# The Analysis of Ordinal Data with Graphs and Odds Ratios

Robin High
Department of Biostatistics
University of Nebraska Medical Center
Omaha, NE

University of Nebraska Medical Center

# Power Calculations for Ordinal Data

The outcome of an ordinal response with k levels; sample size to detect a new treatment affect:

- Patient prognosis
- Behaviors
- Attitudes

Reference:

John Whitehead, (1993)
Sample Size for Ordered Categorical Data.
Statist. Med. 12: 2257 – 2271

For a single binary predictor, the numerator of the score test for testing HO: beta=0 is proportional to the Wilcoxon-Mann-Whitney two-sample test statistic

Paper 209-31
Exploiting the Link Between the Wilcoxon-Mann-Whitney Test and a Simple Odds Statistic
Ralph G. O’Brien, Cleveland Clinic Foundation, Cleveland, OH
John Castelloe, SAS Institute Inc., Cary, NC

# Whitehead formula (1993)

$$N = \frac{3*(A+1)^2 * (z_{\alpha/2} + z_{\beta})^2}{[\ A * (\log(OR))^2 * (\ 1 - \text{SUM}\ p_i^3)\ ]}$$

**A = ratio of sample size of Treatment vs Control**

**OR: odds ratio**

**Compute: $(1 - \text{SUM}\ p_i^3)$ < refer to tables on next slides >**

**where $p_i$ is the average of the response probabilities in k ordinal response columns**

# Power Calculations for Ordinal Data

| Probabilities | Very Good | Good | Moderate | Poor |
|---|---|---|---|---|
| Treatment | ? | ? | ? | ? |
| Control | 0.20 | 0.50 | 0.20 | 0.10 |

known

| Cumulative Probabilities | Very Good | Good | Moderate | Poor |
|---|---|---|---|---|
| Treatment | | 0.850 | | |
| Control | 0.20 | 0.70 | 0.90 | 1.00 |

Treatment Impact

Odds Ratio: [ Very Good or Good vs Moderate or Poor ]

$$= \frac{(0.85 / 0.15)}{(0.70 / 0.30)} = 2.43$$

# Power Calculations for Ordinal Data

| Cumulative Probabilities | Very Good | Good | Moderate | Poor |
|---|---|---|---|---|
| Treatment | 0.378 | 0.850 | 0.956 | 1.00 |
| Control | 0.20 | 0.70 | 0.90 | 1.00 |

Odds Ratio: [ Very Good vs (Good or Moderate or Poor) ]

$$= \frac{(\ p11\ /\ 1\text{-}p11)}{(0.20\ /\ 0.80)} = 2.43$$

| Probabilities | Very Good | Good | Moderate | Poor | |
|---|---|---|---|---|---|
| Treatment | 0.378 | 0.472 | 0.106 | 0.044 | |
| Control | 0.20 | 0.50 | 0.20 | 0.10 | ← known |
| Average | 0.289 | 0.486 | 0.153 | 0.072 | ← $p_i$ |

# John Whitehead formula (1993)

**A = 1: equal group sizes**

**OR = 2.4286 (odds ratio derived from cumulative probs)**

**$(1 – \text{SUM } p_i^3) = 0.857$**

**For $\alpha$ = 0.05** $\quad z_{\alpha/2} = 1.96$

**power = 0.9 (ß=.1)** $\quad z_{ß} = 1.282$

**Result in sample size of:**

**N = 187 ( or 94 in each group)**

# Review: Cumulative Logit Equations

```
Four ordinal response levels (k=4)

Probabilities summed across all four response levels for
treatment and control


T: odds11    = (p11              )  / (p12 + p13 + p14)
C: odds21    = (p21              )  / (p22 + p23 + p24)

T: odds12    = (p11 + p12        )  /  (     p13 + p14)
C: odds22    = (p21 + p22        )  /  (     p23 + p24)

T: odds13    = (p11 + p12 + p13)  /  (           p14)
C: odds23    = (p21 + p22 + p23)  /  (           p24)


Row T:  odds Treatment  =   A     /       B
Row C:  odds Control    =   C     /       D
```

