Rearranging Stata's output for the analysis of epidemiological tables

5ª Reunión de Usuarios Españoles de Stata

Barcelona, 12/09/2012

Aurelio Tobías

IDAEA-CSIC, Barcelona

aurelio.tobias@gmail.com

http://aureliotobias.weebly.com/

idaea

CSIC

Consejo Superior de Investigaciones Científicas

Background

- In epidemiological research we usually investigate the relationship between a binary response (C/nC) and a binary exposure (E/nE)

- Exposure $\xrightarrow{\text{Risk}}$ Outcome

Background

- A major capability in Stata is the analysis of epidemiological tables by using any of the **`epitab`** commands
- These report measures of frequency (proportion or odds), association (risk difference, relative risk, or odds ratio) and impact on public health (attributable risks)

```
. use example 1, clear

Contains data from example1.dta
  obs:             8
 vars:             4
-------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
y               byte   %8.0g       cc         Response
x               byte   %8.0g       yn         Exposure
z               byte   %8.0g       yn         Other
N               int    %8.0g                  Sample size
-------------------------------------------------------------------------------

. list, noobs

  +---------------------------+
  |       y     x     z     N |
  |---------------------------|
  | control    no    no   131 |
  | control    no   yes   234 |
  | control   yes    no   650 |
  | control   yes   yes   295 |
  |    case    no    no   122 |
  |    case    no   yes    78 |
  |    case   yes    no   660 |
  |    case   yes   yes   450 |
  +---------------------------+

. expand N
(2612 observations created)
```

```
. cs y x

                 | Exposure                |
                 |   Exposed   Unexposed   |      Total
-----------------+-------------------------+-----------
           Cases |      1110         200   |       1310
        Noncases |       945         365   |       1310
-----------------+-------------------------+-----------
           Total |      2055         565   |       2620
                 |                         |
            Risk |   .540146    .3539823   |         .5
                 |                         |
                 |      Point estimate     |    [95% Conf. Interval]
                 |-------------------------+------------------------
 Risk difference |         .1861637        |    .1412291    .2310982
      Risk ratio |         1.525912        |    1.355638    1.717574
 Attr. frac. ex. |         .3446544        |    .2623399    .4177835
 Attr. frac. pop |         .2920354        |
                 +-------------------------------------------------
                              chi2(1) =    61.43  Pr>chi2 = 0.0000
```

```
. cc y x

                                                                Proportion
                 |   Exposed   Unexposed  |      Total          Exposed
-----------------+------------------------+------------------------
           Cases |      1110         200  |       1310           0.8473
        Controls |       945         365  |       1310           0.7214
-----------------+------------------------+------------------------
           Total |      2055         565  |       2620           0.7844
                 |                        |
                 |      Point estimate    |    [95% Conf. Interval]
                 |------------------------+------------------------
      Odds ratio |         2.143651       |    1.759248    2.612501
 Attr. frac. ex. |         .5335061       |    .4315754     .617225
 Attr. frac. pop |         .4520548       |
                 +-----------------------------------------------
                              chi2(1) =    61.43  Pr>chi2 = 0.0000
```

```
. cc y x

                 |   Exposed   Unexposed  |      Total
-----------------+------------------------+------------
           Cases |      1110         200  |       1310
        Controls |       945         365  |       1310
-----------------+------------------------+------------
           Total |      2055         565  |       2620
                 |                        |
            Odds |     .54795     1.17460  |
                 |                        |
                 |      Point estimate    |    [95% Conf. Interval]
                 |------------------------+------------------------
      Odds ratio |         2.143651       |    1.759248    2.612501
 Attr. frac. ex. |         .5335061       |    .4315754     .617225
 Attr. frac. pop |         .4520548       |
                 +-------------------------------------------------
                              chi2(1) =    61.43  Pr>chi2 = 0.0000
```

. tabodds y x

x	cases	controls	odds	[95% Conf.	Interval]
no	200	365	0.54795	0.46116	0.65106
yes	1110	945	1.17460	1.07700	1.28105

. tabodds y x, or

x	Odds Ratio	chi2	P>chi2	[95% Conf.	Interval]
no	1.000000	.	.	.	.
yes	2.143651	61.41	0.0000	1.763229	2.606149

. mhodds y x

Maximum likelihood estimate of the odds ratio
Comparing x==1 vs. x==0

Odds Ratio	chi2(1)	P>chi2	[95% Conf.	Interval]
2.143651	61.41	0.0000	1.763229	2.606149

