MSMS 304 - Biostatistics

Effect Measures for Different Study Designs

Ananda Biswas

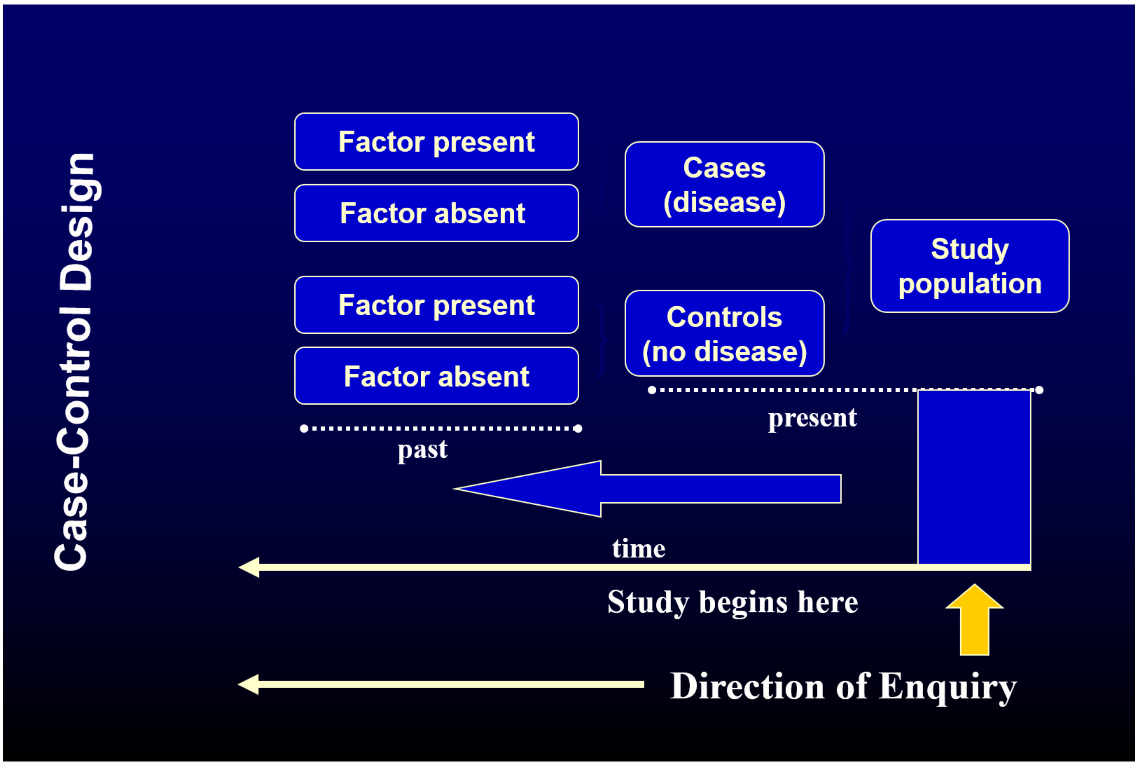
Last updated : November 23, 2025

Contents

1 Effect Measure for Case-Control Design

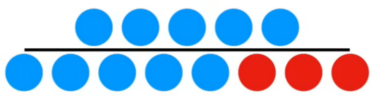
Case-control studies are retrospective studies where we try to determine whether the cases and controls differ in their past exposure to potential risk factors. Here our purpose is to identify associations between exposures and outcomes, especially for rare diseases.

A flowchart of case-control study can be depicted as follows :

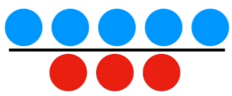


1.1 Odds

(In very simple terms) Probability is the ratio of something happening to everything that could happen.



Odds are the ratio of something happening to something not happening.



Thus for an event A the odds of A or odds in favor of A are

$$\text{odds}(A) = \frac{\Pr(A)}{1 - \Pr(A)} \quad i.e. \quad \frac{\text{probability of occurence of an event}}{\text{probability of non-occurence of the same event}}$$

and of course $\Pr(A) = \dfrac{\text{odds}(A)}{1 + \text{odds}(A)}$. Also $\text{odds}(A^c) = \dfrac{1}{\text{odds}(A)}$.

