VER – Chapter 13
Sample Problems (Not in Text)
Solutions

Evaluating confounding

1. Does atrophic rhinitis increase the risk of pneumonia in pigs?

```
. use "C:\ver\data\pig_adg.dta", clear
. gen ar_pos=(ar>=1) if ar~=.
. cs pn ar_pos, by(sex)

  Sex of the pig |       RR      [95% Conf. Interval]    M-H Weight
-----------------+-------------------------------------------------
          female |   1.364103     .9346469   1.990886      13.60465
        castrate |       1.47     1.017975   2.122743      14.20118
-----------------+-------------------------------------------------
           Crude |   1.417684     1.089216   1.845206
    M-H combined |   1.418187     1.089701   1.845694
-----------------------------------------------------------------
Test of homogeneity (M-H)      chi2(1) =    0.077  Pr>chi2 = 0.7810
```

ar_pos is a risk factor. There is no evidence of confounding by sex (ie the crude and —MH combined RR are virtually identical).

2. Does hoof angle affect the risk of lameness ?

First create the dataset using the Stata editor and save the file as -hoofangle.dta-

```
. cs lame hoof [fweight=count], by(breed)

           breed |       RR      [95% Conf. Interval]    M-H Weight
-----------------+-------------------------------------------------
               0 |    .721519     .3229062   1.612201       7.52381
               1 |   .5132275     .3621156   .7273988      19.89474
-----------------+-------------------------------------------------
           Crude |          2     1.599477   2.500817
    M-H combined |   .5703839     .4038604   .8055699
-----------------------------------------------------------------
Test of homogeneity (M-H)      chi2(1) =    0.680  Pr>chi2 = 0.4094
```

Since the test for interaction is clearly non-significant, it is appropriate to proceed with an evaluation of breed as a confounder, and in this case it is a very strong confounding factor. If it is ignored, low hoof angle appears to be a significant risk factor. Controlling for breed reverses this apparent effect and makes low hoof angle a preventive factor. (Breed is a "distorter variable – see Chapter 13)

3. Is neomycin use a risk factor for Nocardia mastitis?

```
use "C:\ver\data\nocardia.dta", clear
gen large=(numcow>=50) if numcow~=.
```

```
label def yesno 0 no 1 yes
cc casecont dneo, by(large)

           large |       OR      [95% Conf. Interval]    M-H Weight
-----------------+-------------------------------------------------
               0 |   14.29412      2.74241   136.6878        .53125 (exact)
               1 |   10.08333     1.876849   66.80226      .5454545 (exact)
-----------------+-------------------------------------------------
           Crude |     11.368     3.656368   41.25474               (exact)
    M-H combined |   12.16095     4.077007   36.27384
-----------------------------------------------------------------
Test of homogeneity (M-H)      chi2(1) =     0.10  Pr>chi2 = 0.7529

                  Test that combined OR = 1:
                              Mantel-Haenszel chi2(1) =     24.65
                                              Pr>chi2 =    0.0000
```

Since this was a case-control study, we should evaluate associations using odds ratios. Neomycin use is definitely a risk factor and there is no evidence of interaction between neomycin use and herd size. Herd size is probably not a confounder. While it may appear that there is some difference between the crude and adjusted OR (and it is > the 20% "guideline"), it is very small relative to the size of the confidence intervals for the two estimates.

4. Does pneumonia affect average daily gain in pigs?

```
use "C:\ver\data\pig_adg.dta", clear
reg adg pn

      Source |       SS       df       MS              Number of obs =     341
-------------+------------------------------           F(  1,   339) =   10.37
       Model |  50837.5391     1  50837.5391           Prob > F      =  0.0014
    Residual |  1661530.22   339  4901.26909           R-squared     =  0.0297
-------------+------------------------------           Adj R-squared =  0.0268
       Total |  1712367.76   340  5036.37576           Root MSE      =  70.009

------------------------------------------------------------------------------
         adg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          pn |  -24.52151   7.613935    -3.22   0.001    -39.49801      -9.545
       _cons |   532.7742   5.623263    94.74   0.000     521.7133    543.8351
------------------------------------------------------------------------------

. reg adg pn sex

      Source |       SS       df       MS              Number of obs =     341
-------------+------------------------------           F(  2,   338) =   11.64
       Model |  110339.586     2  55169.7929           Prob > F      =  0.0000
    Residual |  1602028.17   338  4739.72832           R-squared     =  0.0644
-------------+------------------------------           Adj R-squared =  0.0589
       Total |  1712367.76   340  5036.37576           Root MSE      =  68.846

------------------------------------------------------------------------------
         adg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          pn |  -24.46469   7.487427    -3.27   0.001    -39.19251   -9.736865
         sex |   26.42023   7.456711     3.54   0.000     11.75282    41.08763
       _cons |   519.6493    6.65588    78.07   0.000     506.5571    532.7415
------------------------------------------------------------------------------
```