Stratified analysis

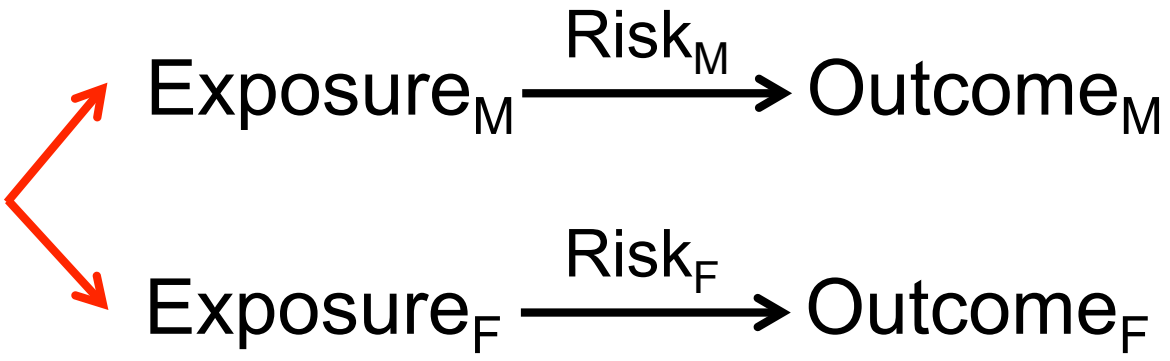
- But ... there could be more variables involved in this relationship

- Exposure → Outcome (Risk?)

Stratified analysis

- Analyse the relationship between the binary response (C/nC) and the binary exposure (E/nE) for each stratum taken by a third variable (M/F)

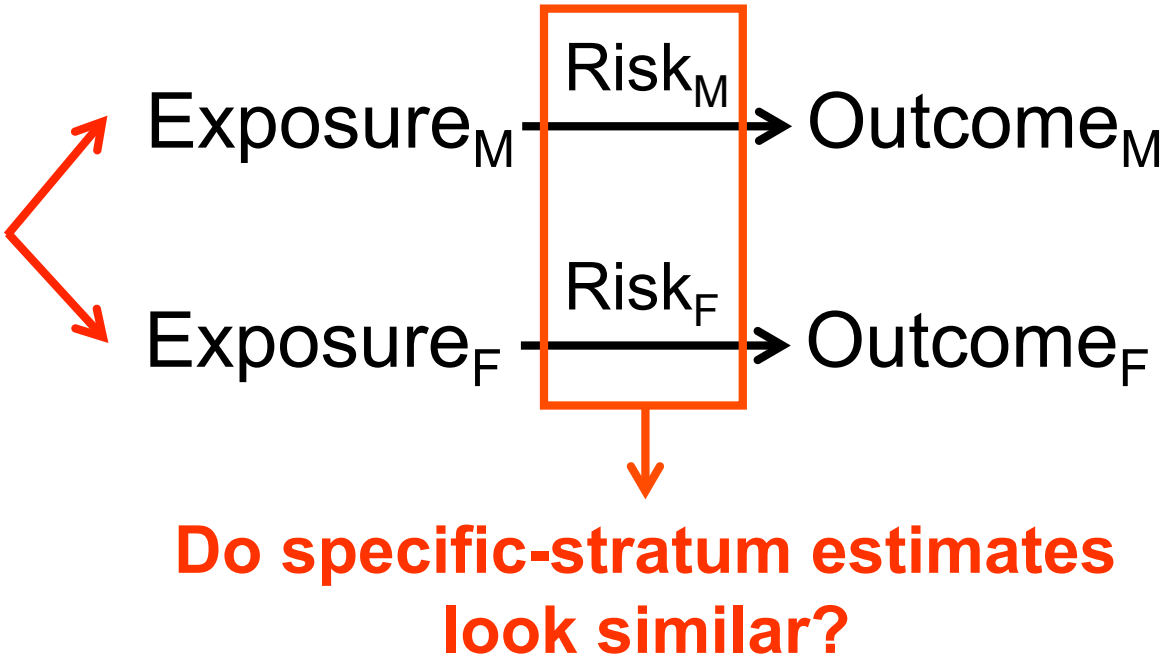
- Third variable



Stratified analysis

- Analyse the relationship between the binary response (C/nC) and the binary exposure (E/nE) for each stratum taken by a third variable (M/F)

