7 MEASURES OF ASSOCIATION AND EFFECT

Nils Toft, Jens Frederik Agger and Jeanett Bruun

7.1 Introduction

Often our concern in an epidemiological study is to assess the chance that an animal that possesses certain traits also has a specific disease. These traits might be intrinsic risk factors such as a specific sex, age or weight or extrinsic risk factors such as housing or management related issues (see Chapter 3). The chance that a healthy animal exposed to a risk factor becomes diseased is defined as the risk of disease. While the risk is a useful summary of the relationship between disease and the risk factor, it is not sufficient to assess the importance of the risk factor. A high risk of developing disease in a group exposed to a certain factor does not implicate the factor as a possible risk factor unless a similar group of animals left unexposed to the factor show a lower risk of developing disease.

In this chapter we present some measures frequently used to determine the association between risk factor and disease as well as methods to assess the effect, i.e. the importance of the risk factor to the study population and target population.

7.2 The 2 × 2 table

To keep things simple we will for the remainder of this chapter focus on situations where data can be presented in a 2×2 table as displayed in Table 7.1. Quite often data of this kind will be obtained from an observational study, i.e. from a cross-sectional, a cohort or a case–control study. Let us here briefly remind ourselves how data are obtained in these different study types.

Table 7.1. *The 2 × 2 table for calculating measures of association and effect*

Risk factor status	*Disease status*		*Total*
	Disease	*No disease*	
Exposed	a	b	$a+b$
Not exposed	c	d	$c+d$
Total	$a+c$	$b+d$	n

In a cross-sectional study our primary interest is an estimate of the prevalence. We obtain a sample from the population under study by randomly selecting n animals (i.e. without regard to exposure or disease status) and afterwards assigning them to one of the four categories defined by Table 7.1 ((1) disease + exposed; (2) no disease + exposed; (3) disease + not exposed; and (4) no disease + not exposed). As we do not follow the animals over a time period, we cannot estimate the incidence risk; however, we can estimate the risk of disease (which is the prevalence among animals in the group) for the exposed ($a/(a+b)$) and the unexposed ($c/(c+d)$) group. This risk is sometimes referred to as the *prevalence risk*. Essentially, it is just the measure of prevalence already introduced in Chapter 6. However, here we add the term *risk* to acknowledge that our focus is the prevalence among animals exposed or unexposed to a certain *risk factor*.

In a cohort study we sample two initially disease-free groups of animals, one group exposed to the suspected risk factor and one group left unexposed. These groups are then followed over time and we can estimate the risk of developing disease (i.e. the incidence risk) for each group using $a/(a+b)$ for the exposed and $c/(c+d)$ for the unexposed. Thus, for a cohort study we have defined the groups row-wise, so we are allowed to make calculations and interpretations row-wise.

In a case–control study we sample column-wise. That is, we have selected a group of diseased (cases) and a group of non-diseased (controls) animals and then subsequently divided them into subgroups with respect to exposure status. This implies that we cannot justify calculating the same kind of risk as in the previous study: having selected column-wise, we cannot calculate measures row-wise, since there is no justification to assume that the proportion of diseased animals in the sample is the same as in the general population. It is, however, possible to calculate the risk of exposure to the risk factor for the diseased ($a/(a+c)$) and not diseased ($b/(b+d)$), but these measures are often of very little direct interest. The risk of exposure may also be calculated in a cross-sectional study, but there is little reason to do so. For case–control

studies we abandon risk as a measure and turn to something else, i.e. odds. Before turning to this alternative we shall, however, give the risk measure some more attention.

7.3 Measures of association

7.3.1 *Risk and relative risk*

In this chapter we use a definition of risk that differs a bit from the one used to define incidence risk in Chapter 6. We define risk (r) as:

$$r = \frac{\text{\# animals with condition (i.e. disease or exposure)}}{\text{\# animals in population (or exposure group)}} \tag{7.1}$$

We will, in general, distinguish between incidence risk, prevalence risk and exposure risk, as mentioned in the previous section. As we assume that data are presented in a 2×2 table (Table 7.1) the above definition is valid for all three cases. However, the methods for calculating incidence risk presented in Chapter 6 should still be used whenever data permit.

