Chapter 3
Introduction to Ratio Measures

Fundamental concepts, definitions, calculating formulae and interpretations of risks and odds as well as the risk ratio and odds ratio are introduced in this chapter. The meta-analytical methods of these two very important and frequently used effect size measures, risk ratio and odds ratio, are covered in Chaps. 4 and 5.

3.1 Introduction

Most of the meta-analyses endeavor to estimate the unknown common population effect size and present the results in forest plots. In this chapter we introduce the effect size measures, namely risk ratio and odds ratio, to study the degree of association between two categorical or binary outcome variables. Here both the intervention (e.g. exposure and non-exposure) and outcome (e.g. cases or non-cases) are categorical.

Effect size

The effect size is the common name to a family of indices that measure the magnitude of a treatment or intervention effect. Depending on the type of study there are various measures that can be used to determine the effect size for the intervention of interest.

The effect size measure depends on the type of outcome variable of interest. For binary (categorical) outcome variables relative risk or risk ratio (RR) and odds ratio (OR) are used as effect size measure.

Effect measures such as a single proportion (of incidences) or difference between two proportions (also called risk difference) are also applicable to binary outcome variables. [See Chaps. 6 and 7 for details.] For continuous outcome variables the effect size is measured by standardised mean difference (SMD) or weighted mean difference (WMD), and correlation coefficient for linear relationship between two quantitative outcome variables. [These are covered in Chaps. 8–10.]

When the outcome of interest is a binary or categorical variable, ratio measures are used to investigate the association between the two variables. The most popular ratio

S. Khan, *Meta-Analysis*, Statistics for Biology and Health,
https://doi.org/10.1007/978-981-15-5032-4_3

measures are the relative risk or risk ratio (RR) and odds ratio (OR). Both the RR and OR are used as effect size measures for binary outcome variables. The choice of a particular ratio measure is a decision of the researchers based on the type of study and objective of the investigation. McNutt et al. 2003 used the RR in cohort studies and clinical trials of common outcomes. Some exploration on the relationship between RR and OR are found in Shrier and Steele 2006 and a conversion formula of RR from OR is provided by Zhang and Yu 1998. Further discussions on the choice of effect measure for epidemiological data are found in Walter 2000 and Barger 2018.

In this chapter we introduce the concept, computational method and interpretation of risk ratio (RR) and odds ratio (OR) as a prelude to meta-analytical methods when they are dealt with as appropriate effect size measures in the forthcoming chapters.

3.2 Relative Risk or Risk Ratio (RR) and Odds Ratio (OR)

The risk ratio (or relative risk) and odds ratio are used to assess the association between two binary (categorical) variables, namely the explanatory variable (factor, e.g. Intervention or Treatment and Control, or Exposure and No Exposure) and outcome variable (Success and Failure, or Cases and No Cases, or Disease and No Disease, or Event and No Event). Although the purpose of the two ratios, RR and OR, is the same, they are not the same, and hence they should not be used as synonymous.

3.2.1 Root Causes for Differences in RR and OR

To appreciate the difference between the RR and OR it is important to carefully understand the difference between proportion and ratio. Although both have numerator and denominator, there are fundamental differences in the definition and interpretation of the two ratios.

Ratio: In mathematics, the ratio is described as the comparison of the size of two quantities of the same unit, which is expressed in terms of times i.e. the number of times the first value contains the second. The ratio is used to compare the quantities of two different categories like the *ratio of men to women* in a population. For example, in a study of 10 mens and 20 womens, the ratio of men to women is $10/20 = 1/2 = 0.5$.

Proportion: Proportion is a mathematical concept, which states the equality of two ratios or fractions. A proportion is the quantity of one category over the total, like the proportion of men out of total people living in a population. For example, in a study of 10 men and 20 women, the *proportion of men to women* is $10/(10 + 20) = 1/3$ or 0.3333, that is, 33.33%.

Odds: In statistics, the odds for or odds of some event reflects the likelihood that the event will take place, while odds against reflects the likelihood that it will not.

An odds is a number that is obtained by dividing one number (e.g. cases or events) by another (e.g. noncases or no events), both measuring the same outcome variable. For example, in a study of 10 men and 20 women, the *odds of men* is 10/20 = 0.5 relative to women.