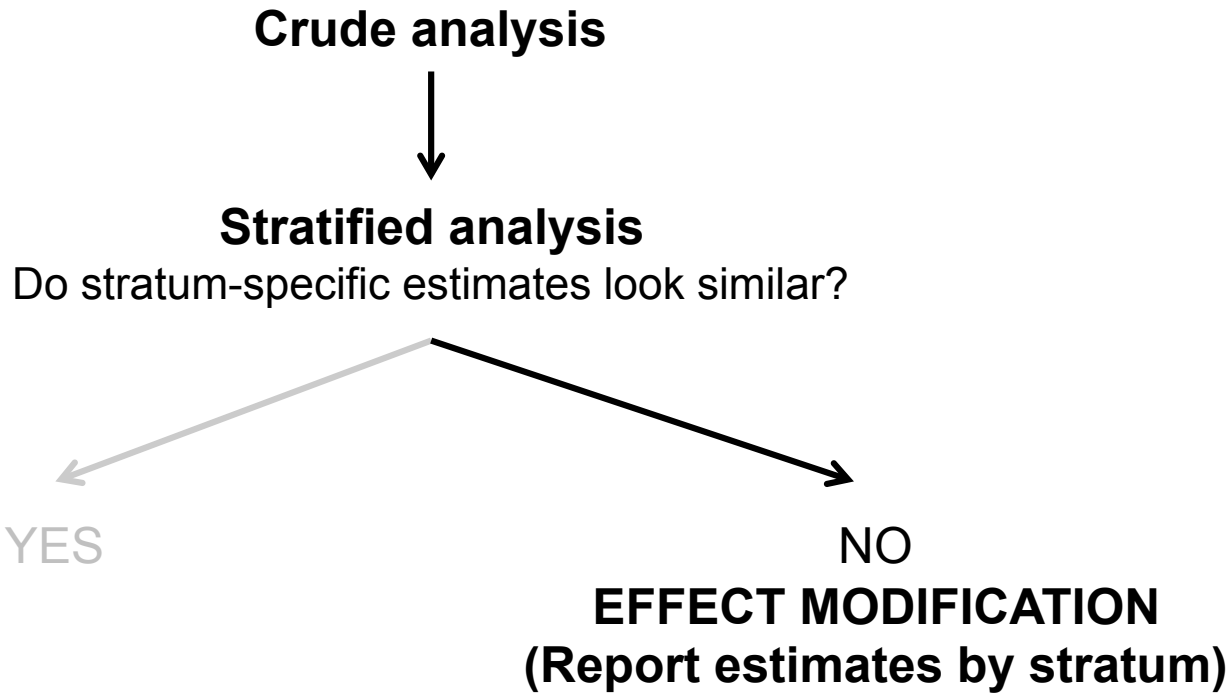
- Third variable



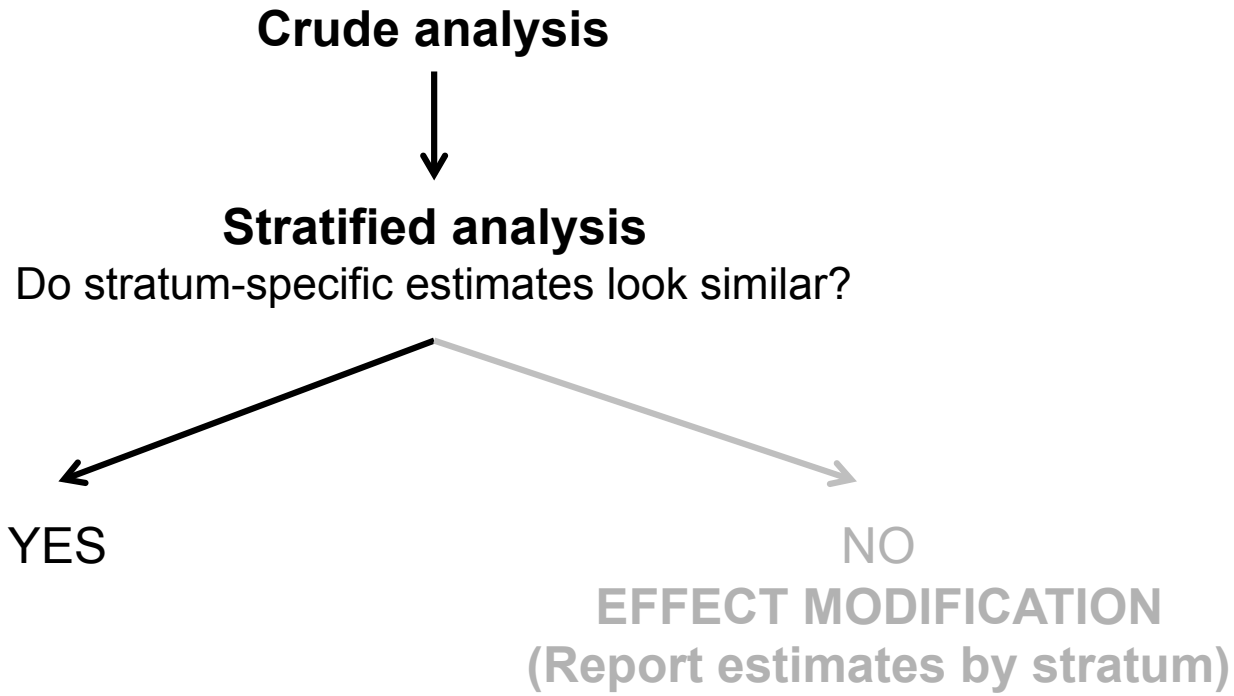
Effect modification

- Variation in the magnitude of measure of effect across stratums of a third variable (subgroups of population)
- Can be tested for by using an statistical test for homogeneity

How to conduct an stratified analysis?

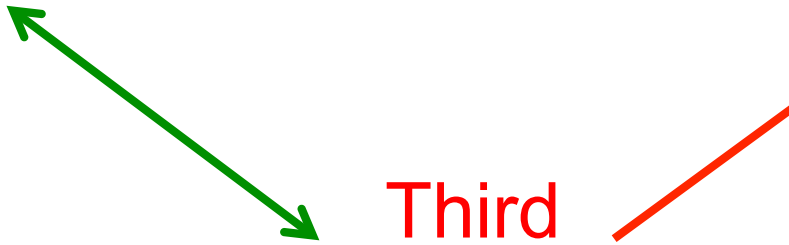


How to conduct an stratified analysis?



