{a_1} + \frac{1}{a_0} + \frac{1}{b_1} + \frac{1}{b_0} \qquad \textit{Eq 6.15}$$

Dann and Koch (2005) have recently reviewed methods of estimating variances (and computing confidence intervals) for ratios of 2 proportions. Methods of estimating the variance of attributable fractions have been reviewed recently (Steenland & Armstrong, 2006).

6.5.2 Significance (hypothesis) testing

Significance (hypothesis) testing is based on the specification of a **null hypothesis** about the population parameter(s). The null hypothesis is usually that there is no association between the factor and the outcome which means that null measures of difference (*eg ID*) will be 0 or that the null ratio measures (*eg IR*) will be 1.

In using this approach, an **alternative hypothesis** is stated and it can be of a one-tailed or 2-tailed nature. For example, if we have disease incidence rates in 2 groups (exposed and non-exposed), the usual 2-tailed hypothesis is that *I* in the exposed group is different than in the non-exposed group (*ie* it could be higher or lower). We are interested in finding out if there is statistical evidence to support a difference in rates that could be in either direction. A one-tailed

hypothesis would be that *I* is higher in the exposed group than in the non-exposed group. We either do not believe that it is possible that *I* could be lower in the exposed group or we have no interest in this possible outcome. (An alternative one-tailed hypothesis would be that the rate is lower, and we are not interested in the possibility of the rate being higher.) In general, the use of one-tailed hypotheses is much harder to justify than the use of 2-tailed hypotheses, so they should be used with caution.

The next step in the hypothesis-testing process is to compute a **test statistic** (*eg* a *t*-statistic, a *Z*-statistic or a χ^2-statistic). From the expected distribution of this test statistic, a **P-value** is determined. The P-value is the probability that the test statistic would be as large or larger (in absolute value) than the computed test statistic, if the null hypothesis were true. A small P-value indicates that, if the null hypothesis were true, it is unlikely (*ie* low probability) that you would obtain a test statistic as large or larger than the one you have obtained. In this case, it is usual to reject the null hypothesis in favour of the alternative.

P-values, while containing useful information, are limited in their ability to convey the full picture about the relationship being evaluated. They are often dichotomised into 'significant' or 'non-significant' based on some arbitrary threshold (usually set at 0.05) but this entails a huge loss of information about the parameter of interest. Knowing that an effect was 'significant' provides neither any indication of the actual probability of observing the test statistic computed, nor information about the magnitude of the effect observed. Reporting the actual P-value solves the first problem but not the second. The second issue will be discussed under confidence intervals (see Section 6.4.3).

Test statistics

There are 4 commonly used types of test statistic for evaluating associations between exposure and disease: Pearson χ^2, exact test statistics, Wald tests and likelihood ratio tests.

Pearson χ^2 is the most commonly used test statistic for the comparison of proportions. For data laid out as shown in Table 6.1, the equation for Pearson χ^2 is:

$$\chi^2 = \sum_{\text{all cells}} \frac{(\text{obs} - \text{exp})^2}{\text{exp}} \qquad \textit{Eq 6.16}$$

where: obs = observed value in each cell of the table, and

exp = expected value for the cell = row total * column total/grand total.

(For example, the expected value for the cell with obs = a_1 is $n_1 * m_1/n$).

The Pearson χ^2 has an approximate χ^2 distribution provided all expected cell values are >1 and 80% (or 3 of 4 entries in a 2X2 table) are >5.

Note A closely related χ^2 statistic, the Mantel-Haenszel χ^2 differs from Pearson χ^2 only by a multiplier of $n/(n-1)$ which is negligible for moderate to large values of n. The Mantel-Haenszel χ^2 is used more commonly in the analysis of stratified data (Chapter 13).

