International Journal of Statistical Sciences *ISSN 1683-5603*
Vol. 20(2), 2020, pp 153-170

Measuring Association between Two Categorical Variables: Revisiting Risk Ratio and Odds Ratio

Shahjahan Khan[1] and Muhammed Ashraf Memon[1,2,3]

[1]Professor, School of Sciences, Centre for Health Research, University of Southern Queensland, Toowoomba, Australia

[2]Sunnybank Obesity Centre & South East Queensland Surgery, Sunnybank, Queensland, Australia

[3]Mayne Academy of Surgery, Medical School, University of Queensland, Brisbane, Australia

[*]Correspondence should be addressed to Shahjahan Khan
(Email: Shahjahan.Khan@usq.edu.au)

[Received July 01, 2020; Accepted September 09, 2020]

Abstract

In many statistical analyses the relative risk or risk ratio (RR) and odds ratio (OR) are two commonly used measures of association between two categorical variables. The risk, and hence the RR, are defined in terms of probability. However, neither odds nor OR is defined in probabilistic term. Although the two ratios share similar mathematical structure, they are not the same. In reality, they are very different and hence should not be used as synonymous or interchangeably. Many epidemiological literatures inappropriately interpret OR as risk (cf. Chen (2010)). The main reasons for this confusion may be due to the fact that both represent `likelihood' of events and that OR and RR share the same numerator, ignoring the other fact that they are measured in different scales (denominators). This paper identifies the differences between the two measures by exploiting their basic definitions, methods of computations, and interpretations of results.

Keywords: Categorical outcomes; Risk and Odds; Risk ratio; Odds ratio; and Logistic regression.

AMS Classification: 62P10; 60E05.

1. Introduction

The use of RR and OR, as measures of the strength of association between two categorical variables, is very common in many public health and epidemiological studies. While it is fine to interpret RR in terms of risk or probability of disease, doing the same for OR is potentially very problematic. Technically OR is defined in terms of odds, unlike RR which is defined in terms of risk. Khan (Khan 2020) provides further details on the differences between RR and OR.

There has been widespread unintentional misuse of OR, especially its inappropriate interpretation as risk or probability of events. Here we present three examples from different epidemiological literatures to highlight necessity of disassociating OR from risk and probability:

a) 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.

b) 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 or 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

c) Referring to (Bland and Altman 2000), (Chen, Cohen, and 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."

These are some of numerous 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.

The main confusion surrounding the RR and OR is most likely rooted in unconsciously mixing up the definition of the risk and odds. This is because people use the terms risk (probability) and odds (not probability) interchangeably in casual usages. Although they measure the same thing - the `likelihood' of a specific outcome or event -- but they do so on different scales. The risk is measured as the ratio of events out of all possible events (all subjects), but odds is measured as a ratio of events to non-events (only subjects of non-event).

If the two events (case and non-case) have equal risk, then their risk is 0.5 (=1/2), that is, '1 case and 1 non-case for every 2 subjects', so that the two outcomes are equally likely. But if the two events have equal odds, then their odds is 1 (=1/1), that is, '1 case and 1 non-case for every 2 subjects'. Therefore, the probability of case 0.5 (one out of two possible outcomes), and the odds of a case 1 (one case against one non-case) are not the same. Equal probability for two events (success and failure) is 0.5, that is, 1 success for every 2 trials. Equal odds is 1, that is, 1 success for every 1 failure. Thus it is wrong to say "The OR may overestimate the risk (or RR) when the event/disease is more common." The fact is that odds can never estimate risk, and vice-versa, and hence there is no question of over or under estimation of one by another. Clearly, there is no statistical justification of using odds as an estimate of risk (or RR as an estimate of OR).

2. Fundamental Differences between Risk and Odds

To appreciate the difference between the RR and OR it is important to carefully understand the difference between proportion and ratio. Although both are fractions proper or improper and as such have numerator and denominator, there are fundamental differences in the definition and interpretation between 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.

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 reflect the 'likelihood' that it will not. An odds is a number that is obtained by dividing one number (eg cases or events) by another (eg non-cases 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 probability. Proportion is often used a synonym as to risk and probability, but not odds. Both odds and risk (probability) has the same numerator but different denominator, that is, they are on different scales.

2.1. 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 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.

To avoid confusion, one should make a clear note that odds 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 (Treatment) and the unexposed (Control) 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.