Confounding

- Distortion in the of measure of effect because of a third variable
- Exposure ⟶ Outcome



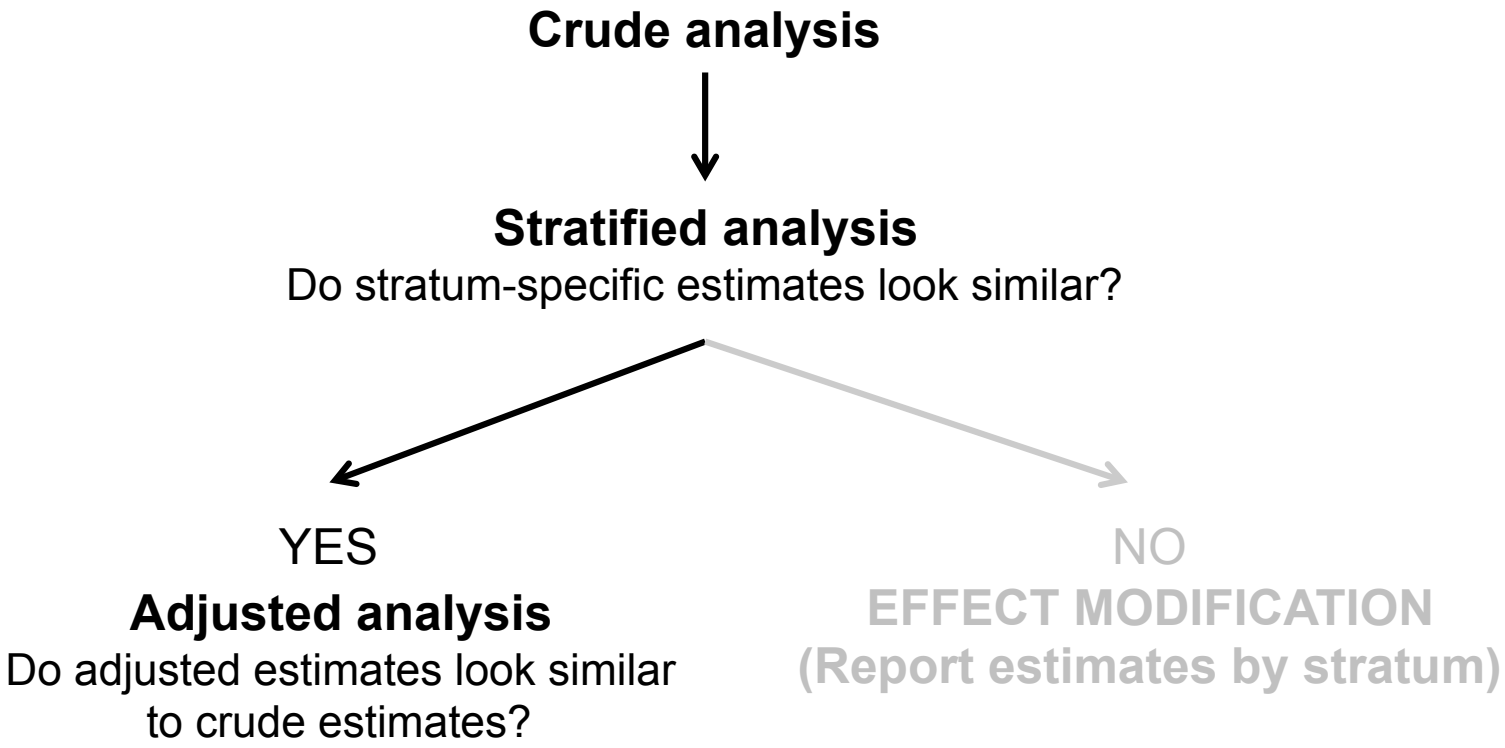
Third variable

- Be associated with outcome, independently of exposure
- Be associated with exposure, without being the consequence of exposure