For example, probability of getting a head in a single toss of a fair coin is 0.5 and so is the probability of getting a tail in a single toss of a fair coin. Thus the odds of getting a heads in a single toss of a fair coin is $\dfrac{0.5}{0.5} = 1$.

1.2 Odds Ratio

Now suppose we have a risk factor X that people either have (X) or do not have ($\overline{X}$) and there is a disease D that people either have (D) or do not have ($\overline{D}$). In particular there are not any intermediate levels of the risk factor or the disease in this setup. A population of people could have these probabilities :

	$\overline{D}$	D
$\overline{X}$	π_{00}	π_{01}
X	π_{10}	π_{11}

Now, $P(D\,|X) = \dfrac{\pi_{11}}{\pi_{10}+\pi_{11}}$ and $P(\overline{D}\,|X) = \dfrac{\pi_{10}}{\pi_{10}+\pi_{11}}$. So $\text{odds}(D\,|X) = \dfrac{\pi_{11}}{\pi_{10}}$.

Similarly, $P(D\,|\overline{X}) = \dfrac{\pi_{01}}{\pi_{00}+\pi_{01}}$ and $P(\overline{D}\,|\overline{X}) = \dfrac{\pi_{00}}{\pi_{00}+\pi_{01}}$. So $\text{odds}(D\,|\overline{X}) = \dfrac{\pi_{01}}{\pi_{00}}$.

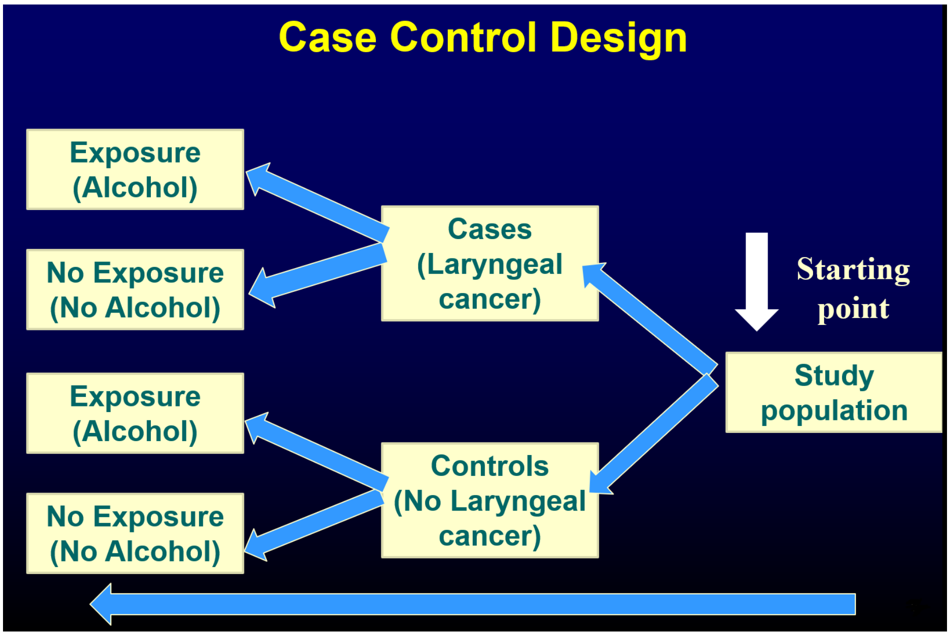
From this we formulate, ***odds ratio***

$$\Lambda = \frac{\text{odds}(D\,|X)}{\text{odds}(D\,|\overline{X})} = \frac{\pi_{11}/\pi_{10}}{\pi_{01}/\pi_{00}} = \frac{\pi_{00}\cdot\pi_{11}}{\pi_{01}\cdot\pi_{10}}.$$

Odds ratio deals with quantifying the associations between two categorical variables. Also it gives us a *direction of association*.

