Models In Epidemiology And Biostatistics
Gordon Hilton Fick

OCCS Example

Breslow & Day 2: The Oxford Childhood Cancer Survey (Breslow & Day, 1980, p. 238; Kneale, 1971; Steward & Kneale, 1970). Data are from a case-control study of childhood leukemia and lymphoid tumors and in utero X-ray exposure (Kneale et al., 1971). [bd2_occs.dta]

[adapted from B&D p239]
"Cases were ascertained as all children under ten years of age in England and Wales who died of cancer (leukaemia or solid tumours) during the period 1954-65. For each of these a neighbourhood control of the same age was selected who was alive and well at the time the case died. Only "traced" pairs, for whom both case and control mothers could be found and interviewed, were analysed. The published data ignore the exact pairing but do preserve the stratification by age and year of birth. Exposure in this example is simply a question of whether or not the study subjects received in utero irradiation, as reported by the mother. The stratification variables were age at death, from 0 to 9 years, and year of birth, from 1944 to 1964. Because of the limited period of case ascertainment, not all 210 possible combinations of these factors appear. For example, among childhood cancer patients born in 1944, only those who died at age 9 are represented. A total of 120 such strata were available. In order to estimate the overall [odds ratio] of obstetric radiation, and to determine whether, and if so how, it varied with age and year, B&D fitted [several models]. Different subsets of these were entered into the regression equation so as to detect particular kinds of trends and patterns in the [odds ratio].

...absolutely no improvements in fit accompanied the addition to the model of either linear or quadratic terms in age. The [model with the quadratic term] would be expected to be particularly sensitive to a peak in [odds ratio] as a function of age. The lack of evidence for any such peak argues against the hypothesis that the age distributions for radiogenic and idiopathic cancers are different. Improvements in radiological technology probably account for the declining effect with year of birth (Bithel & Stewart, 1975)
...there is reasonably strong evidence for a decrease in [odds ratio] with year of birth. Additional improvement in fit occurs when a quadratic term in year is added to the model, which would indicate a degree of curvature in the regression []."