There is no reason why one should not combine the exposed and unexposed groups of a cohort study and calculate the total risk of disease within the sample as $(a + c)/n$ (but be aware that this does not necessarily reflect the risk of disease in the target population). Still, as already mentioned the risk itself is an inadequate measure of the association between disease and exposure status. The risk needs to be compared to the risk in a similar group which has been unexposed. To make this comparison, the *relative risk* RR is defined:

$$\text{RR} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}} = \frac{a(c+d)}{c(a+b)}, \tag{7.2}$$

i.e. the risk in the exposed group relative to the risk in the unexposed group. By definition it only makes sense to calculate the relative risk in a cohort or cross-sectional study, because risk is an inappropriate measure in a case–control study, as we have sampled column-wise. Whenever the value of RR is above 1, there is increased risk of disease for the exposed group, whereas a value below 1 means that exposure to the factor is protective.

By definition, the risk (r) is a probability (proportion), hence the standard error (se) according to general statistical rules (e.g. assuming that the proportion is normal distributed) is given as:

$$\text{se}(r) = \sqrt{\frac{r(1-r)}{n_r}} \tag{7.3}$$

Table 7.2. *Hypothetical data from a study regarding the association between feline hip dysplasia and the exercise options of cats. The letters in the subscripts refer to the notation in Table 7.1*

	Hip dysplasia		*Total cats*
	+	−	
Indoor	$50_{(a)}$	$150_{(b)}$	200
Outdoor	$20_{(c)}$	$780_{(d)}$	800
Total cats	70	930	1000

and the $(1-\alpha)100\%$ confidence interval as:

$$r \pm Z_{1-\alpha/2}\text{se}(r), \tag{7.4}$$

where we use $Z_{1-0.05/2} = 1.96$ to calculate the 95% confidence intervals. In Equation 7.3, we have chosen to write n with a subscript n_r to indicate that the number of animals underlying the estimate depends on the estimate itself (e.g. for the risk of the exposed group, $n_r = a + b$, but for the entire sample population $n_r = a + b + c + d = n$).

The confidence interval for the relative risk is somewhat more complicated to compute. The relative risk can be regarded as the ratio between two approximative normal distributions. This is in itself not a normal distribution. However, it has been shown that the particular ratio defined as the relative risk, when log-transformed follows a normal distribution. On the logarithmic scale Katz *et al*. (1978) showed that the standard error of the relative risk is:

$$\text{se}(\ln \text{RR}) = \sqrt{\frac{1}{a} - \frac{1}{a+b} + \frac{1}{c} - \frac{1}{c+d}} \tag{7.5}$$

Using Equation 7.5 it is straightforward to establish a confidence interval for the log-transformed variable using Equation 7.4, which in turn leads to upper (U_{RR}) and lower (L_{RR}) confidence limits for the relative risk itself:

$$U_{\text{RR}} = \exp(\ln \text{RR} + Z_{1-\alpha/2}\text{se}(\ln \text{RR})) \tag{7.6}$$

$$L_{\text{RR}} = \exp(\ln \text{RR} - Z_{1-\alpha/2}\text{se}(\ln \text{RR})) \tag{7.7}$$

EXAMPLE 7.1. Lack of exercise might be a risk factor in feline hip dysplasia (HD). In Table 7.2 we present data from a hypothetical study regarding the association between HD in cats and their choice of exercise options, i.e. are the cats allowed outdoors or are they kept indoors.

We will use this example throughout the chapter; we will vary the study design accordingly to be able to apply the different measures. For now we consider that the data have been obtained in a cohort study, where the (initially healthy) cats are followed through a period of 5 years.