1.3 Statistical Measure of Risk : Odds Ratio

Suppose we have a case-control study design to answer a research question *Is Laryngeal cancer associated to Alcohol consumption ?* Here is the study flowchart.



After data collection, we have the following contingency table.

Exposure Status	Outcome (Laryngeal Cancer)		Total
	Cases (Yes)	Controls (No)	
Exposed (Alcohol)	a	b	$a+b$
Not exposed (No alcohol)	c	d	$c+d$
Total	$a+c$	$b+d$	N

Table 1: Contingency Table

The primary statistical measure of risk in case-control study design is ***odds ratio*** defined as

$$\frac{\text{Odds of Exposure among Cases}}{\text{Odds of Exposure among Controls}}.$$

From the table above, Odds of Exposure among Cases $= \dfrac{\frac{a}{a+c}}{1-\frac{a}{a+c}} = \dfrac{a}{c}$ and

Odds of Exposure among Controls $= \dfrac{\frac{b}{b+d}}{1-\frac{b}{b+d}} = \dfrac{b}{d}$.

Thus, *odds ratio* $= \dfrac{a/c}{b/d} = \dfrac{ad}{bc}$.

1.4 Examples of Odds Ratio

Example 1.4.1. Consider the following contingency table.

Exposure Status	Laryngeal Cancer	
	Yes	No
Alcohol	160	90
No alcohol	40	110

Odds of alcohol consumption among Laryngeal cancer cases $= \dfrac{160}{40} = 4$.

Odds of alcohol consumption among Non-Laryngeal cancer cases $= \dfrac{90}{110} = 0.82$.

$\therefore$ OR (Odds Ratio) $= \dfrac{4}{0.82} = 4.88$.

Interpretation : (*Remember that case-control study is a retrospective study. So we will comment about the exposure in the past.*) The subjects having Laryngeal cancer (case) had 4.88 times more exposure to Alcohol (exposure) than compared to the subjects who did not have Laryngeal cancer.

Example 1.4.2. Consider the following contingency table.

Physical Exercise	Myocardial Infarction (MI)[1]	
	Yes	No
Yes	27	59
No	73	41

Odds of Physical Exercise among MI cases $= \dfrac{27}{73} = 0.37$.

Odds of Physical Exercise among Non-MI cases $= \dfrac{59}{41} = 1.44$.

$\therefore$ OR (Odds Ratio) $= \dfrac{0.37}{1.44} = 0.257$.

Interpretation : The subjects having MI (case) had 0.257 times less exposure to Physical Exercise than compared to the subjects who did not have MI.

Example 1.4.3. Consider the following contingency table.

Physical Exercise	Myocardial Infarction (MI)	
	Yes	No
No	73	41
Yes	27	59

Odds of no Physical Exercise among MI cases $= \dfrac{73}{27} = 2.704$.

Odds of no Physical Exercise among Non-MI cases $= \dfrac{41}{59} = 0.695$.

$\therefore$ OR (Odds Ratio) $= \dfrac{2.704}{0.695} = 3.891 (\approx 1/0.257)$.

Interpretation : The subjects having MI (case) had 3.891 times more exposure to no Physical Exercise than compared to the subjects who did not have MI.

1.5 Confidence Interval for an Odds Ratio

The confidence interval for any parameter is given by

$$(\text{point estimate} \pm \text{critical value} \times \text{standard error}).$$

For odds ratio, the confidence interval is calculated on the natural log ($\log_e$) scale and then converted back to the original scale.

Steps involved in calculating confidence interval for an odds ratio are as follows :

[1] heart attack in fancy terms

Step I : Calculate the odds ratio from the data.

Step II : Find the natural log *i.e.* $\log_e$ of odds ratio.

Step III : The critical value is from the standard normal distribution : 1.96 for 95% confidence interval, confidence coefficient = 0.95.

Step IV : Calculate Standard error for $ln(OR)$.

