1) EM
2) confoundin

Stratified Tables

- Often, a third measure influences the relationship between the two primary measures (i.e. disease and exposure).
- How do we "remove or control for the effect" of the third measure?
- Issues of causality

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%)

RD −20%pt

Stratified Tables

But, suppose...

	Impact Speed			
	$\leq$ 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% 5%	10% 10%	35% 20%	60% 40%

RD 0% pt -25% pt

How does this affect your inference?

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

RR .5 .5

RD 5% 20%

Stratified tables - Confounding (Simpson's Paradox)

Differences in surgical success between hospitals?

hi

		Death rate	
Hospital	A	63/2100	(3%)
	B	16/800	(2%)

mostly lo

BUT ...

RD

		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%)

-.2

-.3

Explanation: Higher risk individuals are more likely to die AND are more likely to go to hospital A (perhaps it specializes in this type of surgery)

Confounding

"A confounding variable is a variable that is associated with both the disease and the exposure variable." *Rosner (1995)*

"Confounding is the distortion of a disease/exposure association brought about by the association of other factors with both disease and exposure, the latter associations with disease being causal." *Breslow & Day (1980)*

"If any factor either increasing or decreasing the risk of a disease besides the characteristic or exposure under study is unequally distributed in the groups that are being compared with regard to the disease, this itself will give rise to differences in disease frequency in the compared groups. Such distortion, termed confounding, leads to an invalid comparison." *Lilienfeld & Stolley (1994)*

true RD = P(death|A) − P(death|B)

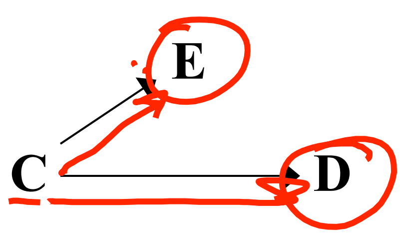
Confounding

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

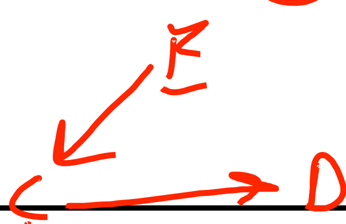
- The implicit assumption is that we want to know if E "causes" D
- A simple, common example from genetics is the linked gene: we discover a gene which appears to be associated with disease … does it cause the disease or is it merely linked to the true causal gene?

Pictorially …

$E \to D$



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



rsky → risky

early death

Adjusting the OR via Stratification

Basic idea

- Compute separate OR for each stratum
- Assess homogeneity of OR's across strata
- Pool OR's: used weighted average
- Global test of pooled OR = 1 H_0: OR = 1
- Different methods of pooling, testing have been proposed. We will focus on Mantel-Haenszel methods
- Same idea for RR and RD

Stratified Contingency Tables - Example

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.

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

```
. cci 33 27 1667 2273
                                                      Proportion
                 |   Exposed   Unexposed  |      Total     Exposed
-----------------+------------------------+------------------------
           Cases |        33          27  |         60      0.5500
        Controls |      1667        2273  |       3940      0.4231
-----------------+------------------------+------------------------
           Total |      1700        2300  |       4000      0.4250
                 |                        |
                 |      Point estimate    |    [95% Conf. Interval]
                 |------------------------+------------------------
      Odds ratio |         1.666533       |   .9677794    2.892948 (exact)
 Attr. frac. ex. |         .399952        |  -.0332934    .6543319 (exact)
 Attr. frac. pop |         .2199736       |
                 +-------------------------------------------------
                               chi2(1) =     3.89  Pr>chi2 = 0.0484
```

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

```
. cci 24 6 776 194
                                                         Proportion
                 |   Exposed   Unexposed  |      Total     Exposed
-----------------+------------------------+----------------------
           Cases |        24           6  |         30      0.8000
        Controls |       776         194  |        970      0.8000
-----------------+------------------------+----------------------
           Total |       800         200  |       1000      0.8000
                 |                        |
                 |      Point estimate    |    [95% Conf. Interval]
                 |------------------------+----------------------
      Odds ratio |                1       |   .3911965    3.033018   (exact)
 Attr. frac. ex. |                0       |   -1.55626    .6702954   (exact)
 Attr. frac. pop |                0       |
                 +-----------------------------------------------
                          chi2(1) =      0.00  Pr>chi2 = 1.0000
```

