STAT 226 Lecture 5
Section 2.1 Type of Studies
Section 2.3.5 Retrospective Studies Can Estimate Prospective Odds Ratios

Yibi Huang
Department of Statistics
University of Chicago

2.1 Probability Structure for Contingency Tables

Two-Way Contingency Tables

X categories	$Y=1$	$Y=2$	$\cdots$	$Y=J$	X margin
	Y categories				
$X=1$	n_{11}	n_{12}	$\cdots$	n_{1J}	n_{1+}
$X=2$	n_{21}	n_{22}	$\cdots$	n_{2J}	n_{2+}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$X=I$	n_{I1}	n_{I2}	$\cdots$	n_{IJ}	n_{I+}
Y margin	n_{+1}	n_{+2}	$\cdots$	n_{+J}	$n=n_{++}$

n_{ij} = count of obs. such that $X=i$ and $Y=j$.

- The subscript + means *summation* over the index it replaces. E.g., when $I=J=2$,

$$n_{i+}=n_{i1}+n_{i2}, \qquad n_{+j}=n_{1j}+n_{2j},$$
$$n_{++}=n_{+1}+n_{+2}=n_{11}+n_{12}+n_{21}+n_{22}$$

- Note n_{i+} = # of obs. such that $X=i$, and hence $(n_{1+}, n_{2+}, \ldots, n_{I+})$ are called the *marginal counts of X*.

Population Parameters of Interest

Suppose units in a population of interest (e.g., all traffic crashes) can be classified on X (e.g., seat belt used or not) and Y (result of crash).

	Y categories				
X categories	$Y = 1$	$Y = 2$	$\cdots$	$Y = J$	X margin
$X = 1$	π_{11}	π_{12}	$\cdots$	π_{1J}	π_{1+}
$X = 2$	π_{21}	π_{22}	$\cdots$	π_{2J}	π_{2+}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$X = I$	π_{I1}	π_{I2}	$\cdots$	π_{IJ}	π_{I+}
Y margin	π_{+1}	π_{+2}	$\cdots$	π_{+J}	$\pi_{++} = 1$

The population parameters of interest may include:

- joint distribution: $\pi_{ij} = \mathrm{P}(X = i, Y = j)$
- marginal distribution of X: $\pi_{i+} = \mathrm{P}(X = i)$
- marginal distribution of Y: $\pi_{+j} = \mathrm{P}(Y = j)$
- conditional distribution of X given Y: $\mathrm{P}(X = i \mid Y = j) = \pi_{ij}/\pi_{+j}$
- conditional distribution of Y given X: $\mathrm{P}(Y = j \mid X = i) = \pi_{ij}/\pi_{i+}$

Joint Distributions of Categorical Random Variables (Review)

Suppose units in a population of interest (e.g., all traffic crashes) can be classified on X (e.g., seat belt used or not) and Y (result of crash).

Let $\pi_{ij} = \mathrm{P}(X = i, Y = j)$. The probabilities $\{\pi_{ij}\}$ form the *joint distribution* of X and Y.

Example. (Hypothetical)

	result of crash (Y)		
seat-belt use (X)	$Y = 1$ (fatal)	$Y = 2$ (nonfatal)	$Y = 3$ (no injury)
$X = 1$ (yes)	$\pi_{11} = 0.01$	$\pi_{12} = 0.50$	$\pi_{13} = 0.20$
$X = 2$ (no)	$\pi_{21} = 0.03$	$\pi_{22} = 0.25$	$\pi_{23} = 0.01$

e.g., $\pi_{13} = P(X = 1, Y = 3) = 0.20$ means in 20% of the traffic crashes, seat-belt was used <u>and</u> had no injury.

Marginal Distributions of Random Variables (Review)

Example. (Hypothetical)

result of crash (Y)

Seat-Belt Use (X)	$Y = 1$ fatal	$Y = 2$ nonfatal	$Y = 3$ no injury	X margin
$X = 1$ (Yes)	$\pi_{11} = 0.01$	$\pi_{12} = 0.50$	$\pi_{13} = 0.20$	
$X = 2$ (No)	$\pi_{21} = 0.03$	$\pi_{22} = 0.25$	$\pi_{23} = 0.01$	
				$\pi_{++} = 1$

- In what percentages of traffic crashes was seat belt used?