Exposure Status	Outcome	
	Cases	Controls
Exposed	a	b
Not Exposed	c	d

For a 2×2 contingency table as above, Standard Error for $ln(OR)$ is given by

$$SE(ln(OR)) = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}.$$

Step V : 95% CI on log scale is $ln(OR) \pm 1.96 \times SE(ln(OR)) = (L, U)$, say.

Step VI : Get the confidence interval limits on the original scale as $\left(e^L, e^U\right)$.

1.6 Examples of CI for an Odds Ratio

Example 1.6.1. In Example (1.4.1), $OR = 4.88$. $\therefore ln(OR) = 1.585$.

$$SE(ln(OR)) = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}} = \sqrt{\frac{1}{160} + \frac{1}{90} + \frac{1}{40} + \frac{1}{110}} = 0.2268.$$

$\therefore (L, U) = (1.585 \pm 1.96 \times 0.2268) = (1.14, 2.03)$.

$\therefore \left(e^L, e^U\right) = (3.127, 7.614)$.

Example 1.6.2. In Example (1.4.2), $OR = 0.257$. $\therefore ln(OR) = -1.357$.

$$SE(ln(OR)) = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}} = \sqrt{\frac{1}{27} + \frac{1}{59} + \frac{1}{73} + \frac{1}{41}} = 0.3034.$$

$\therefore (L, U) = (-1.357 \pm 1.96 \times 0.3034) = (-1.95, -0.76)$.

$\therefore \left(e^L, e^U\right) = (0.142, 0.468)$.

Example 1.6.3. In Example (1.4.3), $OR = 3.891$. $\therefore ln(OR) = 1.359$.

$$SE(ln(OR)) = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}} = \sqrt{\frac{1}{73} + \frac{1}{41} + \frac{1}{27} + \frac{1}{59}} = 0.3034.$$

$\therefore (L, U) = (1.359 \pm 1.96 \times 0.3034) = (0.764, 1.954)$.

$\therefore \left(e^L, e^U\right) = (2.147, 7.057)$.

1.7 Test of Significance for Odds Ratio

An odds of 1 implies occurrence and non-occurrence of the event of interest are equally likely. Odds > 1 implies occurrence of the event is more likely whereas odds < 1 implies non-occurrence of the event is more likely.

Suppose we have an exposure E and an outcome O. An odds ratio of 1 for O implies odds of O are equal for presence and absence of E *i.e.* E does not have any significant influence on occurrence or non-occurrence of O. So odds ratio $= 1$ indicates no association between E and O whereas odds ratio > 1 indicates presence of E increases occurrence of O *i.e.* *exposure to E is a risk factor* and odds ratio < 1 indicates absence of E decreases occurrences of O *i.e.* *non-exposure to E is a protective factor.*

To check the presence of any substantial association between dichotomous outcome O and dichotomous exposure E we test

$$H_0 : \text{Odds Ratio} = 1 \text{ against } H_1 : \text{Odds Ratio} \neq 1.$$

A 95% confidence interval contains the true value of the parameter with confidence coefficient 0.95. If the lower limit of the 95% confidence interval of Odds Ratio is greater than 1, we have great confidence that the true Odds Ratio is greater than 1. Then in light of the sample, we reject H_0 at level of significance $\alpha = 0.05$ and conclude that presence of E is a significant risk factor.

In turn, if the upper limit of the 95% confidence interval of Odds Ratio is less than 1, we have great confidence that the true Odds Ratio is less than 1. Then in light of the sample, we reject H_0 at level of significance $\alpha = 0.05$ and conclude that absence of E is a significant protective factor.

Finally, if the 95% confidence interval contains 1, we fail to reject H_0 in light of the sample in hand and conclude that E is not a statistically significant exposure.

