Models In Epidemiology And Biostatistics
Gordon Hilton Fick

Examples From "Biostatistical Methods in Epidemiology'
by Stephen Newman

Receptor Level And Breast Cancer
[adapted from Newman (2001, p 98 p 126)]

" The data for this example were kindly provided by the Northern Alberta Breast Cancer Registry. This is a population - based registry that collects information on all cases of breast cancer treated in the northern half of the province of Alberta, Canada. After initial treatment, patients are reviewed on an annual basis, or more frequently if necessary. When an annual follow-up appointment is missed, an attempt is made to obtain current information on the patient by corresponding with the patient and the treating physicians. When this fails, a search is made of provincial and national vital statistics records to determine if the patient has died and, if so, of what cause. Due to the intensive methods that are used to ensure follow-up of registrants, it is reasonable to assume that patients who are not known to have died are still alive. The cohort for this example was assembled by selecting a random sample of 199 female breast cancer patients who registered during 1985. Entry into the cohort was restricted to women with either stage I, II, or III disease, thereby excluding cases of disseminated cancer (stage IV). It has been well documented that breast cancer mortality increases as stage of disease becomes more advanced. Another predictor of survival from breast cancer is the amount of estrogen receptor that is present in breast tissue. Published reports show that patients with higher levels of estrogen receptor generally have a better prognosis. Receptor level is measured on a continuous scale, but for the present analysis this variable has been dichotomized into low and high levels using a conventional cutoff value. For this example the maximum length of follow-up was taken to be 5 years and the endpoint was defined to be death from breast cancer. Of the 199 subjects in the cohort, seven died of a cause other than breast cancer. These individuals were dropped from the analysis, leaving a cohort of 192 subjects. Summarily dropping subjects in this manner is methodologically incorrect, but for purposes of illustration this issue will be ignored. Methods for analyzing cohort data when there are losses to follow-up are presented in later chapters " [Newman 2001].

```
. use newman_rl_bc.dta

. lab list
sl:
           1 I
           2 II
           3 III
al:
           0 dead
           1 alive
rl:
           0 Low
           1 High

. table alive rlr stage

----------------------------------------------
          |        Stage and Receptor level
          | ---- I ---    --- II ---    --- III --
 Survival |  Low  High     Low  High     Low  High
----------+-----------------------------------
```

```
     dead |    2     5       9    17      12     9
    alive |   10    50      13    57       2     6
----------------------------------------------------

. logit dead rlevel##i.stage

Logistic regression                               Number of obs     =        192
                                                  LR chi2(5)        =      42.62
                                                  Prob > chi2       =     0.0000
Log likelihood = -92.764362                       Pseudo R2         =     0.1868

------------------------------------------------------------------------------
        dead |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    1.rlevel |   .6931472   .9055385     0.77   0.444    -1.081676     2.46797
             |
       stage |
          II |   1.092747   .5443964     2.01   0.045     .0257499    2.159744
         III |    2.70805   .7055337     3.84   0.000      1.32523    4.090871
             |
rlevel#stage |
       1#II  |    .148966   1.041346     0.14   0.886    -1.892035    2.189967
      1#III  |   .6931472   1.296577     0.53   0.593    -1.848096    3.234391
             |
       _cons |  -2.302585   .4690416    -4.91   0.000     -3.22189    -1.38328
------------------------------------------------------------------------------

. est stor A

. logit dead rlevel i.stage

Logistic regression                               Number of obs     =        192
                                                  LR chi2(3)        =      42.27
                                                  Prob > chi2       =     0.0000
Log likelihood = -92.939847                       Pseudo R2         =     0.1853

------------------------------------------------------------------------------
        dead |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      rlevel |   .9195114   .3954014     2.33   0.020     .1445389    1.694484
             |
       stage |
          II |   1.134549   .4657931     2.44   0.015     .2216116    2.047487
         III |   2.935924   .5856131     5.01   0.000     1.788143    4.083704
             |
       _cons |  -2.366915   .4200155    -5.64   0.000    -3.190131     -1.5437
------------------------------------------------------------------------------

. est stor B

. logit dead rlevel

Logistic regression                               Number of obs     =        192
                                                  LR chi2(1)        =      11.68
                                                  Prob > chi2       =     0.0006
Log likelihood = -108.23407                       Pseudo R2         =     0.0512

------------------------------------------------------------------------------
        dead |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      rlevel |   1.210019   .3529673     3.43   0.001     .5182159    1.901822
       _cons |  -1.293401   .2027501    -6.38   0.000    -1.690784   -.8960176
------------------------------------------------------------------------------

. est stor C
```

```
. lrtest A B

Likelihood-ratio test                                 LR chi2(2)  =      0.35
(Assumption: B nested in A)                           Prob > chi2 =    0.8391

. lrtest B C
Likelihood-ratio test                                 LR chi2(2)  =     30.59
(Assumption: C nested in B)                           Prob > chi2 =    0.0000
```