Probability: Probability is a numerical description of how likely an event is to occur. For example, in a study of 10 men and 20 women, the *probability* of randomly selecting a man is 10/30 = 0.3333 or 33.33%. This is the same as the proportion of men in the study.

Clearly, odds is different from proportion and, or, probability. Proportion is often used a synonymous to risk and probability. Both odds and risk (probability) have the same numerator but different denominator, that is, they are on different scales.

3.2.2 Reasons for Differences Between RR and OR

Conceptually there is a fundamental difference between the risk (proportion) and odds, as the definitions are different, and hence, in general, the RR and OR are not the same. The risk of an intervention is defined as the ratio of number of cases/events relative to the total number of subjects (combining cases and non-cases) in the study, and hence it perfectly resembles probability. But the odds of an intervention is defined as the ratio of number of cases relative to the number of non-cases (excluding number of cases from total number of subjects) in the study, and hence it is different from the notion of probability. In both cases, what is common is the 'likelihood' or 'chance' (but not probability) of happening of cases/events, but relative to two different things. Often the two ratios are mixed and used synonymously by mistake because both ratios have the same numerator and represent some kind of 'likelihood' or 'change' ignoring the fact that the two ratios have totally different denominators. The RR is a ratio of two proportions (or percentages) and OR is a ratio of two odds.

To avoid confusion, make a clear note that odd reflects 'relative likelihood' or better yet 'odds', unlike risk which reflects 'probability'. The value of odds ratio is close to that of the risk ratio only if incidence (number of cases) is very small in both the exposed and the unexposed groups. If the incidence (number of cases) is high in either or both exposed and unexposed groups, then the value of RR is very different from OR.

3.2.3 Why OR is More Appropriate Than RR?

Some people do use the probability ratio, aka the relative risk (RR) to measure the effect of the intervention (X, risk factor) on the outcome (Y, disease). The disadvantage of the RR is that it is not a constant effect of X. The probability ratio changes depending on the value of X. But the OR does not change with a change in X (that is, it is constant with respect to X). The effect of X on the probability of Y has different values depending on the value of X. So if you want to know how X affects Y, odds ratio is the appropriate effect measure.

Odds is not a measure of likelihood of events out of all possible events. It's a ratio of number of events to number of non-events. You can switch back and forth between risk and odds—they will give you different information as they are on different scales. No wonder, the term 'odds' is commonplace, but not always clear, and often used inappropriately. Schmidt and Kohlmann 2008 discussed when to use the odds ratio or the relative risk in the context of epidemiological studies and emphasised that in the absence of meaningful prevalence or incidence data, the OR provides a valid effect measure.

3.2.4 Towards Defining RR and OR

The relative risk (or risk ratio) is defined based on the ratio (proportion) of two *probabilities* (or risks), and odds ratio is defined as the ratio of two *odds*. The understanding of the difference between the two different ratios depends on appreciating the basic difference in the definition of risk and odds.

Probability is the ratio of the number of times event (success) occurred compared/relative to the *total number of trials/subjects*. Probability is a number between 0 and 1. Probability = 0.5 implies success and failure are equally likely.

Odds is the ratio of the number of times success (*event*) occurred compared/relative to the *number of times failure occurred*. Odds is a number between 0 and ∞ (infinity). The two terms (probability and odds) are *related* but not *synonymous*, rather they are very different with the same numerator but different denominator. Equal **odds** is 1, that is, 1 success for every 1 failure.

To explain and illustrate the concepts of risk (probability) and odds, consider the following count data of binary outcomes from an hypothetical experiment on immunization for a particular disease as noted in Table 3.1.

There are two possible outcomes—disease (success/event) with count 'a'or no disease (failure/no event) with count 'b' among the participating '(a + b)' subjects. The numbers in brackets are the observed counts (number of subjects) in the call.

3.2.5 Probability

Probability of disease (success) is the *proportion* of the 'number of patients with disease' (event) relative to the '*total* number of patients (Disease plus no Disease)',

Table 3.1 Incidences of disease data

Disease	No disease	Total
a (20)	b (60)	a + b (80)

that is, P(Disease) = $a/(a+b) = 20/80$ =0.25 and P(No Disease) = $b/(a+b) = 60/80 = 0.75$.

The occurrence of success (Disease) is complementary to the occurrence of failure (No Disease). So, P(Disease) = 1 − P(No Disease).

