Applied Logistic Regression

Week 4

1. **Homework week 3: highlights**
2. Polychotomous independent variable
   - Approaching polychotomous data via contingency tables
   - Reference cell coding
   - Deviation from means coding
   - Example using polychotomous variable
3. Adjusted odds ratios
4. Continuous independent variable
5. Homework

Stanley Lemeshow, Professor of Biostatistics

College of Public Health, The Ohio State University

THE OHIO STATE UNIVERSITY

Suppose now that x has 3 or more levels

- fixed number of outcomes and scale of measurement is nominal.

We must form a set of design variables to represent the categories of the variable.

	CHD		
RACE:	PRESENT	ABSENT	TOTAL
WHITE	5	20	25
BLACK	20	10	30
HISPANIC	15	10	25
OTHER	10	10	20
TOTAL	50	50	100

Now, let us use WHITE as the reference group. Then, the estimated odds ratio for blacks vs. whites is

	CHD	NO CHD
BLACK	20	10
WHITE	5	20

and $\widehat{OR} = \dfrac{20 \times 20}{5 \times 10} = \dfrac{400}{50} = 8.0$

Similarly, for Hispanics, we have

	CHD	NO CHD
HISP	15	10
WHITE	5	20

and $\widehat{OR} = \dfrac{15 \times 20}{5 \times 10} = 6$

Finally, for the "other" group, we have

	CHD	NO CHD
OTHER	10	10
WHITE	5	20

and $\widehat{OR} = \dfrac{10 \times 20}{5 \times 10} = 4$

It is common practice to identify a particular referent group (in this case whites) to which each of the other groups will be compared.

These same estimates of odds ratios may be obtained from a logistic regression program with an appropriate choice of design variables. Specifically,

Race	Design Variables D_1	D_2	D_3
White	0	0	0
Black	1	0	0
Hispanic	0	1	0
Other	0	0	1

← referent group gets all 0's called "reference cell coding"

- Partial method with BMDPLR
- Method used in EGRET, GLIM
- SAS and SYSTAT take the highest coded group as the referent group

To get the same coding as with other methods we would have to reverse the coding to 1 = black, 2 = Hispanic, 3 = other, 4 = white.

Using a logistic regression program with design variables coded as above we get the following:

Variable		$\hat{\beta}$	$\widehat{SE}(\hat{\beta}_i)$	$\hat{\beta}_i / \widehat{SE}(\hat{\beta}_i)$	$\widehat{OR}$
Race (1)	$\hat{\beta}_{11}$	2.079	0.633	3.29	8.0 $= e^{2.079}$
Race (2)	$\hat{\beta}_{12}$	1.792	0.646	2.78	6.0 $= e^{1.792}$
Race (3)	$\hat{\beta}_{13}$	1.386	0.671	2.07	4.0 $= e^{1.386}$
Constant		−1.386	0.500	−2.77	

↑ these coefficients are the log odds ratios

To see that $\hat{\beta}_{11}$ is the log odds ratio of Blacks vs. Whites, we note that: $\ln\left[\widehat{OR}(\text{Black, White})\right] = \hat{g}(\text{Black}) - \hat{g}(\text{White})$

$$= \left(\hat{\beta}_0 + \hat{\beta}_{11}(1) + \hat{\beta}_{12}(0) + \hat{\beta}_{13}(0)\right) - \left(\hat{\beta}_0 + \hat{\beta}_{11}(0) + \hat{\beta}_{12}(0) + \hat{\beta}_{13}(0)\right) = \hat{\beta}_{11}$$

Similarly we can show that:

$$\hat{\beta}_{12} = \text{ log odds ratio for Hispanics vs Whites}$$
$$\hat{\beta}_{13} = \text{ log odds ratio for Others vs Whites}$$

As we know, for 2×2 tables, the standard error is approximated by:

$$\widehat{SE}\left(\ln\left(\widehat{OR}\right)\right) = \left[\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}\right]^{1/2}$$

Consider the 2×2 table we saw earlier

	CHD	NO CHD
BLACK	20	10
WHITE	5	20

$$\widehat{OR} = \frac{20 \times 20}{5 \times 10} = 8; \quad \ln\left(\widehat{OR}\right) = \overset{\hat{\beta}_{11}}{\overset{\downarrow}{2.079}}$$

$$\widehat{SE}\left[\ln\left(\widehat{OR}\right)\right] = \left[\frac{1}{20} + \frac{1}{10} + \frac{1}{5} + \frac{1}{20}\right]^{1/2} = .6325 \longleftarrow \widehat{SE}\left(\hat{\beta}_{11}\right)$$