A 'classic' stratified analysis with year of birth as a potential modifier/confounder reveals modification and gives an indication that the modification is nonlinear. There is a zero cell in the 2x2 table for year 1944.

```
. cc case xray,by(year)

             year |       OR       [95% Conf. Interval]   M-H Weight
------------------+-------------------------------------------------
               44 |          .      .8157625          .            0 (exact)
               45 |    1.86413      .4370265   9.220175     1.735849 (exact)
               46 |   6.086957      1.655264   33.37558     1.277778 (exact)
               47 |   5.403974       1.94783   18.51057     2.157143 (exact)
               48 |   1.837037      .9587835   3.603219     7.765487 (exact)
               49 |   2.298573      1.291701   4.189765     8.855872 (exact)
               50 |   1.815077      1.114283   2.990642     13.49565 (exact)
               51 |   2.208451      1.426582   3.459539     15.30886 (exact)
               52 |   1.608289       1.09168   2.380225     22.19341 (exact)
               53 |   1.139128      .7872242    1.65031     28.75047 (exact)
               54 |   1.241791      .8663844   1.782844     28.95062 (exact)
               55 |   2.180089      1.500615   3.187992     20.76115 (exact)
```

```
              56 |   1.654557      1.154303   2.380164     25.20788 (exact)
              57 |     1.7875      1.150412   2.798637     16.50794 (exact)
              58 |   1.283668      .8268379   1.999771     19.38889 (exact)
              59 |   1.167917      .6518969   2.104466     11.91061 (exact)
              60 |   1.261538      .7046335   2.275107     11.47059 (exact)
              61 |   1.187135      .5792494   2.453088     8.015625 (exact)
              62 |   2.580709       1.03145   7.041538     3.479452 (exact)
              63 |   .8901099       .304439   2.567848         4.55 (exact)
              64 |       8.68      .9789227   403.4644      .390625 (exact)
-----------------+---------------------------------------------------
           Crude |   1.650475      1.477037    1.84433              (exact)
    M-H combined |    1.65431      1.482312   1.846267
-------------------------------------------------------------------
Test of homogeneity (Tarone)   chi2(20) =    35.85  Pr>chi2 = 0.0160

                  Test that combined OR = 1:
                              Mantel-Haenszel chi2(1) =     81.99
                                              Pr>chi2 =    0.0000
. gen cy = case*year

. gen y2 = year*year

. gen cy2 = case*y2

. logit xray i.case##i.year

note: 0.case#44.year != 0 predicts failure perfectly
      0.case#44.year dropped and 28 obs not used

note: 1.case#64.year omitted because of collinearity

Logistic regression                               Number of obs   =      11824
                                                  LR chi2(40)     =     199.09
                                                  Prob > chi2     =     0.0000
Log likelihood = -4463.1587                       Pseudo R2       =     0.0218

------------------------------------------------------------------------------
        xray |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      1.case |   2.161012   1.102318     1.96   0.050     .0005075    4.321516
             |
        year |
          45 |    1.77575   1.363396     1.30   0.193    -.8964566    4.447956
          46 |   .7259616   1.389707     0.52   0.601    -1.997814    3.449737
          47 |   .7549406   1.339521     0.56   0.573    -1.870472    3.380354
          48 |   1.834109   1.284057     1.43   0.153    -.6825964    4.350815
          49 |   1.765116   1.280588     1.38   0.168    -.7447892    4.275022
          50 |   2.000808   1.273925     1.57   0.116    -.4960383    4.497655
          51 |   1.920932   1.272013     1.51   0.131    -.5721683    4.414031
          52 |   2.276298   1.268642     1.79   0.073    -.2101938     4.76279
          53 |   2.350403   1.267095     1.85   0.064    -.1330569    4.833863
          54 |   2.499085   1.266937     1.97   0.049      .015935    4.982236
          55 |   2.216086   1.268739     1.75   0.081    -.2705975    4.702769
          56 |   2.466349   1.267521     1.95   0.052     -.017946    4.950644
          57 |   2.147109   1.271376     1.69   0.091    -.3447428    4.638961
          58 |   2.201834   1.270212     1.73   0.083    -.2877364    4.691404
          59 |   1.734237    1.27668     1.36   0.174    -.7680106    4.236485
          60 |    2.03404   1.277093     1.59   0.111    -.4690153    4.537096
          61 |   2.012592   1.284423     1.57   0.117    -.5048305    4.530014
          62 |   1.433463   1.311749     1.09   0.274    -1.137518    4.004444
          63 |   2.084051   1.303668     1.60   0.110    -.4710915    4.639193
          64 |   .8472979   .7457818     1.14   0.256    -.6144076    2.309003
             |
   case#year |
        0 44 |  (empty)
        1 45 |  -1.538217   1.284566    -1.20   0.231    -4.055921    .9794863
```

```
       1 46  |  -.3548979   1.277232    -0.28   0.781    -2.858227    2.148431
       1 47  |  -.4739031   1.212136    -0.39   0.696    -2.849645    1.901839
       1 48  |  -1.552858   1.145799    -1.36   0.175    -3.798583    .6928669
       1 49  |  -1.328723   1.137329    -1.17   0.243    -3.557848    .9004011
       1 50  |  -1.564884   1.127611    -1.39   0.165     -3.77496    .6451924
       1 51  |  -1.368721   1.122949    -1.22   0.223     -3.56966    .8322188
       1 52  |  -1.685841   1.118543    -1.51   0.132    -3.878145    .5064621
       1 53  |  -2.030749   1.117019    -1.82   0.069    -4.220066    .1585685
       1 54  |  -1.944457   1.116328    -1.74   0.082     -4.13242    .2435051
       1 55  |  -1.381646   1.117533    -1.24   0.216    -3.571971    .8086783
       1 56  |  -1.657478   1.116406    -1.48   0.138    -3.845594    .5306372
       1 57  |  -1.580194    1.12313    -1.41   0.159    -3.781488    .6211005
       1 58  |  -1.911291   1.122866    -1.70   0.089    -4.112068    .2894866
       1 59  |   -2.00579   1.137085    -1.76   0.078    -4.234435    .2228553
       1 60  |   -1.92868   1.137123    -1.70   0.090    -4.157401    .3000406
       1 61  |   -1.98947   1.153202    -1.73   0.084    -4.249705    .2707655
       1 62  |  -1.212948   1.186532    -1.02   0.307    -3.538508    1.112612
       1 63  |  -2.277422    1.20346    -1.89   0.058    -4.636162    .0813169
       1 64  |  (omitted)
             |
       _cons |  -4.281275   1.260333    -3.40   0.001    -6.751483   -1.811068
------------------------------------------------------------------------------
```

Notice that because of the zero cell, Stata flags year 1944 but then drops a coefficient with 1964. This makes interpretation awkward.