2.2. Reasons for professing OR over 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 (Schmidt & Kohlmann, 2008) discussed when to use the odds ratio or the risk ratio 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. Computation and Interpretation of RR

Count data from any study with exposure/intervention versus control as explanatory factor and cases and non-cases as the response/outcomes can be presented in the 2x2 contingency table.

3.1. Calculation of Risk

Consider a two arms intervention with exposure (treatment) and non-exposure (control) resulting in two outcomes, event (case) and non-event (no-case). Let the number of 'cases' in a study of (a+b) subjects in the Treatment group be 'a' and the number of 'no-cases' be 'b'. Hence the risk (or probability) of a 'case' is a/(a+b), but the odds of a 'case' is a/b. Clearly, the risk is synonymous to probability, but odds is not.

In a hypothetical experiment, let a trial vaccine for Covid-19 be tested in a sample population of 300 subjects randomly divided equally into two groups, vaccine or Treatment and non-vaccine or Control. At the end of the trial 25 subjects in the vaccine group and 120 subjects in non-vaccine group were infected with the virus. The count data from the experiment can be displayed in the four cells of a two-way contingency table as in Table 1.

Table 1: Frequency distribution of count data of the Covid-19 vaccine experiment

		Outcomes (Event or Non-event)		
		Event (Cases or infection)	Non-Event (Non-case or no-infection)	Row Total
Intervention (Treatment or Control)	Vaccine (Treatment)	a (25)	b (125)	a+b (150)
	No-vaccine (Control)	c (120)	d (30)	c+d (150)
	Column Total	a+c (145)	b+d (155)	300

Probability

Probability of infection is the *proportion* of the `number of subjects with infection' (cases) relative to the *`total* number of subjects (cases plus no-cases)' in the group. Here the probability of infection in the Treatment group is

P(Infection) = $a/(a+b) = 25/150 = 1/6 = 0.1667$ and P(No-Infection) = $b/(a+b) = 125/150 = 5/6 = 0.8333$. The occurrence of event (Infection) is complementary to the occurrence of non-event (No-Infection), that is, P(Infection) = 1 – P(No-Infection).

Risk

The risk of an event (Infection) in the exposure/treatment group is the same as the probability of the event. This is a proportion of `number of events' relative to the `total of number of events and non-events' in the Treatment group. For the data in Table 3.1, the risk of infection in the Treatment group is calculated as $R_T = a/(a+b) = 25/(25+125) = 25/150 = 1/6 = 0.16667$. The risk of an event can be expressed as percentage. Here the risk of infection is 0.16667 or (16.67%).

3.2 Risk ratio (RR)

The risk of Infection in the Treatment group is $R_T = a/(a+b)$. It's the number of patients in the Treatment group who experienced an Infection out of the total number of patients with and without Infection. Similarly, the risk of Infection in

the Control group is $R_C = c/(c+d)$. Again, it's just the proportion of the number of patients who had the infection relative to the total number of patients with or without infection in the Control group.

The ratio of the above two risks, RR_T in the Treatment group and RR_C in the Control group, is the risk ratio (RR):

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

If the vaccine (Treatment) worked, less subjects would be infected in the Treatment group. Then the relative risk should be smaller than one ($RR<1$), because the risk of having infection is smaller in the Treatment group.

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

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

For the vaccine data in Table 3.1, the risks of infection in the Treatment (R_T) and Control (R_C) groups are

$R_T = a/(a+b) = 25/150 = 1/6 = 0.1667$ (or 25%) and

$R_C = a/(a+b) = 120/150 = 4/5 = 0.80$ (or 80%). So the risk of Infection in the Treatment group is much smaller than that in the Control group.

Then the risk ratio (RR) of Infection in the vaccine group (relative to the Control group) becomes

$$RR = R_T / R_C = \frac{1/6}{4/5} = \frac{5}{24} = 0.2083 \text{ (or 20.83\%)}.$$

Interpretation of RR:

The RR=20.83% is interpreted as "Those who received vaccine (Treatment) had $\frac{5}{24}$ '*times the risk*' compared to those who did not have the vaccine (Control)." Or "The risk of Infection among those who received vaccine (Treatment) was 20.83% '*times as high as the risk*' of Infection among those who did not receive vaccine

(Control)." Since the RR is less than 1 (actually less than 21%), there is (about 79%) less risk of infection in the Treatment group than that in the Control group, and hence the Treatment works.