Confidence intervals for odds ratios may be obtained in the usual way. i.e., we first compute

$$\hat{\beta}_{ij} - 1.96 \times \widehat{SE}\left(\hat{\beta}_{ij}\right) \leq \beta_{ij} \leq \hat{\beta}_{ij} + 1.96 \times \widehat{SE}\left(\hat{\beta}_{ij}\right)$$

and then, exponentiating everything in sight we have

$$e^{\hat{\beta}_{ij} - 1.96\widehat{SE}\left(\hat{\beta}_{ij}\right)} \leq OR \leq e^{\hat{\beta}_{ij} + 1.96\widehat{SE}\left(\hat{\beta}_{ij}\right)}$$

For example, using β_{11} and $\widehat{SE}\left(\hat{\beta}_{11}\right)$

$$2.079 - 1.96(.6325) \leq \beta_{11} \leq 2.079 + 1.96(.6325)$$

$$0.8393 \leq \beta_{11} \leq 3.3187$$

and

$$e^{0.8393} \leq OR \leq e^{3.3187}$$

$$2.315 \leq OR \leq 27.624$$

CIs for the other odds ratios can be obtained similarly.

```
. tab race chd [freq= freq]

           |          chd
      race |         0          1 |     Total
-----------+----------------------+----------
         1 |        20          5 |        25
         2 |        10         20 |        30
         3 |        10         15 |        25
         4 |        10         10 |        20
-----------+----------------------+----------
     Total |        50         50 |       100
```

```
.  logit chd race_2 race_3 race_4 [fweight= freq]

Iteration 0:   log likelihood = -69.314718
Iteration 1:   log likelihood = -62.368156
Iteration 2:   log likelihood = -62.293897
Iteration 3:   log likelihood = -62.293721

Logit estimates                                   Number of obs   =        100
                                                  LR chi2(3)      =      14.04
                                                  Prob > chi2     =     0.0028
Log likelihood = -62.293721                       Pseudo R2       =     0.1013

------------------------------------------------------------------------------
     chd |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  race_2 |   2.079442   .6324524     3.288   0.001       .8398576    3.319026
  race_3 |   1.791759   .6454942     2.776   0.006       .5266141    3.056905
  race_4 |   1.386294   .6708175     2.067   0.039       .0715163    2.701072
   _cons |  -1.386294   .4999961    -2.773   0.006      -2.366269   -.4063201
------------------------------------------------------------------------------
```

```
.  logistic chd race_2 race_3 race_4 [fweight= freq]

Logit estimates                                   Number of obs   =        100
                                                  LR chi2(3)      =      14.04
                                                  Prob > chi2     =     0.0028
Log likelihood = -62.293721                       Pseudo R2       =     0.1013

------------------------------------------------------------------------------
     chd | Odds Ratio   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  race_2 |          8   5.059619      3.288   0.001       2.316037    27.63341
  race_3 |          6   3.872965      2.776   0.006        1.69319    21.26165
  race_4 |          4    2.68327      2.067   0.039       1.074136     14.8957
------------------------------------------------------------------------------

. vce

         |   race_2   race_3   race_4    _cons
---------+------------------------------------
  race_2 |  .399996
  race_3 |  .249996  .416663
  race_4 |  .249996  .249996  .449996
   _cons | -.249996 -.249996 -.249996  .249996
```

Reference cell coding is the one most commonly employed by computer packages and in the literature since it is often of interest to compare various "exposed" groups to a "control" or "unexposed" group.

The "Marginal" method of dealing with categorically defined variables in BMDPLR uses "deviation from means" coding. This expresses effect as the deviation of the group mean from the overall mean.

logit for the group

average logit over all groups

Setting up deviation from means coding for Race, we use the following design variables:

Race (code)	Design Variables D_1	D_2	D_3
White (1)	-1	-1	-1
Black (2)	1	0	0
Hispanic (3)	0	1	0
Other (4)	0	0	1

Use of these design variables results in the following coefficients and SE's:

Variable	$\hat{\beta}_i$	$\widehat{SE}(\hat{\beta}_i)$	$\hat{\beta}_i / \widehat{SE}(\hat{\beta}_i)$
RACE (1)	0.765	0.351	2.18
RACE (2)	0.477	0.362	1.32
RACE (3)	0.072	0.385	0.19
Constant	-0.072	0.219	-0.33

To interpret these coefficients we must first compute the log odds for each of the 4 races:

$$\left.\begin{aligned}
\hat{g}_1 &= \ln\left(\frac{5/25}{20/25}\right) = \ln\left(5/20\right) = -1.38 \\
\hat{g}_2 &= \ln\left(\frac{20/30}{10/30}\right) = \ln\left(20/10\right) = 0.693 \\
g_3 &= \ln\left(\frac{15/25}{10/25}\right) = \ln\left(15/10\right) = 0.405 \\
\hat{g}_4 &= \ln\left(\frac{10/20}{10/20}\right) = \ln\left(10/10\right) = 0
\end{aligned}\right\} \bar{g} = \sum_{i=1}^{4} \hat{g}_i / 4 = -0.072$$

Now,

the estimated coefficient, $\hat{\beta}_1$, for the first design variable is defined, using deviation from means coding, as

$$\hat{g}_2 - \bar{g}$$

In this example,

$$\hat{\beta}_1 = 0.765 = 0.693 - \left(-0.072\right)$$

Similarly,

$$\hat{\beta}_2 = \hat{g}_3 - \bar{g}$$
$$\hat{\beta}_3 = \hat{g}_4 - \bar{g}$$

For this type of coding, $e^{\hat{\beta}_i} = e^{\left(\hat{g}_{i+1} - \bar{g}\right)} = \dfrac{e^{\hat{g}_{i+1}}}{e^{\bar{g}}}$

$$e^{1/4\left[g_1+g_2+g_3+g_4\right]}$$

$$= e^{g_1/4} e^{g_2/4} e^{g_3/4} e^{g_4/4}$$

$$= \left[e^{g_1} e^{g_2} e^{g_3} e^{g_4}\right]^{1/4}$$

= geometric mean

In our example,

$$e^{\hat{\beta}_1} = e^{0.765} = e^{\left(0.693 - \left(-0.072\right)\right)} = \frac{e^{0.693}}{e^{-0.072}} = 2.15$$

The result, 2.15, is not an odds ratio since the values in the numerator and denominator are not the odds for two distinct groups.

- Instead, it is the ratio of the odds of one group to an average odds (in this case the geometric mean).

Now, suppose we would like to compute the odds ratio for Blacks vs. Whites

$$
\begin{aligned}
\ln\left(\widehat{OR}\right) &= g_2 - g_1 \\
&= \hat{\beta}_0 + \hat{\beta}_1 D_{21} + \hat{\beta}_2 D_{22} + \hat{\beta}_3 D_{23} - \left[\hat{\beta}_0 + \hat{\beta}_1 D_{11} + \hat{\beta}_2 D_{12} + \hat{\beta}_3 D_{13}\right] \\
&= \hat{\beta}_0 + \hat{\beta}_1(1) + \hat{\beta}_2(0) + \hat{\beta}_3(0) - \left[\hat{\beta}_0 + \hat{\beta}_1(-1) + \hat{\beta}_2(-1) + \hat{\beta}_3(-1)\right] \\
&= 2\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3
\end{aligned}
$$

and $\widehat{OR}(\text{Blacks, Whites}) = e^{2\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3}$

In our example

$$2\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3 = 2(.765) + .477 + .0719 = 2.0789$$

and $\widehat{OR} = e^{2.0789} = 8,$

which is the same result as we obtained with 0,1 coding.

Now, if we would like to construct a confidence interval for this estimate of the odds ratio we must estimate the variance of the sum of coefficients as follows:

$$\ln\left[\widehat{OR}(\text{Black, White})\right] = 2\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3$$

$$\text{so} \quad \widehat{Var}\left\{\ln\left[\widehat{OR}(\text{Black, White})\right]\right\} = 4\widehat{Var}(\hat{\beta}_1) + \widehat{Var}(\hat{\beta}_2) + \widehat{Var}(\hat{\beta}_3)$$

$$+ 4\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2) + 4\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_3) + 2\widehat{Cov}(\hat{\beta}_2, \hat{\beta}_3)$$

