Open lecture on statistics

ISSN 2234-7658 (print) / ISSN 2234-7666 (online)
https://doi.org/10.5395/rde.2017.42.1.72

RDE Restorative Dentistry & Endodontics

Statistical notes for clinical researchers: Risk difference, risk ratio, and odds ratio

Hae-Young Kim*

Department of Health Policy and Management, College of Health Science, and Department of Public Health Sciences, Graduate School, Korea University, Seoul, Korea

In the previous section, we discussed risk and odds. Both risk and odds can be applied to a cohort study designs based on population. On the other hand, a case-control study is not based on population but designed by separate sampling procedures in the disease group and no disease group. Therefore, there is no denominator to estimate the risk in the entire population and only odds can be obtained in the case-control design. Regarding those study designs, we'll talk about definitions, applicability, difference, and interpretation of risk difference (RD), risk ratio (RR), and odds ratio (OR) as measures of effects in studies with cohort and case-control design.

Definition

1. Risk difference (RD), attributable risk (AR), excessive risk

RD or AR is defined as the difference in risk of a condition such as a disease between an exposed group and an unexposed group (Table 1).

RD = risk of exposure group - risk of unexposed group = (a / [a + c]) - (b / [b + d])

2. Risk ratio or relative risk (RR)

RR is the ratio between the risk of exposed group and unexposed group.

$$RR = \frac{\text{risk of exposed group}}{\text{risk of unexposed group}} = \frac{a / (a + c)}{b / (b + d)}$$

3. Odds ratio (OR)

OR is the ratio between odds of exposed group and unexposed group.

$$OR = \frac{\text{odds of exposed group}}{\text{odds of unexposed group}} = \frac{a / c}{b / d}$$

Table 1. An example of 2*2 cross table showing formulas of risk difference, risk ratio, and odds ratio

	Exposure	No exposure	Total
Disease	a	b	a + b
No disease	c	d	c + d
Total	a + c	b + d	a + b + c + d
Risk (of disease)	a / (a + c)	b / (b + d)	
Odds	a / c	b / d	
Risk difference (RD)			(a / [a + c]) - (b / [b + d])
Risk ratio (RR)			(a / [a + c]) / (b / [b + d])
Odds ratio (OR)			(a / c) / (b / d)

***Correspondence to**
Hae-Young Kim, DDS, PhD.
Associate Professor, Department of Health Policy and Management, College of Health Science, and Department of Public Health Sciences, Graduate School, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul, Korea 02841
TEL, +82-2-3290-5667; FAX, +82-2-940-2879; E-mail, kimhaey@korea.ac.kr

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

©Copyrights 2017. The Korean Academy of Conservative Dentistry.

Risk difference, risk ratio, and odds ratio as measures of effects in cohort design

A cohort study design pursues the effect of exposure such as treatment, prospectively. In the cohort study, we extract an adequate size of a random sample from the target population, then randomly assign the subjects into either the expose group or unexposed group. The effect of exposure is observed as the changes in outcome of interest over time. Risk is easily calculated as the number of persons having the disease in exposed and unexposed groups divided by the number of all the persons in both groups. In the cohort study, we have a clear denominator: the number of persons assigned in the groups. RD and RR are frequently used to assess the association between the exposed and control groups. RD, which is also known as AR or excessive risk, represents the amount of risk, which decreased or increased when the exposure exists compared to that when the exposure is absent. A positive RD value means increased risk and a negative one means decreased risk by the exposure. RR is calculated as the risk of an exposed group divided by the risk of an unexposed group. A RR value of 1 means no difference in risk between groups, and larger or smaller values mean increased or decreased risk in an exposed group compared to the risk in an unexposed group, which can be interpreted that the occurrence of disease is more or less likely in the exposed group, respectively.

In addition, we can also use OR for the same purpose in cohort studies. OR is the ratio of odds of disease in an exposed group and an unexposed group. The interpretation OR is not as intuitive as RR. An OR value of 1 means no difference in odds between groups, and larger value than 1 means increased odds in exposed group, interpreted as a positive association between having disease and having exposure. Contrarily an OR value of smaller than 1 means decreased odds in exposed group which is interpreted as the association between having disease and not having exposure. Though the interpretation of OR is similar to that of RR, they have similar values only when risks of both groups are very low, *e.g.*, $p < 0.1$. Otherwise, they show different values. As seen in Table 2, the values of RR and OR are approximately the same only when risk of both groups are very low ($p < 0.1$, Examples 1 - 5 in Table 2). However, when risks of either one or both groups are not very low ($p > 0.1$), there are considerable discrepancy between RR and OR values (Examples 6 - 14, Table 2). A general rule is that a value of OR is always reflecting larger effect size or stronger association, by showing smaller OR values than corresponding RR values when RR < 1 and larger OR values when RR > 1. In Table 2, we can confirm that all the cases with RR larger than 1 had much larger OR values (Examples 6 - 8 and 10 - 14), and a case with RR smaller than 1 had a smaller OR value than the corresponding RR value (Example 9). Therefore, incorrect interpretation of OR value as RR will lead to an overstatement of the effect by either erroneously increasing or decreasing the true risks. Figure 1 depicts that the differences between OR and RR values get larger as the levels of baseline risk in the control group (I_0) increase.[1] Especially when baseline risk is as large as 0.5, the maximum RR value is confined to 2, while OR value approaches infinity.