Stratified Contingency Tables - Example

Nonsmokers

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

```
. cci 9 21 891 2079
                                                         Proportion
                 |   Exposed   Unexposed  |      Total     Exposed
-----------------+------------------------+------------------------
           Cases |         9          21  |         30      0.3000
        Controls |       891        2079  |       2970      0.3000
-----------------+------------------------+------------------------
           Total |       900        2100  |       3000      0.3000
                 |                        |
                 |      Point estimate    |    [95% Conf. Interval]
                 |------------------------+------------------------
      Odds ratio |                1       |   .4015748    2.288393  (exact)
 Attr. frac. ex. |                0       |  -1.490196    .5630121  (exact)
 Attr. frac. pop |                0       |
                 +-------------------------------------------------
                               chi2(1) =     0.00  Pr>chi2 = 1.0000
```

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 $\neq$ 1.0

Mantel-Haenszel Methods - Example

Lung Cancer data

```
. use "P:\Biostat513_06\drink.dta", clear
. list

     +-------------------------------------+
     | cancer    drink    number    smoke  |
     |-------------------------------------|
  1. |      1        1        24        1  |
  2. |      1        0         6        1  |
  3. |      0        1       776        1  |
  4. |      0        0       194        1  |
  5. |      1        1         9        0  |
  6. |      1        0        21        0  |
  7. |      0        1       891        0  |
  8. |      0        0      2079        0  |
     +-------------------------------------+

. cc cancer drink [freq=number], by(smoke) bd

         Smoker |       OR       [95% Conf. Interval]     M-H Weight
-----------------+-------------------------------------------------
              0 |          1      .4015748   2.288393         6.237 (exact)
              1 |          1      .3911965   3.033018         4.656 (exact)
-----------------+-------------------------------------------------
          Crude |   1.666533      .9677794   2.892949               (exact)
    M-H combined |          1      .5521991   1.810941
-------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(1) =     0.00  Pr>chi2 = 1.0000
Test of homogeneity (B-D)      chi2(1) =     0.00  Pr>chi2 = 1.0000

                   Test that combined OR = 1:
                                Mantel-Haenszel chi2(1) =      0.00
                                                Pr>chi2 =    1.0000
```

Stratified Contingency Tables - Example

EXAMPLE: (Rosner sec 13.5)

A 1985 study identified a group of 518 cancer cases and a group of age- and sex-matched 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  (exact)
 Attr. frac. ex. |         .3892779       |   .2094943    .5282126  (exact)
 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 (exact)
 Attr. frac. ex. |         .5225806       |    .2945527   .6771241 (exact)
 Attr. frac. pop |         .2714705       |
                 +-------------------------------------------------
                               chi2(1) =    15.24  Pr>chi2 = 0.0001
```

Stratified Contingency Tables - Example

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  (exact)
 Attr. frac. ex. |         .2381286       |  -.0887774    .4668978  (exact)
 Attr. frac. pop |          .137909       |
                 +-----------------------------------------------
                               chi2(1) =     2.43  Pr>chi2 = 0.1192
```

Mantel-Haenszel Methods - Example

Passive Smoking data

```
. use "M:\.MyDocs\b513\passive.dta"
. list

     +-----------------------------------+
     | 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
```

Stratified Data - Summary

1. Compute stratum-specific measures
2. Evaluate stratum-specific estimates by a test of homogeneity. Consider test results in light of sample size.
3. If the homogeneity test result is <u>non-significant</u> then consider a common estimate, pooling across all strata
   (a) calculate an overall (common) summary (OR)
   (b) test for significant association H_0: OR = 1
   (c) calculate confidence interval
4. If the homogeneity test result is <u>significant</u> then we are concerned that the ORs vary across strata. We may
   (a) If the direction of association (±) is same and the difference is small in magnitude, then
      - proceed as in 3 above (calculating average summary)
      - report on the test of homogeneity.

   (b) If the direction of the association is different, then
      - report results from test of homogeneity
      - report stratum-specific measures and confidence intervals.
      - does the average make sense at all?

Review

- R x C contingency table
  - Test for homogeneity (Pearson chi-squared)
- Single 2 x 2 table
  - Different sampling schemes
    1. Cohort (row totals fixed)
    2. Case-control (column totals fixed)
    3. Cross-sectional (grand total fixed)
  - Different measures of association

    RD (Designs 1 & 3)

    RR (Designs 1 & 3)

    OR (Designs 1, 2 & 3)
  - Test of association

    Pearson chi-squared

    McNemar's

    Fisher exact

Review

- Series of 2 x 2 tables
  - o Mantel-Haenszel (combined) OR estimate
  - o Mantel-Haenszel test for association

    H_o: OR = 1

    H_a: OR constant, $\neq 1$
  - o Breslow-Day "Score" Test for Homogeneity (Interaction, Effect Modification)