Reverse coding of year of birth [gen y65= 65-year] attempts to resolve the awkward deletion.

```
. logit xray case##y65

note: 0.case#21.y65 != 0 predicts failure perfectly
      0.case#21.y65 dropped and 28 obs not used

note: 1.case#21.y65 omitted because of collinearity

Logistic regression                               Number of obs     =     11,824
                                                  LR chi2(40)       =     199.09
                                                  Prob > chi2       =     0.0000
Log likelihood = -4463.1587                       Pseudo R2         =     0.0218

------------------------------------------------------------------------------
        xray |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      1.case |   2.161012   1.102318     1.96   0.050     .0005075    4.321516
             |
        y65 |
          2  |   1.236753    1.06928     1.16   0.247    -.8589974    3.332504
          3  |   .5861655   1.079118     0.54   0.587    -1.528866    2.701197
          4  |   1.165294    1.04573     1.11   0.265    -.8842993    3.214887
          5  |   1.186743   1.036714     1.14   0.252     -.845179    3.218664
          6  |   .8869392   1.036206     0.86   0.392    -1.143987    2.917865
          7  |   1.354536   1.028226     1.32   0.188    -.6607499    3.369822
          8  |   1.299811   1.029664     1.26   0.207    -.7182926    3.317915
          9  |   1.619051   1.024899     1.58   0.114    -.3897144    3.627817
         10  |   1.368788   1.026406     1.33   0.182    -.6429305    3.380507
         11  |   1.651788   1.024177     1.61   0.107    -.3555623    3.659137
         12  |   1.503105   1.024372     1.47   0.142    -.5046275    3.510838
         13  |      1.429   1.026285     1.39   0.164    -.5824816    3.440482
         14  |   1.073634    1.03045     1.04   0.297    -.9460108    3.093278
         15  |    1.15351   1.032809     1.12   0.264    -.8707574    3.177778
         16  |   .9178184   1.041016     0.88   0.378    -1.122535    2.958172
         17  |   .9868113   1.045281     0.94   0.345    -1.061902    3.035524
         18  |  -.0923572   1.112711    -0.08   0.934    -2.273231    2.088517
         19  |  -.1213363   1.172644    -0.10   0.918    -2.419677    2.177005
         20  |   .9284517    1.14134     0.81   0.416    -1.308534    3.165437
         21  |  -.8472979   .7457818    -1.14   0.256    -2.309003    .6144076
             |
```

```
    case#y65 |
       0 21  |          0  (empty)
       1  2  |  -2.277422    1.20346    -1.89   0.058    -4.636162    .0813169
       1  3  |  -1.212948   1.186532    -1.02   0.307    -3.538508    1.112612
       1  4  |   -1.98947   1.153202    -1.73   0.084    -4.249705    .2707655
       1  5  |   -1.92868   1.137123    -1.70   0.090    -4.157401    .3000406
       1  6  |   -2.00579   1.137085    -1.76   0.078    -4.234435    .2228553
       1  7  |  -1.911291   1.122866    -1.70   0.089    -4.112068    .2894866
       1  8  |  -1.580194    1.12313    -1.41   0.159    -3.781488    .6211005
       1  9  |  -1.657478   1.116406    -1.48   0.138    -3.845594    .5306372
       1 10  |  -1.381646   1.117533    -1.24   0.216    -3.571971    .8086783
       1 11  |  -1.944457   1.116328    -1.74   0.082     -4.13242    .2435051
       1 12  |  -2.030749   1.117019    -1.82   0.069    -4.220066    .1585685
       1 13  |  -1.685841   1.118543    -1.51   0.132    -3.878145    .5064621
       1 14  |  -1.368721   1.122949    -1.22   0.223     -3.56966    .8322188
       1 15  |  -1.564884   1.127611    -1.39   0.165     -3.77496    .6451924
       1 16  |  -1.328723   1.137329    -1.17   0.243    -3.557848    .9004011
       1 17  |  -1.552858   1.145799    -1.36   0.175    -3.798583    .6928669
       1 18  |  -.4739031   1.212136    -0.39   0.696    -2.849645    1.901839
       1 19  |  -.3548979   1.277232    -0.28   0.781    -2.858227    2.148431
       1 20  |  -1.538217   1.284566    -1.20   0.231    -4.055921    .9794863
       1 21  |          0  (omitted)
             |
       _cons |  -3.433978   1.015996    -3.38   0.001    -5.425294   -1.442661
------------------------------------------------------------------------------
```