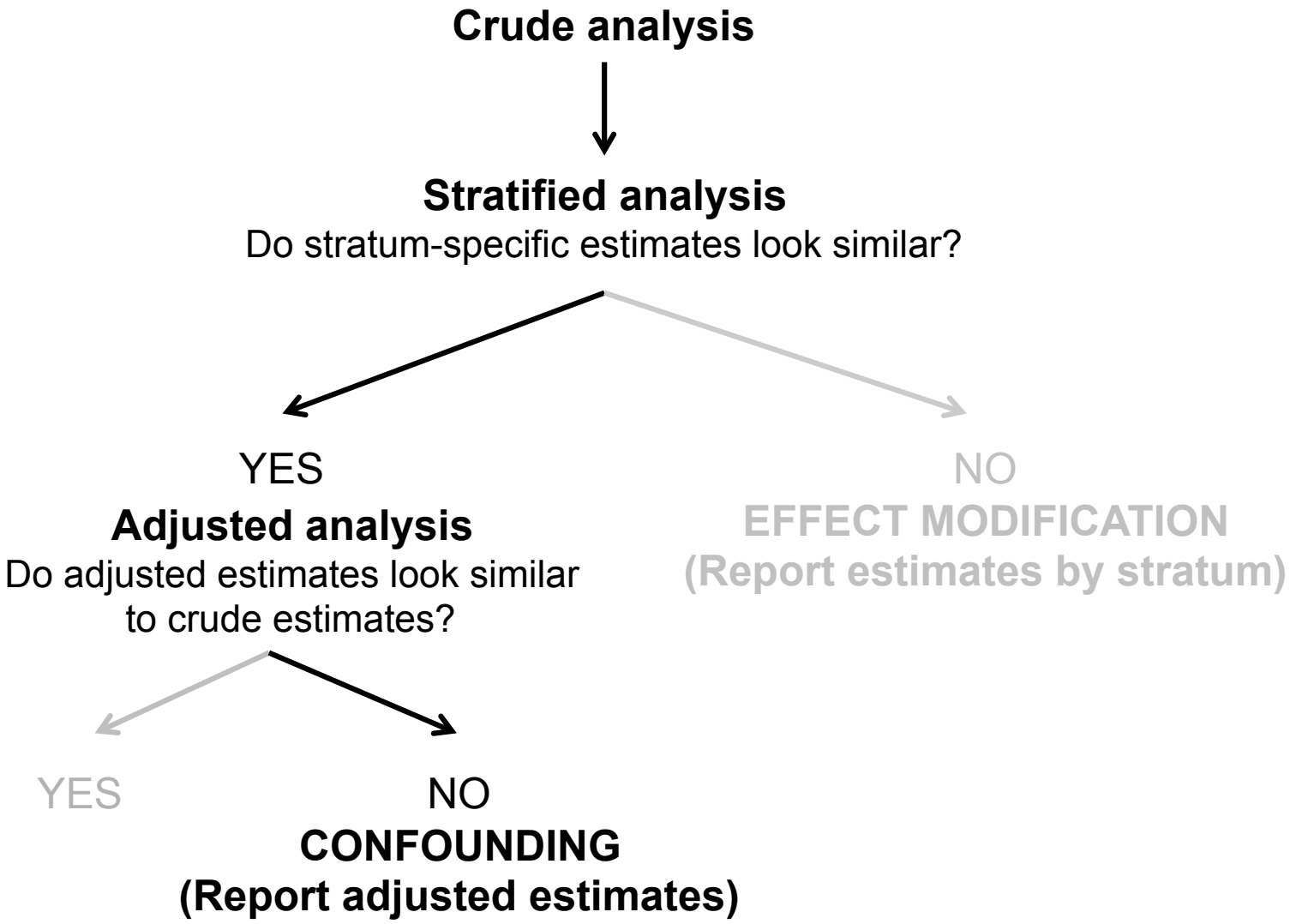
Confounding

- A confounder influences (bias) the effect of a exposure in the same way for everyone
- Needs to be adjusted for
- There is not statistical test to check for confounding, it is usually assessed by comparing crude and adjusted estimates

How to conduct an stratified analysis?



How to conduct an stratified analysis?



Stratified analysis

- Any of the **`epitab`** commands jointly with the **`by()`** option allow to run stratified analyses reporting; specific stratum measures of association, a test of homogeneity, as well as the crude and adjusted estimates
- These allow to epidemiologists to check for effect modification or to assess for confounding
- However … users are still being confused about how Stata reports stratified analyses

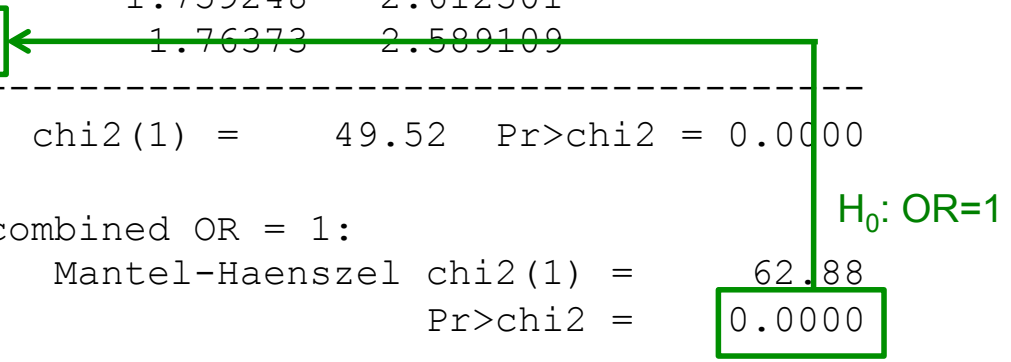
```
. cc y x, by(z)

           Other |       OR       [95% Conf. Interval]   M-H Weight
-----------------+-------------------------------------------------
              no |    1.09029      .8252907   1.440924     50.73576
             yes |   4.576271      3.375437   6.229908     21.76916
-----------------+-------------------------------------------------
           Crude |   2.143651      1.759248   2.612501
    M-H combined |   2.136934       1.76373   2.589109
-----------------------------------------------------------------
Test of homogeneity (M-H)      chi2(1) =    49.52  Pr>chi2 = 0.0000

                    Test that combined OR = 1:
                                Mantel-Haenszel chi2(1) =     62.88
                                                Pr>chi2 =    0.0000
```

```
. cc y x, by(z)

           Other |       OR       [95% Conf. Interval]    M-H Weight
-----------------+------------------------------------------------
              no |    1.09029        .8252907   1.440924      50.73576
             yes |    4.576271       3.375437   6.229908      21.76916
-----------------+------------------------------------------------
           Crude |    2.143651       1.759248   2.612501
    M-H combined |    2.136934        1.76373   2.589109
-----------------------------------------------------------------
Test of homogeneity (M-H)       chi2(1) =    49.52  Pr>chi2 = 0.0000

                    Test that combined OR = 1:
                                Mantel-Haenszel chi2(1) =     62.88
                                                Pr>chi2 =    0.0000
```

