JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Multiple Logistic Regression

John McGready
Johns Hopkins University

JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Section A

Multiple Logistic Regression

In the previous sections of this lecture, we observed a positive association between CHD and positive association between CHD and smoking

Continued

What if smoking is also associated with age?

Age could be a confounder of the smoking-CHD relationship (and vice-versa)

Continued

Can we estimate the age adjusted relationship between CHD and smoking?

Continued

Multiple logistic regression allows us to have more than one predictor in our model

We can also estimate the association between each predictor and $Pr(y = 1)$ controlling for all other predictors

Continued

Here, we need a logistic regression model with two predictors:

$$\log\left(\frac{p}{1-p}\right) = \hat{b}_0 + \hat{b}_1 x_1 + \hat{b}_2 x_2$$

p = Pr(CHD), x_1 = age, x_2 = smoke

Continued

How would we interpret the coefficients from a multiple logistic regression? And the resulting odds ratio estimated?

$$\log\left(\frac{p}{1-p}\right) = \hat{b}_0 + \hat{b}_1 x_1 + \hat{b}_2 x_2$$

p = Pr(CHD), x_1 = age, x_2 = smoke

Continued

x_1 is the age variable

$\hat{b}_1$ is the estimated regression coefficient associated with age

$\hat{b}_1$ estimates the log(OR) for comparing two individuals (groups) who differ by one year in age and are either both smokers or non-smokers

Continued

Write out two equations

- ★ *Obs 1: smoker, k years-old*
- ★ *Obs 2: smoker, k+1 years-old*

Write out two equations

$$log(odds\ 2) = \hat{b}_0 + \hat{b}_1(k+1) + \hat{b}_2$$

$$log(odds\ 1) = \hat{b}_0 + \hat{b}_1 k + \hat{b}_2$$

Continued

Simplify

$$\textit{log(odds 2)} = \hat{b}_0 + \hat{b}_1 k + \hat{b}_1 + \hat{b}_2$$

$$\textit{log(odds 1)} = \hat{b}_0 + \hat{b}_1 k + \hat{b}_2$$

Continued

Subtract

$$log(odds\ 2) = \hat{b}_0 + \hat{b}_1 k + \hat{b}_1 + \hat{b}_2$$

$$log(odds\ 1) = \hat{b}_0 + \hat{b}_1 k + \hat{b}_2$$

Continued

Subtract

$$log(odds\ 2) = \hat{b}_0 + \hat{b}_1 k + \hat{b}_1 + \hat{b}_2$$

$$log(odds\ 1) = \hat{b}_0 + \hat{b}_1 k + \hat{b}_2$$

$$\hat{b}_1$$

Continued

So, $\hat{b}_1$= log(odds of CHD for Obs #2) – log(odds of CHD for Obs #1)

By property of logs:

- $\hat{b}_1 = log \left(\frac{odds\_of\_CHD\_Obs\#2}{odds\_of\_CHD\_Obs\#1} \right)$

Continued

x_1 is the age variable

$e^{\hat{b}_1}$ is the estimated adjusted OR of CHD associated with age, after adjusting for smoking status

- *This $\hat{O}R$ compares two individuals (groups) of the same smoking status where one individual (group) is one year older than the comparison group*

Continued

x_2 is the smoking variable

$\hat{b}_2$ is the estimated regression coefficient associated with smoking

$\hat{b}_2$ estimates the log(OR) for comparing two individuals (groups) of the same age, where one is a smoker and the other is a non-smoker

Continued

x_2 is the smoking variable

$e^{\hat{b}_2}$ is the estimated adjusted OR of CHD associated with smoking, after adjusting for age

- ★ *This $\hat{O}R$ compares two individuals (groups) of the same age where one individual (group) is a smoker and the other is a non-smoker (reference)*

Inference in Multiple Logistic Regression

We can calculate CIs and p-values for each coefficient and hence for each adjusted OR

Each coefficient estimate has its own associated standard error

Stata results

```
. logit  chd age smoke

Iteration 0:   log likelihood = -39.649049
Iteration 5:   log likelihood = -23.931297

Logit estimates                                   Number of obs   =         58
                                                  LR chi2(2)      =      31.44
                                                  Prob > chi2     =     0.0000
Log likelihood = -23.931297                       Pseudo R2       =     0.3964

------------------------------------------------------------------------------
         chd |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .1599163   .0419051     3.816   0.000       .0777838    .2420489
       smoke |   2.396168   .8591632     2.789   0.005       .7122387    4.080096
       _cons |  -8.567143   2.171859    -3.945   0.000      -12.82391   -4.310378
------------------------------------------------------------------------------

.
```

P-Value for Age

P-value for testing:

- $H_o: b_1 = 0$
- $H_a: b_1 \neq 0$

Answering question:

- *After adjusting for smoking status, is there a CHD/age relationship in population?*

```
. logit  chd age smoke

Iteration 0:   log likelihood = -39.649049
Iteration 5:   log likelihood = -23.931297

Logit estimates                                   Number of obs   =         58
                                                  LR chi2(2)      =      31.44
                                                  Prob > chi2     =     0.0000
Log likelihood = -23.931297                       Pseudo R2       =     0.3964

------------------------------------------------------------------------------
     chd |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     age |   .1599163   .0419051      3.816   0.000       .0777838    .2420489
   smoke |   2.396168   .8591632      2.789   0.005       .7122387    4.080096
   _cons |  -8.567143   2.171859     -3.945   0.000      -12.82391   -4.310378
------------------------------------------------------------------------------

.
```

Continued

```
. logit  chd age smoke

Iteration 0:   log likelihood = -39.649049
Iteration 5:   log likelihood = -23.931297

Logit estimates                                   Number of obs   =         58
                                                  LR chi2(2)      =      31.44
                                                  Prob > chi2     =     0.0000
Log likelihood = -23.931297                       Pseudo R2       =     0.3964

------------------------------------------------------------------------------
         chd |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
         age |   .1599163   .0419051      3.816   0.000        .0777838    .2420489
       smoke |   2.396168   .8591632      2.789   0.005        .7122387    4.080096
       _cons |  -8.567143   2.171859     -3.945   0.000       -12.82391   -4.310378
------------------------------------------------------------------------------

.
```

Continued

```
. logit  chd age smoke

Iteration 0:   log likelihood = -39.649049
Iteration 5:   log likelihood = -23.931297

Logit estimates                                   Number of obs   =         58
                                                  LR chi2(2)      =      31.44
                                                  Prob > chi2     =     0.0000
Log likelihood = -23.931297                       Pseudo R2       =     0.3964

------------------------------------------------------------------------------
     chd |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     age |   .1599163   .0419051      3.816   0.000       .0777838    .2420489
   smoke |   2.396168   .8591632      2.789   0.005       .7122387    4.080096
   _cons |  -8.567143   2.171859     -3.945   0.000      -12.82391   -4.310378
------------------------------------------------------------------------------

.
```