These variances and covariances are available from the computer output where the variance/covariance matrix of the coefficients is presented:

$$\begin{array}{cccc} & \hat{\beta}_1 & \hat{\beta}_2 & \hat{\beta}_3 \\ \hat{\beta}_1 & .123 & -.031 & -.040 \\ \hat{\beta}_2 & & .131 & -.044 \\ \hat{\beta}_3 & & & .148 \end{array}$$

In this example,

$$\widehat{Var}\left\{\ln\left[\widehat{OR}\left(\text{Black, White}\right)\right]\right\} = 4\left(.123\right) + .131 + .148$$
$$+4\left(-.031\right) + 4\left(-.040\right) + 2\left(-.044\right) = 0.400$$

and

$$\widehat{SE}\left\{\ln\left[\widehat{OR}\left(\text{Black, White}\right)\right]\right\} = \sqrt{0.400} = 0.633 \ ,$$

which is identical to the standard error we obtained from the computer earlier with 0,1 coding.

This process is never necessary since we can always use reference cell coding.

```
.  logit chd race_2 race_3 race_4 [fweight= freq]

Iteration 0:   log likelihood = -69.314718
Iteration 1:   log likelihood = -62.368156
Iteration 2:   log likelihood = -62.293897
Iteration 3:   log likelihood = -62.293721

Logit estimates                                   Number of obs   =        100
                                                  LR chi2(3)      =      14.04
                                                  Prob > chi2     =     0.0028
Log likelihood = -62.293721                       Pseudo R2       =     0.1013

------------------------------------------------------------------------------
     chd |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  race_2 |   .7650677   .3505944      2.182   0.029       .0779153     1.45222
  race_3 |   .4773856   .3622841      1.318   0.188      -.2326781    1.187449
  race_4 |   .0719205    .384599      0.187   0.852      -.6818797    .8257208
   _cons |  -.0719205   .2188982     -0.329   0.742      -.5009531    .3571121
------------------------------------------------------------------------------
```

```
.  logistic chd race_2 race_3 race_4 [fweight= freq]

Logit estimates                                   Number of obs   =        100
                                                  LR chi2(3)      =      14.04
                                                  Prob > chi2     =     0.0028
Log likelihood = -62.293721                       Pseudo R2       =     0.1013

------------------------------------------------------------------------------
     chd | Odds Ratio   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  race_2 |    2.14914   .7534764      2.182   0.029       1.081031    4.272589
  race_3 |   1.611855   .5839494      1.318   0.188       .7924086    3.278708
  race_4 |    1.07457   .4132786      0.187   0.852       .5056656    2.283526
------------------------------------------------------------------------------

. vce

         |   race_2   race_3   race_4    _cons
---------+------------------------------------
  race_2 |  .122916
  race_3 |  -.03125   .13125
  race_4 | -.039584  -.04375  .147916
   _cons | -.010416  -.00625  .002084  .047916
```

e.g., ICU Study

let

$$y = \begin{cases} 0 \text{ if patient lives to hospital discharge} \\ 1 \text{ if patient dies in hospital} \end{cases}$$

$$x_1 = \begin{cases} 0 \text{ if patient is not in coma at ICU admission or in coma} <48 \text{ hrs} \\ 1 \text{ if patient has been in coma} \geq 48 \text{ hrs at ICU admission} \end{cases}$$

$$x_2 = \begin{cases} 1 \text{ if patient is not on mechanical ventilation or} <24 \text{ hrs} \\ 2 \text{ if MVENT 1-4 days, no PEEP} \\ 3 \text{ if MVENT } \geq 5 \text{ days, or PEEP} \end{cases}$$

$$x_3 = \text{AGE}$$

To handle x_2, we define 2 dummy variables:

level of MVENT	W_1	W_2
1	0	0
2	1	0
3	0	1

This is called the "PARTIAL" method in BMDPLR or "reference cell" parameterization

let

$$\Pr\left(y=1 \middle| x_1, x_2, x_3\right) = \frac{e^{\beta_0+\beta_1 x_1+\beta_2 W_1+\beta_3 W_2+\beta_4 x_3}}{1+e^{\beta_0+\beta_1 x_1+\beta_2 W_1+\beta_3 W_2+\beta_4 x_3}}$$

and we define the

$$\text{logit } f\left(x_1, x_2, x_3\right) = \ln\left\{\frac{\Pr\left(y=1 \middle| x_1, x_2, x_3\right)}{1-\Pr\left(y=1 \middle| x_1, x_2, x_3\right)}\right\}$$

$$= \beta_0 + \beta_1 x_1 + \beta_2 W_1 + \beta_3 W_2 + \beta_4 x_3$$

Fitting a model with these variables gives rise to the following coefficients and estimated standard errors:

VARIABLE		$\hat{\beta}$	$\widehat{SE}(\hat{\beta})$	$\hat{\beta}/\widehat{SE}(\hat{\beta})$
COMA	(x_1)	2.858	0.418	6.8
MVENT	(W_1)	0.962	0.332	2.9
	(W_2)	1.552	0.385	4.0
AGE	(x_3)	0.034	0.009	3.8
CONSTANT		-4.483	0.579	

Consider a patient 50 years of age, who received no *MVENT*

Then

$$x_3 = 50$$

$$W_1 = 0$$

$$W_2 = 0$$

If the patient is <u>not</u> in COMA, then

$$\begin{aligned}\text{logit } f(x_1,x_2,x_3) &= \text{logit } f(0,0,0,50)\\ &= \beta_0+\beta_1(0)+\beta_2(0)+\beta_3(0)+\beta_4(50)\\ &= -4.483+0+0+0+.034(50)\\ &= -2.783\end{aligned}$$

If the patient <u>is</u> in COMA, then

$$\begin{aligned}\text{logit } f(x_1,W_1,W_2,x_3) &= \text{logit } f(1,0,0,50)\\ &= \beta_0+\beta_1(1)+\beta_2(0)+\beta_3(0)+\beta_4(50)\\ &= -4.483+2.858(1)+0+0+.034(50)\\ &= 0.075\end{aligned}$$

Now, the logit difference is

$$g(1,0,0,50) - g(0,0,0,50)$$

$$= 0.075 - (-2.783) = 2.858 = \hat{\beta}_1 = \text{adjusted log-odds ratio}$$

Hence, $e^{\hat{\beta}_1} = e^{2.858} = 17.43 =$ adjusted odds ratio $= \widehat{OR}$

= estimate of odds of dying for patients in COMA compared to those not in COMA, controlling for AGE and MVENT

In the ICU example

MVENT	W_1	W_2		W_1^*	W_2^*
1	0	0		–1	–1
2	1	0	or	1	0
3	0	1		0	1

We would like to compare level 2 of MVENT to level 1 of MVENT.

- Then, as we know, the log of the estimated odds ratio, holding age and coma constant, is the difference in the respective logits. i.e.,

$$\ln\left[\widehat{OR}(2,1)\right] = \left\{\hat{\beta}_0 + \hat{\beta}_1(a) + \hat{\beta}_2(1) + \hat{\beta}_3(0) + \hat{\beta}_4(v)\right\}$$

$$-\left\{\hat{\beta}_0 + \hat{\beta}_1(a) + \hat{\beta}_2(-1) + \hat{\beta}_3(-1) + \hat{\beta}_4(v)\right\}$$

$$= 2\hat{\beta}_2 + \hat{\beta}_3 = 2(.124) + .714 = .962$$

$$\Rightarrow \quad \widehat{OR}(2,1) = e^{.962} = 2.62$$

$$\widehat{Var}\left\{\ln\left[\widehat{OR}\left(2,1\right)\right]\right\} = 4\widehat{Var}\left(\hat{\beta}_2\right) + \widehat{Var}\left(\hat{\beta}_3\right) + 4\widehat{Cov}\left(\hat{\beta}_2, \hat{\beta}_3\right)$$

$$= 4\left(.216\right)^2 + \left(.244\right)^2 + 4\left(-.034\right) = .1102$$

and $\widehat{SE}\left\{\ln\left[\widehat{OR}\left(2,1\right)\right]\right\} = \sqrt{.1102} = .332$

Then, a 95% CI estimate is

$$\ln\left\{\widehat{OR}\left(2,1\right)\right\} - 1.96\widehat{SE}\left[\ln\left\{\widehat{OR}\left(2,1\right)\right\}\right] \le \ln\left\{OR\left(2,1\right)\right\}$$

$$\le \ln\left\{\widehat{OR}\left(2,1\right)\right\} + 1.96\widehat{SE}\left[\ln\left\{\widehat{OR}\left(2,1\right)\right\}\right]$$

$$.962 - 1.96(.332) \le \ln\{OR(2,1)\} \le .962 + 1.96(.332)$$

$$.311 \le \ln\{OR(2,1)\} \le 1.613$$

$$e^{.311} \le OR(2,1) \le e^{1.613}$$

$$1.37 \le OR(2,1) \le 5.02$$

Confidence intervals for $OR(3,1)$ and $OR(2,3)$ can be obtained similarly.