Marginal Distributions of Random Variables (Review)

Example. (Hypothetical)

Seat-Belt Use (X)	result of crash (Y) $Y = 1$ fatal	$Y = 2$ nonfatal	$Y = 3$ no injury	X margin
$X = 1$ (Yes)	$\pi_{11} = 0.01$	$\pi_{12} = 0.50$	$\pi_{13} = 0.20$	$\pi_{1+} = 0.71$
$X = 2$ (No)	$\pi_{21} = 0.03$	$\pi_{22} = 0.25$	$\pi_{23} = 0.01$	
				$\pi_{++} = 1$

- In what percentages of traffic crashes was seat belt used?

  $P(X = 1) = \pi_{1+} = \pi_{11} + \pi_{12} + \pi_{13} = 0.71$

Marginal Distributions of Random Variables (Review)

Example. (Hypothetical)

Seat-Belt Use (X)	result of crash (Y) $Y=1$ fatal	$Y=2$ nonfatal	$Y=3$ no injury	X margin
$X=1$ (Yes)	$\pi_{11}=0.01$	$\pi_{12}=0.50$	$\pi_{13}=0.20$	$\pi_{1+}=0.71$
$X=2$ (No)	$\pi_{21}=0.03$	$\pi_{22}=0.25$	$\pi_{23}=0.01$	$\pi_{2+}=0.29$
				$\pi_{++}=1$

- In what percentages of traffic crashes was seat belt used?

  $P(X=1)=\pi_{1+}=\pi_{11}+\pi_{12}+\pi_{13}=0.71$

- The row sums $\{\pi_{i+}\}$ form the *marginal distribution of X* since

$$\mathrm{P}(X=i)=\sum_j \mathrm{P}(X=i,Y=j)=\sum_j \pi_{ij}=\pi_{i+}.$$

Marginal Distributions of Random Variables (Review)

Example. (Hypothetical)

result of crash (Y)

Seat-Belt Use (X)	$Y = 1$ fatal	$Y = 2$ nonfatal	$Y = 3$ no injury	X margin
$X = 1$ (Yes)	$\pi_{11} = 0.01$	$\pi_{12} = 0.50$	$\pi_{13} = 0.20$	$\pi_{1+} = 0.71$
$X = 2$ (No)	$\pi_{21} = 0.03$	$\pi_{22} = 0.25$	$\pi_{23} = 0.01$	$\pi_{2+} = 0.29$
Y margin	$\pi_{+1} = 0.04$	$\pi_{+2} = 0.75$	$\pi_{+3} = 0.21$	$\pi_{++} = 1$

- In what percentages of traffic crashes was seat belt used?

  $P(X = 1) = \pi_{1+} = \pi_{11} + \pi_{12} + \pi_{13} = 0.71$

- The row sums $\{\pi_{i+}\}$ form the *marginal distribution of X* since

$$\mathrm{P}(X = i) = \sum_j \mathrm{P}(X = i, Y = j) = \sum_j \pi_{ij} = \pi_{i+}.$$

- The column sums $\{\pi_{+j}\}$ form the *marginal distribution of Y*.

$$\mathrm{P}(Y = j) = \sum_i \mathrm{P}(X = i, Y = j) = \sum_i \pi_{ij} = \pi_{+j}.$$

Conditional Distributions (Review)

A conditional distribution of Y given X refers to the probability distribution of Y when we restrict attention to a fixed level of X.

$$P(Y = j \mid X = i) = \frac{P(X = i, Y = j)}{P(X = i)} = \frac{\pi_{ij}}{\pi_{i+}}$$