Smaller the value of the RR weaker is the association (dependence) between the two categorical variables.

3.3. Incidence rate and RR

The incidence rate in a population is defined as the number of events/cases divided by the person-time (or population) at risk. Thus incidence rate is another way to represent the risk at a point of time. The ratio of incidences in exposed group (I_E) divided by that in control/unexposed group (I_C) gives the relative risk (or risk ratio). That is, $RR = \frac{I_E}{I_C}$.

From the data in Table 3.l, it can be shown that $I_E = \frac{a}{a+b}$ and $I_C = \frac{c}{c+d}$. As the incidence rate (risk) in the exposed group increases the value of the RR grows larger.

4. Computation and Interpretation of OR

Odds of an event (Infection) is the `number of patients with *Infection* 'relative to the `number of patients with *no Infection''* in any group. Hence, for the count data in Table 3.1, odds of infection in the Treatment group is Odds (Infection) = $25/125 = 1/5 = 0.20$. Similarly, odds of no-infection in the Treatment group is Odds (No-Infection) = $b/a = 125/25 = 5$. Note that Odds (No-Infection) = $1/Odds(Infection)$. That is, $\frac{1}{1/5} = 5$ for the count data in the above example. Odds of Infection is reciprocal of odds of 'No-Infection ', and vice versa.

Remark: *Sum* of risk of Infection and risk of No-Infection is one. *Product* of odds of Infection and odds of No-Infection is one.

Remark: The odds can also be defined as the ratio of the probability that the event of interest occurs to the probability that it does not. That is, $Odds = \frac{p}{1-p}$, where p is the probability that the event occurs and (1 – p) is the probability that the event does not occur. This definition of odds is used in the specification of

logistic regression model. Although, the above definition of odds involved probability, odds is very different from probability.

4.1. Odds ratio (OR)

The odds of Infection in the Treatment group is $OD_T = a/b$. This is the ratio of the number of Infection divided by number of No-Infection in the Treatment group. The numerator is the same as that of the risk of Infection, but the denominator is different (in fact smaller) from that of the risk. Odds is not a measure of Infection relative to the all possible `events and non-events', rather it is relative to the non-events only. Similarly, the odds of Infection in the Control group is $OD_C = c/d$. This is the ratio of the number of events divided by number of non-events in the Control group.

For the count data in Table 3.1, the odds of Infection in the Treatment (OD_T) and Control (OD_C) groups are:

$$OD_T = a/b = 25/125 = 1/5 \text{ and } OD_C = c/d = 120/30 = 4.$$

So the odds of infection in the Treatment group is 1/5, and the odds of infection in the Control group is 4. Therefore, the odds of infection in the Treatment group is much smaller that in the Control group.

The odds ratio (OR) of Infection in the Treatment group relative to the Control group is then defined as the ratio of the two odds, that it, $OR = \frac{OD_T}{OD_C} = \frac{a/b}{c/d} = \frac{a \times d}{c \times b}$. Then, for the count data in Table 3.1, the odds ratio (OR) of Infection in the Treatment (relative to the Control) group becomes $OR = OD_T / OD_C = \frac{1/5}{4} = \frac{1}{20}$.

4.2. Interpretation of OR

For the above OR = 1/20. one could conclude that the odds of Infection in the vaccine (Treatment) group is 1/20 times the odds of the Control group. The odds of Infection for those who received the vaccine (Treatment) is very small compared to the odds of Infection to those who did not receive vaccine (Control). Thus the vaccine helped reduce the odds of Infection (by a factor of 1/20). This implies that the Treatment (vaccine) works to reduce the Infection.

If $OR = 1$, the odds of Infection in the Treatment group is the same as that in the Control group.

If $1 < OR < \infty$, then the odds of Infection is higher in the Treatment group than in the Control group. As an example, if $OR = 2$ then the odds of Infection in the Treatment group is twice higher than that of the Control group.

If $OR = 0.25$ then the odds of Infection in Treatment group is 0.25 times the odds of Infection in the Control group.

Smaller the value of the OR weaker is the association (dependence) between the two categorical variables.

5. Comparison of RR and OR

Numerical and graphical comparisons between RR and OR are provided in this section.