# Cumulative Logit Model with k=4
# Specify 4 control probs and 1 odds ratio

three equations in the form

$$\text{Odds Ratio} = \text{OR} = \frac{(\ A \ / \ B\ )}{(\ C \ / \ D\ )}$$

Rearrange components of each equation and set equal to 0

( ( 1/OR ) * A * ( D / C ) ) - B = 0

Solve for the unknown k=4 components of A and B (probabilities from the treatment group) under the restriction the sum of these four individual probabilities equals 1

# Ordinal Logistic Regression: Cumulative Logits

```
Row T:  oddsT  =   A  /  B
Row C:  oddsC  =   C  /  D
```

Odds Ratio: $\text{ODRT} = \dfrac{(A / B)}{(C / D)}$ place all items on one side

```
( (1/ODRT) * ( A ) *  ( (      D       ) / ( C ))) -  (      B        ) = 0
```

```
T: odds11 = p11 / (p12 + p13 + p14)  =  A / B
C: odds21 = p21 / (p22 + p23 + p24)  =  C / D
```

First equation comparing response 1 to responses 2,3,4:

```
( (1/ODRT) * (p11) * ( (p22 + p23 + p24) / (p21))) - (p12 + p13 + p14)= 0
```

Derive two more equations from these comparisons (same odds ratio)

```
T: odds12 = (p11 + p12       )  /  (p13 + p14) = A / B
C: odds22 = (p21 + p22       )  /  (p23 + p24) = C / D

T: odds13 = (p11 + p12 + p13)  /  ( p14      ) = A / B
C: odds23 = (p21 + p22 + p23)  /  ( p24      ) = C / D
```

# Verify Treatment Probabilities

Enter known values: p21 p22 p23 and odrt, the odds ratio

```
DATA cprbs;
p21=0.2; p22=0.5; p23=0.2; p24 = 1-(p21 + p22 + p23); * control group;
odrt = 2.4286;    * odds ratio ;
run;

* Four equations, four unknowns: p11 p12 p13 p14 ;
PROC MODEL DATA = cprbs;
(p11+ p12+ p13 +p14) - 1 = 0;
(( 1/odrt )*(p11            )*( (p22 + p23 + p24) / (p21            ))) - (p12 + p13 + p14)= 0;
(( 1/odrt )*(p11 + p12      )*( (      p23 + p24) / (p21 + p22      ))) - (      p13 + p14)= 0;
(( 1/odrt )*(p11 + p12 + p13)*( (            p24) / (p21 + p22 + p23))) - (            p14)= 0;
SOLVE p11 p12 p13 p14 / out=probs ;
RUN; QUIT;

proc print data=probs noobs; run;
```

| p11 | p12 | p13 | p14 | p21 | p22 | p23 | p24 | odrt |
|---|---|---|---|---|---|---|---|---|
| 0.378 | 0.472 | 0.106 | 0.044 | 0.2 | 0.5 | 0.2 | 0.1 | 2.4286 |

# Make an exemplary data set

| Probabilities | Very Good | Good | Moderate | Poor |
|---|---|---|---|---|
| Treatment | 0.378 | 0.472 | 0.106 | 0.044 |
| Control | 0.20 | 0.50 | 0.20 | 0.10 |

**With 188 subjects, analyze the counts from this table**

| Counts | Very Good | Good | Moderate | Poor | Total |
|---|---|---|---|---|---|
| Treatment | 36 | 44 | 10 | 4 | 94 |
| Control | 19 | 47 | 19 | 9 | 94 |

**Cell counts derived from cell probabilities given 94 subjects in each group**

# Analyze Data with PROC LOGISTIC: Compute Power

```
ODS SELECT nobs responseprofile oddsratios;
ODS OUTPUT globaltests=gblT(where=(substr(test,1,10)='Likelihood‘
                                  or substr(test,1,5) ='Score' ));

PROC LOGISTIC DATA =counts ;
FREQ nn;
CLASS trt(ref=‘Control') / PARAM=ref;
MODEL rsp = trt / link=clogit ;
TITLE 'LOGISTIC: cumulative logit';
run;

* power computed from non-central chisquare ;

DATA pwr; SET gblT;
alpha=0.05 ;
c_crit = CINV(1-alpha,df);
powerC = 1-probchi(c_crit,df,ChiSq);
RUN;
```