As there are some small cells, we will include some exact logistic models to check Model B.

Note that exlogistic does not accept factor variables.

```
. gen st2 = (stage ==2)

. gen st3 = (stage ==3)

. exlogistic dead rlevel, condvar(st2 st3) test(prob) coef nolog

Exact logistic regression
                                                      Number of obs =        192
------------------------------------------------------------------------------
        dead |      Coef.       Prob.    Pr<=Prob.     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      rlevel |   .9031172     .011566      0.0232      .0565615    1.744268
------------------------------------------------------------------------------

. exlogistic dead st2, condvar(rlevel st3) test(prob) coef nolog

Exact logistic regression
                                                      Number of obs =        192
------------------------------------------------------------------------------
        dead |      Coef.       Prob.    Pr<=Prob.     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         st2 |   1.120976    .0065697      0.0152       .163657    2.202553
------------------------------------------------------------------------------

. exlogistic dead st3, condvar(rlevel st2) test(prob) coef nolog

Exact logistic regression
                                                      Number of obs =        192
------------------------------------------------------------------------------
        dead |      Coef.       Prob.    Pr<=Prob.     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         st3 |   2.865383    7.29e-08      0.0000      1.645816    4.221508
------------------------------------------------------------------------------
```

Death and Tolbutamide

[adapted from Newman (2001, p 66)]

The University Group Diabetes Program (1970) was a study that was quite controversial when first published. The UGDP study was a randomized controlled trial comparing tolbutamide (a blood sugar-lowering drug) to placebo in patients with diabetes. Long-standing diabetes can cause cardiovascular complications, and this increases the risk of such potentially fatal conditions as myocardial infarction (heart attack), stroke, and renal failure. Tolbutamide helps to normalize blood sugar and would therefore be expected to reduce mortality in diabetic patients. Data from the UGDP study was stratified by age at enrollment, with death from all causes as the study endpoint.

This data is in newman_surv_tolb.dta

Newman suggests there is RD confounding. RR and OR results are less clear. No evidence of modification with any measure. See below.

```
> strat(death,drug,age)
             drug exposed not exposed
age death
0   cases               8           5
    controls           98         115
1   cases              22          16
    controls           76          69

age 0
            exposed  not exposed  odds est
cases     8.0000000   5.00000000 1.6000000
controls 98.0000000 115.00000000 0.8521739
risk est  0.0754717   0.04166667        NA

age 1
            exposed not exposed odds est
cases    22.0000000  16.0000000 1.375000
controls 76.0000000  69.0000000 1.101449
risk est  0.2244898   0.1882353       NA


Stratified analysis for OR by  age
            OR est     lower    upper    p-value
age 0        1.8723   0.52037   7.5212     0.39192
age 1        1.2468   0.57233   2.7647     0.58737
crude        1.5091   0.80138   2.8873     0.18126
adjusted     1.4031   0.76252   2.5820     0.27505
OR Homogeneity test, chi-squared 1 df = 0.34 , p-value = 0.5570604

Stratified analysis for RR by  age
            RR est     lower    upper    p-value
age 0        1.8113   0.61120   5.3679     0.39192
age 1        1.1926   0.67127   2.1188     0.58737
crude        1.4356   0.85102   2.4216     0.18126
adjusted     1.3256   0.79791   2.2021     0.27505
RR Homogeneity test, chi-squared 1 df = 0.45 , p-value = 0.5036593

Stratified analysis for RD by  age
            RD est     lower    upper    p-value
age 0      0.033805 -0.027895 0.095505     0.39192
age 1      0.036255 -0.080920 0.153429     0.58737
crude      0.044620 -0.019294 0.108533     0.18126
adjusted   0.034900 -0.027628 0.097428     0.27505
RD Homogeneity test, chi-squared 1 df = 0 , p-value = 0.966885
```

