Applications In Epidemiology –Paired binary data

Example 4: Matched case control study

213 subjects with a history of acute myocardial infarction (AMI) were ***matched*** by age and sex with one of their siblings who did not have a history of AMI. The prevalence of a particular polymorphism was compared between the siblings

Q: Is there an association between the polymorphism and AMI?

Q: If there is an association then what is the magnitude of the effect?

1

Applications In Epidemiology –Paired binary data

One way to display these data is the following:

	Carrier	Noncarrier	Total
AMI	96	117	213
No AMI	87	126	213
Total	183	243	426

Q: Can't we simply use Pearson's X^2 Test to assess whether this is evidence for an increase in knowledge?

A: NO!!! Pearson's X^2 test assumes that the rows are **independent** samples. In this design the 213 with AMI are genetically related to the 213 w/o AMI. This is an example of **paired binary data**.

2

Applications In Epidemiology –Paired binary data

For paired binary data we display the results as follows:

	AMI		
	carrier	noncarrier	Total
carrier	73	14	87
No AMI noncarrier	23	103	126
Total	96	117	213

This analysis explicitly recognizes the heterogeneity of subjects. The **concordant pairs** (73 and 103) provide no information about the association between AMI and the polymorphism. The information regarding the association is in the **discordant pairs,** 14 and 23.

3

Applications In Epidemiology –Paired binary data

For paired binary data we display the results as follows:

	AMI	
	1	0
No AMI 1	n_{11}	n_{10}
0	n_{01}	n_{00}

p_1 = P(carrier | AMI)

p_0 = P(carrier | No AMI)

$H_0 : p_1 = p_0$

$H_A : p_1 \neq p_0$

The information for testing this hypothesis is contained in the discordant pairs (0,1) and (1,0)

4

Applications In Epidemiology –Paired binary data

Under the null hypothesis we expect equal numbers of 01's and 10's. The McNemar's chi-squared statistic is

$$X^2 =, \left(\frac{n_{10} - M\frac{1}{2}}{\sqrt{M\frac{1}{2}\left(1 - \frac{1}{2}\right)}} \right)^2$$

where $M = n_{01} + n_{10}$. $X^2 \sim \chi^2(1)$ and forms the basis for **McNemar's Test for Paired Binary Responses**.

The odds ratio comparing the odds of carrier in those with AMI to odds of carrier in those w/o AMI is estimated by:

$$\hat{OR} = \frac{n_{01}}{n_{10}}$$

Confidence intervals can be obtained as described in Breslow and Day (1981), section 5.2, or in Armitage and Berry (1987), chapter 16.

Exercise 1: Compute χ^2 and the estimated OR for the AMI paired binary data dataset

	AMI		
	carrier	noncarrier	Total
carrier	73	14	87
No AMI noncarrier	23	103	126
Total	96	117	213

Stratified Tables – Effect Modification

- Often, a third measure influences the relationship between the two primary measures (i.e. disease and exposure).

Example: Effect of seat belt use on accident fatality

	Seat Belt	
Driver	Worn	Not worn
dead	10	20
alive	40	30
Total	50	50
Fatality Rate	10/50 (20%)	20/50 (40%)

Stratified Tables – Effect Modification

But, suppose...

	Impact Speed			
	≤ 40 mph		> 40 mph	
Driver	seat belt		seat belt	
	worn	not	worn	not
dead	3	2	7	18
alive	27	18	13	12
Total	30	20	20	30
Fatality Rate	10%	10%	35%	60%

How does this affect your inference?

➢This is an example of "effect modification" or "interaction".

Effect Modification (aka Interaction)

- Effect modification depends on the effect measure used!

Table x. Rate of fractures over 5 years by age and calcium level in drinking water			
	Age 20 - 35	Age 55 - 80	Overall (pooled)
High calcium	1.1%	11.0%	7.8%
Low calcium	3.3%	13.2%	10.0%
RR	.33	.83	.78
RD	-2.2%	-2.2%	-2.2%

9

Stratified Tables - Confounding

Example: Suppose we are interested in the relationship between lung-cancer incidence and heavy drinking (defined as > 2 drinks per day). We conduct a prospective study where drinking status is determined at baseline and the cohort is followed for 10 years to determine cancer endpoints. We also measure smoking status at baseline.

10

Stratified Contingency Tables - Example

1) Pooled data, not controlling for smoking

	Heavy Drinker		
	Yes	No	
Case	33	27	60
Control	1667	2273	3940
	1700	2300	4000

OR = 1.67

11

Stratified Contingency Tables - Example

2) Stratified by smoking at baseline

Smokers

	Heavy Drinking		
	Yes	No	
Case	24	6	30
Control	776	194	970
	800	200	1000

OR = 1

Nonsmokers

	Heavy Drinking		
	Yes	No	
Case	9	21	30
Control	891	2079	2970
	900	2100	3000

OR = 1

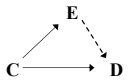
- A higher proportion of heavy drinkers are smokers (800/1700 vs 200/2300)
- A higher proportion of cases are smokers (30/60 vs 970/3940)
- The comparison of heavy drinkers to not-heavy drinkers is really a comparison of smokers to nonsmokers

12

Confounding

A confounder is associated with both the disease and exposure and is not in the causal path between disease and exposure

- The implicit assumption is that we want to know if E "causes" D
- A simple, common example from genetics is the linked gene: we discover a gene which appears to be associated with disease … does it cause the disease or is it merely linked to the true causal gene?
- Pictorially …



An apparent association between E and D is completely explained by C. C is a confounder.