# Selected Output

| Effect | Odds Ratio |
|---|---|
| Treatment vs Control | 2.437 |

| Test | ChiSq | total N | c_crit | power |
|---|---|---|---|---|
| Likelihood Ratio | 10.25 | 188 | 3.84 | 0.893 |
| Score | 10.06 | 188 | 3.84 | 0.887 |

# Compute Treatment Probabilities with Prob(Y LE 2) Specified

```
Whitehead example:
 Treatment effect of Very Good or Good increases the cumulative probability
 from 0.7 to 0.85

* Five equations, five unknowns: estimate p11 p12 p13 p14 odrt (the odds ratio)

PROC MODEL DATA = cprbs;
(p11 + p12)                  - 0.85 = 0;
(p11 + p12 + p13 + p14) - 1         = 0;

((1/odrt)*(p11           ) *  ((  p22+p23+p24) /
          (p21           ) ) ) - (p12+p13+p14) = 0;

((1/odrt)*(p11+p12       ) *  ((      p23+p24) /
          (p21+p22       ) ) ) - (    p13+p14) = 0;

((1/odrt)*(p11+p12+p13) *    ((           p24) /
          (p21+p22+p23)) ) - (            p14) = 0;

SOLVE p11 p12 p13 p14 odrt /out=root1a ;
RUN; QUIT;
```

# Compute cell probabilities, derive counts based on N=94, analyze with LOGISTIC, SELECTED Output

| Effect | Odds Ratio |
|---|---|
| Treatment vs Control | 2.437 |

| Test | ChiSq | total N | c_crit | power |
|---|---|---|---|---|
| Likelihood Ratio | 10.25 | 188 | 3.84 | 0.893 |
| Score | 10.06 | 188 | 3.84 | 0.887 |

# Ordinal Logistic Regression: Cumulative Logits

Given:

Odds Ratio = 1.75 and probabilities entered in row 2 for Control

Objective: Compute probabilities in row 1 for the outcomes of Treatment

| Probabilities | Very Good | Good | Moderate | Poor |
|---|---|---|---|---|
| Treatment | 0.4286 | 0.3747 | 0.1370 | 0.0597 |
| Control | 0.3 | 0.4 | 0.2 | 0.1 |

Make exemplary data set for various sample sizes and compute power for each

# Power of OLR: cumulative logits

alpha=0.05  odds ratio = 1.75

| Total N | Total N Whitehead | Power | Estimated Odds Ratio |
|---|---|---|---|
| 120 | 117 | 0.378 | 1.74 |
| 160 | 148 | 0.456 | 1.72 |
| 200 | 195 | 0.565 | 1.74 |
| 300 | 284 | 0.727 | 1.73 |
| 350 | 356 | 0.820 | 1.76 |
| 400 | 404 | 0.865 | 1.76 |
| 450 | 448 | 0.897 | 1.75 |
| 500 | 493 | 0.922 | 1.75 |

# Three Ordinal Response Categories
# Treatment vs Control, two odds ratios

| Counts | Great | So-So | Bad |
|---|---|---|---|
| Treatment | ? | ? | ? |
| Control | 0.2 | 0.33 | 0.47 |

```
odrtA = Great vs So-So/Bad = 1.75
odrtB = Great/So-So vs Bad = 2.25
```

