Measurement of Risk

Saengsit Kritsadee, ATM.Dr., Ph.D.

Measurement of Risk

1. Risk Factor
2. Major Risk Factors
3. Potential impact of intervention program

- Relative Risk (biological significance)
- Attributable Risk (public health significance)

1. Relative Risk

$$\text{Relative risk (R)} = \frac{IE}{INE} \qquad (1)$$

1.1 Prospective or Cohort Study

R =Ie/Io

(Exposure factor)	(Disease)		(Total)
	(Present)	(Absent)	
(Present)	a	c	a+c
(Absent)	b	d	b+d
(Total)	a+b	c+d	a+b+c+d

IE = a/(a+b)

INE = c/(c+d)

$$R = \frac{a/(a+b)}{c/(c+d)}$$

$$= \frac{a/\ (c+d)}{c/\ (a+b)}$$

1.2 Unmatched case-control study

Odds ratio / Relative risk

(Disease)	(Exposure factor) (Present)	(Exposure factor) (Absent)	(Total)
(Present)	a	c	a+c
(Absent)	b	d	b+d
(Total)	a+b	c+d	a+b+c+d

Odds Ratio

$$= \frac{ad}{bc}$$

(rare disease)

1.3 Matched case-control study

(Matched)

(cases)	(controls)		(total)
	+ (factor present)	- (factor absent)	
+ (factor present)	a	b	a+b
- (factor absent)	c	d	c+d
(total)	a+c	b+d	a+b+c+d

2.1 Exposure attributable risk , Attributable risk

1. Absolute attributable risk in the exposed

   $$ARE = IE - INE$$

2. Attributable risk percent in the exposed

   $$AR\%E = \frac{IE - INE}{IE} \times 100\% \quad \text{หรือ} = \frac{R - 1}{R} \times 100\%$$

2.2 Population attributable risk

1. Absolute attributable risk in the population

   ARP = IP – INP

2. Attributable risk percent in the population

   $$AR\%P = \frac{Pe(R-1) \times 100\%}{1 + Pe(R-1)}$$

   เมื่อ R=relative risk = IE/INE

Odds ratio and 95% confidence interval

Lung cancer

	yes	no	total
smoker	a	b	a+b
Non-smoker	c	d	c+d
total	a+c	b+d	a+b+c+d

E = exposed, E = not exposed, D = diseased, D = non diseased

Odd ratio = OR = odds of smoking among those w/lung cancer

$$= \frac{(a/c)}{(b/d)} = \frac{ad}{bc}$$

Odds ratio and 95% confidence interval

- If there is no association between smoking and lung cancer, the odds of smoking are the same for those with and those without lung cancer, and thus OR = 1
- If the odds of smoking are greater among those with than those without lung cancer, and thus OR >1.
- If the odds of smoking are smaller among those with than those without lung cancer, and thus OR <1.
- The more different from 1 the OR, the stronger the association.

Example: Case-Control Study

Lung cancer

	yes	no	total
smoker	30	40	70
Non-smoker	10	120	130
total	40	160	200

Odd ratio = OR = odds of smoking among those w/lung cancer

$$= \frac{ad}{bc} = \frac{30\text{x}120}{40\text{x}10} = 9$$

The odds of smoking are 9 times as high among people with lung cancer as among people without lung cancer.

Odds ratio and 95% confidence interval

A large sample confidence interval for the Odds Ratio can be calculated in two steps.

1. Calculate the confidence interval for ln (OR)

   $\ln(OR) \pm 1.96 \sqrt{1/a + 1/b + 1/c + 1/d}$

2. Exponential the interval endpoints.

Example:

1. $\ln(9) \pm 1.96 \sqrt{1/30 + 1/40 + 1/10 + 1/120} = 2.2 \pm 0.8$ or (1.4,3.0)
2. 95% confidence interval for the OR: ($e^{1.4}$, $e^{3.0}$) or (4.1, 20.1). We can be 95% confident that the true OR lies between 4.1 and 20.1.

The Relative Risk

$$RR = \frac{\text{probability of developing lung cancer among smokers}}{\text{probability of developing lung cancer among non-smokers}} = \frac{a/n_E}{c/n_E}$$

If there is no association between smoking and lung cancer, the probability of developing lung cancer is the same for smokers and non-smokers, and thus $RR = 1$.

If the probability of developing lung cancer is greater among smokers than non-smokers, then $RR > 1$.

If the probability of developing lung cancer is smaller among smokers than non-smokers, then $RR < 1$.