In such as study we can calculate the incidence risk for each group as well as the relative risk. The incidence risk (I_{risk}) of HD for the exposed (indoor) and unexposed (outdoor) is:

$$I^{\text{I}}_{\text{risk}} = \frac{50}{200} = 0.25, \qquad I^{\text{O}}_{\text{risk}} = \frac{20}{800} = 0.025$$

The relative risk (RR) is:

$$\text{RR} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}} = \frac{I^{\text{I}}_{\text{risk}}}{I^{\text{O}}_{\text{risk}}} = \frac{0.25}{0.025} = 10$$

This implies that the incidence risk of HD (i.e. the risk that an initially healthy cat develops HD during a 5 year period) among cats kept indoors, is 10 times higher than among cats allowed outdoors. However, had the data in Table 7.2 originated from a cross-sectional study, then RR = 10 should be interpreted as a 10 times higher risk of finding a cat with HD among cats kept indoors, compared to cats allowed outdoors.

To calculate the confidence interval for the relative risk, we use Equation 7.5 to calculate the standard error of the log-transformed relative risk:

$$\text{se}(\ln \text{RR}) = \sqrt{\frac{1}{a} - \frac{1}{a+b} + \frac{1}{c} - \frac{1}{c+d}} = \sqrt{\frac{1}{50} - \frac{1}{200} + \frac{1}{20} - \frac{1}{800}} = 0.25$$

Applying this to Equations 7.6 and 7.7 we obtain the following 95% confidence interval for the relative risk:

$$U_{\text{RR}} = \exp(\ln \text{RR} + Z_{1-0.05/2}\text{se}(\ln \text{RR})) = \exp(\ln 10 + 1.96 \times 0.25) = 16.4$$

$$L_{\text{RR}} = \exp(\ln \text{RR} - Z_{1-0.05/2}\text{se}(\ln \text{RR})) = \exp(\ln 10 - 1.96 \times 0.25) = 6.1$$

i.e. a 95% CI: [6.1; 16.4].

7.3.2 *Odds and odds ratio*

A compulsive gambler knows that the risk, i.e. probability, is not the only measure of chance. There is an alternative called the *odds*:

$$\text{odds} = \frac{\text{\# animals with disease (or exposure)}}{\text{\# animals without disease (or unexposed)}} \tag{7.8}$$

If P is the probability of disease (or exposure), then there is the following connection between odds and P:

$$\text{odds} = \frac{P}{1-P}$$

The interpretation of odds is somewhat less intuitive than that of a risk.

The motivation to apply a measure such as odds lies not in the odds themselves but more in the ratio between two different odds, the *odds ratio*. For a cohort study, the odds ratio (OR_{CO}) for disease is estimated by the

ratio between the odds of the exposed group and the odds of the unexposed group as:

$$\mathrm{OR_{CO}} = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{ad}{bc}, \tag{7.9}$$

i.e. the odds of disease in the exposed group relative to the odds of disease in the unexposed group.

For a case–control study, however, the odds are defined in the columns as the ratio between exposed and unexposed for the diseased group and non-diseased group respectively. Hence, the odds ratio for a case–control study ($\mathrm{OR_{CC}}$) is defined as:

$$\mathrm{OR_{CC}} = \frac{\frac{a}{c}}{\frac{b}{d}} = \frac{ad}{bc}, \tag{7.10}$$

which is the odds of being exposed to the risk factor in the group of diseased animals relative to the odds of being exposed in the group of non-diseased animals.

Note that even though the odds ratios of the case–control and the cohort study are defined using different odds, the odds ratios are calculated in the same way. In a cross-sectional study both interpretations of the odds ratio are possible, but usually Equation 7.9 is preferred. Thus, it is possible to compare the results of different study designs using the odds ratios. Hence, from now on we will refer to the odds ratio as simply OR and omit the subscripts.