3.2.6 Risk

The risk of an event (Disease) in the Treatment group is the probability of the event. This is a proportion of 'number of events' relative to the 'total of number of events and no-events' in the Treatment group. For the data in Table 3.1, the risk of the event in the Treatment group is calculated as $R_T = a/(a+b) = 20/(20+60) = 20/80 = 1/4$ or (25%), that is P(Disease) = 1/4. Then P(No Disease) = 60/80 = 3/4. Hence P(Disease) + P(No Disease) = $1/4 + 3/4 = 1$.

3.2.7 Odds

Odds of disease (success) is the *ratio* of the 'number of patients with *disease*' relative to the 'number of patients with *no disease*', that is, Odds (Disease) = $a/b = 20/60 = 1/3 = 0.33$. Similarly, Odds (No Disease) = $b/a = 60/20 = 3$.

Note that Odds (No Disease) = $1/Odds(Disease)$. That is, $3 = \frac{1}{1/3}$ for the count data in the above example. Odds of Disease is reciprocal of odds of 'No Disease', and vice versa.

Remark: *Sum* of risk of Disease and risk of 'No Disease' is one. *Product* of odds of Disease and odds of 'No Disease' is one.

Odds ranges from 0 to ∝(infinity).
Odds (Disease) = 1 implies that success (Disease) and failure (No Disease) are equally likely.
Odd (Disease) > 1 implies that success (Disease) is *more* likely than failure (No Disease).
Odd (Disease) < 1 implies that success (Disease) is *less* likely than failure (No Disease).

Both relative risk or risk ratio (RR) and odds ratio (OR) assesses or measures *association* between two categorical variables, namely a binary outcome (or response) variable (Y) and a binary predictor (or explanatory) variable (X). Sometimes these two ratios are wrongly used interchangeably. They are not the same, and shouldn't be confused because they're actually defined and interpreted very differently. So it's important to keep them separate and to be precise in the language used in the interpretation of the two measures. The obvious difference is appreciated due to the fact that the odds ratio is the ratio of two *odds* whereas the risk ratio is the ratio of two *risks*.

3.2.8 Calculations of RR and OR

The concepts of RR and OR, and their differences, are explained using the following 2 × 2 contingency table representing an outcome variable (Y with two levels, event and non-event) and an exposure variable or intervention (X with two levels, treatment and control). The values in each of the cells in the two-way table represent the counts (frequencies) (Table 3.2).

3.3 Relative Risk (RR)

The probability (or risk) of an event (Disease) in the Treatment group is $R_T = a \big/ (a + b)$. This is a proportion of 'number of events' relative to the 'total of number of events and non-events' in the Treatment group. It's the number of patients in the Treatment group who experienced an event (Disease) out of the total number of patients with and without event (Disease and No Disease) in the Treatment group. This is to say that if a patient was treated (vaccinated), what is the probability (or risk) of having the disease (event)?

Similarly, the probability (or risk) of an event (Disease) in the Control group is $R_C = c \big/ (c + d)$. Again, it's just the proportion of the number of patients who had the disease (event) relative to the total number of patients with or without disease in the Control group. Although each of these probabilities (i.e., risks) is itself a

Table 3.2 Incidences of disease and vaccination data

		Y = Outcome variable		
		Event (e.g. disease)	Non event (e.g. no disease)	Row total
X = Exposure variable	Treatment (e.g. vaccinated)	a (20)	b (60)	a + b (80)
	Control (e.g. unvaccinated)	c (80)	d (20)	c + d (100)
	Column total	a + c (100)	b + d (80)	180

proportion, none of them is the risk ratio. The risk of having an event (Disease) for the subjects in the Treatment group needs be compared to that in the Control group to measure the effect of the Treatment.

The ratio of the above two probabilities (risks), RR_T and RR_C, is the relative risk or risk ratio:

$$RR = R_T \big/ R_C = \frac{a\big/(a+b)}{c\big/(c+d)}.$$

So the RR is the ratio of the probability (risk) of the event (Disease) in the Treatment group relative to that in the Control group.

If the Treatment worked (i.e., less subjects had disease in the Treatment group), the relative risk should be smaller than one ($RR < 1$), since the risk of having disease (event) should be smaller in the Treatment group.

If the relative risk is 1, that is, $RR = 1$, the Treatment (vaccination) made no difference at all.