P-value for testing:

- H_o: $b_2 = 0$
- H_a: $b_2 \neq 0$

Answering question:

- *After adjusting for age, is there a CHD/smoking relationship in population?*

```
. logistic  chd age smoke

Logit estimates                                   Number of obs   =         58
                                                  LR chi2(2)      =      31.44
                                                  Prob > chi2     =     0.0000
Log likelihood = -23.931297                       Pseudo R2       =     0.3964

------------------------------------------------------------------------------
     chd | Odds Ratio   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     age |   1.173413    .049172      3.816   0.000       1.080889    1.273856
   smoke |   10.98101   9.434481      2.789   0.005        2.03855    59.15117
------------------------------------------------------------------------------
```

Continued

```
. logistic  chd age smoke

Logit estimates                                   Number of obs   =         58
                                                  LR chi2(2)      =      31.44
                                                  Prob > chi2     =     0.0000
Log likelihood = -23.931297                       Pseudo R2       =     0.3964

------------------------------------------------------------------------------
     chd | Odds Ratio   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     age |   1.173413    .049172      3.816   0.000       1.080889    1.273856
   smoke |   10.98101   9.434481      2.789   0.005        2.03855    59.15117
------------------------------------------------------------------------------
```

In a sample of 58 individuals, a multiple logistic regression was performed to estimate the relationship between CHD evidence and an individuals' age and smoking status

Both age ($p < .001$) and smoking ($p = .005$) were found to be significant predictors of CHD

Continued

The adjusted OR associated with a one year increase in age was 1.17 (95% CI 1.08 to 1.27)

The adjusted OR associated with smoking was 11.0 (95% CI 2.0–59.2)

Age a Confounder

The adjusted OR associated with smoking was 11.0

When we estimated the OR between CHD and smoking, without adjusting for age, our estimate was 4.7

Coefficients, from a multiple logistic regression, estimate the magnitude of each predictor/outcome relationship after adjusting for all other predictors in the model

The estimates can be exponentiated to get associated adjusted odds ratios

JOHNS HOPKINS
BLOOMBERG
SCHOOL of PUBLIC HEALTH

Section A

Practice Problems

1. On the following slide, you will find regression output which relates the probability that a mother breast-feeds her child to the age of the child (in months) and the child's sex. The results come from a sub-sample of mother-child pairs in the Nepal Nutrition Intervention Study.

Continued

```
. logit

Logit estimates                                   Number of obs   =        472
                                                  LR chi2(2)      =     375.97
                                                  Prob > chi2     =     0.0000
Log likelihood = -132.02134                       Pseudo R2       =     0.5874

------------------------------------------------------------------------------
      bf |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
sex_chld |  -.3110831   .3158853
agechldc |  -.1867791   .0176001
   _cons |  -.9197549   .2251506
------------------------------------------------------------------------------
```

That Darn Cat!!!!!!!!!!!

Here sex_chld is the sex of the child (1= female, 0 =male), and agechldc is the age of the child in months.

Continued

a. Write out the logistic regression equation described by this output.

b. What is the interpretation of the coefficient of sex_chld?

c. Give a 95% CI for the age adjusted odds ratio for the odds of being breast-fed for female children as compared to male children.

Continued

d. Is child's age (statistically) significantly associated with the odds of breast-feeding even after adjusting for sex?

Continued

e. Suppose we are comparing two male children: One is 23 months old; the other is 13 months old. What is the estimated odds ratio of being breast- fed for the 23-month-old compared to the 13-month-old? Give a 95% confidence interval for this estimate.

JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Section A

Practice Problem Solutions

1. On the following slide, you will find regression output which relates the probability that a mother breast-feeds her child to the age of the child (in months) and the child's sex. The results come from a sub-sample of mother-child pairs in the Nepal Nutrition Intervention Study.

Continued

```
. logit

Logit estimates                                   Number of obs   =        472
                                                  LR chi2(2)      =     375.97
                                                  Prob > chi2     =     0.0000
Log likelihood = -132.02134                       Pseudo R2       =     0.5874

------------------------------------------------------------------------------
      bf |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
sex_chld |  -.3110831   .3158853
agechldc |  -.1867791   .0176001
   _cons |  -.9197549   .2251506
------------------------------------------------------------------------------
```

That Darn Cat!!!!!!!!!!!

Here sex_chld is the sex of the child (1= female, 0 =male), and agechldc is the age of the child in months.

Continued

a. Write out the logistic regression equation described by this output.

$$\log\left(\frac{p}{1-p}\right) = -.92 + (-.31)x_1 + (-.19)x_2$$

Where p = probability a child is breast-fed

x_1 = sex, x_2 = age

Continued

b. What is the interpretation of the coefficient of sex_chld?

It is the estimated difference in the log odds of being breast-fed for a female child when compared to a male child of the same age (the adjusted log OR).

Continued

c. Give a 95% CI for the adjusted odds ratio for the odds of being breastfed for female children as compared to male children.

Continued

c. Same old same old! Start with a 95% CI for sex_chld (b_1):

$$\hat{b}_1 \pm 2SE(\hat{b}_1)$$

-.31+2*.32

(-.95, .33)

Continued

c. Exponentiate to get the 95% CI!

So the 95% CI for the OR is as follows:

$$(e^{(-.95)}, e^{(.33)}) \rightarrow (0.36, 1.39)$$

Continued

d. Is child's age (statistically) significantly associated with the odds of breast-feeding even after adjusting for sex?

Continued

d. Compute the test statistic:

$$z = \frac{\hat{b}_2}{se(\hat{b}_2)} = \frac{-0.19}{.018} == -10.6!$$

So $p < .001$, which for most researchers' α levels would be considered statistically significant!

Continued

e. Suppose we are comparing two male children: One is 23 months old; the other is 13 months old. What is the estimated odds ratio of being breast- fed for the 23-month-old compared to the 13-month-old? Give a 95% CI for this estimate.