Adding sex to the model improves the predictive ability of the model substantially (the R^2 goes from 2% to 6%), but it is not a confounder since it does not appear to affect the coefficient for pn. (As in a Mantel-Haenszel procedure, using a "guideline" of a 20% change is one way to evaluate the magnitude of the difference between the crude and adjusted estimates.)

Evaluating interaction and confounding

5. More on Nocardia mastitis

```
use "C:\ver\data\nocardia.dta", clear
label def yesno 0 no 1 yes /* this fixes a problem with the dataset that
                              was put on the web site */
. cc casecont dneo, by(dclox)

Cloxacillin used |       OR      [95% Conf. Interval]    M-H Weight
-----------------+-------------------------------------------------
              no |   29.33333     5.748126    274.188      .3703704 (exact)
             yes |        1.5     .2112196   12.35273      1.111111 (exact)
-----------------+-------------------------------------------------
           Crude |     11.368     3.656368   41.25474               (exact)
    M-H combined |   8.458333     3.020121   23.68892
-----------------------------------------------------------------
Test of homogeneity (M-H)      chi2(1) =     6.44  Pr>chi2 = 0.0112

                  Test that combined OR = 1:
                              Mantel-Haenszel chi2(1) =     22.51
                                              Pr>chi2 =    0.0000
```

There is very strong evidence of interaction between neomycin use and cloxacillin use. Consequently, it is not appropriate to try to compute a single summary measure for the effect of neomycin.

6. Risk factors for pneumonia in pigs

The data for this are already available in question 1. There was no evidence of interaction between sex and ar_pos (the stratum-specific RR are very close). Consequently, we conclude that the effect of ar_pos is the same in both sexes.

7. Are pastured horses at higher risk of Potomac horse fever?

First create the dataset using the Stata editor and save the file as -potomac.dta-

```
use "C:\ver\ch13\teach\potomac.dta", clear
cs sc past [fweight=count], or


                | past                    |
                |   Exposed   Unexposed   |     Total
----------------+-------------------------+----------
          Cases |        20           5   |        25
       Noncases |        50         105   |       155
----------------+-------------------------+----------
          Total |        70         110   |       180
                |                         |
```

```
           Risk |  .2857143    .0454545  |  .1388889
                |                        |
                |      Point estimate    |  [95% Conf. Interval]
                |------------------------+----------------------
Risk difference |          .2402597      |  .1274998    .3530197
     Risk ratio |          6.285714      |  2.472523    15.97971
Attr. frac. ex. |          .8409091      |  .5955548    .9374206
Attr. frac. pop |          .6727273      |
     Odds ratio |              8.4       |  3.071856    22.82809  (Cornfield)
                +-----------------------------------------------
                             chi2(1) =    20.65  Pr>chi2 = 0.0000
```

(a) Measures of association

All of the following statements are based on the assumption that you are satisfied that the apparent relationship between exposure to pasture and seroconversion are "real" and causal in nature.

The **risk ratio** of 6.3 means that horses exposed to pasture were 6.3 times as likely to seroconver during the summer as were horses not exposed to pasture.

The **odds ratio** of 8.4 means that the odds of seroconversion was 8.4 times higher in horses exposed to pasture than in those not exposed. *Note* In this fictitious dataset, Potomac horse fever is not really a rare disease in that it affected 25/180 = 14% of horses over the summer. Consequently, the odds ratio does not provide a good estimate of the risk ratio.

The **risk difference** of 0.24 means that if you prevented pasture exposure in 100 horses (that otherwise would go out to pasture), you would prevent 24 cases of Potomac Horse Fever.

The **attributable fraction (exposed)** of 0.84 means that if you prevented pasture exposure in 100 horses (that otherwise would go out to pasture), you would prevent 84% of the cases of Potomac Horse Fever that would have occurred in this group of horses.