Do stratum-specific estimates look similar?

- Misleading p-value

```
. cc y x, by(z)

           Other |       OR       [95% Conf. Interval]    M-H Weight
-----------------+-------------------------------------------------
              no |    1.09029      .8252907   1.440924      50.73576
             yes |   4.576271      3.375437   6.229908      21.76916
-----------------+-------------------------------------------------
           Crude |   2.143651      1.759248   2.612501
    M-H combined |   2.136934       1.76373   2.589109
-----------------------------------------------------------------
Test of homogeneity (M-H)      chi2(1) =    49.52  Pr>chi2 = 0.0000

                   Test that combined OR = 1:
                               Mantel-Haenszel chi2(1) =     62.88
                                                Pr>chi2 =    0.0000
```



- Misleading p-value

```
. cc y x, by(z)

           Other |       OR       [95% Conf. Interval]   M-H Weight
-----------------+-------------------------------------------------
              no |    1.09029      .8252907   1.440924     50.73576
             yes |   4.576271      3.375437   6.229908     21.76916
-----------------+-------------------------------------------------
           Crude |   2.143651      1.759248   2.612501
    M-H combined |   2.136934       1.76373   2.589109
-----------------------------------------------------------------
Test of homogeneity (M-H)      chi2(1) =    49.52  Pr>chi2 = 0.0000

                    Test that combined OR = 1:
                                Mantel-Haenszel chi2(1) =     62.88
                                                Pr>chi2 =    0.0000
```

- Misleading p-values
- No p-values for specific-stratum and crude estimates
- Missing the relative change between the crude and adjusted estimates

```
. mhodds y x, by(z)

Maximum likelihood estimate of the odds ratio
Comparing x==1 vs. x==0
by z

------------------------------------------------------------------------------
          z | Odds Ratio       chi2(1)       P>chi2     [95% Conf. Interval]
------------+-----------------------------------------------------------------
         no |   1.090290         0.40       0.5294       0.83276    1.42746
        yes |   4.576271       110.14       0.0000       3.34905    6.25319
------------------------------------------------------------------------------

    Mantel-Haenszel estimate controlling for z
    ------------------------------------------------------------
     Odds Ratio    chi2(1)        P>chi2        [95% Conf. Interval]
    ------------------------------------------------------------
       2.136934      62.88        0.0000         1.763202   2.589884
    ------------------------------------------------------------

Test of homogeneity of ORs (approx): chi2(1)    =   48.86
                                     Pr>chi2    =  0.0000
```

```
. mhodds y x, by(z)

Maximum likelihood estimate of the odds ratio
Comparing x==1 vs. x==0
by z

------------------------------------------------------------------------------
          z | Odds Ratio       chi2(1)       P>chi2     [95% Conf. Interval]
------------+-----------------------------------------------------------------
         no |   1.090290         0.40       0.5294       0.83276    1.42746
        yes |   4.576271       110.14       0.0000       3.34905    6.25319
------------------------------------------------------------------------------

    Mantel-Haenszel estimate controlling for z
    ------------------------------------------------------------
     Odds Ratio    chi2(1)        P>chi2        [95% Conf. Interval]
    ------------------------------------------------------------
       2.136934      62.88        0.0000         1.763202   2.589884
    ------------------------------------------------------------

Test of homogeneity of ORs (approx): chi2(1)    =    48.86
                                     Pr>chi2    =   0.0000
```

Do stratum-specific estimates look similar?