Continued

e. We need to first find the 95% CI for $10 b_2$. Start with the 95% CI for b_2:

$$\hat{b}_1 + 2SE(\hat{b}_1)$$

$-0.19 \pm 2*(.018)$

(-0.23, -.15)

Then, multiply the endpoints by 10:

(-2.3, -1.5)

Continued

e. Now, exponentiate to get the 95% CI for $10b_2$!

$$(e^{(-2.3)}, e^{(-1.5)})$$

(0.10, 0.22)

JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Section B

Statistical Interaction and Logistic Regression

Between January, 1974 and May,1984, a double-blinded randomized trial of patients with primary biliary cirrhosis (PBC) of the liver was conducted at the Mayo clinic

A total of 312 patients were randomized to either receive the drug D-penicillamine (DPCA) or a placebo

Patients were followed until they died from PBC or until censoring

Continued

Data, as it appears in Stata

```
. list  death drug sex in 1/10

     +------------------------+
     | death   drug      sex |
     |------------------------|
  1. |     0      0   Female |
  2. |     0      1   Female |
  3. |     0      0     Male |
  4. |     0      1   Female |
  5. |     0      0   Female |
     |------------------------|
  6. |     0      1   Female |
  7. |     0      1   Female |
  8. |     0      1   Female |
  9. |     0      0   Female |
 10. |     0      1   Female |
     +------------------------+
```

Continued

Logistic regression of death as predicted by treatment (1 = DPCA, 0 = placebo)

```
. logistic death drug

Logistic regression                               Number of obs   =        312
                                                  LR chi2(1)      =       0.15
                                                  Prob > chi2     =     0.6946
Log likelihood = -209.98341                       Pseudo R2       =     0.0004

------------------------------------------------------------------------------
       death | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        drug |   1.094982   .2531334     0.39   0.695     .6960329    1.722599
------------------------------------------------------------------------------
```

Continued

Logistic regression of death as predicted by sex (1 = female, 0 = male)

```
. logistic death sex

Logistic regression                               Number of obs   =        312
                                                  LR chi2(1)      =       7.34
                                                  Prob > chi2     =     0.0067
Log likelihood = -206.39105                       Pseudo R2       =     0.0175

------------------------------------------------------------------------------
       death | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |   .3788755   .1378462    -2.67   0.008     .1856966     .773017
------------------------------------------------------------------------------
```

Continued

Is is possible that sex confounds the relationship between death and the treatment?

- *Sex would have to be related to both death AND treatment*
- *Sex should not be related to treatment—why?*

Continued

Tabulation of sex and treatment group

```
. tab drug sex, row chi2

+----------------+
| Key            |
|----------------|
|   frequency    |
| row percentage |
+----------------+

           |          sex
      drug |      Male     Female |     Total
-----------+----------------------+----------
         0 |        15        139 |       154
           |      9.74      90.26 |    100.00
-----------+----------------------+----------
         1 |        21        137 |       158
           |     13.29      86.71 |    100.00
-----------+----------------------+----------
     Total |        36        276 |       312
           |     11.54      88.46 |    100.00

          Pearson chi2(1) =   0.9634   Pr = 0.326
```

Continued

What should happen when we perform a multiple logistic regression of death on both treatment and sex?

How should the adjusted estimates of the relationship between death and treatment and death and sex compare to the unadjusted estimates?

Continued

Results of logistic regression

```
. logistic death sex drug

Logistic regression                               Number of obs   =        312
                                                  LR chi2(2)      =       7.40
                                                  Prob > chi2     =     0.0247
Log likelihood = -206.36138                       Pseudo R2       =     0.0176

------------------------------------------------------------------------------
       death | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         sex |   .3806878   .1387036    -2.65   0.008     .1863947    .7775068
        drug |    1.05873   .2480293     0.24   0.808     .6689195    1.675702
------------------------------------------------------------------------------
```

Continued

So, a recap

	Odds Ratio of Death (95% CI)	
	Unadjusted	**Adjusted**
Treatment (1 =drug)	1.09 (0.69 - 1.72)	1.06 (0.67 - 1.68)
Sex (1= female)	0.38 (0.19 - 0.77)	0.38 (0.19 - 0.78)

Continued

It appears that the treatment is not associated with death (either an increase or decrease) after taking into account sampling variability

Females have a lower odds of death

Because it is a relatively large randomized trial, sex does not confound the relationship between death and treatment

Continued

However, is it possible that the death/treatment relationship is different for men and women?

The previous regression model with both treatment and sex as predictors "assumes" the same death/treatment relationship for each sex

Suppose we wanted allow for different estimates of the relationship between death and treatment depending on sex

We could run two logistic regressions, death on treatment for males and females separately, and report the results separately

Let's examine death/treatment relationship in women

```
. logistic death  drug if sex==1

Logistic regression                               Number of obs   =        276
                                                  LR chi2(1)      =       0.00
                                                  Prob > chi2     =     0.9748
Log likelihood = -182.33361                       Pseudo R2       =     0.0000

------------------------------------------------------------------------------
       death | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        drug |   .9921735    .246971    -0.03   0.975     .6091262    1.616099
------------------------------------------------------------------------------
```

Continued

Let's examine death/treatment relationship in men

```
. logistic death  drug if sex==0

Logistic regression                               Number of obs   =         36
                                                  LR chi2(1)      =       0.65
                                                  Prob > chi2     =     0.4192
Log likelihood = -23.730647                       Pseudo R2       =     0.0136

------------------------------------------------------------------------------
       death | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        drug |       1.75   1.215138     0.81   0.420       .448739    6.824679
------------------------------------------------------------------------------
```

Continued

Overall unadjusted odds ratio of death, drug compared to placebo

- *1.09 (95% CI 0.69 – 1.72)*

Sex specific odds ratio of death, drug compared to placebo

- *Females: 0.99 (95% CI 0.61–1.61)*
- *Males: 1.75 (95% CI 0.45–6.82)*

Does this Suggest Effect Modification?

Coefficient estimates for treatment "look different" for males and females

But are they really different? We would need to do a formal hypothesis test/estimation of the difference comparing elevation coefficient across models (a pain!)—although we could note crossover in CIs

Is there anyway this could be done without running separate regressions?

More Elegant Solution!

Add an interaction term to the model which already includes treatment (x_1) and sex (x_2) as predictors

Add $x_3 = x_1 * x_2$

New model

$$\log\left(\frac{p}{1-p}\right) = \hat{b}_0 + \hat{b}_1 x_1 + \hat{b}_2 x_2 + \hat{b}_3 x_3$$

Where p is probability of death

What is value of x_3 for Females? Males? (recall $x_2 = 1$ if subject is female, 0 if subject is male)

Continued

Model for males ($x_2 = 0$)

$$\log\left(\frac{p}{1-p}\right) = \hat{b}_0 + \hat{b}_1 x_1$$

So $\hat{b}_1$ estimates the association between death and treatment in males—it is the log odds ratio of death for men on the drug compared to men not on the drug

Continued

Model for females ($x_2=1$)

$$\log\left(\frac{p}{1-p}\right) = \hat{b}_0 + \hat{b}_1 x_1 + \hat{b}_2 x_2 + \hat{b}_3 x_1$$

$$\log\left(\frac{p}{1-p}\right) = \hat{b}_0 + \hat{b}_2 x_2 \boxed{+ (\hat{b}_1 + \hat{b}_3) x_1}$$

So $(\hat{b}_1 + \hat{b}_3)$ estimates the association between death and treatment in females—it is the log odds ratio of death for women on the drug compared to women not on the drug

Continued

Why Does this Work?

