Models In Epidemiology And Biostatistics
Gordon Hilton Fick
Session 2 : Stratified Analysis

We wish to study disease-exposure relationships

We are concerned that such a study may require the consideration of other characteristics

Let us focus on the odds ratio to measure this disease-exposure relationship

Further, let us suppose that

age group and gender may be involved.

Describing the variables involved

E – exposure D – disease

absence : label has a “bar” $\bar{E}$ or $\bar{D}$

presence : label has no “bar” E or D

age: Y - young; O - old

gender: F - female; M - male

2 Probabilities : 2 Odds

	Exposed	Unexposed
Cases	p_{E}	$p_{\bar{\mathrm{E}}}$
Controls	$1-p_{\mathrm{E}}$	$1-p_{\bar{\mathrm{E}}}$

$p_{\mathrm{E}} = \mathrm{P}(\mathrm{D} \mid \mathrm{E}) =$ probability of disease given exposure

$p_{\bar{\mathrm{E}}} = \mathrm{P}(\mathrm{D} \mid \bar{\mathrm{E}}) =$ probability of disease given no exposure

$\dfrac{p_{\mathrm{E}}}{1-p_{\mathrm{E}}} =$ the odds of disease given exposure

$\dfrac{p_{\bar{\mathrm{E}}}}{1-p_{\bar{\mathrm{E}}}} =$ the odds of disease given no exposure

Odds Ratio

$$\text{OR} = \frac{\frac{p_{\text{E}}}{1-p_{\text{E}}}}{\frac{p_{\bar{\text{E}}}}{1-p_{\bar{\text{E}}}}}$$

The odds ratio is the odds of disease given exposure divided by the odds of disease given the absence of exposure

The Data

	Exposed	Unexposed
Cases	a	b
Controls	c	d

$\frac{a}{a+c}$ estimates p_{E} $\frac{a}{c}$ estimates $\frac{p_{\mathrm{E}}}{1-p_{\mathrm{E}}}$

$\frac{b}{b+d}$ estimates $p_{\bar{\mathrm{E}}}$ $\frac{b}{d}$ estimates $\frac{p_{\bar{\mathrm{E}}}}{1-p_{\bar{\mathrm{E}}}}$

Estimates and Population Characteristics ?

$\frac{ad}{bc}$ estimates the odds ratio OR or we write:

$$\hat{\mathrm{OR}} = \frac{ad}{bc}$$

For the first part of this session, let us suppose that we are 'viewing' the population odds ratios. Lets leave the sampling error issues aside for now.

Many Odds Ratios

We have the 'crude' odds ratio.

We can consider 2 odds ratios: one for each age group: Y, O

We can consider 2 odds ratios: one for each gender: F, M

We can consider 4 odds ratios: one for each of the 4 groups determined by age and gender: YF, OF, YM, OM

Many Two by Two Tables

So we could [and should] consider 9 (!) 2x2 tables:

The 'crude' 2x2 table

The two 2x2 tables stratified on age

The two 2x2 tables stratified on gender

The four 2x2 tables stratified on both age and gender.

How do we make sense of all of these tables?

We need a disciplined strategy

We may end up with a very simple set of interpretations.

We could end up with a rather elaborate set of interpretations. The elaborate nature of the findings may be the most important part of the research.

Avoiding the simple (with justification(s)) may be critical to the science.

Notation

We will be looking at a number of different scenarios. Here are some abbreviations:

OR – the (population) odds ratio

C - Crude (no stratification)

A – Age alone (two strata)

G – Gender alone (two strata)

AG – Age/Gender (four strata)

X – means ' a number not under consideration'

Simplest situation: Neither age nor gender is relevant

All 9 odds ratios are the same.

The odds of disease for those exposed is 4 times the odds of disease for those unexposed. Neither age nor gender appear to be involved. We could 'report' the crude.

		OR
C		4
A:	Y	4
A:	O	4
G:	F	4
G:	M	4
AG:	YF	4
AG:	OF	4
AG:	YM	4
AG:	OM	4

Gender modifies; Age is irrelevant

The Crude OR and the age specific ORs are not relevant.

The odds ratio is 6 for the males while the odds ratio is 0.25 for the females. We could 'report' the gender specific ORs.

		OR
C		X
A:	Y	X
A:	O	X
G:	F	0.25
G:	M	6
AG:	YF	0.25
AG:	OF	0.25
AG:	YM	6
AG:	OM	6

Age modifies: Gender is irrelevant

The Crude OR and the gender specific ORs are not relevant.

The odds ratio is 0.5 for the young while the odds ratio is 4 for the old. We could 'report' the age specific ORs.

		OR
C		X
A:	Y	0.5
A:	O	4
G:	F	X
G:	M	X
AG:	YF	0.5
AG:	OF	4
AG:	YM	0.5
AG:	OM	4

Gender confounds: Age is irrelevant

We needed to stratify on gender but we did not need to stratify on age. The OR common to both genders could be reported.

	OR
C	1
A: Y	1
A: O	1
G: F	5
G: M	5
AG: YF	5
AG: OF	5
AG: YM	5
AG: OM	5

Age confounds: Gender is irrelevant

		OR
C		0.5
A:	Y	3
A:	O	3
G:	F	0.5
G:	M	0.5
AG:	YF	3
AG:	OF	3
AG:	YM	3
AG:	OM	3

We needed to stratify on age but we did not need to stratify on gender. The OR common to both age groups could be reported.

Age confounds and gender confounds the age confounding

		OR
C		1
A:	Y	0.25
A:	O	0.25
G:	F	1
G:	M	1
AG:	YF	5
AG:	OF	5
AG:	YM	5
AG:	OM	5

We need to stratify on both age and gender in order to see the correct form of age confounding.

The OR common to all 4 strata could be reported.

Age and Gender confound : Gender confounds the age confounding and vice versa

		OR
C		1
A:	Y	1
A:	O	1
G:	F	1
G:	M	1
AG:	YF	4
AG:	OF	4
AG:	YM	4
AG:	OM	4

We need to stratify on both age and gender in order to see the confounding.

The OR common to all 4 strata could be reported.

Gender modifies; Age confounds the gender modification

		OR
C		X
A:	Y	X
A:	O	X
G:	F	1
G:	M	1
AG:	YF	0.25
AG:	OF	0.25
AG:	YM	6
AG:	OM	6

We see gender modification only by stratifying on age as well.

We can report that the OR common to both female groups is 0.25 while the OR common to both male groups is 6.

Age modifies: Gender confounds the age modification

		OR
C		X
A:	Y	0.75
A:	O	0.75
G:	F	X
G:	M	X
AG:	YF	0.5
AG:	OF	4
AG:	YM	0.5
AG:	OM	4

We see age modification only by stratifying on gender as well.

We can report that the OR common to both young groups is 0.5 while the OR common to both old groups is 4.

Age and Gender both modify: age modification depends on gender and vice versa

		OR
C		X
A:	Y	X
A:	O	X
G:	F	X
G:	M	X
AG:	YF	0.5
AG:	OF	4
AG:	YM	6
AG:	OM	0.25

We need to stratify on both age and gender here.

We need to report all 4 age-gender specific ORs.

Age and Gender both modify: age modification does not depend on gender

		OR
C		X
A:	Y	X
A:	O	X
G:	F	X
G:	M	X
AG:	YF	0.5
AG:	OF	4
AG:	YM	1
AG:	OM	8

We need to stratify on both age and gender.

We would report all 4 ORs but the situation is slightly simpler than the last scenario. More on this in coming classes.