In some cases, **exact probabilities** for test statistics can be computed based on the distribution of the data. In these instances, the P-values are derived directly from the permutations of the data rather than by relying on an assumed distribution (*eg* normal or χ^2) for the test statistic. For example, an exact test statistic for a 2X2 table (*eg* testing the significance of an *RD* or an *RR*) can be obtained from the hypergeometric distribution. First, the hypergeometric probability of

every possible table with the same row and column totals as the observed data is computed. Fisher's exact P-value is the sum of the probabilities of all tables with equal or smaller hypergeometric probabilities than the observed table. In general, computation of exact statistics is computationally demanding so, historically, they usually have been used for relatively small datasets where approximations based on large numbers of observations are unsatisfactory. With recent advances in computing, this limitation rarely applies.

Wald statistics are appropriate provided the sample size is moderate to large (see guideline for Pearson χ^2 above). The general formula for a Wald statistic is computed as:

$$Z_{Wald} = \frac{\theta - \theta_0}{\mathrm{SE}(\theta)} \qquad \textit{Eq 6.17}$$

where SE(θ) is the estimated standard error of θ, and θ_0 is the value of θ specified in the null hypothesis (this is often zero). Under the null hypothesis, a Wald statistic is assumed to have a normal distribution (or a χ^2 distribution for the square of the statistic).

Likelihood ratio tests (***LRT***) are based on the likelihood of a parameter (θ). The likelihood of a parameter [L(θ)] is the probability (density) of obtaining the observed data, if θ is the true value of the population parameter. A likelihood ratio (*LR*) compares the likelihood of the estimated θ with the likelihood of θ_0 (the value of θ specified in the null hypothesis). An *LRT* is computed as follows and, provided the sample size is reasonably large, it has an approximate χ^2 distribution.

$$LRT = -2(\ln LR) = -2\left(\frac{\ln \mathrm{L}(\theta)}{\ln \mathrm{L}(\theta_0)}\right) \qquad \textit{Eq 6.18}$$

Note In some cases it is possible to derive an exact probability for an *LRT* rather than rely on the χ^2 approximation. In general, *LRT*s are superior to Wald tests. *LRT*s are discussed further in Chapter 16.

6.5.3 Confidence intervals

Confidence intervals (CIs) reflect the level of uncertainty in point estimates and indicate the expected range of values that a parameter might have. Although a CI covers a range of possible values for an estimated parameter, values close to the centre of the range are much more likely than those at the ends of the range. While we use an estimated SE and a specific percentile of a test statistic distribution to compute a CI, a CI generally conveys more information than simply presenting a point estimate of a parameter and its P-value because it clearly shows a range of likely values for the population parameter. Specifically, a 95% CI means that if we were to repeat the study an infinite number of times under the same conditions and create a CI for each study, 95% of these CIs would contain the true parameter value.

If the 95% CI includes the null value (*eg* 1 for *RR*, *IR* or *OR* , 0 for *RD*, *ID*), it suggests that the parameter is not statistically significant from the null at a P-value of 0.05. However, this surrogate significance test is an 'under-use' of CI because it doesn't fully use all the information contained in the CI.

Computing confidence intervals

As with hypothesis tests, CIs can be computed using either exact probability distributions or large sample approximations. Exact CIs are based on the exact probabilities of the distributions

underlying the parameter (binomial for proportions, Poisson for rates and hypergeometric for odds ratios). They are generally employed when dealing with relatively small sample sizes although increasing computer power has made the computation of exact CIs for most measures of association feasible for moderate to large sample sizes.

CIs based on large sample approximations have the following general formulae.

The confidence interval of a difference measure (θ) is:

$$\theta \pm Z_{\alpha}\sqrt{\text{var}(\theta)} \qquad \textit{Eq 6.19}$$

where var(θ) is the large sample approximate estimate of the variance of θ.

As noted above, for ratio measures, the variance is computed on the log scale so the general formula for a confidence interval of $\ln\theta$ is:

$$\ln\theta \pm Z_{\alpha}\sqrt{\text{var}(\ln\theta)} \qquad \textit{Eq 6.20}$$

and for θ it is:

$$\theta * \exp(\pm Z_{\alpha}\sqrt{\text{var}(\ln\theta)}) \qquad \textit{Eq 6.21}$$

Because the CI is computed on the log scale, it is symmetrical about $\ln\theta$, but not about θ.