The more different from 1 the RR, the stronger the association.

Example: Cohort Study

Lung cancer

	yes	no	total
smoker	30	40	70
Non-smoker	10	120	130
total	40	160	200

$$RR = \frac{10/698}{2/1302} = 9.3$$

The probability of developing lung cancer is 9 times as high for smokers as for non-smokers.

Note: For rare diseases (such as all cancers) the odds ratio approximates the relative risk well. This is important since the relative risk is the measure we care really interested in.

in our example $OR = \frac{10 \times 1300}{688 \times 2} = 9.4$

Thus, if the above data had come from a case-control study the OR we would have calculated in this case would have been a close approximation of the RR.

This would have affected our interpretation of the OR. Instead of saying “the odds of smoking are 9.4 times as high among people with lung cancer as among people without lung cancer” we could have said “the probability of developing lung cancer is 9.4 times as high for smokers as for non-smokers”, the latter being the interpretation of the relative risk.

Relative risk and 95% confidence interval

A large sample confidence interval for the relative risk can be calculated in two steps.

1. Calculate the confidence interval for ln (RR)

$$\ln(RR) \pm 1.96 \sqrt{\frac{1-(a/n_E)}{a} + \frac{1-(c/n_E)}{c}}$$

2. Exponential the interval endpoints.

Example:

$$\ln(9.3) \pm 1.96 \sqrt{\frac{1-(10/698)}{10} + \frac{1-(2/1302)}{2}}$$

95% confidence interval for the RR: $(e^{0.7}, e^{3.7})$ or (2.0, 40.4). We can be 95% confident that the true RR lies between 2.0 and 40.4.

Interpreting 95% Confidence Intervals

Probabilistic definition:

In repeated sampling, from a normally distributed population, 95% of all confidence intervals will in the long run include the true value.

Practical definition:

We are 95% confident that the true value is included in the one confidence interval we calculated.

Interpreting 95% Confidence Intervals

For OR and RR:

1 is the null value, meaning that if OR and RR are 1 there is no association between the exposure and the outcome. If the 95% confidence interval does not include 1 we can say that, at the 0.05 level of significance, the exposure is a significant risk factor of the outcome.

In the above example we found a RR of 9.3 with 95% confidence interval (2.0, 40.4).

What does this tell us?

We are 95% confidence that the RR is somewhere between 2 and 40 (more likely in the center of the interval than in the tails). 1 is not included in the interval and the interval is heavily skewed (meaning that it includes more values much different from 1 than values close to 1).

Thus, even though the confidence interval is very wide and we cannot be sure what the true value of the RR is, we can be quite confident that the RR is, we can be quite confident that the RR is considerably greater than 1, and that smoking is indeed a risk factor of lung cancer.

Note

The width of a confidence interval depends on the sample size and on the distribution of the study subjects in the exposure and outcome groups.

- If the sample size is small the confidence interval tends to be wide.
- If only few individuals are exposed or diseased, the confidence interval tends to be wide even if the sample size is large.

And example of the second point is the confidence interval for the RR calculated above. Only 12 people have hung cancer and therefore the confidence interval is wide even though the total sample size is 2000.

Note

If we use a confidence interval to determine whether an exposure is a significant risk factor of the outcome we should not only check whether 1 is included in the interval.

Example [ ] confidence interval in study A

[ ] confidence interval in study B

OR = 1

Study A: Is exposure E a risk factor of the disease?

Study B: Is exposure F a risk factor of the disease?

Note that the confidence interval in study A is very narrow (probably due to a very large sample size), whereas the confidence interval in study B is very wide.

Based on statistical significance alone we would conclude that exposure E is a risk factor, but exposure F is not.

However, we can be 95% confident that the OR of E is near 1, whereas the OR of F is likely to be much larger than 1 (even though the confidence interval includes 1, most values in the interval are much greater than 1).

Thus, a better interpretation of the confidence intervals would be:

Exposure E appears to only minimally increase the risk of developing the disease (if at all), but exposure F seems to have a big effect on the outcome.

Reference

- ไพบูลย์ โล่ห์สุนทร ระบาดวิทยา ภาควิชาเวชศาสตร์ป้องกัน คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย 2540
- Annette Bachand, Introduction to Epidemiology: Colorado State University, Department of Environmental Health 1998
- Annette Rossignol, Epidemiology class note: Oregon State University, Department of Public Health 2001
- Kenneth J.Rothman and Sander Greenland (1998) Modern Epidemiology, Lippincott Williams and Wilkins, USA.