Example. (Hypothetical)

seat-belt use (X)	result of crash (Y): $Y = 1$ fatal	$Y = 2$ nonfatal	$Y = 3$ no injury	X margin
$X = 1$ (yes)	$\pi_{11} = 0.01$	$\pi_{12} = 0.50$	$\pi_{13} = 0.20$	$\pi_{1+} = 0.71$
$X = 2$ (no)	$\pi_{21} = 0.03$	$\pi_{22} = 0.25$	$\pi_{23} = 0.01$	$\pi_{2+} = 0.29$
Y margin	$\pi_{+1} = 0.04$	$\pi_{+2} = 0.75$	$\pi_{+3} = 0.21$	$\pi_{++} = 1$

- $P(Y = 1 \mid X = 1) = \frac{0.01}{0.71} = 0.014$ $\Rightarrow$ Among crashes with seat belt used, only 1.4% resulted in fatal injury
- $P(Y = 1 \mid X = 2) = \frac{0.03}{0.29} = 0.103$; $\Rightarrow$ Among crashes with no seat belt use, 10.3% resulted in fatal injury

Conditional distributions of Y given X:

seat-belt use (X)	result of crash (Y) $Y=1$ fatal	$Y=2$ nonfatal	$Y=3$ no injury	total
$X=1$ (yes)	$\frac{0.01}{0.71}=0.014$	$\frac{0.50}{0.71}=0.704$	$\frac{0.20}{0.71}=0.282$	1
$X=2$ (no)	$\frac{0.03}{0.29}=0.103$	$\frac{0.25}{0.29}=0.862$	$\frac{0.01}{0.29}=0.034$	1

Conditional distributions of X given Y: $\mathrm{P}(X=i\,|\,Y=j)=\dfrac{\pi_{ij}}{\pi_{+j}}$

seat-belt use (X)	result of crash (Y) $Y=1$ fatal	$Y=2$ nonfatal	$Y=3$ no injury
$X=1$ (yes)	$\frac{0.01}{0.04}=0.25$	$\frac{0.50}{0.75}=0.667$	$\frac{0.20}{0.21}=0.282$
$X=2$ (no)	$\frac{0.03}{0.04}=0.75$	$\frac{0.25}{0.75}=0.333$	$\frac{0.01}{0.21}=0.034$
total	1	1	1

Conditional distributions of Y given X:

seat-belt use (X)	result of crash (Y) $Y=1$ fatal	$Y=2$ nonfatal	$Y=3$ no injury	total
$X=1$ (yes)	$\frac{0.01}{0.71}=0.014$	$\frac{0.50}{0.71}=0.704$	$\frac{0.20}{0.71}=0.282$	1
$X=2$ (no)	$\frac{0.03}{0.29}=0.103$	$\frac{0.25}{0.29}=0.862$	$\frac{0.01}{0.29}=0.034$	1

Conditional distributions of X given Y: $\mathrm{P}(X=i\,|\,Y=j)=\dfrac{\pi_{ij}}{\pi_{+j}}$

seat-belt use (X)	result of crash (Y) $Y=1$ fatal	$Y=2$ nonfatal	$Y=3$ no injury
$X=1$ (yes)	$\frac{0.01}{0.04}=0.25$	$\frac{0.50}{0.75}=0.667$	$\frac{0.20}{0.21}=0.282$
$X=2$ (no)	$\frac{0.03}{0.04}=0.75$	$\frac{0.25}{0.75}=0.333$	$\frac{0.01}{0.21}=0.034$
total	1	1	1

Interpret $P(X=2\,|\,Y=1)=P(X=\text{no seat-belt}\,|\,Y=\text{fatal})=0.75$.

Among all fatal traffic crashes, 75% of them didn't wear a seat-belt.

Independence (Review)

X and Y are said to be *independent*

- if the conditional distribution of Y given X doesn't change with the level of X,
- or if the conditional distribution of X given Y doesn't change with the level of Y

The two conditions are equivalent.

Independence (Review)

X and Y are said to be *independent*

- if the conditional distribution of Y given X doesn't change with the level of X,
- or if the conditional distribution of X given Y doesn't change with the level of Y

The two conditions are equivalent.