1. The RR and OR are comparable in magnitude when the event/infection studied is rare or very uncommon in both exposed and unexposed groups.
2. The OR is larger than the RR when the event/infection is more common.
3. In case-control studies, risks and RR can't be calculated but OR can be calculated and use as an approximation of RR only if event/infection 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, $Odds(Event) = \dfrac{1}{Odds(No - Event)}$.
6. If the OR for an event (Infection) deviates from 1 substantially, the OR of its non-event (no Infection) will also deviate from 1 substantially, although in the opposite direction.

It can be shown that the ratio of risk ratios of events/cases (RR_E) and non-events/non-cases (RR_{NE}) is the same as the odds ratio of events/cases (OR_E) in the intervention/treatment group.

From the cell counts in the contingency table (Table 3.1), define the following:

Risk ratio of events/cases in the intervention/treatment group relative to the control/placebo group is

RR Event (Treatment| Control), $RR_E = \frac{a/(a+b)}{c/(c+d)} = \frac{a}{c} \times \frac{c+d}{a+b}$.

Similarly, relative risk of non-events/no-cases in the intervention/treatment group relative to the control/placebo group is RR Non Event (Treatment| Control), $RR_{NE} = \frac{b/(a+b)}{d/(c+d)} = \frac{d}{b} \times \frac{c+d}{a+b}$.

Now, ratio of the above two risk ratios for events and non-events is

$$\frac{RR_E}{RR_{NE}} = \frac{\frac{a}{c} \times \frac{c+d}{a+b}}{\frac{b}{d} \times \frac{c+d}{a+b}} = \frac{a/c}{b/d} = \frac{ad}{bc}$$

which is same as the odds ratio of event (OR_E) in the treatment group relative to the control group, that is, $OR_E = \frac{a/b}{c/d} = \frac{ad}{bc}$.

Remark: The OR of event is the reciprocal of OR of non-event. If the event (Infection) is rare then the OR is closer (or similar) to the RR.

The values of RR and OR depend on the counts in each of the cells in the contingency table (eg Table 3.1). Obviously, the changes in the values of cells a (No of Cases in Treatment group), b (No of Non Cases in Treatment group), c (No of Cases in Control group) and d (No of Non Cases in Control group) will change both RR and OR. The value of RR and OR are equal if and only if

$\frac{a \times d}{b \times c} = \frac{a/(a+b)}{c/(c+d)}$ or $\frac{a}{c} \times \frac{d}{b} = \frac{a}{c} \times \frac{(c+d)}{(a+b)}$, that is, $\frac{d}{b} = \frac{(c+d)}{(a+b)}$.

Thus RR and OR are exactly equal if and only if both a and c are 0. But this is not a realistic situation. However, the values of RR and OR are very close if a and c are very small. This is called the rare disease assumption.

The condition under which OR>RR is $\frac{d}{b} > \frac{(c+d)}{(a+b)}$ or $ad > bc$; otherwise OR<RR.

For the example in Table 3.1, we produced the following graphs for different

values of a and c keeping a+b = c+d = 150 fixed to display the relationship between OR and RR if Figures 1-4.

