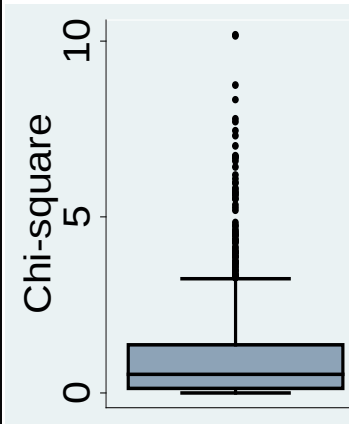


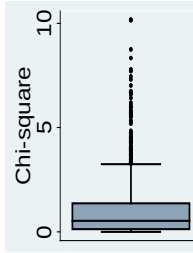
Repeated Measures, Lecture 3, Part 2

*Charles E. McCulloch,
Division of Biostatistics,
Dept of Epidemiology and Biostatistics,
UCSF*



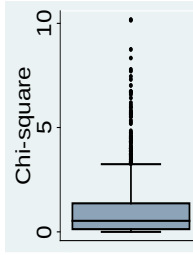
Last time

1. More on **mixed** and **xtgee**
2. Examples
3. Robust standard errors

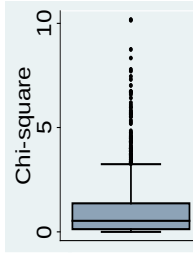


Outline

1. Binary outcomes
2. Changing the link function
3. Modeling practice
4. Summary



Binary outcomes/logistic regression



- Now let's take a look at the use of xtgee for clustered logistic regression. I took the Georgia babies data set and dichotomized it as to whether birthweight was above or below 3000 grams.
- What options do we use now for family, link and correlation structure?

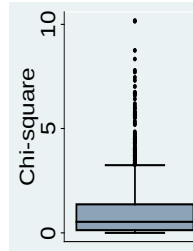
family = binomial

link = logit

corr = exchangeable

Binary outcomes: low birthweight

```
xtgee lowbirth birthord initage, i(momid) family(binomial)
```



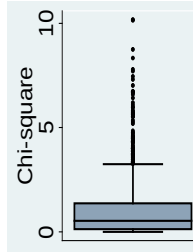
GEE population-averaged model		Number of obs	=	1000
Group variable:	momid	Number of groups	=	200
Link:	logit	Obs per group: min	=	5
Family:	binomial	avg	=	5.0
Correlation:	exchangeable	max	=	5
		Wald chi2(2)	=	11.30
Scale parameter:	1	Prob > chi2	=	0.0035

lowbirth	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
birthord	-.0829363	.0390214	-2.125	0.034	-.159417 - .0064557
initage	-.089028	.0337755	-2.636	0.008	-.1552267 - .0228293
_cons	1.267884	.6036077	2.101	0.036	.0848346 2.450933

How does this compare to the continuous data fit?

OK to assume exchangeable?

Binary outcomes: robust option



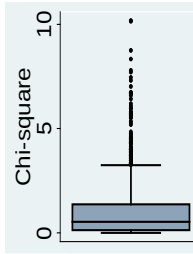
```
xtgee lowbirth birthord initag, i(momid) family(bino) robust
```

```
GEE population-averaged model
Group variable:          momid      Number of obs      =      1000
Link:                   logit        Number of groups   =      200
Family:                 binomial      Obs per group: min =      5
Correlation:           exchangeable   avg              =     5.0
                                           max              =      5
                                           Wald chi2(2)      =     10.64
Scale parameter:        1            Prob > chi2         =     0.0049
```

(Std. Err. adjusted for clustering on momid)

	Semi-robust					
lowbirth	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
+-----						
birthord	-.0829363	.0384829	-2.16	0.031	-.1583614	-.0075113
initage	-.089028	.0341776	-2.60	0.009	-.1560149	-.0220412
_cons	1.267884	.6099252	2.08	0.038	.0724524	2.463315

Binary outcomes: low birthweight



```
. xtgee lowbirth initage birthord, i(momid) link(logit) family(binomial) robust ef
```

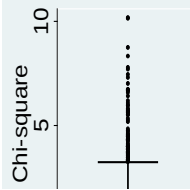
```
GEE population-averaged model                Number of obs      =       1000
Group variable:                             momid             Number of groups   =       200
Link:                                         logit               Obs per group: min =        5
Family:                                     binomial              avg =       5.0
Correlation:                               exchangeable         max =        5
                                                Wald chi2(2)         =      10.64
Scale parameter:                            1                   Prob > chi2         =      0.0049
```

(Std. Err. adjusted for clustering on momid)

lowbirth			Robust				
		Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
-----+-----							
initage		.9148199	.0312663	-2.60	0.009	.8555464	.9781999
birthord		.9204098	.03542	-2.16	0.031	.8535413	.9925168
_cons		3.553325	2.167263	2.08	0.038	1.075142	11.74368