Note A CI for *OR* that is based on the Taylor series approximation of the variance is sometimes referred to as Woolf's approximation. An approximation of an exact CI (although it seems illogical that such an entity can exist) for *OR* is Cornfield's approximation (Cornfield, 1956). Computation of this CI is an iterative process and it is used less now that it is possible to directly compute exact confidence intervals.

Example 6.2 presents a variety of point estimates and CIs for parameters computed in Example 6.1.

6.6 MULTIVARIABLE ESTIMATION OF MEASURES OF ASSOCIATION

(**Note** skip this section unless you have some familiarity with confounding (Chapter 13) and linear and logistic regression (Chapters 14-16)). The concept of using adjusted estimates of *RR* or *OR* was introduced in Section 6.3.2 to allow for estimates of AF_p to be adjusted for known or suspected confounders. We often want to use multivariable models to simultaneously control for several potential confounders. If we want to estimate an adjusted *OR*, this is straightforward because *OR*s can be derived directly from a logistic regression model.

Multivariable estimation of *RR*s is more difficult. Procedures based on a generalised linear model or a Poisson regression model are discussed briefly in Section 18.4.1 Recently, a method based on ordinary logistic regression which computes the *RR* as the ratio of the sum of the predicted probabilities of the outcome assuming all animals were exposed to the comparable value assuming all animals were not exposed has been published (Kleinman & Norton, 2009). It is relatively easy to implement and the author's simulation studies suggest that the procedure is reliable.

A method of computing adjusted *RD*s based on ordinary linear regression with robust SEs (see Section 20.5.4) has been proposed (Cheung, 2007). These SEs may be modified for small sample situations. The author's simulation results suggest the method is reliable.

Example 6.2 Confidence intervals for measures of association

The following table presents a variety of CIs computed for some of the measures of association computed in Example 6.1

			CI	
Measure of effect	**Point estimate**	**Type of CI**	**Lower bound**	**Upper bound**
ID	0.122	direct	0.066	0.177
IR	2.354	exact	1.719	3.190
RD	0.290	exact	0.186	0.393
RR	1.952	exact	1.587	2.402
OR	3.346	exact	2.108	5.329
		Woolf's (Taylor series)	2.157	5.192
		Cornfield's	2.161	5.181
		Test based	2.188	5.117

Direct or exact CIs were computed for *ID*, *IR*, *RD* and *RR*. A variety of CIs were computed for *OR* for comparison purposes. The exact CIs are the widest, followed by Woolf's and Cornfield's approximations (which were similar). The test-based CI was the narrowest and these are not recommended for general use.

Note We have shown 3 significant digits in this example, but we need to remind ourselves that fewer (*eg* 1 or 2) decimal places might better represent the underlying variability of our study data as shown under 'practical interpretation' in Example 6.1.

REFERENCES

Cheung YB. A modified least-squares regression approach to the estimation of risk difference Am J Epidemiol. 2007; 166: 1337-44.

Cornfield J, Halperin M, Moore F. Some statistical aspects of safety testing the Salk poliomyelitis vaccine. Public Health Rep. 1956;71(10):1045-56.

Dann RS, Koch GG. Review and evaluation of methods for computing confidence intervals for the ratio of two proportions and considerations for non-inferiority clinical trials J Biopharm Stat. 2005; 15: 85-107.

Greenland S, Robins JM. Conceptual problems in the definition and interpretation of attributable fractions Am J Epidemiol. 1988; 128: 1185-97.

Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic research: principles and quantitative methods. New York: John Wiley and Sons; 1982.

Kleinman LC, Norton EC. What's the Risk? A simple approach for estimating adjusted risk measures from nonlinear models including logistic regression Health Serv Res. 2009; 44: 288-302.

Rockhill B, Newman B, Weinberg C. Use and misuse of population attributable fractions Am J Public Health. 1998; 88: 15-9.

Rothman K, Greenland S, Lash T. Modern Epidemiology, 3rd Ed. Lippincott Williams & Wilkins: Philadelphia; 2008.

Steenland K, Armstrong B. An overview of methods for calculating the burden of disease due to specific risk factors Epidemiology. 2006; 17: 512-9.