Proof. By the definition of conditional probability $P(Y = j \mid X = i) = \dfrac{P(X = i, Y = j)}{P(X = i)}$, we can see

$$P(Y = j \mid X = i) = P(Y = j) \iff P(X = i, Y = j) = P(X = i)P(Y = j),$$

which implies

$$P(X = i \mid Y = j) = \frac{P(X = i, Y = j)}{P(Y = j)} = \frac{P(X = i)P(Y = j)}{P(Y = j)} = P(X = i).$$

Example. If the conditional distributions of $Y|X$ are like

	result of crash (Y)		
seat-belt use (X)	$Y = 1$ fatal	$Y = 2$ nonfatal	$Y = 3$ no injury
$X = 1$ (yes)	0.04	0.75	0.21
$X = 2$ (no)	0.04	0.75	0.21

or if the conditional distributions of $X|Y$ are like

	result of crash (Y)		
seat-belt use (X)	$Y = 1$ fatal	$Y = 2$ nonfatal	$Y = 3$ no injury
$X = 1$ (yes)	0.71	0.71	0.71
$X = 2$ (no)	0.29	0.29	0.29

then seat-belt use and the severity of traffic crashes are independent.

Summary

	Y categories				
X categories	$Y = 1$	$Y = 2$	$\cdots$	$Y = J$	X margin
$X = 1$	π_{11}	π_{12}	$\cdots$	π_{1J}	π_{1+}
$X = 2$	π_{21}	π_{22}	$\cdots$	π_{2J}	π_{2+}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$X = I$	π_{I1}	π_{I2}	$\cdots$	π_{IJ}	π_{I+}
Y margin	π_{+1}	π_{+2}	$\cdots$	π_{+J}	$\pi_{++} = 1$

- joint distribution: $\pi_{ij} = \mathrm{P}(X = i, Y = j)$
- marginal distribution of X: $\pi_{i+} = \mathrm{P}(X = i)$
- marginal distribution of Y: $\pi_{+j} = \mathrm{P}(Y = j)$
- conditional distribution of X given Y: $\mathrm{P}(X = i|Y = j) = \dfrac{\pi_{ij}}{\pi_{+j}}$
- conditional distribution of Y given X: $\mathrm{P}(Y = j|X = i) = \dfrac{\pi_{ij}}{\pi_{i+}}$

Type of Studies

Types of Studies

Many types of studies result in data in the form of a contingence table.

The analysis and the conclusion can be drawn depend on *how the study is done*.

Example (Prenatal Vitamin and Autism)

Researchers wanted to study whether mothers used prenatal vitamins during the three months before pregnancy (periconceptional period) affects whether the children had autism.

Model:

	Child		
Mother	Autism	No Autism	Total
Vitamin	π_{11}	π_{12}	π_{1+}
No Vitamin	π_{21}	π_{22}	π_{2+}
Total	π_{+1}	π_{+2}	$\pi_{++} = 1$

Data:

	Child		
Mother	Autism	No Autism	Total
Vitamin	n_{11}	n_{12}	n_{1+}
No Vitamin	n_{21}	n_{22}	n_{2+}
Total	n_{+1}	n_{+2}	$n = n_{++}$

One Sample Study

In a one-sample study, randomly sample n mother-child pairs and classify each according to whether the mom took vitamin and whether the child has autism.

In a one-sample study, all joint, marginal, and conditional probabilities can be estimated

- joint:

$$\widehat{\pi}_{ij} = \frac{n_{ij}}{n_{++}},$$

- marginal:

$$\widehat{\pi}_{i+} = \frac{n_{i+}}{n_{++}}, \quad \widehat{\pi}_{+j} = \frac{n_{+j}}{n_{++}},$$

- conditional:

$$\widehat{\mathrm{P}(Y = j|X = i)} = \frac{n_{ij}}{n_{i+}}, \quad \widehat{\mathrm{P}(X = i|Y = j)} = \frac{n_{ij}}{n_{+j}}$$

Drawbacks of One-Sample Study

- If autism is rare, n_{+1} would be small, estimation of $\mathrm{P}(X = i|Y = 1) = P(\text{vitamin}\,|\,\text{autism})$ won't be accurate. To get enough autism cases, the overall sample size must be huge.
- We might not be interested in all of the joint, marginal or conditional prob.