So $\hat{b}_3$ is an estimate of the difference in the relationship between death and treatment for males and females

It allows for the estimated “effect” of treatment on death to be different depending on the sex of the subject

If there is no difference in the death/treatment relationship between men and women, then the true value of b_3 would be 0

Testing the following hypothesis:

- H_o: $b_3 = 0$
- H_a: $b_3 \neq 0$

Is sometimes called a "test of interaction"

To create interaction term, use generate command:

*generate interact = drug*sex*

Continued

Results from Stata

```
. generate interact=drug*sex

. logit  death drug sex interact

Iteration 0:   log likelihood = -210.06047
Iteration 1:   log likelihood = -206.06515
Iteration 2:   log likelihood = -206.06426

Logit estimates                                   Number of obs   =        312
                                                  LR chi2(3)      =       7.99
                                                  Prob > chi2     =     0.0462
Log likelihood = -206.06426                       Pseudo R2       =     0.0190

------------------------------------------------------------------------------
       death |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        drug |   .5596158   .6943637     0.81   0.420    -.8013121    1.920544
         sex |  -.6481958   .5464266    -1.19   0.236    -1.719172    .4227806
    interact |  -.5674731   .7376325    -0.77   0.442    -2.013206    .8782601
       _cons |   .1335314    .517549     0.26   0.796     -.880846    1.147909
------------------------------------------------------------------------------
```

Continued

Results from Stata

```
. generate interact=drug*sex

. logit  death drug sex interact

Iteration 0:   log likelihood = -210.06047
Iteration 1:   log likelihood = -206.06515
Iteration 2:   log likelihood = -206.06426

Logit estimates                                   Number of obs   =        312
                                                  LR chi2(3)      =       7.99
                                                  Prob > chi2     =     0.0462
Log likelihood = -206.06426                       Pseudo R2       =     0.0190

------------------------------------------------------------------------------
       death |      Coef.   Std. Err.       z     P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        drug |   .5596158   .6943637     0.81   0.420    -.8013121    1.920544
         sex |  -.6481958   .5464266    -1.19   0.236    -1.719172    .4227806
    interact |  -.5674731   .7376325    -0.77   0.442    -2.013206    .8782601
       _cons |   .1335314    .517549     0.26   0.796     -.880846    1.147909
------------------------------------------------------------------------------
```

$\hat{b}_1$ → drug, $\hat{b}_2$ → sex, $\hat{b}_3$ → interact, $\hat{b}_0$ → _cons

Model for males ($x_2 = 0$)

$$\log\left(\frac{p}{1-p}\right) = 0.13 + 0.56x_1$$

The estimated relationship between death and treatment for males is positive. The estimated odds ratio of death for males on treatment relative to males on placebo is $e^{(.56)} = 1.75$.

Continued

Model for females ($x_2 = 1$)

$$\log\left(\frac{p}{1-p}\right) = \hat{b}_0 + 0.56x_1 - 0.65 - 0.57x_1$$

$$\log\left(\frac{p}{1-p}\right) = 0.13 - 0.65 + (0.56 - 0.57)x_1$$

$$\log\left(\frac{p}{1-p}\right) = -0.52 + (-.01)x_1$$

The estimated relationship between death and treatment for females is negative. The estimated odds ratio of death for females on treatment relative to females on placebo is $e^{(.01)} = 0.99$.

Continued

Confidence interval is trickier—must appeal to Stata and *lincom* command

```
. lincom drug + interact

 ( 1)  drug + interact = 0

------------------------------------------------------------------------------
       death |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |  -.0078573   .2489192    -0.03   0.975    -.4957299    .4800154
------------------------------------------------------------------------------
```

The p-value suggests we do not need interaction term

```
. generate interact=drug*sex

. logit  death drug sex interact

Iteration 0:   log likelihood = -210.06047
Iteration 1:   log likelihood = -206.06515
Iteration 2:   log likelihood = -206.06426

Logit estimates                                   Number of obs   =        312
                                                  LR chi2(3)      =       7.99
                                                  Prob > chi2     =     0.0462
Log likelihood = -206.06426                       Pseudo R2       =     0.0190

------------------------------------------------------------------------------
       death |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        drug |   .5596158   .6943637     0.81   0.420    -.8013121    1.920544
         sex |  -.6481958   .5464266    -1.19   0.236    -1.719172    .4227806
    interact |  -.5674731   .7376325    -0.77   0.442    -2.013206    .8782601
       _cons |   .1335314    .517549     0.26   0.796     -.880846    1.147909
------------------------------------------------------------------------------
```

Can be done with logistic command too

```
. logistic  death drug sex interact

Logistic regression                               Number of obs   =        312
                                                  LR chi2(3)      =       7.99
                                                  Prob > chi2     =     0.0462
Log likelihood = -206.06426                       Pseudo R2       =     0.0190

------------------------------------------------------------------------------
       death | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        drug |       1.75   1.215137     0.81   0.420     .4487398    6.824668
         sex |   .5229885   .2857748    -1.19   0.236     .1792144    1.526199
    interact |   .5669563   .4182054    -0.77   0.442     .1335598    2.406709
------------------------------------------------------------------------------

. lincom drug + interact

 ( 1)  drug + interact = 0

------------------------------------------------------------------------------
       death | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   .9921735    .246971    -0.03   0.975     .6091261    1.616099
------------------------------------------------------------------------------
```

JOHNS HOPKINS BLOOMBERG SCHOOL *of* PUBLIC HEALTH

Section B

Practice Problems: Another Example—Statistical Interaction and Logistic Regression

Recall the results from a fictitious data set comparing two treatments for a fatal disease as to the impact of each on reducing deaths (lecture two)

- *The data presented is stratified by two age groups, "young" and "old"*

Continued

The data:

	YOUNG		OLD	
	Surgery	Drug	Surgery	Drug
Died	100	200	200	100
Survived	200	100	100	200
Totals	300	300	300	300

Continued

Surgery is better for this group!

	YOUNG		OLD	
	Surgery	Drug	Surgery	Drug
Died	100	200	200	100
Survived	200	100	100	200
Totals	300	300	300	300

33% (100/300) of those who had surgery died, as compared to 67% (200/300) of those taking the drug!

(RR = .50, OR = .25)

Continued

Taking the drug is better for this group!