Myocardial Infarction and OC
[adapted from Newman]

```
. use newman_MI_OC.dta

. table nmi noc strata

-------------------------------------------------------------------------------
          |                           strata and OC
          | --- 1 --- --- 2 --- --- 3 --- --- 4 --- --- 5 --- --- 6 --- --- 7 ---
       MI |    Y    N    Y    N    Y    N    Y    N    Y    N    Y    N    Y    N
----------+--------------------------------------------------------------------
     Case |         1    2    5   11    8    1   13    1   32    8   53    3   20
  Control |   38  281   35  221   22  112   12  318   15  249    8  125    2  155
-------------------------------------------------------------------------------

------------------------------
          |    strata and OC
          | --- 8 --- --- 9 ---
       MI |    Y    N    Y    N
----------+-------------------
     Case |        42    3   31
  Control |    1   96    2   50
------------------------------

. tab strata,gen(st)

     strata |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |        320       16.19       16.19
          2 |        263       13.31       29.50
          3 |        153        7.74       37.25
          4 |        344       17.41       54.66
          5 |        297       15.03       69.69
          6 |        194        9.82       79.50
          7 |        180        9.11       88.61
          8 |        139        7.03       95.65
          9 |         86        4.35      100.00
------------+-----------------------------------
      Total |      1,976      100.00

. exlogistic mi oc,condvar(st2 st3 st4 st5 st6 st7 st8 st9) nolog coef test(prob)

Exact logistic regression
                                                Number of obs =       1976
-------------------------------------------------------------------------------
          mi |      Coef.       Prob.    Pr<=Prob.     [95% Conf. Interval]
-------------+-----------------------------------------------------------------
          oc |   1.029089    .0000659      0.0001      .4841107    1.559759
-------------------------------------------------------------------------------

. logit mi oc i.strata

Logistic regression                             Number of obs     =      1,976
                                                LR chi2(9)        =     270.32
                                                Prob > chi2       =     0.0000
Log likelihood =  -583.6447                     Pseudo R2         =     0.1880

------------------------------------------------------------------------------
          mi |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          oc |   1.035944   .2584133     4.01   0.000     .5294633    1.542425
             |
      strata |
          2  |   2.139551   1.073126     1.99   0.046     .0362628    4.242839
          3  |    3.71001   1.032558     3.59   0.000     1.686234    5.733786
```

```
          4  |   2.737486   1.039655     2.63   0.008     .6997996    4.775172
          5  |   3.800612   1.019936     3.73   0.000     1.801575     5.79965
          6  |   5.084603    1.01507     5.01   0.000     3.095103    7.074103
          7  |   3.998539   1.028263     3.89   0.000     1.983181    6.013896
          8  |   5.114426   1.021144     5.01   0.000      3.11302    7.115832
          9  |   5.473006   1.027787     5.33   0.000      3.45858    7.487432
             |
       _cons |  -5.959912   1.004416    -5.93   0.000    -7.928531   -3.991294
------------------------------------------------------------------------------
```