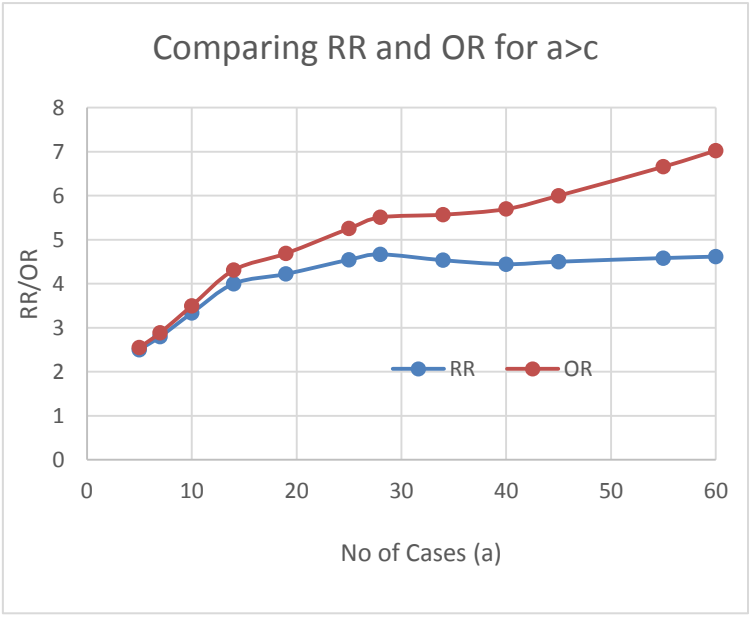

Figure 1: Graph of OR and RR when a>c

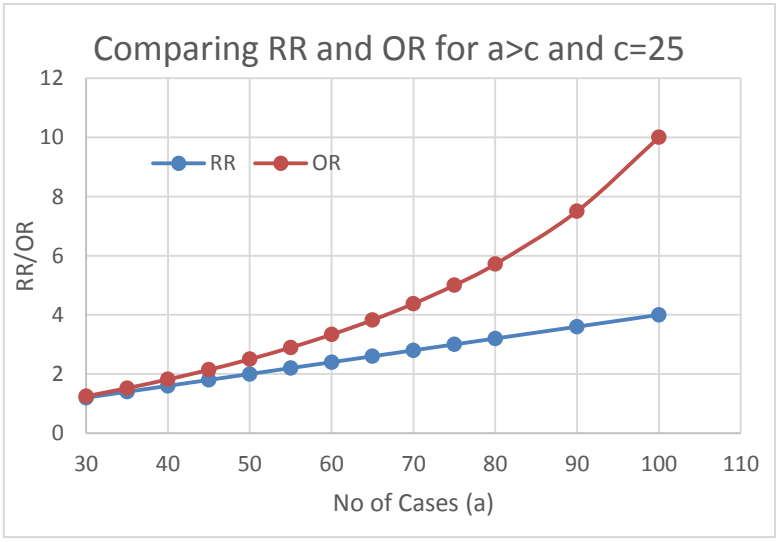

Figure 2: Graph of OR and RR when a>c and c is fixed at 25

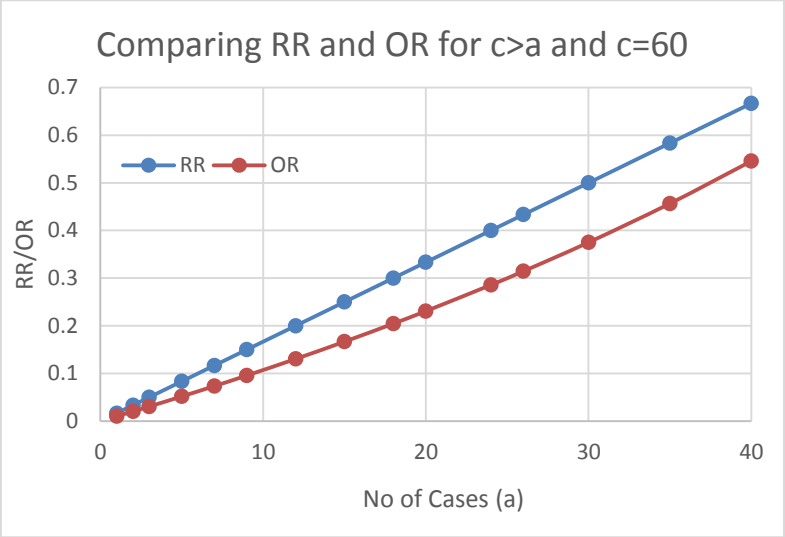

Figure 3: Graph of OR and RR when c>a and c is fixed at 60

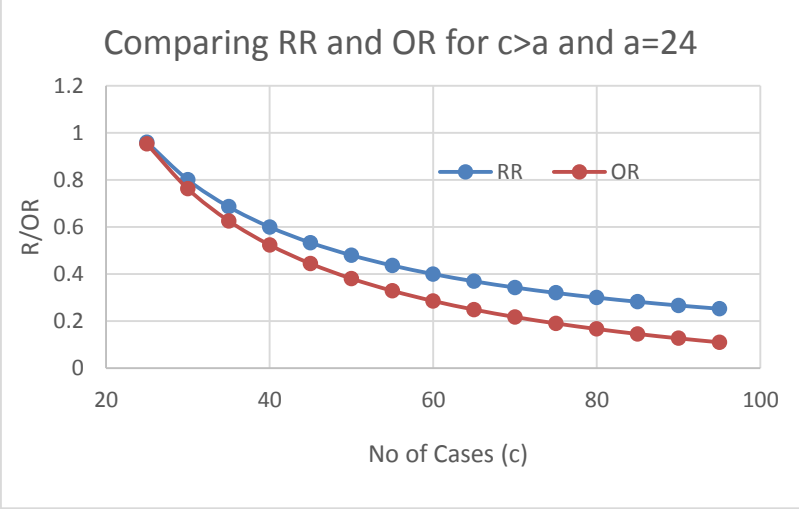

Figure 4: Graph of OR and RR when c>a and a is fixed at 24

For this example, the OR is larger than RR when a >c and c is fixed, and OR is smaller than RR if c>a and a is fixed.