	YOUNG		OLD	
	Surgery	Drug	Surgery	Drug
Died	100	200	200	100
Survived	200	100	100	200
Totals	300	300	300	300

67% (200/300) of those who had surgery does, as compared to 37% (100/300) of those who took the drug!

(RR = 2, OR = 4)

Continued

What happens when we combine tables?

	Combined Young and Old	
	Surgery	**Drug**
Died	300	300
Survived	300	300
Totals	600	600

Surgery and drug groups have identical proportions dying! (RR = OR = 1, 50% in each group!)

Continued

$$\log\left(\frac{p}{1-p}\right) = \hat{b}_0 + \hat{b}_1 x_1 + \hat{b}_2 x_2 + \hat{b}_3 x_3$$

Where p = probability of death

x_1 = treatment (1 if surgery, 0 if drug therapy)

x_2 = age group (1 if older, 0 if younger)

x_3 = treatment X age

Model for younger patients ($x_2 = 0$)

$$\log\left(\frac{p}{1-p}\right) = \hat{b}_0 + \hat{b}_1 x_1$$

So $\hat{b}_1$ estimates the association between death and surgery (as compared to drug therapy) in younger patients—it is the log odds ratio of death for younger patients who have surgery compared to younger patients who get drug therapy

Continued

Model for older patients ($x_2 = 1$)

$$\log\left(\frac{p}{1-p}\right) = \hat{b}_0 + \hat{b}_1 x_1 + \hat{b}_2 x_2 + \hat{b}_3 x_1$$

$$\log\left(\frac{p}{1-p}\right) = \hat{b}_0 + \hat{b}_2 x_2 + (\hat{b}_1 + \hat{b}_3) x_1$$

So $(\hat{b}_1 + \hat{b}_3)$ estimates the association between death and surgery (as compared to drug therapy) in older patients—it is the log odds ratio of death for older patients who have surgery compared to older patients who get drug therapy

Here are the regression coefficients estimated by the computer:

```
. logit  death age treatment interact

Iteration 0:   log likelihood = -831.77662
Iteration 1:   log likelihood = -763.91077
Iteration 2:   log likelihood =   -763.817
Iteration 3:   log likelihood =   -763.817

Logit estimates                                   Number of obs   =       1200
                                                  LR chi2(3)      =     135.92
                                                  Prob > chi2     =     0.0000
Log likelihood =   -763.817                       Pseudo R2       =     0.0817

------------------------------------------------------------------------------
       death |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -1.386294   .1732051    -8.00   0.000     -1.72577   -1.046819
   treatment |  -1.386294   .1732051    -8.00   0.000     -1.72577   -1.046819
    interact |   2.772589    .244949    11.32   0.000     2.292498     3.25268
       _cons |   .6931472   .1224745     5.66   0.000     .4531016    .9331928
------------------------------------------------------------------------------
```

Continued

The interaction term estimates the difference in the log odds of death (surgery versus drug) between younger and older patients

```
. logit  death age treatment interact

Iteration 0:   log likelihood = -831.77662
Iteration 1:   log likelihood = -763.91077
Iteration 2:   log likelihood =   -763.817
Iteration 3:   log likelihood =   -763.817

Logit estimates                                   Number of obs   =       1200
                                                  LR chi2(3)      =     135.92
                                                  Prob > chi2     =     0.0000
Log likelihood =   -763.817                       Pseudo R2       =     0.0817

------------------------------------------------------------------------------
       death |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -1.386294   .1732051    -8.00   0.000     -1.72577   -1.046819
   treatment |  -1.386294   .1732051    -8.00   0.000     -1.72577   -1.046819
    interact |   2.772589    .244949    11.32   0.000     2.292498     3.25268
       _cons |   .6931472   .1224745     5.66   0.000     .4531016    .9331928
------------------------------------------------------------------------------
```

Model for younger patients ($x_2 = 0$)

$$\log\left(\frac{p}{1-p}\right) = 0.69 + (-1.39x_1)$$

The estimated relationship between death and surgery (as compared to drug therapy) for younger patients is negative. The estimated odds ratio of death for younger patients getting surgery relative to younger patients on drug therapy is $e^{(-1.39)} = 0.25$.

Continued

Model for older patients (x_2 = 1)

$$\log\left(\frac{p}{1-p}\right) = 0.69 + (-1.39)x_1 + (-1.39) + 2.77x_1$$

$$\log\left(\frac{p}{1-p}\right) = 0.13 - 1.39 + (-1.39 + 2.77)x_1$$

$$\log\left(\frac{p}{1-p}\right) = -1.26 + 1.38x_1$$

The estimated relationship between death and surgery (as compared to drug therapy) for older patients is positive. The estimated odds ratio of death for younger patients getting surgery relative to younger patients on drug therapy is $e^{(-1.38)}$ = 4.0.

Continued

Results using logistic command

```
. logistic  death treatment age interact

Logistic regression                               Number of obs   =       1200
                                                  LR chi2(3)      =     135.92
                                                  Prob > chi2     =     0.0000
Log likelihood =   -763.817                       Pseudo R2       =     0.0817

------------------------------------------------------------------------------
       death | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   treatment |        .25   .0433013    -8.00   0.000     .1780359    .3510528
         age |        .25   .0433013    -8.00   0.000     .1780359    .3510528
    interact |         16   3.919184    11.32   0.000     9.899632    25.85955
------------------------------------------------------------------------------

. lincom treatment+interact

 ( 1)  treatment + interact = 0

------------------------------------------------------------------------------
       death | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |          4   .6928203     8.00   0.000     2.848574    5.616845
------------------------------------------------------------------------------
```

Results, as they might be presented in a journal (remember, though, this example is fictitious)

Table 1: Logistic Regression Results						
		Bivariate (Unadjusted) Results			Multivariate Results **	
Predictor		Odds Ratio	95% CI		Odds Ratio	95% CI
Treatment	Drug	1.00			1.00	
	Surgery	1.00	(0.79,1.25)		0.25	(0.18,0.35) *
Age	Young	1.00			1.00	
	Old	1.00	(0.79,1.25) *		0.25	(0.18, 0.35) *
* p < .05						
** Multivariate Analysis Included Interaction Between Treatment and Age: See Table 1A for details						

Continued

Interactions can be detailed in a second, separate table

Table 1A: Age Specific Death/Treatment Asssociations

		Odds Ratio of Death Surgery to Drug	95% Confidence Interval
Age	Young	0.25	(0.18,0.35)
	Old	4.00	(2.85,5.61)

JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Section C

Predictors of Death among Patients with Severe Sepsis: An Example of Multiple Logistic Regression

Many predictors

$$\log\left(\frac{p}{1-p}\right) = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + \ldots.$$

Which we will estimated by:

$$\log\left(\frac{p}{1-p}\right) = \hat{b}_0 + \hat{b}_1 x_1 + \hat{b}_2 x_2 + \hat{b}_3 x_3 + \ldots.$$

Patients with Sepsis

Predicators of death for patients with severe sepsis (blood poisoning) undergoing surgery

- *Can we predict the risk of death (y) from five predcitors (x's) in one model?*

A sample of 106 patients who had surgery with severe sepsis was collected (Pine, et al.)