Prospective v.s. Retrospective Study

Suppose we want to study the association of some disease and some risk factor (exposed, unexposed).

In a **prospective** study, the two samples are the exposed and the unexposed.

		Disease	No Disease	Total
Sample 1 →	Exposed	n_{11}	n_{12}	n_{1+}
Sample 2 →	Unexposed	n_{21}	n_{22}	n_{2+}

In a **retrospective** study, the two samples are the diseased and no-diseased.

	Sample 1 ↓ Disease	Sample 2 ↓ No Disease
Exposed	n_{11}	n_{12}
Unexposed	n_{21}	n_{22}
Total	n_{+1}	n_{+2}

Example (Prenatal Vitamin and Autism – Prospective Designs)

Study 1A (Cohort Study): randomly sample 200 moms who had taken prenatal vitamins during the periconceptional period and 200 mothers who didn't, and see if their children have autism at age 5.

Study 1B (Randomized experiment): randomly split 400 women to two groups. Given women in the treatment group prenatal vitamins until they get pregnant and give placebo to those in the control group until they get pregnant, and see if their children have autism at age 5.

Example (Prenatal Vitamin and Autism – Prospective Designs)

Both Study 1A and 1B are prospective.

		Autism	No Autism	Total
Sample 1 →	Vitamin	n_{11}	n_{12}	n_{1+}
Sample 2 →	No Vitamin	n_{21}	n_{22}	n_{2+}

- Both n_{1+}, n_{2+} are fixed (at 200)
- Can estimate the probabilities P(autism | vitamin) and P(autism | no vitamin)
- Drawback: number of diseased cases n_{11} and n_{21} are very small if the disease is rare. unless the sample sizes n_{1+}, n_{2+} are very big (> 1000 or even > 10000)

Example (Prenatal Vitamin and Autism – Retrospective Design)

Study 2 (Retrospective): randomly sample 200 children age 3-5 with autism and 200 children age 3-5 with typical development, and see if their mother took prenatal vitamins during the periconceptional period.

	Sample 1 $\downarrow$ Autism	Sample 2 $\downarrow$ No Autism
Vitamin	n_{11}	n_{12}
No Vitamin	n_{21}	n_{22}
Total	n_{+1}	n_{+2}

- Both n_{+1}, n_{+2} are fixed (at 200)
- Only $\mathrm{P}(\text{vitamin} \mid \text{autism})$ and $\mathrm{P}(\text{vitamin} \mid \text{no autism})$ are estimable.
- Advantage: number of disease cases n_{11} and n_{21} can be large without making the overall sample size too big.
- Drawback: $\mathrm{P}(\text{autism} \mid \text{vitamin or not})$ is not estimable

Properties of Prospective Studies

In a prospective study,

		Disease	No Disease	Total
Sample 1 $\rightarrow$	Exposed	n_{11}	n_{12}	n_{1+}
Sample 2 $\rightarrow$	Unexposed	n_{21}	n_{22}	n_{2+}

we can estimate

$$\pi_1 = P(\text{disease} \mid \text{exposed}), \text{ by } \widehat{\pi}_1 = \frac{n_{11}}{n_{1+}}, \text{ and}$$
$$\pi_2 = P(\text{disease} \mid \text{unexposed}) \text{ by } \widehat{\pi}_2 = \frac{n_{21}}{n_{2+}}.$$

Hence, the difference of proportions $\pi_1 - \pi_2$ and relative risk π_1/π_2 are both estimable.

Properties of Retrospective Studies

In a retrospective study,

	Sample 1 $\downarrow$ Disease	Sample 2 $\downarrow$ No Disease
Exposed	n_{11}	n_{12}
Unexposed	n_{21}	n_{22}
Total	n_{+1}	n_{+2}

only

$$\tau_1 = P(\text{exposed} \mid \text{disease})$$

$$\tau_2 = P(\text{exposed} \mid \text{no disease})$$

are estimable, but they are not of interest.