With this coding, both terms involving y65=21 [year=44] are dropped. A more natural circumstance.

The 2x2 table with the zero cell can be studied using a 'classic' approach and an 'exact' approach.

```
. cc xray case if year==44,exact

                 | case                    |             Proportion
                 |   Exposed   Unexposed   |      Total     Exposed
-----------------+-------------------------+-------------------------
           Cases |         3           0   |          3       1.0000
        Controls |        25          28   |         53       0.4717
-----------------+-------------------------+-------------------------
           Total |        28          28   |         56       0.5000
                 |                         |
                 |      Point estimate     |    [95% Conf. Interval]
                 |-------------------------+-------------------------
      Odds ratio |                .        |    .8157625           . (Cornfield)
 Attr. frac. ex. |                .        |   -.2258469           . (Cornfield)
 Attr. frac. pop |                .        |
                 +-------------------------------------------------
                                  1-sided Fisher's exact P = 0.1182
                                  2-sided Fisher's exact P = 0.2364

 Note: Exact confidence levels not possible with zero count cells.

. exlogistic xray case if year==44,test(prob) nolog
note: CMLE estimate for case is +inf; computing MUE

Exact logistic regression                       Number of obs =         56
                                                Model prob.   =   .1181818
                                                Pr <= prob.   =     0.2364
------------------------------------------------------------------------------
        xray | Odds Ratio       Prob.    Pr<=Prob.     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        case |   4.082205*   .1181818      0.2364      .4225252       +Inf
------------------------------------------------------------------------------
(*) median unbiased estimates (MUE)
```

For a discussion of median unbiased estimation, see [for example] the paper: "An Estimate of the Odds

Ratio That Always Exists" by Michael Parzen , Stuart Lipsitz , Joseph Ibrahim , and Neil Klar [2002] Journal of Computational and Graphical Statistics, Volume 11, Number 2, Pages 420–436

```
. logit xray case year y2 cy cy2

Logistic regression                                     Number of obs   =      11852
                                                        LR chi2(5)      =     147.55
                                                        Prob > chi2     =     0.0000
Log likelihood = -4492.8123                             Pseudo R2       =     0.0162

------------------------------------------------------------------------------
        xray |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        case |   23.70231   9.102694     2.60   0.009     5.861357    41.54326
        year |   1.622869   .2713843     5.98   0.000     1.090966    2.154773
          y2 |  -.0147354   .0024866    -5.93   0.000    -.0196091   -.0098617
          cy |  -.8189415    .335467    -2.44   0.015    -1.476445   -.1614382
         cy2 |   .0071771   .0030814     2.33   0.020     .0011377    .0132166
       _cons |  -46.61945   7.383641    -6.31   0.000    -61.09112   -32.14778
------------------------------------------------------------------------------

. logit xray case year cy

Logistic regression                                     Number of obs   =      11852

Log likelihood = -4522.5276                             Pseudo R2       =     0.0096

------------------------------------------------------------------------------
        xray |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        case |    2.22213   .7077249     3.14   0.002     .8350145    3.609245
        year |   .0173936   .0099927     1.74   0.082    -.0021917     .036979
          cy |  -.0316978   .0129821    -2.44   0.015    -.0571422   -.0062534
       _cons |  -3.125853   .5465336    -5.72   0.000    -4.197039   -2.054667
------------------------------------------------------------------------------
```