6. OR in logistic regression model

When the response/outcome of a study is a binary variable logistic regression method is used to model data, the dependent variable of this model is expressed in terms of the log of odds ratio, and then attempts are made to predict it based on the values of the predictors (exposure variables).

In many studies the relative risk (RR) is used to measure the effect of the intervention (X, Vaccine/Intervention) on the outcome (Y, Infected or not). The disadvantage of the RR is that it is not a constant effect of X. The effect of X on the probability of Y changes for different values of X. The probability ratio changes depending on the value of X. But the OR does not change with a change in X.

In the logistic regression, the key phrase is constant effect. In regression models, we often want a measure of the unique effect of each X on Y. If we try to express the effect of X on the likelihood of a categorical Y having a specific value through a probability, the effect is not constant. What that means is there is no way to express in one number how X affects Y in terms of a probability.

For this obvious reason, although people would love to use probabilities because of its intuitive appeal, it is just not possible to describe that effect of X on Y in a constant number. So when we are required to communicate that effect to a research audience, we must use odds ratios. This reinforces the need for the research community to properly understand and interpret OR.

The logistic regression model is based on the logit function $Logit(p) = ln\left(\frac{p}{1-p}\right)$ which is the natural logarithm of the odds $\left(\frac{p}{1-p}\right)$, where p is the probability of event (Infection) of the outcome of interest. The logit function is referred to as the *link function* because it links probability of the binary outcome (Y) to the linear function of the predictor variables (X).

The goal of logistic regression is to find the best fitting model to describe the relationship between the binary outcome (response) variable and a set of independent predictors or explanatory variables.

For k predictors, the logistic regression is defined as

$ln\left(\frac{p}{1-p}\right) = \ln(odds) = b_0 + b_1X_1 + +b_2X_2 + ... + b_kX_k$, where b_i 's are the regression coefficients of the predictors, X_i's. From the logistic regression model, we could write

$p = \frac{1}{1+e^{-\left(b_o + \sum_{i=1}^{k} b_i X_i\right)}}$ which enables us to find the probability of the event/case for the given values of the predictors.

For only one predictor the regression coefficient (b_1) represents the estimated increase in the log odds of the outcome per unit increase in the value of the exposure (predictor). In other words, the exponential function of the regression coefficient (e^{b1}) is the odds ratio associated with a one-unit increase in the exposure (predictor).

As an example, consider the data in Table 2 to fit logistic regression of outcome on exposure.

Table Parameter estimates by logistic regression for the data in Table 2

Variables in the Equation							
		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Treatment	-2.996	.299	100.086	1	.000	.050
	Constant	4.605	.483	90.760	1	.000	100.000
a. Variable(s) entered on step 1: Treatment.							

Here the regression coefficient (b_1 = -2.996) implies that the estimated decrease in the log odds of the outcome per unit increase in the value of the intervention. In other words, the exponential function of the regression coefficient e^{b1} = 0.05 (rounded from 0.049986) is the odds ratio associated with a one-unit increase in the exposure. Thus the results are expressed and interpreted in terms of odd ratio or log of odds ratio. Once again the exposure (vaccine) is significant as the p-value of the test on the slope is 0.

7. Conversion of OR to RR

There is no need to convert OR to RR if OR is properly interpreted and understood. However, some researchers prefer 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 able to interpret results based on OR.

(Greenland and Holland 1991) proposed the conversion formula of RR from OR as

$RR = \dfrac{OR}{1 - I_C + I_C \times OR}$ and it was later reported by (Zhang and Yu 1998).

Recall, I_E is the *incidence rate* in the exposed (vaccine) group and I_C is that in control group.

It is interesting to note that $(1 - I_C) = 1 - \dfrac{c}{c+d} = \dfrac{d}{c+d} = I_C^0$ 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, $(1 - I_E) = 1 - \dfrac{a}{a+b} = \dfrac{b}{a+b} = I_E^0$, and hence I_E and I_E^0 add to one.