The **population attributable risk** is not computed in the output shown above. It is the difference between the risk in the population (0.139) and the risk in the non-exposed group (0.045) so it equals 0.094. This means that if you prevented pasture exposure in 100 horses randomly chosen from the study population, you would prevent about 9 cases of Potomac Horse Fever.

The **attributable fraction (population)** of 0.67 means that if you prevented pasture exposure in 100 randomly chosen horses, you would prevent 67% of the cases of Potomac Horse Fever that would have occurred in this group of horses.

(b) Is pasture a confounder?

Pasture use can not be a confounder for sex because it can not be antecedent (ie occur before) sex. No calculations are needed to determine this.

(c) Is sex a confounder for the pasture --> seroconversion relationship?

```
. cs sc past [fweight=count], by(sex)

             sex |       RR      [95% Conf. Interval]    M-H Weight
-----------------+-------------------------------------------------
               0 |   5.666667     1.751628   18.33215           1.5
               1 |        7.5     1.432664   39.26251      .3333333
-----------------+-------------------------------------------------
           Crude |   6.285714     2.472523   15.97971
    M-H combined |          6     2.224288   16.18496
-------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(1) =    0.079  Pr>chi2 = 0.7788
```

There is no evidence that sex is a confounding variable (ie the crude and M-H adjusted risk ratios are very close).

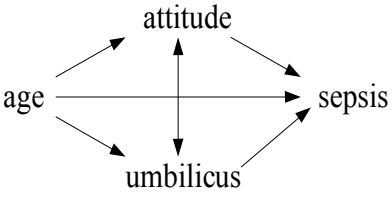
8. Factors that affect ADG in pigs

```
gen sexpn=sex*pn
reg adg sex pn sexpn

      Source |       SS       df       MS              Number of obs =     341
-------------+------------------------------           F(  3,   337) =    7.83
       Model |  111606.691     3  37202.2302           Prob > F      =  0.0000
    Residual |  1600761.07   337  4750.03285           R-squared     =  0.0652
-------------+------------------------------           Adj R-squared =  0.0569
       Total |  1712367.76   340  5036.37576           Root MSE      =   68.92

------------------------------------------------------------------------------
         adg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |   30.64352   11.07188     2.77   0.006     8.864817    52.42223
          pn |  -20.62575   10.55605    -1.95   0.052    -41.38979    .1382915
       sexpn |  -7.742968   14.99165    -0.52   0.606    -37.23197    21.74604
       _cons |   517.5513   7.803708    66.32   0.000     502.2012    532.9014
------------------------------------------------------------------------------
```

The interaction term for sex*pn is completely non-significant so we can conclude that the effect of pn on adg is the same in both sexes.

9. Factors that are related to the probability of septicemia

The first thing to do is to draw a causal diagram that you feel represents the relationships among the variables.



This is a bit different than previous exercises since we are not evaluating risk factors for a disease, but rather clinical signs that are indicative of a condition. The diagram at the left might adequately represent the relationships we are going to investigate. While -age- is not a "cause" of either a swollen umbilicus or attitude, the age of the calf at admission might influence the presence/absence of either of those other signs.

First, dichotomise -attd- and -age- (In reality, we could leave -attd- as a three level variable, but everything we have done so far has been with 2 level variables, so lets stick with that approach.). In this example I have chosen to dichotomise age at 7 days (0-6, 7+). This cutpoint was chosen rather

arbitrarily as the probability of sepsis appeared to decline initially but level off about 6-7 days. Other cut-points could be justified. Attitude was dichotomised by putting "bright" and "depressed" together. The alternative (putting "depressed" and "comatose" together would have left a very small reference group (there were only 28 "bright" calves).

```
. use "c:\ver\data\calf.dta", clear
. egen age_c2=cut(age), at(0,7,99)
(1 missing value generated)
. egen attd_c2=cut(attd), at(0,2,99)
(6 missing values generated)
```

(***Note*** without the -icodes- option on the -egen cut- command, the new variables are given values equal to the left hand end of the category.)