If it's above 1, that is, $RR > 1$ then the Treatment group actually had a higher risk (i.e., more subjects had disease in the vaccination group) than that in the Control group.

Using the count data in the above contingency table we calculate the risk of disease for the Treatment and Control groups as follows:

$R_T = {}^{a}\!\big/_{(a+b)} = {}^{20}\!\big/_{80} = 1\big/4$ (or 25%) and
$R_C = {}^{c}\!\big/_{(c+d)} = {}^{80}\!\big/_{100} = 4\big/5$(or 80%).

Then the relative risk (RR) of the Treatment (relative to the Control) becomes

$$RR = R_T \big/ R_C = \frac{1\big/4}{4\big/5} = \frac{5}{16} \text{ (or 31.25\%)}.$$

Since the RR is much less than 1, the Treatment reduced the risk of disease in the exposed/intervention group.

3.3.1 Interpretations of RR

Because RR is a ratio and expresses how many times more probable the outcome (Disease) is in the Treatment group, the simplest way to interpret the RR is to use the phrase "times the risk" or "times as high as" compared to those in the Control group.

If you are interpreting a risk ratio, you will always be correct by saying: "Those who received vaccine (Treatment) had RR '*times the risk*' compared to those who did not have the vaccine (Control)." Or "The risk of Disease among those who received vaccine (Treatment) was RR '*times as high as*' the risk of Disease among those who did not receive vaccine (Control)."

For the above example RR = 0.3125. Thus, the risk of disease for those who received the vaccine (Treatment) is RR = 0.3125 'times the risk' of disease for those who did not receive vaccine (Control). Since the RR is less than 1 (actually less than 1/3), there is less risk of disease in the Treatment group than that in the Control group, and hence the Treatment works. Smaller the value of RR weakest is the association between the two categorized variables.

For inference (confidence interval and hypothesis test) on RR, use the log transformation Ln(RR) with its approximate variance,

$$\mathrm{Var}[\mathrm{Ln}(\mathrm{RR})] = \frac{1}{a} - \frac{1}{(a+b)} + \frac{1}{c} - \frac{1}{(c+d)}$$

and standard error

$$\mathrm{SE}[\mathrm{Ln}(\mathrm{RR})] = \sqrt{\frac{1}{a} - \frac{1}{(a+b)} + \frac{1}{c} - \frac{1}{(c+d)}}.$$

For the data in the above example, the standard error becomes

$$\mathrm{SE}[\mathrm{Ln}(\mathrm{RR})] = \sqrt{\frac{1}{20} - \frac{1}{(20+60)} + \frac{1}{80} - \frac{1}{(80+20)}} = \sqrt{0.04} = 0.2.$$

3.4 Odds Ratio (OR)

The odds of event (Disease) in the Treatment group is $\mathrm{OD_T}$ = a/b. This is the ratio of the number of events (Disease) relative to the number of non-events (No Disease) in the Treatment group. The numerator is the same as that of the probability, but the denominator here is different (in fact smaller). It's not a measure of events (Disease) relative to the all possible 'events and non events', rather it is relative to the non-events (No Disease) only.

Similarly, the odds of event (Disease) in the Control group is $\mathrm{OD_C}$ = c/d. This is the ratio of the number of events divided by number of non-events in the Control group.

The odds ratio (OR) of the Treatment group relative to the Control group is then defined as the ratio of the two odds, that it,

$$OR = \frac{OD_T}{OD_C} = \frac{a/b}{c/d} = \frac{a \times d}{c \times b}.$$

Now using the count data in the contingency table we calculate the odds for the Treatment and Control groups as follows:

$$OD_T = {}^{a}\!/_{b} = {}^{20}\!/_{60} = 1/3 \text{ and } OD_C = {}^{c}\!/_{d} = {}^{80}\!/_{20} = 4.00.$$

Then the odds ratio (OR) of the Treatment (against the Control) becomes

$$OR = OD_T / OD_C = \frac{1/3}{4} = \frac{1}{12}.$$

Since the OR is much less than 1, the Treatment has reduced the odds of disease in the exposed/intervention group.

3.4.1 Interpretations of OR

For the above example OR = 1/12 = 0.0833.

Thus, the odds ratio of event (Disease) for the vaccine (Treatment) group is 1/12 (or 0.0833). The odds of event (Disease) for those who received the vaccine (Treatment) is OR = 0.0833 'times the odds' of event to those who did not receive vaccine (Control).

