

## Data set: Pain 3×3 cross over Trial Data

From the summary table (on our website), we may rearrange the data into the long format data (in table 1) with the following STATA codes:

```
*read in the summary table
infile _000 _100 _010 _001 _110 _101 _011 _111 Total using
c:/data/crossover33.raw,clear

*create the dataset from the summary table
*generate group indicator
gen Group=_n
label define group 1 "PLH" 2 "PHL" 3 "LPH" 4 "LHP" 5 "HPL" 6 "HLP"
label value Group group

gen rel1= _100+ _110+ _101+ _111
gen rel2= _010+ _110+ _011+ _111
gen rel3= _001+ _101+ _011+ _111

gen trt1=0
replace trt1=1 in 3/4
replace trt1=2 in 5/6
gen trt2=0
replace trt2=1 in 1
replace trt2=2 in 2
replace trt2=2 in 4
replace trt2=1 in 6
gen trt3=0
replace trt3=2 in 1
replace trt3=1 in 2
replace trt3=2 in 3
replace trt3=1 in 5

drop _000 _100 _010 _001 _110 _101 _011 _111
reshape long rel trt, i(Group) j(visit)

label define tr 0 "P" 1 "L" 2 "H"
label value trt tr

*generate treatment group indicators
gen low=0
replace low=1 if trt==1
gen high=0
replace high=1 if trt==2

*generate STATA file
save c:\data\cross.dta, replace
*generate a text for SAS
outfile Group visit Total rel trt low high using c:\data\cross.txt,
nolab replace
```

Table 1 Long format data for cross over trial

<i>visit</i>	<i>trt</i>	<i>Group</i>	<i>rel</i>	<i>Total</i>
1	P	PLH	2	15
2	L	PLH	13	15
3	H	PLH	12	15
1	P	PHL	5	16
2	H	PHL	13	16
3	L	PHL	13	16
1	L	LPH	10	15
2	P	LPH	5	15
3	H	LPH	13	15
1	L	LHP	10	12
2	H	LHP	10	12
3	P	LHP	2	12
1	H	HPL	9	14
2	P	HPL	4	14
3	L	HPL	10	14
1	H	HLP	12	14
2	L	HLP	5	14
3	P	HLP	4	14

Outcome: Binomial response, percent of patients experienced pain relief ( $rel/Total$ )

Covariate: Drug level, P=Placebo, L=Low, H=High

Scientific question: Does different treatment relive patients' pain?

## 1 Logistic Regression by assuming independent responses

- Model:

$$\log it(E(Y_{ij})) = \log\left(\frac{P(Y_{ij} = 1)}{1 - P(Y_{ij} = 1)}\right) = \beta_0 + \beta_1 I(Low) + \beta_2 I(High)$$

- STATA Command:

glm outcomes covariates, fam(bin N) link(logit)

- SAS Procedure:

PROC LOGISTIC DATA= data name;

MODEL events/trials=covariates;

RUN;

• STATA Output:

```
. glm rel low high, f(bin Total) l(logit)

Residual df =          15                No. of obs =          18
Pearson X2   = 17.62081                Deviance   = 17.01392
Dispersion  =  1.17472                Dispersion =  1.134261
```

Binomial (N=Total) distribution, logit link

rel	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
low	1.959839	.3427433	5.718	0.000	1.288074 2.631603
high	2.468734	.3666004	6.734	0.000	1.75021 3.187257
_cons	-1.067841	.2471428	-4.321	0.000	-1.552232 -.5834497

```
. glm rel low high, f(bin Total) l(logit) eform
```

```
Residual df =          15                No. of obs =          18
Pearson X2   = 17.62081                Deviance   = 17.01392
Dispersion  =  1.17472                Dispersion =  1.134261
```

Binomial (N=Total) distribution, logit link

rel	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
low	7.098182	2.432854	5.718	0.000	3.625797 13.89603
high	11.80749	4.328629	6.734	0.000	5.755813 24.2219

• SAS Output:

```
LIBNAME lab "c:\data";

DATA cross;
INFILE 'c:\data\cross.txt';
INPUT Group visit Total rel trt low high;
RUN;

PROC LOGISTIC DATA=cross;
MODEL rel/Total= low high;
RUN;
```

The LOGISTIC Procedure  
Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Chi-Square	Pr > ChiSq
Intercept	1	-1.0678	0.2471	18.6688	<.0001
low	1	1.9598	0.3427	32.6966	<.0001
high	1	2.4687	0.3666	45.3482	<.0001

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits
low	7.098	3.626 13.896
high	11.807	5.756 24.221