13

Exercise 2: In each case, decide whether this is an example of confounding or effect modification

a) Two hospitals are compared with respect to the rate death following a particular type of surgery. Here are the data … is risk group a confounder or effect modifier?

		Death rate	
High risk			
Hospital	A	57/1500	(3.8%)
	B	8/200	(4%)
Low risk			
Hospital	A	6/600	(1%)
	B	8/600	(1.3%)

b) A randomized clinical trial is conducted to determine if a new drug can increase levels of HDL cholesterol among men and women. Using the mean difference as a measure of effect, is sex a confounder or effect modifier?

	Mean HDL		
	Women	Men	All
New Drug	38.9	45.2	40.2
Placebo	39.2	39.1	39.2

14

Exercise 2: In each case, decide whether this is an example of confounding or effect modification

c) Researchers at the International Agency for Research on Cancer in France found that women infected with both HPV and HSV-2 were nearly three times more likely to get cervical cancer compared to women with only HPV infection.

Does HSV-2 confound or modify the effect of HPV on cervical cancer?

d) If the mother took antidepressant medication during the first trimester, without accounting for other possible influences, children had roughly twice the risk of having autism. The researchers then compared siblings in families where the mother used antidepressants in one pregnancy but not the other. This helped account for all of the factors that make siblings similar — their shared genetics and environment. In the sibling matchup, the children had essentially the same risk for autism, ADHD and poor fetal growth whether they were exposed to antidepressants in the womb or not.

Do genetic factors confound or modify the effect of antidepressants on autism?

15

Adjusting the OR via Stratification

Basic idea

- Compute separate OR for each stratum
- Assess homogeneity of OR's across strata (Is there EM?)
- Pool OR's: used weighted average (Adjust for confounding)
- Global test of pooled OR = 1 (Is there association, after adjustment)
- Different methods of pooling, testing have been proposed. We will focus on Mantel-Haenszel methods
- Same idea for RR and RD

16

Stratified Contingency Tables - Example

EXAMPLE: (Rosner sec 13.5)

A 1985 study identified a group of 518 cancer cases and 518 controls by mail questionnaire. The main purpose of the study was to look at the effect of passive smoking on cancer risk. In the study passive smoking was defined as exposure to the cigarette smoke of a spouse who smoked at least one cigarette/day for at least 6 months. One potential confounding variable was smoking by the test subjects themselves since personal smoking is related to both cancer risk and having a spouse that smokes. Therefore, it was important to control for personal smoking before looking at the relationship between passive smoking and cancer risk.

Stratified Contingency Tables - Example

1) Pooled data, not controlling for personal smoking

	Passive smoking		
	Yes	No	
Case	281	228	509
Control	210	279	489
	491	507	998

```
. cci 281 228 210 279
                                                        Proportion
                |   Exposed   Unexposed  |      Total      Exposed
----------------+------------------------+------------------------
          Cases |         281         228 |         509      0.5521
       Controls |         210         279 |         489      0.4294
----------------+------------------------+------------------------
          Total |         491         507 |         998      0.4920
                |                        |
                |    Point estimate      | [95% Conf. Interval]
                |------------------------+------------------------
     Odds ratio |        1.637406        |    1.265013    2.119599
Attr. frac. ex. |        .3892779        |    .2094943    .5282126
Attr. frac. pop |        .2149059        |
                +-------------------------------------------------
                          chi2(1) =    15.00  Pr>chi2 = 0.0001
```


Stratified Contingency Tables - Example

2) Stratified by personal smoking

Nonsmokers

	Passive smoking		
	Yes	No	
Case	120	111	231
Control	80	155	235
	200	266	466

```
. cci 120 111 80 155
                                                        Proportion
                |   Exposed   Unexposed  |      Total      Exposed
----------------+------------------------+------------------------
          Cases |         120         111 |         231      0.5195
       Controls |          80         155 |         235      0.3404
----------------+------------------------+------------------------
          Total |         200         266 |         466      0.4292
                |                        |
                |    Point estimate      | [95% Conf. Interval]
                |------------------------+------------------------
     Odds ratio |        2.094595        |    1.41754    3.097165
Attr. frac. ex. |        .5225806        |    .2945527    .6771241
Attr. frac. pop |        .2714705        |
                +-------------------------------------------------
                          chi2(1) =    15.24  Pr>chi2 = 0.0001
```

Smokers

	Passive smoking		
	Yes	No	
Case	161	117	278
Control	130	124	254
	291	241	532

```
. cci 161 117 130 124
                                                        Proportion
                |   Exposed   Unexposed  |      Total      Exposed
----------------+------------------------+------------------------
          Cases |         161         117 |         278      0.5791
       Controls |         130         124 |         254      0.5118
----------------+------------------------+------------------------
          Total |         291         241 |         532      0.5470
                |                        |
                |    Point estimate      | [95% Conf. Interval]
                |------------------------+------------------------
     Odds ratio |        1.312558        |    .9184614    1.875813
Attr. frac. ex. |        .2381286        |   -.0887774    .4668978
Attr. frac. pop |        .137909         |
                +-------------------------------------------------
                          chi2(1) =     2.43  Pr>chi2 = 0.1192
```


Stratified Contingency Tables

Q: How can we combine the information from both tables to obtain an overall test of significance that takes account of the stratification?