Since this is much less than 1, there is much less odds of event (Disease) in the Treatment group, and hence the Treatment works. This implies that the Treatment (vaccine) works to reduce the incidences of events (Disease).

For inference (confidence interval and hypothesis test) on OR, use the log transformation Ln(OR) with approximate variance,

$$\text{Var[Ln(OR)]} = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}$$

and standard error

$$\text{SE[Ln(OR)]} = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}.$$

For the data in the above example, the standard error becomes

$$\text{SE[Ln(OR)]} = \sqrt{\frac{1}{20} + \frac{1}{60} + \frac{1}{80} + \frac{1}{20}} = \sqrt{0.129167} = 0.359398.$$

Avoid division by zero

In many cases a slightly *amended* estimator of OR is used by adding 0.5 to each cell count to avoid division by 0 as suggested by Agresti, A 1996, p. 25 and Soukri M.M 1999, p. 49. Hence, for the above example the two amended odds become

$$OD_T^* = {}^{(a+0.5)}\!/_{(b+0.5)} = {}^{(20.5)}\!/_{(60.5)} = 0.3388 \text{ and}$$

$$OD_C^* = {}^{(c+0.5)}\!/_{(d+0.5)} = {}^{(80.5)}\!/_{(20.5)} = 3.9268.$$

Then the amended odds ratio (OR*) of the Treatment (against the Control) becomes

$$OR^* = OD_T^* / OD_C^* = \frac{0.3388}{3.9268} = 0.0863.$$

Obviously, this amended $OR^* = 0.0863$ is slightly different from the original $OR = 0.0833$.

Comment: All statistical packages have this option to adjust for continuity correction. MetaXL allows a number of choices to be added to zero including 0.5.

3.4.2 Properties of OR

If $OR = 1$, the odds of event (Disease) in the Treatment group is the same as that in the Control group.

If $1 < OR < \infty$, then the odds of event (Disease) is higher in the Treatment group than in the Control group. As an example, if $OR = 4$ then the odds of evert (Disease) in the Treatment group is four 'times the odds' of event (Diseases) in the Control group. So the subjects in the Treatment group is 4 times more likely to have event (Disease) than the Control group.

If $OR = 0.25$ then the odds of event (Disease) in Treatment group is 0.25 'times the odds' of event (Disease) in the Control group.

Smaller the value of the OR weaker is the association (dependence) between the two categorical variables, exposure variable (X—Treatment or Control) and outcome variable (Y—Event and No Event).

For example, when $OR = 4$ then the association/dependence of (Exposure and Outcome variables, X and Y) is higher than when $OR = 2$ (or less). Similarly, $OR = 0.50$ indicates more dependence between (Explanatory and Outcome variables) than $OR = 0.25$ (or less).

3.5 Comparison of RR and OR

1. The RR and OR are comparable in magnitude when the event/disease studied is rare or very uncommon in both exposed and unexposed groups.
2. The OR > RR when the event/disease is more common. But OR should not be viewed as risk, it is a ratio of odds.

3. In case-control studies, risks and RR can't be calculated but OR can be calculated and use as an approximation of RR if event/disease is uncommon in the population.
4. The OR can be used to describe results of both case-control and prospective cohort studies.
5. One advantage of OR is that it is not dependent on whether we focus on the event's occurrence or its failure. That is, the OR is symmetric to which outcome level is of interest, but RR is not symmetric. That is, $Odds(Event) = \frac{1}{Odds(No-Event)}$.
6. If the OR for an event (success) deviates from 1 substantially, the OR of its non-event (failure) will also deviate from 1 substantially, although in the opposite direction.

3.5.1 *When OR is Equal (or Close) to RR?*

If the event (Disease) is rare, in both the exposed and unexposed groups, then the OR is closer (or equal) to the RR.

Consider the following modified count data with rare (or very small) number of events (2 for the Treatment group and 5 for the Control group) to illustrate how/when OR is closer to the RR (Table 3.3).