## 2 Marginal Logistic Regression

- Model:

$$\text{Mean model: } \text{logit}(E(Y_{ij})) = \log\left(\frac{P(Y_{ij} = 1)}{1 - P(Y_{ij} = 1)}\right) = \beta_0^* + \beta_1^* I(\text{Low}) + \beta_2^* I(\text{High})$$

$$\text{Variance: } \text{Var}(Y_{ij}) = E(Y_{ij})(1 - Y_{ij})$$

Correlation: choose from exchangeable, AR1, unstructured for GEE

- STATA Command

xtgee outcomes covariates, i(id) fam(bin N) link(logit) corr(?) robust

- SAS Procedure

PROC GENMOD DATA= data name;

CLASS ID;

MODEL events/trials=covariates / dis=bin link=logit;

REPEATED SUBJECT=ID/TYPE=? (IND, CS,AR,UN)

RUN;

- STATA Output

Exchangeable Correlation (model based standard error):

```
. xtgee rel low high, i(Group) f(bin Total) l(logit) corr(exc)
```

```
GEE population-averaged model
Group variable:          Group      Number of obs      =      18
Link:                   logit      Number of groups   =       6
Family:                 binomial   Obs per group: min =       3
Correlation:           exchangeable  avg                =      3.0
Scale parameter:       1          max                =       3
                          Wald chi2(2)    =      43.26
                          Prob > chi2     =      0.0000
```

```
-----+-----
      rel |      Coef.   Std. Err.      z    P>|z|      [95% Conf. Interval]
-----+-----
      low |  1.959839   .3799878     5.158  0.000   1.215076   2.704601
      high |  2.468734   .40631      6.076  0.000   1.672381   3.265087
      _cons | -1.067841   .2471428    -4.321  0.000  -1.552232  -.5834497
-----+-----
```

```
. xtcorr
```

Estimated within-Group correlation matrix R:

```
      c1      c2      c3
r1  1.0000
r2 -0.2293  1.0000
r3 -0.2293 -0.2293  1.0000
```

**Exchangeable Correlation (robust standard error):**

. xtgee rel low high, i(Group) f(bin Total) l(logit) corr(exc) robust

Iteration 1: tolerance = 7.502e-16

```
GEE population-averaged model          Number of obs      =      18
Group variable:                        Group              Number of groups   =      6
Link:                                   logit              Obs per group: min =      3
Family:                                 binomial           avg                =      3.0
Correlation:                            exchangeable       max                =      3
                                           Wald chi2(2)       =      87.20
Scale parameter:                        1                 Prob > chi2        =      0.0000
```

(standard errors adjusted for clustering on Group)

rel	Coef.	Semi-robust Std. Err.	z	P> z	[95% Conf. Interval]	
low	1.959839	.4749766	4.13	0.000	1.028902	2.890776
high	2.468734	.2666482	9.26	0.000	1.946113	2.991355
_cons	-1.067841	.177592	-6.01	0.000	-1.415915	-.7197666

. xtcorr

Estimated within-Group correlation matrix R:

	c1	c2	c3
r1	1.0000		
r2	-0.2293	1.0000	
r3	-0.2293	-0.2293	1.0000

• **SAS Output**

**Exchangeable Correlation (model based standard error):**

```
PROC GENMOD DATA=cross;
CLASS Group;
MODEL rel/Total= low high /DIST=bin LINK=logit;
REPEATED subject=Group /TYPE=cs MODELSE;
run;
```

Working Correlation Matrix

	Col1	Col2	Col3
Row1	1.0000	-0.2293	-0.2293
Row2	-0.2293	1.0000	-0.2293
Row3	-0.2293	-0.2293	1.0000

Analysis Of GEE Parameter Estimates  
Model-Based Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr >  Z
Intercept	-1.0678	0.2679	-1.5928	-0.5428	-3.99	<.0001
low	1.9598	0.4118	1.1526	2.7670	4.76	<.0001
high	2.4687	0.4404	1.6056	3.3319	5.61	<.0001
Scale	1.0838	.	.	.	.	.

Exchangeable Correlation (robust standard error):

```
PROC GENMOD DATA=cross;
CLASS Group;
MODEL rel/Total= low high /DIST=bin LINK=logit;
REPEATED subject=Group /TYPE=cs CORRW;
RUN;
```

Working Correlation Matrix

	Col1	Col2	Col3
Row1	1.0000	-0.2293	-0.2293
Row2	-0.2293	1.0000	-0.2293
Row3	-0.2293	-0.2293	1.0000

Analysis Of GEE Parameter Estimates  
Empirical Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr >  Z
Intercept	-1.0678	0.1621	-1.3856	-0.7501	-6.59	<.0001
low	1.9598	0.4336	1.1100	2.8097	4.52	<.0001
high	2.4687	0.2434	1.9916	2.9458	10.14	<.0001