Two 2x2 Tables : 'Hypothetical' Examples From Newman Chapter Two

2.2(a) is ct1 .... 2.2(f) is ct7
There are illustrations of various forms of OR/RR modification and OR/RR confounding.
strat could be used to show RD forms

```
. use newman_2_2.dta

. cc D E [fw=ct1], by(F)

                F |       OR       [95% Conf. Interval]   M-H Weight
------------------+--------------------------------------------------
                0 |        3.5      1.872804   6.567155            6 (exact)
                1 |        3.5      2.266193     5.4132           12 (exact)
------------------+--------------------------------------------------
            Crude |        3.5      2.461396   4.980692              (exact)
    M-H combined |        3.5      2.495869    4.90811
-------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(1) =     0.00  Pr>chi2 = 1.0000

                   Test that combined OR = 1:
                               Mantel-Haenszel chi2(1) =     54.36
                                                Pr>chi2 =    0.0000

. cs D E [fw=ct1], by(F)

                F |       RR       [95% Conf. Interval]   M-H Weight
------------------+--------------------------------------------------
                0 |        1.75     1.332997   2.297455           20
                1 |        1.75     1.443616   2.121409           40
------------------+--------------------------------------------------
            Crude |        1.75     1.495513   2.047793
    M-H combined |        1.75     1.495513   2.047793
-------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(1) =    0.000  Pr>chi2 = 1.0000

. cc D E [fw=ct2], by(F)

                F |       OR       [95% Conf. Interval]   M-H Weight
------------------+--------------------------------------------------
                0 |        3.5      1.872804   6.567155            6 (exact)
                1 |          6      3.751301   9.643365            8 (exact)
------------------+--------------------------------------------------
            Crude |   4.928571      3.410608   7.135158              (exact)
    M-H combined |   4.928571      3.459339   7.021808
-------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(1) =     2.06  Pr>chi2 = 0.1517

                   Test that combined OR = 1:
                               Mantel-Haenszel chi2(1) =     82.88
                                                Pr>chi2 =    0.0000

. cs D E [fw=ct2], by(F)
```

```
                F |       RR       [95% Conf. Interval]   M-H Weight
------------------+------------------------------------------------
                0 |        1.75     1.332997   2.297455           20
                1 |           2     1.664974    2.40244           40
------------------+------------------------------------------------
            Crude |    1.916667     1.646392    2.23131
    M-H combined |    1.916667     1.646392    2.23131
-------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(1) =    0.636  Pr>chi2 = 0.4251

. cc D E [fw=ct3], by(F)

                F |       OR       [95% Conf. Interval]   M-H Weight
------------------+------------------------------------------------
                0 |        3.5      2.036928   6.062955            8 (exact)
                1 |          6      3.417617   10.54801     5.333333 (exact)
------------------+------------------------------------------------
            Crude |   4.928571      3.410608   7.135158              (exact)
    M-H combined |        4.5      3.113227   6.504505
-------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(1) =     2.06  Pr>chi2 = 0.1514

                   Test that combined OR = 1:
                               Mantel-Haenszel chi2(1) =     68.95
                                                Pr>chi2 =    0.0000

. cs D E [fw=ct3], by(F)

                F |       RR       [95% Conf. Interval]   M-H Weight
------------------+------------------------------------------------
                0 |        1.75     1.414588   2.164942     26.66667
                1 |           2     1.557839   2.567659     26.66667
------------------+------------------------------------------------
            Crude |    1.916667     1.646392    2.23131
    M-H combined |       1.875     1.589871   2.211264
-------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(1) =    0.660  Pr>chi2 = 0.4165

. cc D E [fw=ct4], by(F)

                F |       OR       [95% Conf. Interval]   M-H Weight
------------------+------------------------------------------------
                0 |          6      2.671724   14.38874            3 (exact)
                1 |          6      3.410538   10.86564            6 (exact)
------------------+------------------------------------------------
            Crude |   3.596154      2.516361   5.145841              (exact)
    M-H combined |          6      3.855339   9.337701
-------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(1) =     0.00  Pr>chi2 = 1.0000