Low birthweight: mixed effects model



```
. meglm lowbirth initage birthord ||momid:, link(logit) family(binomial) or
```

```
Mixed-effects GLM                                Number of obs      =       1000
Family:                                binomial
Link:                                  logit
Group variable:                            momid

Number of groups      =        200
Obs per group: min    =         5
                  avg  =        5.0
                  max  =         5
```

```
Integration method: mvaghermite                    Integration points =         7
Wald chi2(2)              =        11.90
Prob > chi2                =        0.0026

Log likelihood = -588.05261
```

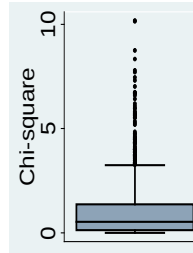
lowbirth	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
initage	.8808697	.0404965	-2.76	0.006	.804969 .9639271
birthord	.8872647	.0500711	-2.12	0.034	.7943597 .9910354
_cons	6.00727	4.979094	2.16	0.031	1.183486 30.49236

momid	
var(_cons)	2.595213 .5419826 3.907845

```
LR test vs. logistic regression: chi2 = 0.00008
```

ORs are further from 1 for mixed effects models, though p-values are usually similar. Slightly different interpretation (subject specific).

Changing the link function



What kind of model would this command fit?

```
xtgee bweight birthord initage, i(momid) link(log)
```

GEE population-averaged model		Number of obs	=	1000
Group variable:	momid	Number of groups	=	200
Link:	log	Obs per group: min	=	5
Family:	Gaussian	avg	=	5.0
Correlation:	exchangeable	max	=	5
		Wald chi2(2)	=	30.56
Scale parameter:	324595.5	Prob > chi2	=	0.0000

bweight	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
birthord	.0147553	.0031742	4.648	0.000	.0085339	.0209767
initage	.008179	.0027336	2.992	0.003	.0028212	.0135368
_cons	7.862211	.0503139	156.263	0.000	7.763598	7.960825

Interpretations:

Changing the link function

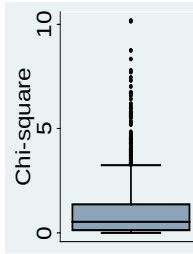
The log link implies we are modeling the log of the mean birthweight. The distribution has defaulted to normal. So our model is

$$\ln(\text{mean birthweight}) = \beta_0 + \beta_1 \text{birthord} + \beta_2 \text{initage}$$

or

$$\text{mean birthweight} = \exp(\beta_0 + \beta_1 \text{birthord} + \beta_2 \text{initage})$$

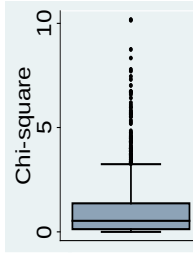
Therefore the mean birthweight for a second birth at an initial age of 22 would be $\exp(\beta_0 + \beta_1 * 2 + \beta_2 * 22)$. Similarly for a third birth and the same initial age, it would be $\exp(\beta_0 + \beta_1 * 3 + \beta_2 * 22)$.



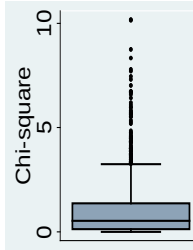
Changing the link function

If we divide the value for birth 3 by the value for birth 2 then the β_0 and $\beta_2 \cdot 22$ terms cancel out leaving just $\exp(\beta_1 \cdot 3 - \beta_1 \cdot 2) = \exp(\beta_1)$. So that is the ratio increase in birthweight when you increase birth order by 1 and hold initial weight constant.

So with each increase of birth order by one the mean birthweight is predicted to change by $\exp(\beta_1) \times 100\%$, or in our case by $\exp(0.0148) = 1.015 \times 100\% = 101.5\%$ or an increase of 1.5%.



Changing the link function



You can ask Stata to exponentiate for you using the `eform` option:

```
xtgee bweight birthord initage, i(momid) link(log) eform
```

```
GEE population-averaged model
Group variable:                momid
Link:                          log
Family:                        Gaussian
Correlation:                   exchangeable

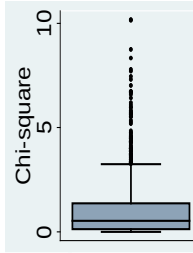
Number of obs      =      1,000
Number of groups   =       200
Obs per group:
    min            =          5
    avg            =         5.0
    max            =          5

Wald chi2(2)       =      30.56
Prob > chi2        =      0.0000

Scale parameter:      324595.5
```

bweight	exp(b)	Std. Err.	z	P> z	[95% Conf. Interval]	
birthord	1.014865	.0032214	4.65	0.000	1.00857	1.021198
initage	1.008213	.0027561	2.99	0.003	1.002825	1.013629
_cons	2597.258	130.6783	156.26	0.000	2353.357	2866.437

Changing the link function



What kind of model would this command fit?