This is an observational study

Outcome of interest—death

$$y = \begin{cases} 1 & \text{Death} \\ 0 & \text{Survive} \end{cases}$$

Continued

Patients with Sepsis: The Data

Potential predictors

- *Shock* (x_1)
- *Malnutrition* (x_2)
- *Alcohol use* (x_3)
- *Age* (x_4)
- *Bowel infarction* (x_5)

Continued

Patients with Sepsis: The Data

Shock, malnutrition, alcohol, and bowel infarction are all recorded as binary

Each is coded “1” or “0” in a dummy variable

Age is continuous and measured in years

Continued

Snippet from data set

ID	Death	Malnut	Shock	Alcohol	Age (yrs)	Bowel Inf
1	0	0	0	0	56	0
2	1	0	0	1	66	0
3	0	0	0	0	20	0
4	0	1	0	0	50	0
5	0	0	0	1	36	1
6	1	0	1	0	85	1

Continued

Snippet from data set

ID	Death	Malnut	Shock	Alcohol	Age (yrs)	Bowel Inf
1	0	0	0	0	56	0
2	1	0	0	1	66	0
3	0	0	0	0	20	0
4	0	1	0	0	50	0
5	0	0	0	1	36	1
6	1	0	1	0	85	1

Continued

Snippet from data set

ID	Death	Malnut	Shock	Alcohol	Age (yrs)	Bowel Inf
1	0	0	0	0	56	0
2	1	0	0	1	66	0
3	0	0	0	0	20	0
4	0	1	0	0	50	0
5	0	0	0	1	36	1
6	1	0	1	0	85	1

Twenty-one of the 106 (19.8 %) patients died following surgery

The average age of the patients is 51 years and the age range in the sample is from 17 to 94 years

Continued

Patients with Sepsis: The Data

We want to answer some questions:

- ★ *How is death following surgery for patients with severe sepsis associated with these five potential predictors?*
- ★ *Do certain predictors confound each other's relationship with death?*
- ★ *Can we estimate risk (proportion of patients who will die) given patient characteristics at time of surgery?*

Can employ (multiple) logistic regression

Patients with Sepsis: The Model

First step—estimate unadjusted associations between death and each predictor using simple logistic regression models

The data is entered as follows

	death	shock	malnut	alcohol	age	infarc
1.	0	0	0	0	56	0
2.	0	0	0	0	80	0
3.	0	0	0	0	61	0
4.	0	0	0	0	26	0
5.	0	0	0	0	53	0
6.	1	0	1	0	87	0
7.	0	0	0	0	21	0
8.	1	0	0	1	69	0
9.	0	0	0	0	57	0
10.	0	0	1	0	76	0

Example—death and shock

```
. logistic death shock

Logistic regression                                 Number of obs   =        106
                                                    LR chi2(1)      =      13.55
                                                    Prob > chi2     =     0.0002
Log likelihood = -45.988289                         Pseudo R2       =     0.1284

------------------------------------------------------------------------------
       death | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       shock |   13.66667   10.22543     3.49   0.000     3.153493    59.22885
------------------------------------------------------------------------------
```

Continued

Second step—fit a logistic regression with all five predictors

$$\mathbf{log}\left(\frac{p}{1-p}\right)=\hat{b}_0+\hat{b}_1x_1+\hat{b}_2x_2+\hat{b}_3x_3+\hat{b}_3x_4+\hat{b}_5x_5.$$

$x_1 - x_5$ defined as before

p = Pr(y = 1), the probability of death

Continued

Results from Stata

```
. logistic  death shock malnut alcohol age infarc

Logit estimates                                   Number of obs   =        106
                                                  LR chi2(5)      =      49.13
                                                  Prob > chi2     =     0.0000
Log likelihood =  -28.19964                       Pseudo R2       =     0.4656

------------------------------------------------------------------------------
   death | Odds Ratio   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
   shock |   30.94636   32.74522      3.244   0.001       3.889816    246.2011
  malnut |   2.568156   1.766134      1.372   0.170       .6671843    9.885465
 alcohol |   18.39333   15.99545      3.349   0.001        3.34524    101.1332
     age |   1.086349   .0300663      2.993   0.003        1.02899    1.146906
  infarc |   6.368962   6.266786      1.882   0.060       .9258118    43.81417
------------------------------------------------------------------------------
```

Continued

Note—at this point, we could review other potential models, leaving out non-statistically significant predictors from the previous model, or including interactions, etc.

Model building is part art, part science

Frequently, the results of the unadjusted and adjusted analyses are presented in one table

Not only is this a concise summary, it allows for side-by-side comparisons of the unadjusted and adjusted estimates for each predictor which helps give a sense of confounding amongst the predictors

Continued

Table of results

Results from Logistic Regression
Outcome is Death (yes/no)

Predictor	Undadjusted Odds Ratio (95% CI)	Adjusted Odds Ratio (95% CI)
Shock	13.7 (3.2 , 59.2)	30.9 (3.9, 246.2)
Malnutrition	3.4 (1.3, 9.0)	2.6 (0.7, 9.9)
Alcohol	5.5 (1.9 , 15.8)	18.4 (3.4, 101.1)
Infarction	5.6 (1.7, 18.4)	6.4 (0.9, 43.8)
Age (yrs)	1.05 (1.02,1.08)	1.08 (1.03, 1.15)

Continued

Table of results

Results from Logistic Regression
Outcome is Death (yes/no)

Predictor	Undadjusted Odds Ratio (95% CI)	Adjusted Odds Ratio (95% CI)
Shock	13.7 (3.2 , 59.2)	30.9 (3.9, 246.2)
Malnutrition	3.4 (1.3, 9.0)	2.6 (0.7, 9.9)
Alcohol	5.5 (1.9 , 15.8)	18.4 (3.4, 101.1)
Infarction	5.6 (1.7, 18.4)	6.4 (0.9, 43.8)
Age (yrs)	1.05 (1.02,1.08)	1.08 (1.03, 1.15)

Continued

Table of results

Results from Logistic Regression
Outcome is Death (yes/no)