Alternatively, had we used reference cell coding, then it can be shown that:

$$\ln\left(\widehat{OR}(2,1)\right) = \hat{\beta}_2 \qquad \widehat{SE}\left[\ln\left\{\widehat{OR}(2,1)\right\}\right] = \widehat{SE}\left(\hat{\beta}_2\right)$$

$$\ln\left(\widehat{OR}(3,1)\right) = \hat{\beta}_3 \qquad \widehat{SE}\left[\ln\left\{\widehat{OR}(3,1)\right\}\right] = \widehat{SE}\left(\hat{\beta}_3\right)$$

$$\ln\left(\widehat{OR}(2,3)\right) = \hat{\beta}_2 - \hat{\beta}_3 \qquad \widehat{SE}\left[\ln\left\{\widehat{OR}(2,3)\right\}\right] = \left[\widehat{Var}\left(\hat{\beta}_2\right) + \widehat{Var}\left(\hat{\beta}_3\right) - 2\widehat{Cov}\left(\hat{\beta}_2, \hat{\beta}_3\right)\right]^{1/2}$$

this can be avoided by simply running the program again using a different reference group

Hence the deviation from average logit design variables is computationally much more complex than the referent cell method for the estimation of odds ratios.

Under the assumption that the logit is linear in the continuous variable, x, we have $g(x) = \beta_0 + \beta_1 x$

change in log odds ratio for an increase of "1" unit in x

i.e., $g(x+1) = \beta_0 + \beta_1(x+1) = \beta_0 + \beta_1 x + \beta_1$

$$g(x) = \beta_0 + \beta_1 x$$

$\Rightarrow g(x+1) - g(x) = \beta_1$ for any value of x

Most often the value of "1" will not be biologically very interesting.

e.g.,
- increased risk for 1 additional year of age
- increased risk for 1 additional mmHg in systolic blood pressure
- increased risk for 1 additional mg/100 ml of cholesterol?

A change of <u>5</u> years or <u>20</u> mmHg or <u>10</u> mg/100 ml may be more meaningful.

In other instances, a change of 1 unit might be too much.

e.g., If we had an index ranging from 0 to 1, a change of .01 might be more meaningful.

We develop a method for point and interval estimation for any arbitrary change of c units in the covariate.

The log-odds ratio for a change of c units in x is:

$$g(x+c)-g(x)=c\beta_1$$

and the odds ratio is obtained as

$$OR(x+c,x)=e^{c\beta_1}$$

This is estimated as

$$\widehat{OR}(x+c,x)=e^{c\hat{\beta}_1}$$

$$\widehat{Var}\left[\ln\left\{\widehat{OR}(x+c,x)\right\}\right]=c^2\widehat{Var}\left(\hat{\beta}_1\right)$$

and

$$\widehat{SE}\left[\ln\left\{\widehat{OR}(x+c,x)\right\}\right]=c\widehat{SE}\left(\hat{\beta}_1\right)$$

Our 95% CI is obtained as

$$e^{c\hat{\beta}_1 - 1.96c\widehat{SE}\left(\beta_1\right)} \le OR\left(x + c, x\right) \le e^{c\hat{\beta}_1 + 1.96c\widehat{SE}\left(\hat{\beta}_1\right)}$$

The choice of *c* should be stated clearly in all tables and calculations.

- Use of rounded values of *c* (such as 5 or 10 or 20) is a recommended strategy.
- We must keep in mind that we are attempting to provide the reader with a clear indication of how the risk of the outcome being present changes with the variable in question.

In the AGE, CHD example the estimated logit was

$$\hat{g}(AGE) = -5.31 + 0.11 \times AGE$$

The estimated odds ratio for an increase of 10 years of age is

$$\widehat{OR}(a+10,a) = e^{10(0.111)} = 3.03$$

$\Rightarrow$ for every increase of 10 years of age, the odds of CHD goes up 3.03 times

The 95% confidence interval is

$$e^{10(0.111)-1.96(10)(0.024)} \le OR(a+10,a) \le e^{10(0.111)+1.96(10)(0.024)}$$

$$1.90 \le OR(a+10,a) \le 4.86$$