The parameter of interest, π_1 and π_2, are not estimable, and neither are $\pi_1 - \pi_2$ or π_1/π_2.

Most Important Property of the Odds Ratio

		Y (e.g., disease)		
		1 (Disease)	2 (No Disease)	X margin
X	1 Exposed	π_{11}	π_{12}	π_{1+}
	2 Unexposed	π_{21}	π_{22}	π_{2+}
	Y margin	π_{+1}	π_{+2}	π_{++}

Prospective study:

$$\pi_1 = P(\text{Disease} \mid \text{Exposed}) = \frac{\pi_{11}}{\pi_{1+}}$$

$$\pi_2 = P(\text{Disease} \mid \text{Unexposed}) = \frac{\pi_{21}}{\pi_{2+}}$$

Retrospective study:

$$\tau_1 = P(\text{Exposed} \mid \text{Disease}) = \frac{\pi_{11}}{\pi_{+1}}$$

$$\tau_2 = P(\text{Exposed} \mid \text{No Disease}) = \frac{\pi_{12}}{\pi_{+2}}$$

Most Important Property of the Odds Ratio

		Y (e.g., disease)		
		1 (Disease)	2 (No Disease)	X margin
X	1 Exposed	π_{11}	π_{12}	π_{1+}
	2 Unexposed	π_{21}	π_{22}	π_{2+}
	Y margin	π_{+1}	π_{+2}	π_{++}

Prospective study:

$$\pi_1 = P(\text{Disease} \mid \text{Exposed}) = \frac{\pi_{11}}{\pi_{1+}}$$

$$\pi_2 = P(\text{Disease} \mid \text{Unexposed}) = \frac{\pi_{21}}{\pi_{2+}}$$

Retrospective study:

$$\tau_1 = P(\text{Exposed} \mid \text{Disease}) = \frac{\pi_{11}}{\pi_{+1}}$$

$$\tau_2 = P(\text{Exposed} \mid \text{No Disease}) = \frac{\pi_{12}}{\pi_{+2}}$$

$$\boxed{\theta = \frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)} = \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}} = \frac{\tau_1/(1-\tau_1)}{\tau_2/(1-\tau_2)}}$$

Odds ratio treats the rows and columns symmetrically, i.e., it does not distinguish X and Y.

Odds Ratio & Retrospective Studies

In a retrospective study, even though the parameter of interest,

$$\pi_1 = P(\text{disease} \mid \text{exposed}), \text{ and}$$
$$\pi_2 = P(\text{disease} \mid \text{unexposed})$$

are not estimable, and neither are $\pi_1 - \pi_2$ or the relative risk π_1/π_2.

However, the odds ratio $\dfrac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)}$ are estimable since it is also equal to $\dfrac{\tau_1/(1-\tau_1)}{\tau_2/(1-\tau_2)}$ and τ_1 and τ_2 are estimable.

A Case Control Study Example (p.32 in ICDA 2ed)

- **cases**: 262 young and middle-aged women (age < 69) admitted to 30 coronary care units in northern Italy with acute heart attack during a 5-year period
- **controls**: each of the 262 cases above was matched with two control patients admitted to the same hospitals with other acute disorders[1].

	Heart Attack (Y)	
Ever Smoker (X)	Cases	Controls
Yes	172	173
No	90	346
Total	262	519

- This is a *retrospective* (“look into the past”) study

[1] *Source*: A. Gramenzi et al., *J. Epidemiol. Community Health*, 43:214-217, 1989.

In the case-control study, the marginal totals for "heart attack or not" are fixed, we can estimate

$$\tau_1 = \text{P(smoker | heart attack) and}$$
$$\tau_2 = \text{P(smoker | no heart attack)}$$

		heart attack (Y)	
	$P(X \mid Y)$	Yes	No
smoker (X)	Yes	τ_1	τ_2
	No	$1 - \tau_1$	$1 - \tau_2$

but

$$\pi_1 = \text{P(heart attack | smoker) and}$$
$$\pi_2 = \text{P(heart attack | nonsmoker).}$$

are not estimable from such a study.