Predictor	Undadjusted Odds Ratio (95% CI)	Adjusted Odds Ratio (95% CI)
Shock	13.7 (3.2 , 59.2)	30.9 (3.9, 246.2)
Malnutrition	3.4 (1.3, 9.0)	2.6 (0.7, 9.9)
Alcohol	5.5 (1.9 , 15.8)	18.4 (3.4, 101.1)
Infarction	5.6 (1.7, 18.4)	6.4 (0.9, 43.8)
Age (yrs)	1.05 (1.02,1.08)	1.08 (1.03, 1.15)

Continued

After controlling for age, alcohol use, malnutrition, and whether or not the patient experienced a bowel infarction, patients who experience shock have and estimated 30.9 the odds of death compared to those who do not experience shock (95% CI 2.9 – 246.2)

After controlling for age, shock, malnutrition, and whether or not the patient experienced a bowel infarction, patients who used alcohol had 18.4 times the odds of death (95% CI 3.4 to 101.1) compared to those who did not use alcohol

Continued

After controlling for shock, malnutrition, alcohol usage and whether or not the patient experienced a bowel infarction, a statistically significant relationship was found between the risk of death and age

The odds ratio of death associated with a one year increase in age, all other factors being equal, was 1.08 (95% CI 1.02 to 1.14)

Continued

Odds ratios give an estimate of relative odds of outcome—can help us assess risk factors

However, odds ratios are neither direct comparisons of risk, nor do they tell us anything about the actual risk of death for different subsets of patients with different characteristics at the time of study

As this is not a case-control study, we are allowed to estimate risk and relative risk via the sample—how can we do this with logistic regression results?

Continued

We can actually estimate the logistic regression equation (using the *logit* command)

```
. logit death shock malnut alcohol age infarc

(iteration information deleted by John)

Logit estimates                                   Number of obs   =        106
                                                  LR chi2(5)      =      49.13
                                                  Prob > chi2     =     0.0000
Log likelihood =  -28.19964                       Pseudo R2       =     0.4656

------------------------------------------------------------------------------
       death |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       shock |   3.432255   1.058128     3.24   0.001     1.358362    5.506149
      malnut |   .9431883   .6877051     1.37   0.170     -.404689    2.291066
     alcohol |   2.911988   .8696332     3.35   0.001     1.207538    4.616438
         age |   .0828228   .0276765     2.99   0.003     .0285779    .1370677
      infarc |   1.851436   .9839573     1.88   0.060    -.0770843    3.779957
       _cons |  -8.675357   2.226694    -3.90   0.000     -13.0396   -4.311116
------------------------------------------------------------------------------
```

Continued

Our estimated equation (multiple logistic regression)

$$log\left(\frac{p}{1-p}\right) = -8.67 + 3.43 \times SHOCK + 0.94 \times MALNUTRION + 2.91 \times ALCOHOL$$
$$+ 0.082 \times AGE + 1.9 \times INFARCTION$$

We can use this to estimate the log odds of death for any group of patients with any combination of values for the five predictors

Continued

By the formulation of logistic regression:

$$RISK(\hat{p}) = \frac{ODDS}{1+ODDS} = \frac{e^{(\log(ODDS))}}{1+e^{(\log(ODDS))}}$$

Translate equation back into (estimated) probability function

$$\hat{p} = \frac{e^{\hat{b}_0 + \hat{b}_1 x_1 + \hat{b}_2 x_2 + \hat{b}_3 x_3 + \hat{b}_4 x_4 + \hat{b}_5 x^5}}{1 + e^{\hat{b}_0 + \hat{b}_1 x_1 + \hat{b}_2 x_2 + \hat{b}_3 x_3 + \hat{b}_4 x_4 + \hat{b}_5 x^5}}$$

Continued

So for example, the estimated log odds of death for a 50 year old patient with sepsis who uses alcohol, but is not in shock, not malnourished, and does not have infarction at the time of surgery is given by . . .

$$log\left(\frac{p}{1-p}\right) = -8.67 + 3.43\times 0 + 0.94\times 0 + 2.91\times 1$$

$$+\,0.082\times 50 + 1.85\times 0 =$$

$$= -8.67 + 2.91 + 4.1$$

$$= -1.66$$

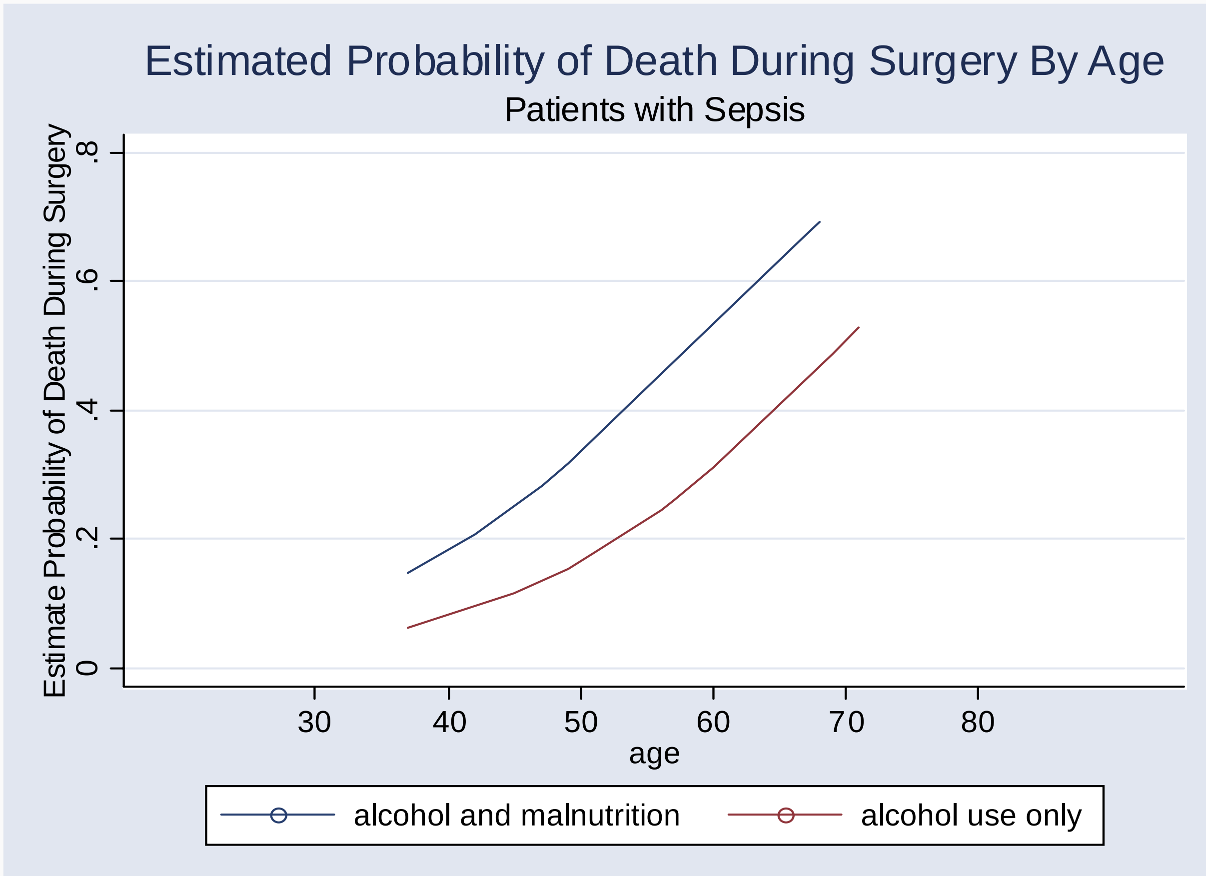
Continued

So the estimate proportion (probability, risk) of death during surgery for this group of patients is given by . . .