- Another misleading p-value

```
. mhodds y x, by(z)

Maximum likelihood estimate of the odds ratio
Comparing x==1 vs. x==0
by z

------------------------------------------------------------------------------
          z | Odds Ratio       chi2(1)        P>chi2     [95% Conf. Interval]
------------+-----------------------------------------------------------------
         no |   1.090290         0.40        0.5294       0.83276    1.42746
        yes |   4.576271       110.14        0.0000       3.34905    6.25319
------------------------------------------------------------------------------

    Mantel-Haenszel estimate controlling for z
    ------------------------------------------------------------
     Odds Ratio    chi2(1)        P>chi2     [95% Conf. Interval]
    ------------------------------------------------------------
       2.136934      62.88        0.0000       1.763202   2.589884
    ------------------------------------------------------------

Test of homogeneity of ORs (approx): chi2(1)    =   48.86
                                     Pr>chi2    =  0.0000
```

- Another misleading p-value
- Missing the crude estimate, and the relative change between the crude and adjusted estimates

Rearranging the output

- Rearrange the output of the **`epitab`** commands from the scheme of a classic epidemiological analysis
  - Specific stratum estimates, with 95%CI and p-values
  - Test of homogeneity (to check for effect modification)
  - Crude and adjusted estimates, with 95%CI and p-values
  - Relative change between them (to assess for confounding)

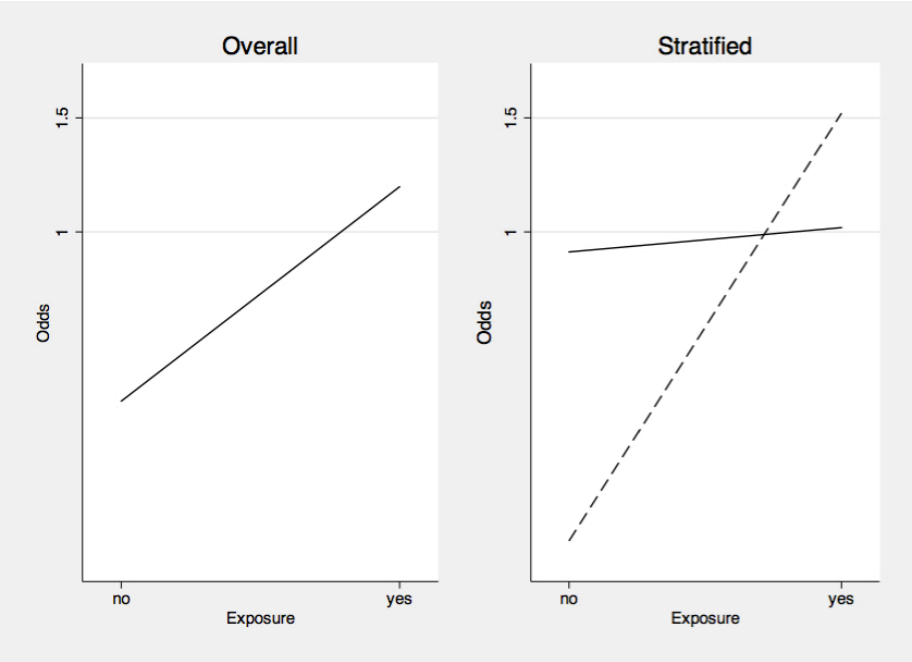
```
. mymhodds y x, by(z)

Stratified analysis with test for effect modification
------------------------------------------------------------------------------
 z          |  Odds Ratio   [95% Conf. Interval]       chi2(1)     P>chi2
------------+-----------------------------------------------------------------
 no         |     1.09029    .832758    1.42746           0.40     0.5294
 yes        |     4.57627    3.34905    6.25319         110.14     0.0000
------------------------------------------------------------------------------
Test of homogeneity of ORs (approx): chi2(1) = 48.86 , P>chi2 = 0.0000


Adjusted analysis with assessment for confounding
------------------------------------------------------------------------------
            |  Odds Ratio   [95% Conf. Interval]       chi2(1)     P>chi2
------------+-----------------------------------------------------------------
 Crude      |     2.14365    1.76323    2.60615          61.41     0.0000
 Adjusted   |     2.13693     1.7632    2.58988          62.88     0.0000
------------------------------------------------------------------------------
Adjusted/crude relative change = -0.31%
```

```
. use example2, clear

. expand N
(2612 observations created)

. mymhodds y x, by(z)

Stratified analysis with test for effect modification
------------------------------------------------------------------------------
 z          |  Odds Ratio   [95% Conf. Interval]       chi2(1)     P>chi2
------------+-----------------------------------------------------------------
 no         |     5.57126    3.45525    8.98315          63.13     0.0000
 yes        |     5.61665    3.42494    9.21089          59.58     0.0000
------------------------------------------------------------------------------
Test of homogeneity of ORs (approx): chi2(1) =  0.00 , P>chi2 = 0.9815


Adjusted analysis with assessment for confounding
------------------------------------------------------------------------------
            |  Odds Ratio   [95% Conf. Interval]       chi2(1)     P>chi2
------------+-----------------------------------------------------------------
 Crude      |     2.14365    1.76323    2.60615           61.41     0.0000
 Adjusted   |     5.59398    3.96622    7.88980          122.56     0.0000
------------------------------------------------------------------------------
Adjusted/crude relative change = 160.96%
```