                   Test that combined OR = 1:
                               Mantel-Haenszel chi2(1) =     71.76
                                                Pr>chi2 =    0.0000

. cs D E [fw=ct4], by(F)

                F |       RR       [95% Conf. Interval]   M-H Weight
------------------+------------------------------------------------
                0 |         1.5     1.261892   1.783036           30
                1 |           4      2.55285   6.267506           10
------------------+------------------------------------------------
            Crude |       2.125     1.718801   2.627194
    M-H combined |       2.125     1.759657   2.566197
-------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(1) =   23.219  Pr>chi2 = 0.0000
```

```
. cc D E [fw=ct5], by(F)

               F |       OR       [95% Conf. Interval]   M-H Weight
-----------------+-------------------------------------------------
               0 |          6      2.874738   13.65971            4 (exact)
               1 |   1.588235      .7145961   3.807349     5.666667 (exact)
-----------------+-------------------------------------------------
           Crude |   .8717949      .6216959   1.222378              (exact)
    M-H combined |   3.413793      2.053635   5.674807
-------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(1) =     6.31  Pr>chi2 = 0.0120

                   Test that combined OR = 1:
                               Mantel-Haenszel chi2(1) =     25.00
                                                Pr>chi2 =    0.0000

. cs D E [fw=ct5], by(F)

               F |       RR       [95% Conf. Interval]   M-H Weight
-----------------+-------------------------------------------------
               0 |         1.5     1.316269   1.709378           40
               1 |         1.5     .7643306   2.943752     6.666667
-----------------+-------------------------------------------------
           Crude |    .9230769     .7636443   1.115796
    M-H combined |         1.5     1.294012   1.738778
-------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(1) =    0.000  Pr>chi2 = 1.0000

. cc D E [fw=ct6], by(F)

               F |       OR       [95% Conf. Interval]   M-H Weight
-----------------+-------------------------------------------------
               0 |   2.333333      1.378717   3.943053           10 (exact)
               1 |          6       3.51979   10.59745          7.2 (exact)
-----------------+-------------------------------------------------
           Crude |   3.559524      2.557155   4.976222              (exact)
    M-H combined |   3.868217      2.730159    5.48067
-------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(1) =     6.78  Pr>chi2 = 0.0092

                   Test that combined OR = 1:
                               Mantel-Haenszel chi2(1) =     61.19
                                                Pr>chi2 =    0.0000

. cs D E [fw=ct6], by(F)

               F |       RR       [95% Conf. Interval]   M-H Weight
-----------------+-------------------------------------------------
               0 |         1.4     1.128055   1.737504     33.33333
               1 |           4     2.580629   6.200039           12
-----------------+-------------------------------------------------
           Crude |    2.228571     1.785376   2.781785
    M-H combined |    2.088235      1.69844   2.567489
-------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(1) =   21.616  Pr>chi2 = 0.0000

. cc D E [fw=ct7], by(F)

               F |       OR       [95% Conf. Interval]   M-H Weight
-----------------+-------------------------------------------------
               0 |   2.333333      1.378717   3.943053           10 (exact)
               1 |          6       3.51979   10.59745          7.2 (exact)
               2 |   5.444444      3.237062   9.237116         6.75 (exact)
-----------------+-------------------------------------------------
           Crude |   3.361111      2.619619   4.314823              (exact)
    M-H combined |   4.312457       3.24303   5.734539
```

```
------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(2) =     8.34  Pr>chi2 = 0.0155

                  Test that combined OR = 1:
                              Mantel-Haenszel chi2(1) =    108.85
                                                Pr>chi2 =    0.0000

. cs D E [fw=ct7], by(F)

                F |       RR       [95% Conf. Interval]   M-H Weight
------------------+--------------------------------------------------
                0 |         1.4     1.128055   1.737504     33.33333
                1 |           4     2.580629   6.200039           12
                2 |    2.333333     1.881421   2.893794         22.5
------------------+--------------------------------------------------
           Crude |      2.0625     1.773177    2.39903
    M-H combined |    2.169533     1.859165   2.531714
------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(2) =   23.726  Pr>chi2 = 0.0000
```