Although it is possible to calculate OR for all kinds of studies it is still very important to realise that there is a difference in the interpretation of OR for case–control studies compared to cohort and cross-sectional studies. The reason is that we cannot make inference row-wise in a case–control study, because the data were sampled column-wise. So in a cohort study, $\mathrm{OR} = 3$ means that the odds of getting disease is three times greater for the group exposed to the factor in question compared to the unexposed group. For the case–control study, $\mathrm{OR} = 3$ implies that odds of having been exposed to the factor is three times greater for the diseased group than for the non-diseased group of animals. Hence, to interpret the OR of a case–control study in terms of exposure being a risk factor, you need to be sure of the causality between exposure and outcome.

Another reason for adopting odds ratio is a matter of convenience in calculation (e.g. in logistic regression or analysis, see Chapter 13). Still, whenever possible, i.e. in cohort or cross-sectional studies, one should always calculate the relative risk because of the more intuitive interpretation of parameters expressed in terms of probabilities.

The odds themselves are rarely quoted in a study, because they are of limited interest. Thus, we will ignore the calculation of confidence intervals for these and only focus on the approach for the odds ratio. Woolf (1955) showed that a logarithmic transformation would yield a better approximation to a normal distribution with a standard error:

$$\text{se}(\ln \text{OR}) = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}} \tag{7.11}$$

Confidence limits for the OR itself are derived analogously to the relative risk (Equation 7.6):

$$U_{\text{OR}} = \exp(\ln \text{OR} + Z_{1-\alpha/2}\text{se}(\ln OR)) \tag{7.12}$$

$$L_{\text{OR}} = \exp(\ln \text{OR} - Z_{1-\alpha/2}\text{se}(\ln OR)) \tag{7.13}$$

EXAMPLE 7.2. Consider the table from Example 7.1 (Table 7.2). The odds ratio for these data is:

$$\text{OR} = \frac{ad}{bc} = \frac{50 \times 780}{150 \times 20} = 13$$

The standard error of the log-transformed OR (Equation 7.11) is:

$$\text{se}(\ln \text{OR}) = \sqrt{\frac{1}{50} + \frac{1}{150} + \frac{1}{20} + \frac{1}{780}} = 0.28$$

which gives a 95% confidence interval (Equations 7.12 and 7.13):

$$U_{\text{OR}} = \exp(\ln \text{OR} + Z_{1-0.05/2}\text{se}(\ln \text{OR})) = \exp(\ln 13 + 1.96 \times 0.28) = 22.5$$

$$L_{\text{OR}} = \exp(\ln \text{OR} - Z_{1-0.05/2}\text{se}(\ln \text{OR})) = \exp(\ln 13 - 1.96 \times 0.28) = 7.5$$

i.e. a 95% CI: [7.5; 22.5].

If we still assume that data are from a cohort study (and ignore that we normally would not calculate OR in such a case) we can interpret OR = 13 as odds being 13 times greater for developing HD among cats kept indoors compared to cats allowed outside.

However, if we assume that the data are from a case–control study, then we should interpret OR = 13 as odds of being kept indoors being 13 times greater among cats with HD than among those that do not develop HD. It is important to keep this distinction in mind when interpreting odds ratios.

The calculations in Examples 7.1 and 7.2 are good illustrations of the use of OR as an approximation of RR whenever the disease in question is rare. Referring to Table 7.1, the numbers defined by a and c must be small, because the prevalence ($(a + c)/n$) is low. Hence, it follows that:

$$a + b \approx b \tag{7.14}$$

$$c + d \approx d \tag{7.15}$$

Table 7.3. *Data for Example 7.3 to illustrate the need for the disease to be rare in both case and control groups when approximating the relative risk by the odds ratio*

Risk factor status	*Disease status*		*Total*
	Disease	*No disease*	
Exposed	3	3	6
Not exposed	5	989	994
Total	8	992	1000

where $\approx$ means approximately equal to. Combining Equations 7.9 and 7.2 gives:

$$\text{OR} = \frac{ad}{bc} \approx \frac{a(c+d)}{(a+b)c} = \text{RR} \tag{7.16}$$