Is attitude an indicator of sepsis, and is age a confounding variable for this relationship?

```
. cs sepsis attd_c2, by(age_c2)

          age_c2 |       RR      [95% Conf. Interval]    M-H Weight
-----------------+-------------------------------------------------
               0 |   1.808696     1.143266   2.861433       5.47619
               7 |    1.75641     .9381198   3.288468      4.306748
-----------------+-------------------------------------------------
           Crude |   1.932174     1.313575   2.842089
    M-H combined |   1.785678     1.225425   2.602074
-----------------------------------------------------------------
Test of homogeneity (M-H)      chi2(1) =    0.006  Pr>chi2 = 0.9400
```

Sepsis is more common in calves that are comatose, and there is some limited evidence that age may be a confounder. There is no indication that the effect of being comatose is age-dependent (ie interaction). We will investigate this potential confounding effect of age later.

Is having a swollen umbilicus an indicator of sepsis, and is age a confounding variable for this relationship?

```
. cs sepsis umb, by(age_c2)

          age_c2 |       RR      [95% Conf. Interval]    M-H Weight
-----------------+-------------------------------------------------
               0 |   1.933638     1.234679   3.028284      5.202381
               7 |   1.915459      1.08379   3.385326      5.240506
-----------------+-------------------------------------------------
           Crude |   1.932016     1.327225   2.812398
    M-H combined |   1.924516     1.339222   2.765606
-----------------------------------------------------------------
Test of homogeneity (M-H)      chi2(1) =    0.001  Pr>chi2 = 0.9790
```

Sepsis is more common in calves that have a swollen umbilicus and there is evidence that age is a confounder. There is no indication that the effect of having a swollen umbilicus is age-dependent (ie interaction).

Is there a relationship between age and sepsis that is independent of attitude?

```
. cs sepsis age_c2 if attd_c2==0
```

```
                 | age_c2                |
                 |   Exposed   Unexposed |     Total
-----------------+-----------------------+----------
           Cases |        27          23 |        50
        Noncases |       110          41 |       151
-----------------+-----------------------+----------
           Total |       137          64 |       201
                 |                       |
            Risk |  .1970803     .359375 |  .2487562
                 |                       |
                 |      Point estimate   |  [95% Conf. Interval]
                 |-----------------------+----------------------
 Risk difference |        -.1622947      | -.2974085   -.0271809
      Risk ratio |         .5483973      |   .342628    .8777438
 Prev. frac. ex. |         .4516027      |  .1222562     .657372
 Prev. frac. pop |         .3078088      |
                 +-------------------------------------------
                           chi2(1) =     6.15  Pr>chi2 = 0.0132
```

Old calves were approximately ½ as likely to be septicemic as young calves (or alternatively, young calves were twice as likely to be septicemic). This suggests that there is a direct connection between age and sepsis. (*Note* We won't bother evaluating whether the effect of age was independent of having a swollen umbilicus, since there was absolutely no evidence of confounding in that relationship).

Were old calves more or less likely to be comatose than young calves?

```
. cs attd_c2 age_c2
```

```
                 | age_c2                |
                 |   Exposed   Unexposed |     Total
-----------------+-----------------------+----------
           Cases |        26          20 |        46
        Noncases |       137          64 |       201
-----------------+-----------------------+----------
           Total |       163          84 |       247
                 |                       |
            Risk |  .1595092    .2380952 |  .1862348
                 |                       |
                 |      Point estimate   |  [95% Conf. Interval]
                 |-----------------------+----------------------
 Risk difference |         -.078586      | -.1856167    .0284446
      Risk ratio |         .6699387      |  .3982454    1.126988
 Prev. frac. ex. |         .3300613      |  -.126988    .6017546
 Prev. frac. pop |         .2178138      |
                 +-------------------------------------------
                           chi2(1) =     2.26  Pr>chi2 = 0.1328
```

Old calves were only about 2/3 as likely to be comatose as young calves, but this risk ratio was of questionable significance.

Were old calves more or less likely to show signs of a swollen umbilicus?

```
. cs umb age_c2
```

```
                 | age_c2                |
                 |   Exposed   Unexposed |     Total
```

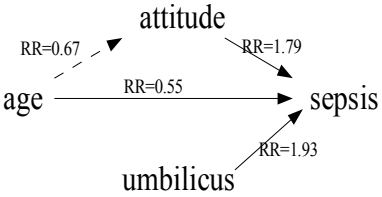
```
-----------------+------------------------+----------
           Cases |        36          19  |        55
        Noncases |       122          65  |       187
-----------------+------------------------+----------
           Total |       158          84  |       242
                 |                        |
            Risk |  .2278481    .2261905  |  .2272727
                 |                        |
                 |      Point estimate    |  [95% Conf. Interval]
                 |------------------------+----------------------
 Risk difference |          .0016576      | -.1091657    .1124809
      Risk ratio |          1.007328      |  .6179081    1.642171
 Attr. frac. ex. |          .0072751      | -.6183636    .3910499
 Attr. frac. pop |          .0047619      |
                 +-----------------------------------------------
                           chi2(1) =     0.00  Pr>chi2 = 0.9766
```

There was no connection between age and showing signs of a swollen umbilicus.