From the data in Table 3.1 it can be seen that

$I_E = \dfrac{a}{a+b} = R_E(T)$, risk of event in the exposure/treatment group, and

$I_C = \dfrac{c}{c+d} = R_E(C)$, risk of event in the unexposed/control group.

Therefore, the risk ratio of event in the treatment group relative to the control group is

$$RR_E(T \mid C) = \frac{R_E(T)}{R_E(C)} = \frac{I_E}{I_C}.$$

Similarly, the risk ratio of non-event in the treatment group relative to the control group is

$RR_{NE}(T \mid C) = \dfrac{R_{NE}(T)}{R_{VE}(C)} = \dfrac{(1-I_E)}{(1-I_C)}$, where $R_{NE}(T) = (1-I_E)$, `risk' of non-event in the treatment group, and $(1-I_C) = R_{NE}(C)$, `risk' of non-event in the control group

Lemma: The RR of event is equal to the product of the RR of non-event and OR of event.

Proof:

The odds ratio of event in the treatment group relative to the control group is $OR_E(T \mid C) = \dfrac{a/b}{c/d} = \dfrac{ad}{bc}$ and the risk ratio of non-event relative to the control group is given by $RR_{NE}(T \mid C) = \dfrac{R_{NE}(T)}{R_{VE}(C)} = \dfrac{\frac{b}{a+b}}{\frac{d}{c+d}} = \dfrac{b(c+d)}{d(a+b)}$.

The product of the two becomes,

$OR_E(T \mid C) \times RR_{NE}(T \mid C) = \dfrac{ad}{bc} \times \dfrac{b(c+d)}{d(a+b)} = \dfrac{\frac{a}{a+b}}{\frac{c}{c+d}} = RR_E(T \mid C)$ which is the risk ratio of event in the treatment group relative to the control group.

Hence $RR_E(T \mid C) = RR_{NE}(T \mid C) \times OR_E(T \mid C)$.

8. Concluding remarks

Although both risk and odds represent `likelihood of events', they are very different because they are measured in different scales. Hence the RR is different from the OR, and they should not be used interchangeably. While interpreting the results, the RR should be interpreted in terms of risk and OR should be interpreted in terms of odds.

In practice, both RR and OR are used to measure the association between two categorical variables. Depending on the study design and objective of the study, researcher may select the appropriate measure to determine the association between two categorical outcome variables. However, if someone wants to know how any exposure (eg vaccine) affects the outcome (eg Infection), OR is the best effect size measure.

When the outcome is not rare in the population, if the OR is used to estimate the RR it will overstate the effect of the treatment on the outcome measure. The OR will be greater than the RR if the relative risk is greater than one and less than the relative risk otherwise.

It is essential to make a clear distinction between RR and OR and use anyone of them consistently without interchanging or mixing between them. Describing risk in terms of probability is perfectly appropriate but that is not the case with odds. The interpretation of OR must be in terms of odds, not risk. When OR is converted to RR using appropriate method, valid interpretations can be made in terms of risk, if needed.

Acknowledgments: The authors are thankful to anonymous referees for valuable comments and suggestions to improve the presentation of the paper.

References

[1] Bland, J Martin, and Douglas G Altman (2000). The odds ratio. *BMJ* no. 320 (7247):1468. doi: 10.1136/bmj.320.7247.1468.

[2] Boston University, School of Public Health (2020). The Difference Between "Probability" and "Odds" 2020 [cited 26 May 2020]. Available from http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_Confidence_Intervals/BS704_Confidence_Intervals10.html.

[3] Chen, Henian, Patricia Cohen, and Sophie Chen (2010). How Big is a Big Odds Ratio? Interpreting the Magnitudes of Odds Ratios in Epidemiological Studies. Communications in Statistics - Simulation and Computation no. 39 (4):860-864. doi: 10.1080/03610911003650383.

[4] Greenland, Sander, and Paul Holland (1991). Estimating Standardized Risk Differences from Odds Ratios. Biometrics no. 47 (1):319-322. doi: 10.2307/2532517.

[5] Khan, S. (2020). Meta-Analysis Methods for Health and Experimental Studies. Singapore: Springer Nature.

[6] Merrill, Ray M. (2021). Introduction to epidemiology. Eighth edition. ed. Burlington, Massachusetts: Jones & Bartlett Learning.

[7] Zhang, Jun, and Kai F. Yu. (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*, The Journal of the American Medical Association no. 280 (19):1690. doi: 10.1001/jama.280.19.1690.