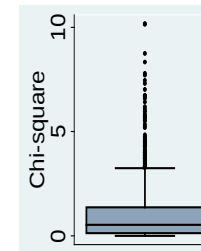
```
xtgee lowbirth birthord initage,  
      i(momid) family(bino) link(log)
```

The math is exactly the same as the normal distribution model. The coefficient for birth order will tell us the ratio increase in the mean value, which for a binomial model is the probability of the outcome.

Ratio increase in the probability is known as the (drum roll)

Relative risk.

Changing the link function (cont.)

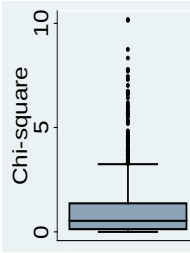


To fit a relative risk model for binary data *with or without clustering* (think of the same question using cross-sectional data) some people recommend the following model. Why should this work?

```
xtgee lowbirth birthord initage,  
      i(momid) family(poisson) robust
```

Log link doesn't constrain probabilities to be within (0,1). So change the distribution, but use robust to fix the incorrect SEs due to the Poisson model.

Changing the link function



Output for binary outcome, log link:

```
xtgee lowbirth birthord initage, i(momid) family(poisson) robust eform
```

```
GEE population-averaged model
Group variable:          momid      Number of obs      =      1,000
Link:                    log         Number of groups   =       200
Family:                  Poisson     Obs per group:
Correlation:             exchangeable      min =          5
                                           avg  =          5.0
                                           max  =          5
                                           Wald chi2(2)    =      10.44
Scale parameter:        1           Prob > chi2        =      0.0054
```

(Std. Err. adjusted for clustering on momid)

lowbirth		IRR	Robust Std. Err.	z	P> z	[95% Conf. Interval]
-----+-----						
birthord		.9500198	.0227377	-2.14	0.032	.9064839 .9956465
initage		.9426217	.0219578	-2.54	0.011	.900553 .9866557
_cons		1.197149	.4797687	0.45	0.653	.5457846 2.625882

Modeling practice 1

Epileptics were randomly allocated to a placebo or an anti-seizure drug (Progabide) group. The number of seizures was recorded during a baseline period and for four periods after beginning treatment.

Is the drug effective at reducing the number of seizures? GEE or mixed model? What are the family, link, correlation structure, and predictors?

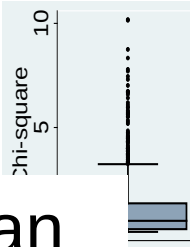
GEE/mixed = GEE (easier), robust available

family = Poisson

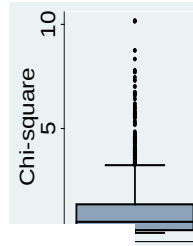
link = log

corr = exchangeable with robust

predictors = treatment, period, interaction



Modeling practice 2



A quality improvement program was designed to reduce prescription of antibiotics for antibiotic resistant infections in emergency departments. 16 hospitals were randomized to the program or control group. Is the program effective? GEE or mixed model? What are the family, link, correlation structure, and predictors?

GEE/mixed = mixed (not enough clusters)

family = binomial (yes/no antibiotic)

link = logit or log

corr = random hospital effect

predictors = treatment, before/after, interaction

Modeling practice 3

The OAI is a longitudinal cohort study of osteoarthritis. ~5000 participants were studied yearly for 5 years. Is sex, presence of a lesion on the MRI, or race associated with the change in pain over time? Pain is assessed using a numerical scale called the WOMAC. GEE or mixed model? What are the family, link, correlation structure, and predictors?

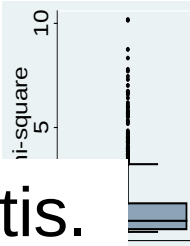
GEE/mixed = GEE (easier), robust available

family = Gaussian

link = identity

corr = exchangeable with robust

predictors = sex, MRI lesion, race, time, race X time interaction, sex X time interaction



Summary

- Hierarchical data structures are common.
- Lead to correlated data.
- Ignoring the correlation can be a serious error.
- `xtgee` can handle single level of clustering and a variety of outcome types.
- `mixed` can handle multiple levels for normally distributed data.
- Not discussed in class: `meologit` and `mepoisson` can handle multiple levels of clustering for binary and Poisson outcomes. Random effects models and robust SEs are also available for time-to-event data.
- GEE methods have the advantage of robust SEs.
- Mixed effects models have the advantage of being able to handle more general types of designs and partition variability.