A: Mantel-Haenszel Methods – assesses association between disease and exposure after controlling for one or more confounding variables.

Notation:

	E	$\overline{E}$	
D	a_i	b_i	$(a_i + b_i)$
$\overline{D}$	c_i	d_i	$(c_i + d_i)$
	$(a_i + c_i)$	$(b_i + d_i)$	N_i

where i = 1,2,…,K is the number of strata.

Mantel-Haenszel Methods

(1) **Test of effect modification** (heterogeneity, interaction)

Ho: $OR_1 = OR_2 = \ldots = OR_K$

Ha: not all stratum-specific OR's are equal

(2) **Estimate the common odds ratio**

The Mantel-Haenszel estimate of the odds ratio assumes there is a **common** odds ratio:

$$OR_{pool} = OR_1 = OR_2 = \ldots = OR_K$$

To estimate the common odds ratio we take a weighted average of the stratum-specific odds ratios:

MH estimate: $\hat{O}R_{pool} = \sum_{i=1}^{K} w_i \cdot \hat{O}R_i$

(3) **Test of common odds ratio**

H_o: common odds ratio is 1.0

H_a: common odds ratio ≠ 1.0

Mantel-Haenszel Methods - Example

	case	passive	number	smoke
1.	1	1	120	0
2.	1	0	111	0
3.	0	1	80	0
4.	0	0	155	0
5.	1	1	161	1
6.	1	0	117	1
7.	0	1	130	1
8.	0	0	124	1

```
. cc case passive [freq=number], by(smoke) bd

Personal Smoking |       OR      [95% Conf. Interval]    M-H Weight
-----------------+-------------------------------------------------
               0 |   2.094595      1.41754   3.097165      19.05579 (exact)
               1 |   1.312558     .9184614   1.875813      28.59023 (exact)
-----------------+-------------------------------------------------
           Crude |   1.637406     1.265013   2.119599               (exact)
   M-H combined |   1.625329     1.263955   2.090024
-----------------+-------------------------------------------------
Test of homogeneity (M-H)      chi2(1) =     3.27  Pr>chi2 = 0.0706
Test of homogeneity (B-D)      chi2(1) =     3.27  Pr>chi2 = 0.0704

                 Test that combined OR = 1:
                         Mantel-Haenszel chi2(1) =     14.42
                                  Pr>chi2 =    0.0001
```


Exercise 3

Based on the abundance of specific bacterial genera, the human gut microbiota can be divided into two relatively stable groups (enterotypes) that might play a role in personalized nutrition. We studied these simplified enterotypes as prognostic markers for successful body fat loss on two different diets. A total of 62 participants with increased waist circumference were randomly assigned to receive a New Nordic Diet (NND) high in fiber/wholegrain or an Average Danish Diet (ADD) for 26 weeks. At enrollment, participants were grouped into two discrete enterotypes by their relative abundance of Prevotella spp. divided by Bacteroides spp. (P/B ratio) obtained by quantitative PCR analysis. Among individuals with high P/B the NND resulted in a 3.15 kg larger body fat loss compared to ADD whereas virtually no difference (0.88 kg) was observed among individuals with low P/B. Consequently, a 2.27 kg difference in responsiveness to the diets were found between the high and low P/B groups. In summary, subjects with high P/B-ratio appeared more susceptible to lose body fat on diets high in fiber and wholegrain than subjects with a low P/B-ratio.

a) Which of the following best describes the design of this study?

Cross-sectional survey

Case-control study

Prospective cohort

b) Identify the role of diet, weight loss, and P/B ratio using one of the following terms – Outcome, Exposure, Effect modifier, Confounder

Review

- R x C contingency table
  - o Test for homogeneity (Pearson chi-squared)
- Single 2 x 2 table
  - o Different sampling schemes
    - 1.Cohort (row totals fixed)
    - 2.Case-control (column totals fixed)
    - 3.Cross-sectional (grand total fixed)
  - o Different measures of association
    - RD (Designs 1 & 3)
    - RR (Designs 1 & 3)
    - OR (Designs 1, 2 & 3)
  - o Test of association
    - Pearson chi-squared
    - McNemar's

Review

- Series of 2 x 2 tables
  - o Effect Modification
  - o Confounding
  - o Stratified analysis
    - Breslow-Day "Score" Test for Homogeneity (Interaction, Effect Modification)
    - Mantel-Haenszel (combined) OR estimate
    - Mantel-Haenszel test for association (H_o: OR = 1)