Table 2. Comparison of risk difference, risk ratio, and odds ratio based on risks (p) and odds of two competitive groups (assume n = 1,000 per group)

	No. of event		Risk (p)		Odds		Risk difference	Risk ratio	Odds ratio
Example	Control	Tx.	Control (1)	Tx. (2)	Control (3)	Tx. (4)	(2) - (1)	(2) / (1)	(4) / (3)
1	1	2	0.001	0.002	0.001	0.002	0.001	2.000	2.000
2	5	10	0.005	0.010	0.005	0.010	0.005	2.000	2.000
3	10	20	0.010	0.020	0.010	0.020	0.010	2.000	2.000
4	15	30	0.015	0.030	0.015	0.031	0.015	2.000	2.067
5	50	100	0.050	0.100	0.053	0.111	0.050	2.000	2.096
6	100	200	0.100	0.200	0.111	0.250	0.100	2.000	2.252
7	200	400	0.200	0.400	0.250	0.667	0.200	2.000	2.668
8	200	700	0.200	0.700	0.250	2.333	0.500	3.500	9.333
9	500	200	0.500	0.200	1.000	0.250	-0.300	0.400	0.250
10	500	600	0.500	0.600	1.000	1.500	0.100	1.200	1.500
11	500	700	0.500	0.700	1.000	2.333	0.200	1.400	2.333
12	500	990	0.500	0.990	1.000	99.00	0.490	1.980	99.00
13	900	950	0.900	0.950	9.000	19.00	0.050	1.060	2.111
14	998	999	0.998	0.999	499.0	999.0	0.001	1.001	2.002

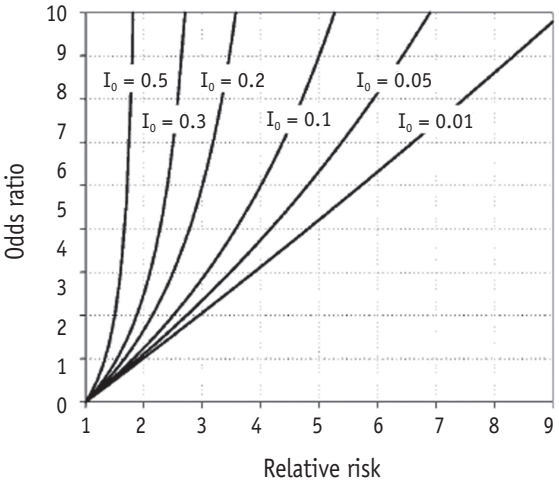


Figure 1. Relationship between odds ratio and relative risk at various levels of baseline risks in the control group (I_0 = 0.5, 0.3, 0.2, 0.1, 0.05, and 0.01).[1] I_0, baseline risk of control group.

OR has been used as a very popular estimate of effect in epidemiological studies. As the logistic regression has been frequently used in multivariate assessment of binary outcomes, OR which is the exponentiated regression coefficient from logistic regression has been popular, too. The logistic regression has a computational advantage that the convergence is efficient because the related logit link can convert risk (p) values, confined from 0 to 1, into log odds values ranging from negative infinity to positive infinity. Fortunately, lots of life-threatening diseases tend to have very low risk (or prevalence), *e.g.*, lower than 0.1, therefore use of OR can be justified as a good estimator of RR. However, when we analyze data of prevalent diseases such as dental caries or periodontitis, we need to be careful not to interpret the strong association by OR as if it is by RR. Because the OR value is far from 1 than the corresponding RR value when the disease is not rare, to avoid possible mistake of overstating the effect, the resulting OR value can be converted into RR using following equation only when baseline risk can be appropriately assumed:

$$RR = \frac{OR}{1 - I_0 * (1 - OR)}$$, where I_0 is baseline risk of control group.[2]

When the outcome is not rare, Poisson regression or log-binomial model are preferred to obtain RR instead of logistic regression.

Odds ratio as the measure of effects in case-control designs

When we are interested in a disease that is very rare, implementing a study with a cohort design is disadvantageous because not only it requires a long observation period and high cost but also it is very difficult to get sufficient information on occurrence of the disease. Case-control design offers a more efficient alternative in such a situation. In case-control design, subjects were selected from disease and no disease groups separately. The sample size of control group without disease is determined as appropriate, only considering the sample size of disease group. Therefore, there is no denominator which is needed to calculate a risk because information on the entire population is not collected. Computing both risk (p) and RR is impossible in the case-control design.