For the above count data, the number of events (Disease) for both the Treatment and Control groups are small, say 2 and 5 respectively, and the RR of the Treatment group is $RR_T = \frac{a/(a+b)}{c/(c+d)} = \frac{2/80}{5/100} = \frac{2\times100}{80\times5} = \frac{1}{2}$ (or 50%) and the OR of the Treatment group is $OR_T = \frac{a/b}{c/d} = \frac{2/78}{5/95} = \frac{2\times95}{78\times5} = \frac{19}{39}$ (or 0.48) which is closer to the RR of 50%.

Furthermore, if the number of events (Disease) for both the Treatment and Control groups are even smaller, say 1 for each group, that is, (a* = 1, b* = 79, c* = 1, d* = 99) then

$RR_T = \frac{a*/(a*+b*)}{c*/(c*+d*)} = \frac{1/80}{1/100} = \frac{1\times100}{80\times1} = \frac{5}{4}$ (or 125.00%) and
$OR_T = \frac{a*/b*}{c*/d*} = \frac{1/79}{1/99} = \frac{1\times99}{79\times1} = \frac{99}{79}$ (or 125.31) which is much closer (almost equal) to the RR of 125%.

Table 3.3 Revised incidences of disease and vaccination data

		Y = Outcome variable		
		Event (e.g. disease)	Non event (e.g. no disease)	Row total
X = Exposure variable	Treatment (e.g. vaccinated)	a (**2**)	b (**78**)	a + b (80)
	Control (e.g. unvaccinated)	c (**5**)	d (**95**)	c + d (100)
	Column total	a + c (7)	b + d (173)	180

Thus for very small number of events (Disease), in both the exposed and unexposed groups, the OR is not much different from the RR.

3.5.2 Incidence and Prevalence

Incidence and prevalence are very commonly used terms in epidemiology and public health. They may be related but very different.

Prevalence (also called prevalence rate) indicates the probability that a member of the population *has* a given condition at a point in time. So, it is the actual number of cases alive, with the disease either during a period of time (period prevalence) or at a particular date in time (point prevalence).

Incidence (also called incidence rate) is a measure of the occurrence of *new cases* of disease *during a span of time*.

The relationship between incidence and prevalence depends greatly on the natural history of the disease state being reported. In the case of a corona pandemic, the incidence may be high but not contribute to much growth of prevalence because of the high, spontaneous rate of disease resolution.

The changes in the values of RR and OR with the changes in the value of the incidence rate (I_0) at a time is presented in a graph found in Schmidt and Kohlmann 2008. In Fig. 1 of this paper the changes in the values of RR and OR with the changes in the incidence rate is displayed. Lowest (almost diagonal) line with incidence rate 0.01 represents the equality of RR and OR. The value of OR grows larger and larger as the incidence rate increases. But the value of RR becomes smaller and smaller as the incidence rate increases.

3.6 Conversion of OR to RR

There is no need to convert OR to RR if OR is properly interpreted and understood. However, some researchers try to convert OR to RR, may be to make it easily understandable to the non-specialised readers. But every researcher in the epidemiology and public health areas requires to understand odds and OR, and be prepared to interpret results based on OR.

Greenland and Holland 1991 and Zhang and Yu 1998 independently proposed a popular conversion formulas of OR to RR given by

$$RR = \frac{OR}{1 - I_C + I_C \times OR}.$$

It is interesting to note that $(1 - I_C) = 1 - \frac{c}{c+d} = \frac{d}{c+d} = I_C^0$ which is the non-incidence rate in the control group. Clearly, sum of I_C and I_C^0 is one. Similarly, for the intervention/treatment group, $(1 - I_E) = 1 - \frac{a}{a+b} = \frac{b}{a+b} = I_E^0$, and hence I_E and I_E^0 add to one.

3.7 Misuse and Misinterpretation of OR

There has been widespread unintentional misuse of odds ration (OR), especially its inappropriate interpretation as risk or probability of events. Readers may note the following examples in the epidemiological and public health literatures where *odds* is inappropriately used interchangeably with *probability* and/or risk and fails to acknowledge that odds is different from risk.

In the eight edition of the book (Merrill, 2021) notes OR as the 'relative probabilities' of disease in case-control studies without recognising that 'probabilities' are different from odds. It also recommends that odds ratios are generally interpreted as if they were risk ratios when the outcome occurs relatively infrequently (<10%). However, this 'rare disease' assumption is very infrequent.