1.8 Examples of Test of Significance for Odds Ratio

Example 1.8.1. In Example (1.6.1), 95% CI of Odds Ratio was $(3.127, 7.614)$. Lower limit of the CI $= 3.127 > 1$; so we reject H_0 at 5% level of significance and conclude that *exposure to Alcohol is a significant risk factor in developing Laryngeal cancer.*

Example 1.8.2. In Example (1.6.2), 95% CI of Odds Ratio was $(0.142, 0.468)$. Upper limit of the CI $= 0.468 < 1$; so we reject H_0 at 5% level of significance and conclude that *exposure to Physical Exercise is a significant protective factor against developing Myocardial Infraction.*

Example 1.8.3. In Example (1.6.3), 95% CI of Odds Ratio was $(2.147, 7.057)$. Lower limit of the CI $= 2.147 > 1$; so we reject H_0 at 5% level of significance and conclude that *non-exposure to Physical Exercise is a significant risk factor in developing Myocardial Infraction.*

1.9 An Exercise

? In a case-control study, 200 Esophageal cancer[2] cases and 775 controls were studied. Among the 200 cases it was observed that 96 patients were consuming alcohol and among controls 109 individuals were consuming alcohol. Assess the appropriate risk measure.

[2] cancer that starts in the Esophagus, the muscular tube that connects throat to stomach

From the given, we have the following contingency table.

Exposure Status	Outcome (Esophageal Cancer)		Total
	Cases (Yes)	Controls (No)	
Exposed (Alcohol)	96	109	205
Not exposed (No alcohol)	104	666	770
Total	200	775	975

Odds of alcohol consumption among Esophageal cancer cases $= \dfrac{96}{104} = 0.92$.

Odds of alcohol consumption among controls $= \dfrac{109}{666} = 0.16$.

$\therefore$ OR (Odds Ratio) $= \dfrac{0.92}{0.16} = 5.75$.

Interpretation : The subjects having Esophageal cancer had 5.75 times more exposure to Alcohol than compared to the subjects who did not have Esophageal cancer.

To calculate a 95% confidence interval for odds ratio, we proceed as follows.

$OR = 5.75$. $\therefore ln(OR) = 1.75$.

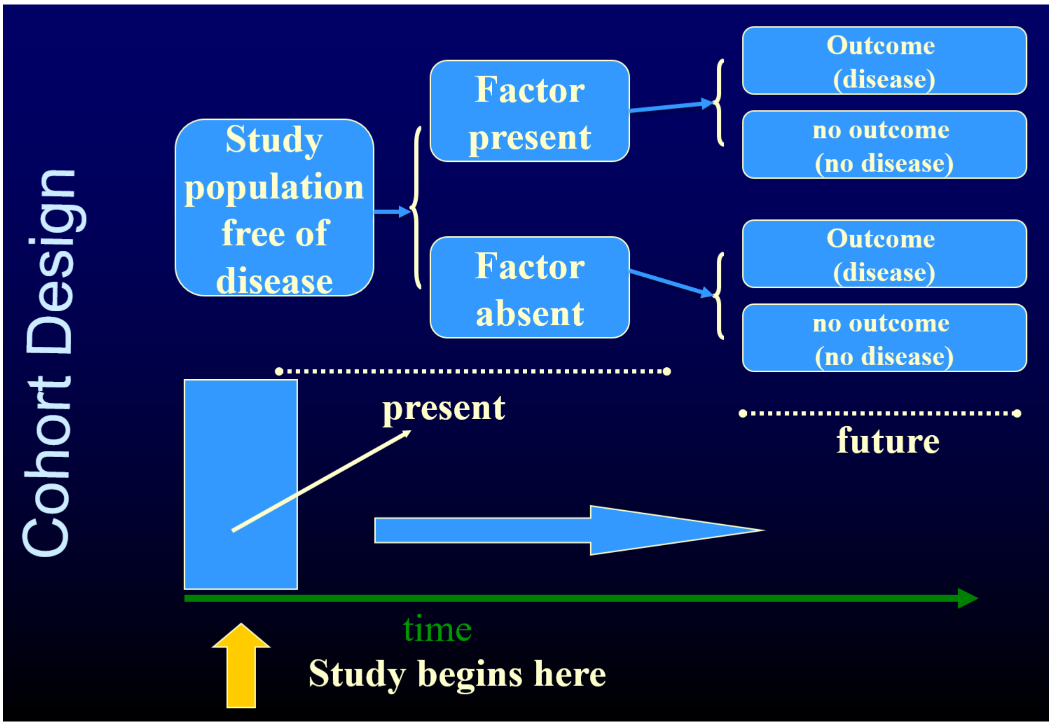
$$SE(ln(OR)) = \sqrt{\frac{1}{96} + \frac{1}{109} + \frac{1}{104} + \frac{1}{666}} = 0.1752$$