However, the assumption of rare must be true for both the exposed and the unexposed group as the next example shows:

EXAMPLE 7.3. Consider the data in Table 7.3. Here, we have a case where disease is present in only eight out of 1000 cases. But the RR and OR, respectively, are:

$$\text{RR} = \frac{\frac{3}{6}}{\frac{5}{994}} = 99$$

$$\text{OR} = \frac{3 \times 989}{3 \times 5} = 197,$$

i.e. the OR is nearly twice as big as RR because there is a relatively high occurrence of disease within the exposed group even though the overall occurrence is rare. Note that the data in Table 7.3 are an example of a poor study design with only six exposed and eight cases in a total of 1000 study units.

7.4 Measures of effect

Through relative risk or odds ratio it is possible to establish the relative importance or association of the risk factor to the disease. However, this does not tell us the overall importance of that risk factor. To achieve this we must have a measure that combines the relative risk with the proportional occurrence of the risk factor (i.e. the prevalence of the risk factor). One possible measure is the *attributable risk*. The intuitive motivation is that there is a risk of developing disease even if the animal is left unexposed. The difference in the risk of

developing disease between the exposed group and the unexposed group is the *attributable risk* (AR):

$$\mathrm{AR} = \frac{a}{a+b} - \frac{c}{c+d} \tag{7.17}$$

Note that this is an absolute measure. It might be easier to use the relative equivalent, i.e. the fraction (proportion) of the total risk in the exposed group that refers to the subjects being exposed. This measure is called the *attributable fraction* (AF):

$$\mathrm{AF} = \frac{\frac{a}{a+b} - \frac{c}{c+d}}{\frac{a}{a+b}} = \frac{\mathrm{AR}}{\frac{a}{a+b}} = \frac{\mathrm{RR}-1}{\mathrm{RR}} \tag{7.18}$$

The last part of the formula suggests that an approximation can be used in case–control studies. The result is the *estimated attributable fraction* ($\mathrm{AF_{Est}}$):

$$\mathrm{AF_{Est}} = \frac{\mathrm{OR}-1}{\mathrm{OR}} \tag{7.19}$$

The precautions regarding the validity of the formula are essentially the same as the one underlying the substitution of odds ratio for relative risk, i.e. if OR is a poor estimate of RR, then $\mathrm{AF_{Est}}$ is a poor estimate of AF. In a case where it is possible to calculate AF directly, we should do so.

It is possible to calculate confidence intervals for AF, but the formula is rather tedious and is omitted. The interested reader is instead referred to Woodward (1999) for details.

EXAMPLE 7.4. Turn again to the data from Example 7.1 (Table 7.2) and assume that the data originate from a cohort study. Then Equation 7.17 gives the attributable risk (AR):

$$\mathrm{AR} = \frac{a}{a+b} - \frac{c}{c+d} = \frac{50}{200} - \frac{20}{800} = 0.225,$$

i.e. the risk among indoor cats which may be attributed to them being indoors is 0.225 out of a total risk of 0.25. The attributable fraction (AF) is given from Equation 7.18 as:

$$\mathrm{AF} = \frac{\mathrm{RR}-1}{\mathrm{RR}} = \frac{10-1}{10} = 0.9,$$

thus 90% of the HD among indoor cats is due to their being kept indoors (and hence deprived of proper exercise).

The estimated attributable fraction ($\mathrm{AF_{Est}}$) which should only be used in a case–control study is given here for comparison:

$$\mathrm{AF_{Est}} = \frac{\mathrm{OR}-1}{\mathrm{OR}} = \frac{13-1}{13} = 0.92,$$

which is only slightly different from the true AF.