# Three Ordinal Response Categories treatment vs control, two odds ratios

```
DATA cprbs3;
p31 = .2 ; p32 = .33; p33 = 1 - (p31 + p32); * enter control probabilities ;
odrtA = 1.75 ;
odrtB = 2.25 ;   * enter desired odds ratios;
run;

PROC MODEL DATA = cprbs3;
(p11+ p12+ p13) - 1 = 0;
( (1/odrtA) * ((p11    )*(p32+p33)/(p31    )) ) - (p12+p13)  = 0;
( (1/odrtB) * ((p11+p12)*(    p33)/(p31+p32)) ) - (    p13)  = 0;

solve p11 p12 p13 /  out=root3 ;  * solve for treatment probabilities;
RUN; QUIT;
```

| odrtA | odrtB | p11 | p12 | p13 | p21 | p22 | p23 |
|---|---|---|---|---|---|---|---|
| 1.75 | 2.25 | 0.30435 | 0.41295 | 0.28271 | 0.2 | 0.33 | 0.47 |

With PROC LOGISTIC power calculations

| totalN | power | OR_1 | OR_2 |
|---|---|---|---|
| 320 | 0.899 | 1.766 | 2.255 |

# Three Ordinal Response Categories
# Two treatments, Two odds ratios

| Counts | Great | So-So | Bad |
|---|---|---|---|
| Treatment 1 | ? | ? | ? |
| Treatment 2 | ? | ? | ? |
| Control | 0.2 | 0.33 | 0.47 |

```
Treatment 1 vs Control: odrtA = 1.6
Treatment 2 vs Control: odrtB = 1.3
```

# Power with 3 Ordinal Categories
# Two Treatments vs Control, Two Odds Ratios

```
* x=1: treatment 1;      x=2: Treatment 2 ;     * x=3: Control;

DATA cprbs3;
p31 = .2 ; p32 = .33; p33 = 1 - (p31 + p32); * enter control probs ;
odrt1 = 1.6 ;
odrt2 = 1.3 ;   * enter desired odds ratios, treatments vs control;
run;

Six equations, six unknowns

PROC MODEL DATA = cprbs3;
* Treatment 1 vs control;
(p11+ p12+ p13) - 1 = 0;
( (1/odrt1) * ((p11    )*(p32+p33)/(p31    )) ) - (p12+p13)  = 0;
( (1/odrt1) * ((p11+p12)*(    p33)/(p31+p32)) ) - (    p13)  = 0;

* Treatment 2 vs control;
(p21+ p22+ p23) - 1 = 0;
( (1/odrt2) * ((p21    )*(p32+p33)/(p31    )) ) - (p22+p23)  = 0;
( (1/odrt2) * ((p21+p22)*(    p33)/(p31+p32)) ) - (    p23)  = 0;

solve p11 p12 p13
      p21 p22 p23 /  out=root3 ;

RUN; QUIT;
```

# Results

| | Y=1 | Y=2 | Y=3 |
|---|---|---|---|
| Treatment 1 | 0.286 | 0.358 | 0.357 |
| Treatment 2 | 0.245 | 0.349 | 0.406 |
| Control | 0.200 | 0.330 | 0.470 |

alpha=0.05

| total N | power_ likelihood_ ratio | OR_1 | OR_2 |
|---|---|---|---|
| 600 | 0.621 | 1.60 | 1.30 |
| 675 | 0.673 | 1.60 | 1.31 |
| 750 | 0.717 | 1.60 | 1.30 |
| 900 | 0.803 | 1.60 | 1.30 |

# Ordinal Logistic Regression: Adjacent Logits

```
Adjacent logit  (for respective pairs of adjacent cells)

Row T:  oddsT  =      A   /  B
Row C:  oddsC  =      C   /  D
Odds Ratio = ODRT = ( A / B ) / ( C / D )

columns
 1 and 2; ((1/odrt) * (p11 * p22) / p21 ) - p12 = 0
 2 and 3; ((1/odrt) * (p12 * p23) / p22 ) - p13 = 0
 3 and 4; ((1/odrt) * (p13 * p24) / p23 ) - p14 = 0
```

$\boxed{\texttt{p11 + p12 + p13 + p14 = 1}}$

| Probabilities | Very Good | Good | Moderate | Poor |
|---|---|---|---|---|
| Treatment | 0.438 | 0.389 | 0.130 | 0.043 |
| Control | 0.3 | 0.4 | 0.2 | 0.1 |