$$RISK(\hat{p}) = \frac{ODDS}{1+ODDS} = \frac{e^{-1.66}}{1+e^{-1.66}} \approx \frac{0.19}{1.19} \approx 0.16$$

Continued

Possible graphical display



JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Section C

Practice Problems

1. On the next slide, you will see both the unadjusted and adjusted odds ratio for the association between death and five predictors amongst a sample of 106 surgery patients with sepsis. Compare the unadjusted and adjusted odds ratios relating death to alcohol use at the time of surgery. Do the results suggest the death/alcohol relationship is confounded by the other predictors?

Continued

Table of results

Results from Logistic Regression
Outcome is Death (yes/no)

Predictor	Undadjusted Odds Ratio (95% CI)	Adjusted Odds Ratio (95% CI)
Shock	13.7 (3.2 , 59.2)	30.9 (3.9, 246.2)
Malnutrition	3.4 (1.3, 9.0)	2.6 (0.7, 9.9)
Alcohol	5.5 (1.9 , 15.8)	18.4 (3.4, 101.1)
Infarction	5.6 (1.7, 18.4)	6.4 (0.9, 43.8)
Age (yrs)	1.05 (1.02,1.08)	1.08 (1.03, 1.15)

Continued

2. Use the results to estimate the odds ratio of death for a group of 60-year-old malnourished patients in shock at the time of surgery compared to 40-year-old malnourished patients in shock at the time of surgery.

Continued

3. Recall the estimated equation (multiple logistic regression).

$$log\left(\frac{p}{1-p}\right) = -8.67 + 3.43 \times SHOCK + 0.94 \times MALNUTRION + 2.91 \times ALCOHOL$$
$$+ 0.082 \times AGE + 1.9 \times INFARCTION$$

Continued

Recall we estimated the risk of death for 50-year-old patients using alcohol at the time of surgery at 17% (these patients were not malnourished, in shock or with infarction at time of surgery).

Use the equation to estimate the risk for comparable 50-year-old patients who did not use alcohol at the time of surgery.

JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Section C

Practice Problem Solutions

1. On the next slide, you will see both the unadjusted and adjusted odds ratio for the association between death and five predictors amongst a sample of 106 surgery patients with sepsis. Compare the unadjusted and adjusted odds ratios relating death to alcohol use at the time of surgery. Do the results suggest the death/alcohol relationship is confounded by the other predictors?

Continued

Table of results

Results from Logistic Regression
Outcome is Death (yes/no)

Predictor	Undadjusted Odds Ratio (95% CI)	Adjusted Odds Ratio (95% CI)
Shock	13.7 (3.2 , 59.2)	30.9 (3.9, 246.2)
Malnutrition	3.4 (1.3, 9.0)	2.6 (0.7, 9.9)
Alcohol	5.5 (1.9 , 15.8)	18.4 (3.4, 101.1)
Infarction	5.6 (1.7, 18.4)	6.4 (0.9, 43.8)
Age (yrs)	1.05 (1.02,1.08)	1.08 (1.03, 1.15)

Continued

The unadjusted estimated odds ratio for alcohol use is 5.5 as compared to the adjusted odds ratio estimate of 18.4.

The adjusted estimate is larger then the unadjusted estimate—but to see “how different” these estimates actually are after taking into account sampling variation, let’s look at the corresponding CI’s.

The lower endpoint for the unadjusted odds ratio is 1.9 as compared to 3.4 for the adjusted odds ratio

So it does appear that the relationship between alcohol use and death in these patients is decidedly larger when taking into account the other predictors—some evidence of confounding.

Continued

Use the results to estimate the odds ratio of death for a group of 60-year-old malnourished patients in shock at the time of surgery compared to 40-year-old malnourished patients in shock at the time of surgery.

As the only difference in terms of the five predictors between these two groups of patients is age, we only need to focus on the adjusted results for age.

The adjusted odds ratio for a one year difference in age is 1.08. We have two groups of persons who differ by 20 years, how can we compute the odds ratio comparing these two groups?

Continued

It's best to go back to the coefficient scale where everything is linear—the corresponding regression coefficient for age is ln(1.08) = .082.

This is the estimated difference in log odds of death for a one year difference in age. To get the corresponding difference in the log odds of death for a 20-year difference, we would take .082*20 = 1.64.

To convert this to the odds ratio scale, exponentiate this result (i.e., take $e^{(1.64)}$: this gives the resulting odds ratio estimate of 5.5—this is the estimated odds ratio of death comparing two otherwise comparable groups of patients who differ in age by 20 years (older relative to younger).

Continued

3. Recall the estimated equation (multiple logistic regression)

$$log\left(\frac{p}{1-p}\right) = -8.67 + 3.43 \times SHOCK + 0.94 \times MALNUTRION + 2.91 \times ALCOHOL$$

$$+ 0.082 \times AGE + 1.9 \times INFARCTION$$

Continued

Recall we estimated the risk of death for 50 year old patients using alcohol at the time of surgery at 16% (these patients were not malnourished, in shock or with infarction at time of surgery).

Use the equation to estimate the risk for comparable 50-year-old patients who did not use alcohol at the time of surgery.

Continued

The estimated log odds of death for a 50-year-old patient with sepsis who does not use alcohol, and is not in shock, not malnourished, and does not have infarction at the time of surgery is given by . . .

$$log\left(\frac{p}{1-p}\right) = -8.67 + 3.43\times 0 + 0.94\times 0 + 2.91\times 0$$

$$+ 0.082\times 50 + 1.85\times 0 =$$

$$= -8.67 + 4.1$$

$$= -4.57$$

Continued

So the estimate proportion (probability, risk) of death during surgery for this group of patients is given by . . .

$$RISK(\hat{p}) = \frac{ODDS}{1+ODDS} = \frac{e^{-4.57}}{1+e^{-4.57}} \approx \frac{0.01}{1.01} \approx 0.01$$