Was there any indication of an association between attitude and swollen umbilicus

```
. cs umb attd_c2, by(age_c2)

          age_c2 |       RR      [95% Conf. Interval]    M-H Weight
-----------------+-------------------------------------------------
               0 |   1.554656     .6842461   3.532289      2.975904
               7 |   .8516129     .3668099   1.977167      4.936306
-----------------+-------------------------------------------------
           Crude |   1.119318     .6303399   1.987615
    M-H combined |   1.116038     .6211064   2.005359
-----------------------------------------------------------------
Test of homogeneity (M-H)      chi2(1) =    1.023  Pr>chi2 = 0.3119
```

There is no substantive evidence of a connection between attitude and swollen umbilicus.



A summary of the relationships is shown here. The association between age and attitude is questionable (P=0.13). Calves that were young, comatose and had a swollen umbilicus had the highest risk of being septicemic.

10. Interaction in the calf dataset

(a) We have already evaluated multiplicative interaction between -age_c2- and -umb- in the previous question. The risk ratios for -umb- in young and old calves were 1.93 and 1.92 respectively. Clearly, there is no evidence of interaction on a multiplicative scale.

(b) To evaluate interaction on an additive scale, we need to obtain the risks of sepsis for each of the 4 possible combinations of the factors. Before proceeding though, I will recode -age_c2- so that young calves (the high risk group) are coded 1 and old calves are coded 0. (This will make -age_c2- a risk factor instead of a preventive factor which will make the assessment of interaction easier to follow.)

```
. recode age_c2 0=1 1=0
(age_c2: 253 changes made)

. cs sepsis umb if age_c2==0

                 | swollen umbilicus [0=no,|
                 | 1=yes]                  |
                 |   Exposed   Unexposed   |     Total
-----------------+-------------------------+----------
           Cases |        13          23   |        36
        Noncases |        23          99   |       122
-----------------+-------------------------+----------
           Total |        36         122   |       158
                 |                         |
            Risk |  .3611111    .1885246   |  .2278481
                 |                         |
                 |      Point estimate     |  [95% Conf. Interval]
                 |-------------------------+----------------------
 Risk difference |         .1725865        |  .0010189    .3441541
      Risk ratio |         1.915459        |   1.08379    3.385326
 Attr. frac. ex. |         .4779319        |  .0773122    .7046075
 Attr. frac. pop |         .1725865        |
                 +-------------------------------------------------
                           chi2(1) =     4.71  Pr>chi2 = 0.0301

. cs sepsis umb if age_c2==1

                 | swollen umbilicus [0=no,|
                 | 1=yes]                  |
                 |   Exposed   Unexposed   |     Total
-----------------+-------------------------+----------
           Cases |        13          23   |        36
        Noncases |         6          42   |        48
-----------------+-------------------------+----------
           Total |        19          65   |        84
                 |                         |
            Risk |  .6842105    .3538462   |  .4285714
                 |                         |
                 |      Point estimate     |  [95% Conf. Interval]
                 |-------------------------+----------------------
 Risk difference |         .3303644        |   .091205    .5695237
      Risk ratio |         1.933638        |  1.234679    3.028284
 Attr. frac. ex. |         .4828402        |  .1900727      .66978
 Attr. frac. pop |          .174359        |
                 +-------------------------------------------------
                           chi2(1) =     6.55  Pr>chi2 = 0.0105
```

The risk estimates are:

				Risk Difference
		No	Yes	
Age	Old	0.19	0.36	0.17
	Young	0.35	0.68	
Risk Difference		0.16		

Under the assumption of no interaction, we would expect a risk of 0.19 + 0.16 + 0.17 = 0.52 in calves that were young and had a swollen umbilicus. However, the observed risk was 0.68, suggesting synergy on the additive scale. (Note We have not done any evaluation of the statistical significance of the difference between the values of 0.52 and 0.68)

11. Using matching to control confounding

This question does not lend itself to a printed solution ... we will discuss it in class.