In a cross-sectional study or whenever the sample is representative of the target population it is possible to estimate, not only the effect, but also the importance of the risk factor in the population. This is achieved by comparing the risk in the population to the risk in the unexposed group using the *population attributable risk* (PAR):

$$\text{PAR} = \frac{a+c}{n} - \frac{c}{c+d} = \frac{a+b}{n}\text{AR} \tag{7.20}$$

which is completely analogous to the AR defined in Equation 7.17. PAR denotes the risk in the population that can be attributed to the risk factor. Again it might be easier to interpret the importance when expressed as a fraction of the total risk. Hence, we introduce the *population attributable fraction* (PAF):

$$\text{PAF} = \frac{\frac{a+c}{n} - \frac{c}{c+d}}{\frac{a+c}{n}} \tag{7.21}$$

i.e. the proportion of cases in the population that are due to the risk factor. It is possible to estimate PAF using the population odds ratio (not defined here, see Table 7.6). However, we omit these calculations because the assumptions underlying this approximation are rarely fulfilled.

EXAMPLE 7.5. Assume for the time being that the data in Example 7.1 (Table 7.2) are from a cross-sectional study. This allows us to calculate the measures of effect defined as population attributable risk (PAR) and population attributable fraction (PAF) (Equations 7.20 and 7.21):

$$\text{PAR} = \frac{a+c}{n} - \frac{c}{c+d} = \frac{70}{1000} - \frac{20}{800} = 0.045 \tag{7.22}$$

$$\text{PAF} = \frac{\frac{a+c}{n} - \frac{c}{c+d}}{\frac{a+c}{n}} = \frac{\frac{70}{1000} - \frac{20}{800}}{\frac{70}{1000}} = 0.64. \tag{7.23}$$

Hence, the risk of HD in the population that may be attributed to indoor cats is 0.045. Hence, we would expect the total risk of HD in the population ($70/1000 = 0.07$) to drop 0.045 if all cats had access to outdoor exercise. (Note that this suggests that the overall risk would fall to the outdoor level (0.025).) The PAF of 0.64 implies that 64% of the HD cases in the population are due to some cats being housed indoors.

7.5 Summary of measures

To conclude this chapter we briefly summarise the measures of association, with respect to strength, effect and importance with the emphasis on the appropriate use in different types of study. For cross-sectional studies the

Table 7.4. *The applicable measures of association, effect and importance in a cross-sectional study*

Summations

		D		*Total*
		+	−	
E	+	a	b	$a+b$
	−	c	d	$c+d$
Total		$a+c$	$b+d$	n

D = diseased
E = exposed

Measure of association		
Relative risk	$\mathrm{RR} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}$	The relative risk is a measure of the *prevalence* risk
Odds ratio	$\mathrm{OR} = \frac{ad}{bc}$	Interpret in rows, but use RR whenever possible
Population relative risk*	$\mathrm{RR}_{\mathrm{pop}} = \frac{\frac{a+c}{n}}{\frac{c}{c+d}}$	Note that the risks are not independent, as c and d are used in both risks
Population odds ratio*	$\mathrm{OR}_{\mathrm{pop}} = \frac{d(a+c)}{c(b+d)}$	Again, use $\mathrm{RR}_{\mathrm{pop}}$ whenever possible
Measure of effect		
Attributable risk	$\mathrm{AR} = \frac{a}{a+b} - \frac{c}{c+d}$	This is an absolute measure of effect
Attributable fraction	$\mathrm{AF} = \frac{\mathrm{AR}}{\frac{a}{a+b}}$	This is a relative measure of effect
	$= \frac{\mathrm{RR}-1}{\mathrm{RR}}$	
Measure of importance		
Population attributable risk	$\mathrm{PAR} = \frac{a+c}{n} - \frac{c}{c+d}$	This is an absolute measure of importance
	$= \frac{a+b}{n}\mathrm{AR}$	
Population attributable fraction	$\mathrm{PAF} = \frac{a+c}{n} - \frac{\frac{c}{c+d}}{\frac{a+c}{n}}$	This is a relative measure of importance

*Not mentioned in this chapter.

summary is given in Table 7.4, for cohort studies in Table 7.5, and for case–control studies in Table 7.6. Some elements in the tables that have been excluded from the previous presentation, are included in the summary for completeness.