However, odds and OR, ratio of odds in exposed and unexposed group, can be easily obtained in this design. When the disease is very rare, we can use OR as an approximation of RR. Also different size of no disease group is applicable as the study design is needed. Table 3 shows the value of RR can be easily changed by only arbitrarily changing the size of no disease group from 1.91 to 1.98, while OR shows consistency having the same value of 2. Therefore, you should not use risk or RR inappropriately in studies with case-control design. When the outcome of interest has very low prevalence, OR calculated in case-control studies can be used as an approximation of RR.

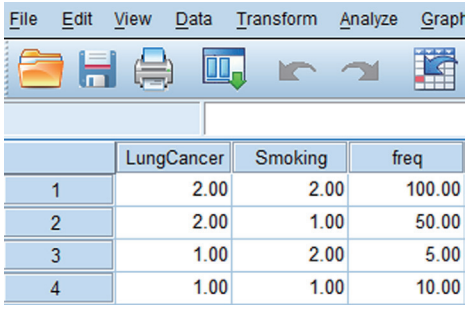
https://doi.org/10.5395/rde.2017.42.1.72

Table 3. Comparison of odds ratio (OR) and relative risk (RR) in two case-control designs with different size of no disease group

	Disease	No disease	Total		Disease	No disease	Total
Exposure	10	100	110	Exposure	10	500	510
No exposure	5	100	105	No exposure	5	500	505
	OR = (10/100) / (5/100) = 2.00				OR = (10/500) / (5/500) = 2.00		
	RR = (10/110) / (5/105) = 1.91				RR = (10/510) / (5/505) = 1.98		

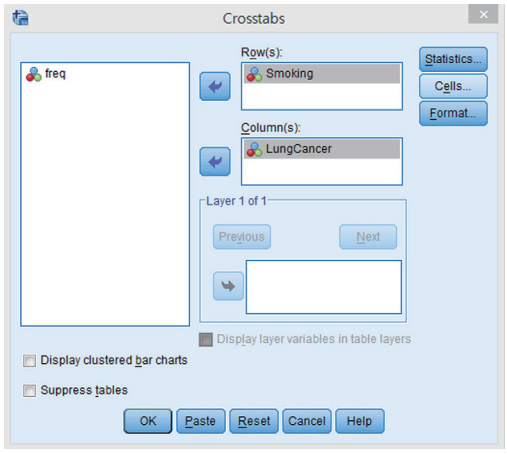
The procedure of obtaining OR and RR using IBM SPSS Statistics for Windows Version 23.0 (IBM Corp., Armonk, NY, USA)

(a) Data (weight by freq)

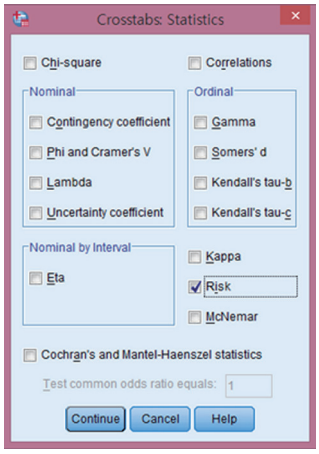


	LungCancer	Smoking	freq
1	2.00	2.00	100.00
2	2.00	1.00	50.00
3	1.00	2.00	5.00
4	1.00	1.00	10.00

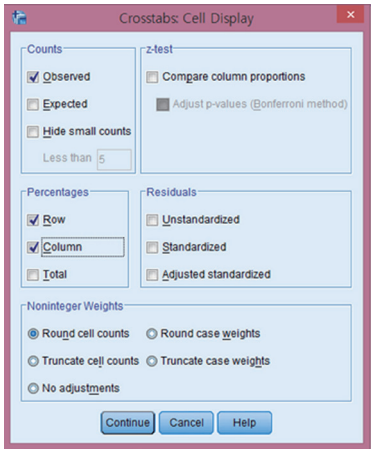
(b) Analyze-Descriptive Statistics-Crosstabs



(b-1) Statistics



(b-2) Cells



(c) Crosstable

Smoking * LungCancer Crosstabulation

			LungCancer		
			1.00	2.00	Total
Smoking	1.00	Count	10	50	60
		% within Smoking	16.7%	83.3%	100.0%
		% within LungCancer	66.7%	33.3%	36.4%
	2.00	Count	5	100	105
		% within Smoking	4.8%	95.2%	100.0%
		% within LungCancer	33.3%	66.7%	63.6%
Total		Count	15	150	165
		% within Smoking	9.1%	90.9%	100.0%
		% within LungCancer	100.0%	100.0%	100.0%

(d) OR (95% CI) and RR (95% CI)

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Smoking (1.00 / 2.00)	4.000	1.297	12.332
For cohort LungCancer = 1.00	3.500	1.255	9.761
For cohort LungCancer = 2.00	.875	.775	.988
N of Valid Cases	165		

References

1. Schmidt CO, Kohlmann T. When to use the odds ratio or the relative risk? *Int J Public Health* 2008;53:165-167.
2. Zhang J, Yu KF. What is the relative risk? A method to directly estimate risk ratios in cohort studies of common outcomes. *J Am Med Assoc* 1998;280:1690-1691.

https://doi.org/10.5395/rde.2017.42.1.72