$\therefore (L, U) = (1.75 \pm 1.96 \times 0.1752) = (1.41, 2.09)$.

$\therefore$ 95% confidence interval of Odds Ratio is $\left(e^L, e^U\right) = (4.1, 8.09)$. So we reject H_0 at 5% level of significance and conclude that *exposure to Alcohol is a significant risk factor in developing Esophageal cancer.*

2 Effect Measure for Cohort Design

A cohort design is an observational design for comparing individuals with a known risk factor or exposure with others without the risk factor or exposure.
A flowchart of a prospective cohort design is depicted as follows :



2.1 Relative Risk

Relative Risk or Risk Ratio is defined as

$$\frac{\text{Incidence rate in Exposed } (R_1)}{\text{Incidence rate in Not Exposed } (R_0)}.$$

- Relative Risk $= 1 \Rightarrow$ Exposure is not associated with Outcome.
- Relative Risk $> 1 \Rightarrow$ Exposure is a risk factor for the Outcome.
- Relative Risk $< 1 \Rightarrow$ Exposure is a protective factor against the Outcome.

2.2 Statistical Measure of Risk : Relative Risk / Risk Ratio

Consider the following contingency table.

Exposure Status	Outcome		Total
	Cases	Controls	
Exposed	a	b	$a+b$
Not Exposed	c	d	$c+d$

The primary statistical measure of risk in cohort study design is ***relative risk***.
From the above table,

$$\text{incidence rate in Exposed } (R_1) = \frac{a}{a+b}$$

and

$$\text{incidence rate in Not Exposed } (R_0) = \frac{c}{c+d}.$$

Thus,

$$\textit{relative risk} = \frac{a/(a+b)}{c/(c+d)} = \frac{a \cdot (c+d)}{c \cdot (a+b)}.$$

2.3 Examples of Relative Risk

Example 2.3.1. Suppose we conducted a cohort study to answer a research question *Does HIV infection increase risk of developing TB among a population of drug users ?* We have the following contingency table.

Exposure	Outcome		Total
	Tuberculosis	No Tuberculosis	
HIV^+	8	207	215
HIV^-	1	288	289

Incidence rate of TB patients among $HIV^+ = \frac{8}{215} = 0.0372$

Incidence rate of TB patients among $HIV- = \frac{1}{289} = 0.0035$

$\therefore$ Relative Risk $= \frac{0.0372}{0.0035} = 10.62 \approx 11.$

Interpretation : (*Remember that in a prospective cohort study, we will comment about the outcome in the future.*) HIV^+ patients are having ≈ 11 times higher risk of developing Tuberculosis than compared to HIV^- patients (reference group).

Example 2.3.2. Consider the following data from an RCT.

Group	Reduction in Chest Pain		Total
	Yes	No	
Propranolol	162	58	220
Nifedipine	33	150	183
Total	195	208	403

Incidence rate of Reduction of Chest Pain among Propranolol $= \frac{162}{220} = 0.736$

Incidence rate of Reduction of Chest Pain among Nifedipine $= \frac{33}{183} = 0.18$

$\therefore$ Relative Risk $= \dfrac{0.736}{0.18} = 4.089.$


Interpretation : Subjects receiving Propranolol have 4.089 times higher reduction in chest pain as compared to the subjects receiving Nifedipine (reference group).

2.4 Confidence Interval for a Relative Risk

The confidence interval for any parameter is given by

$$(\text{point estimate} \pm \text{critical value} \times \text{standard error}).$$

For relative risk, the confidence interval is calculated on the natural log ($\log_e$) scale and then converted back to the original scale.
Steps involved in calculating confidence interval for a relative risk are as follows :

Step I : Calculate the relative risk from the data.