Table 7.5. *The applicable measures of association, effect and importance in a cohort study*

Summations

		D		*Total*
		+	−	
E	+	a	b	$a+b$
	−	c	d	$c+d$
Total		$a+c$	$b+d$	n

D = diseased
E = exposed

Measure of association		
Relative risk	$\mathrm{RR} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}$	The relative risk is a measure of the *incidence* risk
Odds ratio	$\mathrm{OR} = \frac{ad}{bc}$	Interpret in rows, but use RR whenever possible
Population relative risk*	$\mathrm{RR}_{\mathrm{pop}} = \frac{\frac{a+c}{n}}{\frac{c}{c+d}}$	Use only when the cohorts are sampled so the ratio between exposed and un-exposed is representative of the target population
Measure of effect		
Attributable risk	$\mathrm{AR} = \frac{a}{a+b} - \frac{c}{c+d}$	This is an absolute measure of effect
Attributable fraction	$\mathrm{AF} = \frac{\mathrm{AR}}{\frac{a}{a+b}} = \frac{\mathrm{RR}-1}{\mathrm{RR}}$	This is a relative measure of effect
Measure of importance		
Population attributable risk	$\mathrm{PAR} = \frac{a+c}{n} - \frac{c}{c+d} = \frac{a+b}{n}\mathrm{AR}$	An absolute measure of importance. Use only when the prevalence of disease in the sample is representative of the population prevalence
Population attributable fraction	$\mathrm{PAF} = \frac{\frac{a+c}{n} - \frac{c}{c+d}}{\frac{a+c}{n}}$	A relative measure of importance. Use only when the prevalence of disease in the sample is representative of the population prevalence

*Not mentioned in this chapter.

There are, however, also measures that we have chosen to ignore both in the presentation and in the tables. One such measure is the *incidence rate ratio*, i.e. the ratio between the incidence rates in the exposed and unexposed groups (see Chapter 6 for a definition of incidence rate). The use of an incidence rate

Table 7.6. *The applicable measures of association, effect and importance in a case–control study*

Summations

		D		
		+	−	
E	+	a	b	D = diseased E = exposed
	−	c	d	
Total		$a+c$	$b+d$	n

Measure of association		
Odds ratio	$\mathrm{OR} = \frac{ad}{bc}$	Interpret in columns
Population odds ratio	$\mathrm{OR}_{\mathrm{pop}} = \frac{d(a+c)}{c(b+d)}$	Use only when the controls are sampled so the ratio between exposed and unexposed is representative of the target population
Measure of effect		
Estimated attributable fraction	$\mathrm{AF}_{\mathrm{Est}} = \frac{\mathrm{OR}-1}{\mathrm{OR}}$	Use when OR is a good approximation of RR, i.e. when the disease is rare in both the exposed and unexposed groups
Measure of importance		
Estimated population attributable fraction*	$\mathrm{PAF}_{\mathrm{Est}} = \frac{\mathrm{OR}_{\mathrm{pop}}-1}{\mathrm{OR}_{\mathrm{pop}}}$	Use only when $\mathrm{OR}_{\mathrm{pop}}$ may be calculated

*Not mentioned in this chapter.

ratio is essentially that of a relative risk, but in terms of rates. This makes it somewhat less straightforward to interpret. Furthermore, the associated statistical tests for significance and calculation of confidence intervals are rather difficult compared to the relative risk. Other such measures might exist, but, in general, we recommend that you rely on the measures presented here. For a more thorough discussion and presentation of alternatives see, e.g. Kleinbaum *et al*. (1982).

Whenever there is serious doubt about the study design, one can always use the odds ratio. However, as we have emphasised in the text, the interpretation might differ depending on the study design used. As a starting point, each of the Tables 7.4–7.6 presents the 2×2 table with the summations, i.e. whether the study design allows column-wise, row-wise or both kinds of summation.