# Compute Adjacent Logit Treatment Probabilities

```
Enter known control probabilities and desired odds ratio

DATA cntr_probs;
p21 = 0.3; p22 = 0.4; p23 = 0.2 ; p24 = 1 - (p21 + p22 + p23);
odrt = 1.5;
RUN;
* Solve for p11 p12 p13 p14 (four unknowns with four equations) ;

PROC MODEL DATA = cntr_probs;
(p11 + p12 + p13 + p14) - 1 = 0;
*p=3;  ((1/odrt) * (p11 * p22) / p21 ) - p12 = 0;
       ((1/odrt) * (p12 * p23) / p22 ) - p13 = 0;
*p=4;  ((1/odrt) * (p13 * p24) / p23 ) - p14 = 0;

SOLVE p11 p12 p13 p14 / out=root2 ;
RUN; QUIT;
```

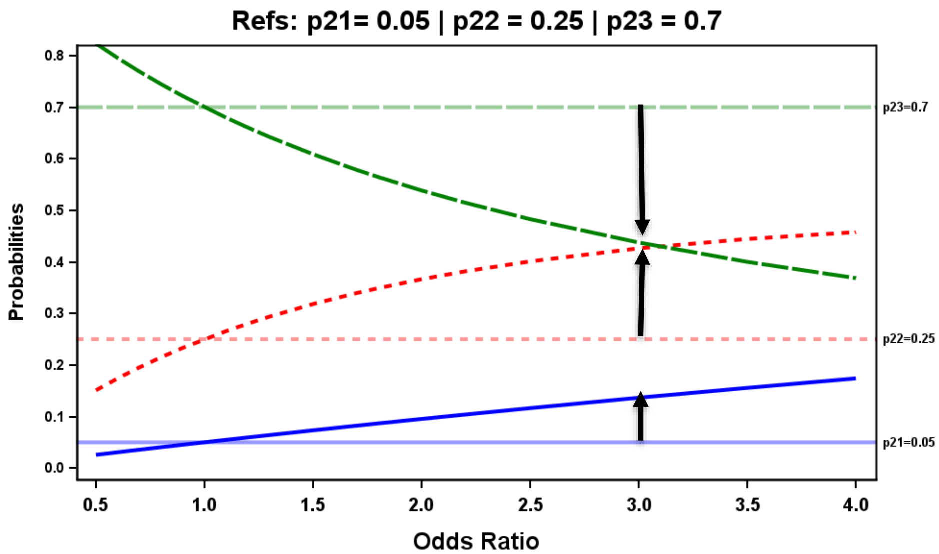
# Power of Logistic Regression Adjacent Logits

Alpha=0.05, Odds Ratio = 1.5

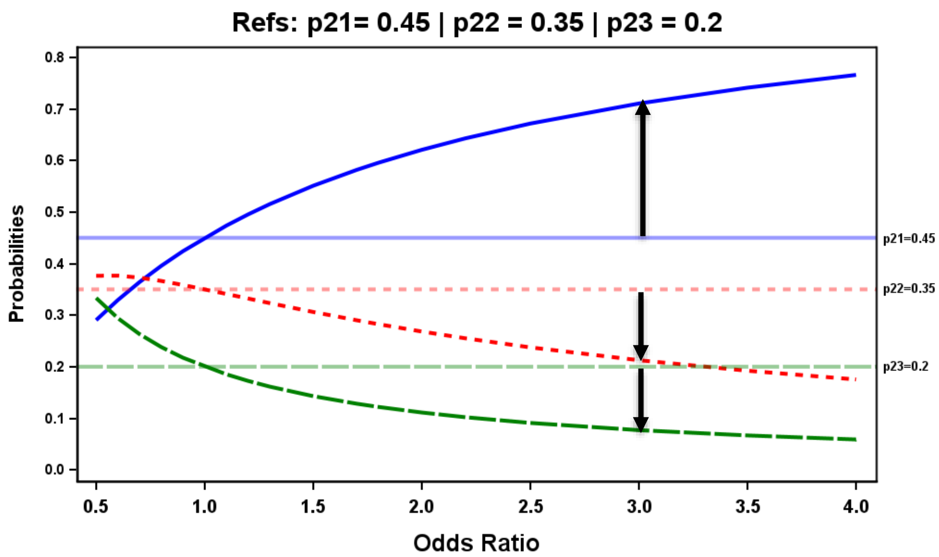
| total N | Power | Estimated Odds Ratio |
|---|---|---|
| 100 | 0.484 | 1.55 |
| 150 | 0.619 | 1.52 |
| 200 | 0.743 | 1.52 |
| 250 | 0.842 | 1.53 |
| 300 | 0.900 | 1.53 |