In discussing the difference between "Probability" and "Odds" the (Boston University, 2020) notes,that the odds are defined as the *probability* that the event will occur divided by the probability that the event will not occur. Then illustrates OR with the following hypothetical pilot study on pesticide exposure and breast cancer and comes up with OR = (7/10) / (6/57) = 6.65. Interestingly, neither the numerator nor the denominator of the OR here is a probability, and that is correct because they should be odds. Yet, the definition above defines OR as ratio of probabilities.

	Diseased	**Non-diseased**
Pesticide Exposure	7	10
Non-exposed	6	57

Referring to (Bland & Altman, 2000), (Chen, Cohen, & Chen, 2010) report, "Odds ratio (OR) originally was proposed to determine whether the *probability* of an event (or disease) is the same or differs between the two groups, generally a high-risk group and a low-risk group."

3.8 Conclusions

The most popular ratio measures to investigate the association between two categorical variables, the RR and OR, are covered in this chapter. Meta-analyses based these effect size measures are presented in the upcoming chapters. In conducting meta-analysis for the RR or OR the log transformation of the ratio is essential. However, the final results for forest plots must be expressed in the original ratio scale by inverse/back (exponential) transformation.

Schmidt and Kohlmann 2008 note that the direct computation of RR is feasible if meaningful prevalence or incidence data is available. Cross-sectional data may serve to calculate RR from prevalence data. Cohort study designs allow for the direct calculation of RR from incidence.

The situation is more complicated for case-control studies. If meaningful prevalence or incidence data are not available, the OR provides a valid effect size measure. It describes the ratio of disease odds given exposure status. The OR for a given exposure is routinely obtained within logistic models while controlling for confounders. The availability of this approach in standard statistical software largely explains the popularity of this measure. However, it does not have as intuitive interpretation as the RR. Often people wrongly describe an OR of “2” in terms of a “double risk” of developing a disease given exposure.

Often selection of RR or OR depends on the study objective or design and choice or priority of the researcher. However, if you want to know how any exposure (e.g. smoking) affects the outcome (e.g. cancer), odds ratio is the best effect size measure.

There are some other ratio measures, such as proportion, which are covered separately in forthcoming chapters. They include single proportion and difference between two proportions (also known as risk difference).

References

Barger MK (2018) When does the odds ratio not equal the relative risk, and why should you care?(Report). J Midwifery Women’s Health 63(6):648. https://doi.org/10.1111/jmwh.12919

Bland JM, Altman DG (2000) The odds ratio. Bmj, 320(7247), 1468. https://doi.org/10.1136/bmj.320.7247.1468

Boston University, So PH (2020) The Difference Between “Probability” and “Odds”. Retrieved from http://sphweb.bumc.bu.edu/otlt/MPHModules/BS/BS704_Confidence_Intervals/BS704_Confidence_Intervals10.html

Chen H, Cohen P, Chen S (2010) How Big is a Big Odds Ratio? Interpreting the Magnitudes of Odds Ratios in Epidemiological Studies. Communications in Statistics–Simulation and Computation, 39(4), 860–864. https://doi.org/10.1080/03610911003650383

Greenland S, Holland P (1991) Estimating standardized risk differences from odds ratios. Biometrics 47(1):319–322. https://doi.org/10.2307/2532517

McNutt L-A, Wu C, Xue X, Hafner JP (2003) Estimating the relative risk in cohort studies and clinical trials of common outcomes.(medical research)(Author Abstract). Am J Epidemiol 157(10):940. https://doi.org/10.1093/aje/kwg074

Merrill RM (2021) Introduction to Epidemiology: Johan & Bartlett Learning.

Schmidt C, Kohlmann T (2008) When to use the odds ratio or the relative risk? Int J Public Health 53(3):165–167. https://doi.org/10.1007/s00038-008-7068-3

Shrier I, Steele R (2006) The relationship between odds ratios and relative risks: 2744. Med Sci Sports Exerc 38(Supplement):S528. https://doi.org/10.1249/00005768-200605001-03078

Walter SD (2000) Choice of effect measure for epidemiological data. J Clin Epidemiol 53(9):931–939. https://doi.org/10.1016/S0895-4356(00)00210-9

Zhang J, Yu KF (1998) What's the relative risk? A method for correcting the odds ratio in cohort studies of common outcomes.(statistical expression of risk in medical research). JAMA (Journal of the American Medical Association) 280(19):1690. https://doi.org/10.1001/jama.280.19.1690