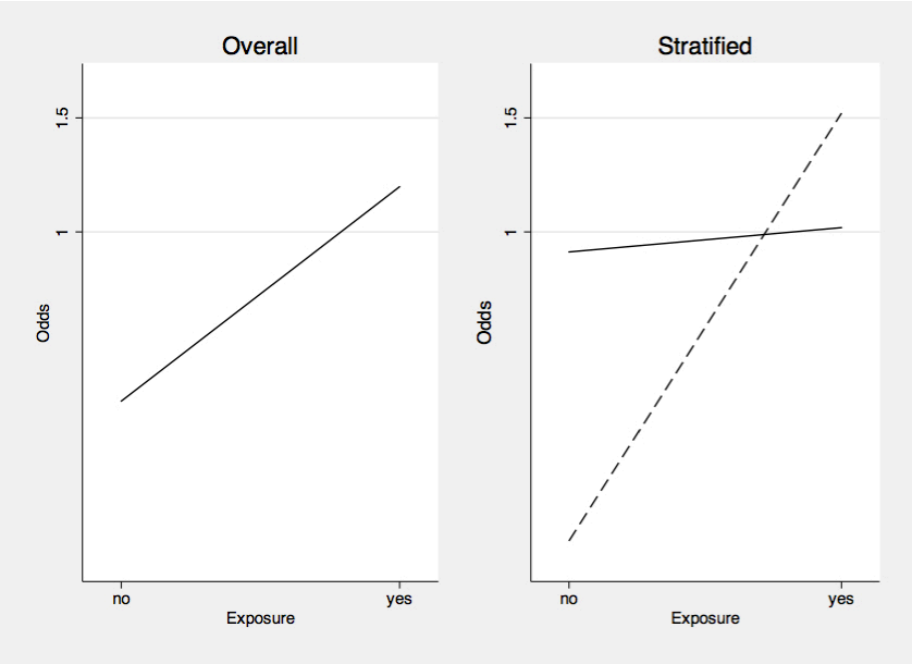
Rearranging the output

- Rearrange the output of the **`epitab`** commands from the scheme of a classic epidemiological analysis
  - Specific stratum estimates, with 95%CI and p-values
  - Test of homogeneity (to check for effect modification)
  - Crude and adjusted estimates, with 95%CI and p-values
  - Relative change between them (to assess for confounding)
- Furthermore, graphs are always helpful … and nice!

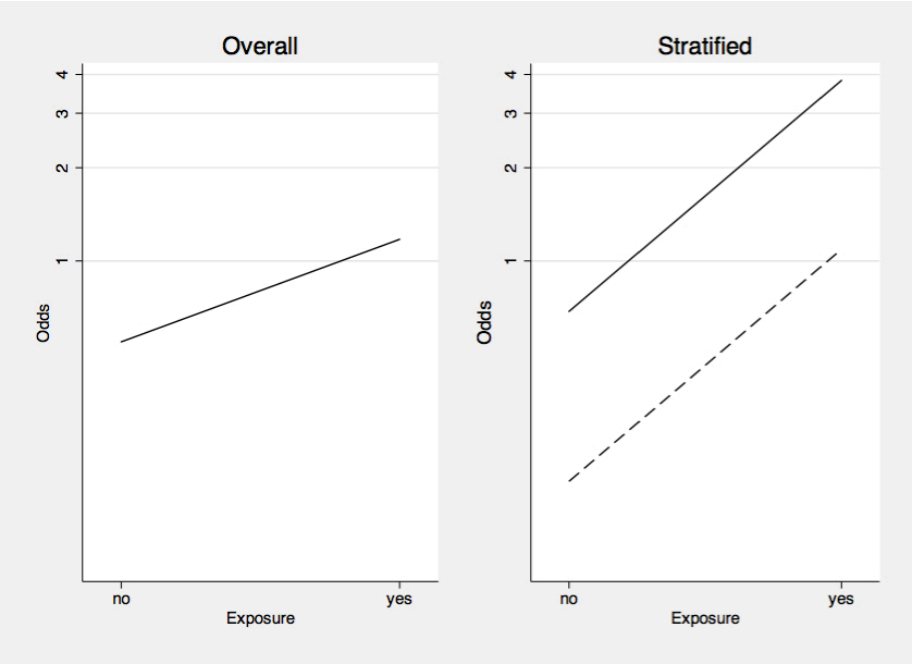
```
. use example1, clear
. expand N
. mymhodds y x, by(z) graphs
```



```
. use example1, clear
. expand N
. mymhodds y x, by(z) graphs
```




```
. use example2, clear
. expand N
. mymhodds y x, by(z) graphs
```



Conclusions

- Stata is mainly used in epidemiological research thanks to the **`epitab`** commands
- The **`cc`** command must also display the odds
- **`epitab`** commands with capabilities to run stratified analysis with the **`by()`** option (**`cs`**, **`cc`**, **`ir`**, and **`mhodds`**) must rearrange their output to address effect modification and confounding in a easiest way
- **`diagt`** (SJ4-4) command should also be included in **`epitab`** to cover all types of study designs in epidemiological research

Rearranging Stata's output for the analysis of epidemiological tables

5ª Reunión de Usuarios Españoles de Stata

Barcelona, 12/09/2012

Aurelio Tobías

IDAEA-CSIC, Barcelona

aurelio.tobias@gmail.com

http://aureliotobias.weebly.com/

idaea

CSIC

Consejo Superior de Investigaciones Científicas