Step II : Find the natural log *i.e.* $\log_e$ of relative risk.

Step III : The critical value is from the standard normal distribution : 1.96 for 95% confidence interval, confidence coefficient = 0.95.

Step IV : Calculate Standard error for $ln(RR)$.

Exposure Status	Outcome		Total
	Cases	Controls	
Exposed	a	b	$a+b$
Not Exposed	c	d	$c+d$

For a 2×2 contingency table as above, Standard Error for $ln(RR)$ is given by

$$SE(ln(RR)) = \sqrt{\frac{b}{a(a+b)} + \frac{d}{c(c+d)}}.$$

Step V : 95% CI on log scale is $ln(RR) \pm 1.96 \times SE(ln(RR)) = (L, U)$, say.

Step VI : Get the confidence interval limits on the original scale as $\left(e^L, e^U\right)$.

2.5 Examples of CI for a Relative Risk

Example 2.5.1. In Example (2.3.1), $RR = 10.62$. $\therefore ln(RR) = 2.263$.

$$SE(ln(RR)) = \sqrt{\frac{b}{a(a+b)} + \frac{d}{c(c+d)}} = \sqrt{\frac{207}{8(8+207)} + \frac{288}{1(1+288)}} = 1.057.$$

$\therefore (L, U) = (2.263 \pm 1.96 \times 1.057) = (0.191, 4.335).$

$\therefore \left(e^L, e^U\right) = (1.21, 76.32).$

Example 2.5.2. In Example (2.3.2), $RR = 4.089$. $\therefore ln(RR) = 1.408$.

$$SE(ln(RR)) = \sqrt{\frac{b}{a(a+b)} + \frac{d}{c(c+d)}} = \sqrt{\frac{58}{162(162+58)} + \frac{150}{33(33+150)}} = 0.163.$$

$\therefore (L, U) = (1.408 \pm 1.96 \times 0.163) = (1.09, 1.73)$.

$\therefore \left(e^L, e^U\right) = (2.97, 5.64)$.

2.6 An Exercise

In a cohort of 1000 individuals using the bed-nets during the sleep, 3 were infected with Malaria and among the cohort of 800 individuals who were not using the bed-nets, 30 were diagnosed with Malaria. Assess the risk of getting Malaria among the individuals who were not using the bed-nets during sleep when compared to the individuals using the bed-nets during the sleep.

From the given, we have the following contingency table.

Bed-nets	Malaria		Total
	Yes	No	
No	30	770	800
Yes	3	997	1000
Total	195	208	403

Incidence rate of Malaria among individuals not using bed-nets $= \dfrac{30}{800} = 0.038$

Incidence rate of Malaria among individuals using bed-nets $= \dfrac{3}{1000} = 0.003$

$\therefore$ Relative Risk $= \dfrac{0.038}{0.003} = 12.67$.

Interpretation : Individuals not using bed-nets have 12.67 times higher risk of Malaria as compared to the individuals using bed-nets.

To calculate 95% confidence interval for relative risk, we proceed as follows.

$RR = 12.67$. $\therefore ln(RR) = 2.54$.

$$SE(ln(RR)) = \sqrt{\frac{770}{30(30+770)} + \frac{997}{3(3+997)}} = 0.6.$$

$\therefore (L, U) = (2.54 \pm 1.96 \times 0.6) = (1.36, 3.72)$.

$\therefore$ 95% confidence interval for relative risk is $\left(e^L, e^U\right) = (3.9, 41.3)$.

3 Relation Between Odds Ratio and Relative Risk

$$\text{Relative Risk} = \frac{\text{Odds Ratio}}{1-p} + p \times \text{ Odds Ratio}$$

where p is the proportion of the outcome in the non-exposed group.

The relationship implies that the magnitude of Odds Ratio and that of Relative Risk are similar only when p is low.