- (π_1, π_2) cannot be computed from (τ_1, τ_2)
- If we just want to know if heart attack is independent of smoking, testing $\pi_1 = \pi_2$ is equivalent to testing $\tau_1 = \tau_2$.

Case-Control Study About Smoking & Heart Attack Revisit

	Heart Attack (Y)	
Smoker (X)	Cases	Controls
Yes	172	173
No	90	346
Total	262	519

Recall

$$\pi_1 = \mathrm{P}(\text{heart attack} \mid \text{smoker}),$$
$$\pi_2 = \mathrm{P}(\text{heart attack} \mid \text{nonsmoker}),$$
$$\tau_1 = \mathrm{P}(\text{smoker} \mid \text{heart attack}),$$
$$\tau_2 = \mathrm{P}(\text{smoker} \mid \text{no heart attack}),$$

Want π_1, π_2, but only got $\widehat{\tau}_1 = \frac{172}{262}, \widehat{\tau}_2 = \frac{173}{519}$. Neither $\pi_1 - \pi_2$ nor π_1/π_2 is estimable.

However, the odds ratio $\theta = \frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)}$ is estimable from $\widehat{\tau}_1$ and $\widehat{\tau}_2$ since

$$\widehat{\theta} = \frac{\widehat{\pi}_1/(1-\widehat{\pi}_1)}{\widehat{\pi}_2/(1-\widehat{\pi}_2)} = \frac{\widehat{\tau}_1/(1-\widehat{\tau}_1)}{\widehat{\tau}_2/(1-\widehat{\tau}_2)} = \frac{172 \times 346}{173 \times 90} \approx 3.844$$

Conclusion: Odds of heart attack for smokers estimated to be about 3.8 times the odds for non-smokers.

If $\pi_1, \pi_2 \approx 0$ (heart attack was rare), then $\theta \approx$ relative risk, can conclude that risk of heart attack is ≈ 3.8 times as high for smokers as for non-smokers.

	Heart Attack (Y)	
Ever Smoker (X)	Cases	Controls
Yes	172	173
No	90	346
Total	262	519

$$\widehat{\theta} = \frac{172 \times 346}{173 \times 90} = 3.844$$

$$\log \widehat{\theta} = \log(3.84) = 1.3466$$

$$\text{SE}(\log \widehat{\theta}) = \sqrt{\frac{1}{172} + \frac{1}{90} + \frac{1}{173} + \frac{1}{346}} \approx 0.160$$

95% CI for $\log \theta : 1.3466 \pm 1.96(0.160) \approx (1.033, 1.660)$

95% CI for $\theta : (e^{1.033}, e^{1.660}) \approx (2.81, 5.26)$

Interpretation: With 95% confidence, the odds of having a heart attack for smokers is 2.81 to 5.26 times as large for smokers as for nonsmokers

Different Ways to Interpret an Odds Ratio

	Heart Attack (Y)	
Smoker (X)	Cases	Controls
Yes	172	173
No	90	346
Total	262	519

Recall

$$\pi_1 = \mathrm{P}(\text{heart attack} \mid \text{smoker}),$$
$$\pi_2 = \mathrm{P}(\text{heart attack} \mid \text{nonsmoker}),$$
$$\tau_1 = \mathrm{P}(\text{smoker} \mid \text{heart attack}),$$
$$\tau_2 = \mathrm{P}(\text{smoker} \mid \text{no heart attack}),$$

- $\theta = \dfrac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)} = 3.84$: The odds of having a heart attack for ever smokers were 3.84 times as large as for those who have never smoked
- $\theta = \dfrac{\tau_1/(1-\tau_1)}{\tau_2/(1-\tau_2)} = 3.84$: The odds of being an ever smoker for those who have had a heart attack were 3.84 times as large as for those who never have a heart attack