# What does the odds ratio imply about the treatment effect?

Horizontal lines represent reference category probabilities

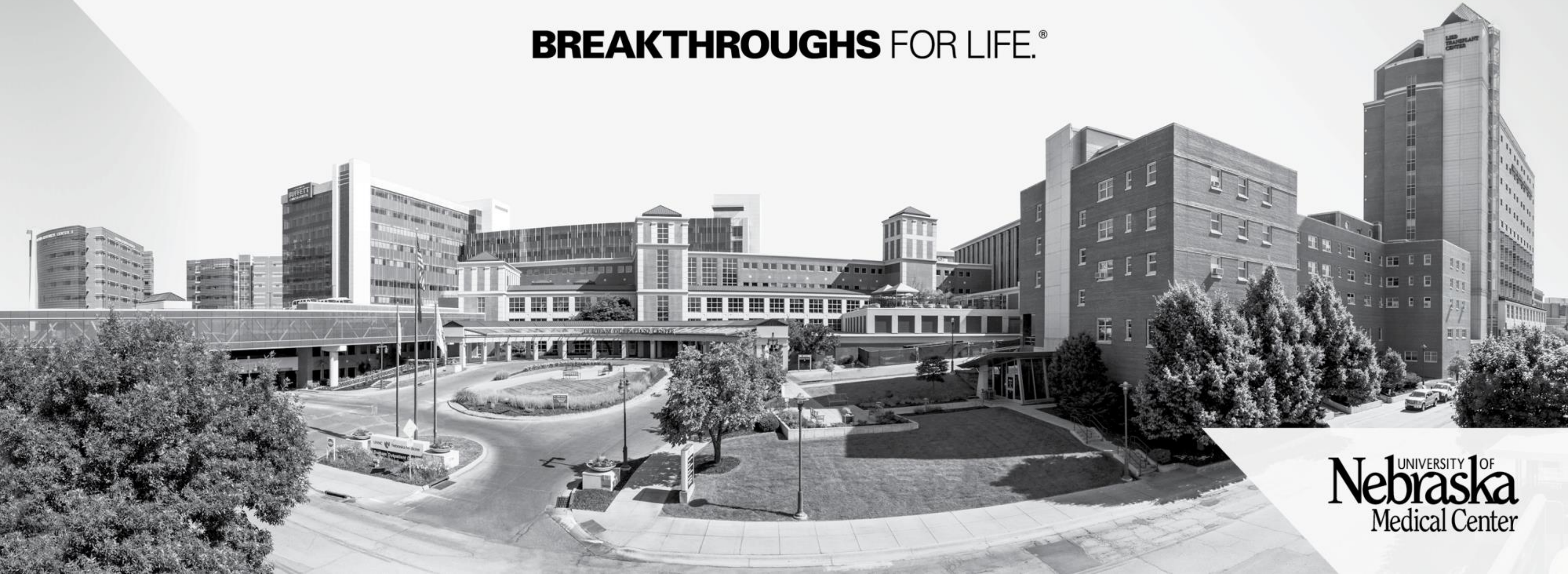



Odds ratio = 3



University of Nebraska Medical Center℠

BREAKTHROUGHS FOR LIFE.®



University of Nebraska Medical Center