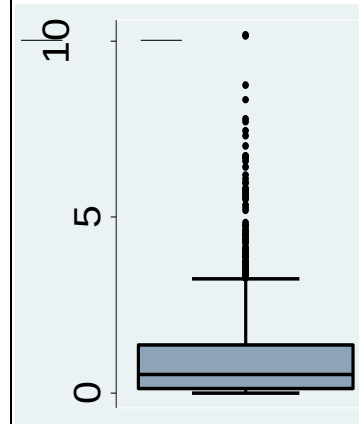


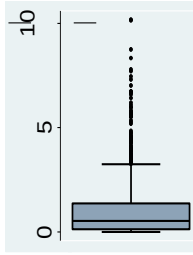
Repeated Measures, Part 2

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

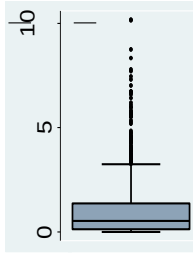


May, 2015

Outline



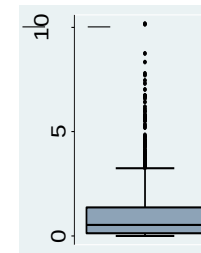
1. Two overarching analysis approaches to accommodating correlation
2. Longitudinal designs
3. GEE (`xtgee`) methods
4. Mixed models (`mixed`)
5. Model checking
6. Summary



Analyses for correlated data

- Generalized estimating equations (GEEs)
 - Stata `xtgee`
- Mixed effects (ME) models
 - Stata `meglm`
 - Or specialized routines for different outcome types
 - Numeric (`mixed`)
 - Binary (`meologit`, `meprobit`)
 - Count (`mepoisson`)
 - Ordered categorical (`meologit`, `meoprobit`)
- *Interpretations are virtually the same for corresponding correlated and non-correlated analyses (linear/linear, logit/logit)*

Analyses for correlated data

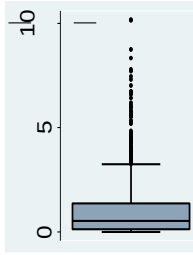


- Generalized estimating equations (GEEs)
 - Only works for one level of clustering (e.g., patients).
 - Requires largish number of clusters (ideally > 50) and works best when the number of clusters $\gg$ number of observations per cluster.
 - Does not require modeling of the correlation when using “robust” standard errors.
 - No overall measure of model fit like likelihood or R^2 .
 - Less robust than mixed effects models to possible bias due to missing data or dropout.

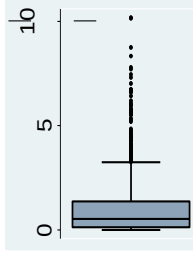
Analyses for correlated data

● Mixed models

- Can handle a wider variety of data structures. For example times within patients within doctors or compartments within knees within time within subjects.
- Requires modeling the correlation.
- Gives additional information about the correlation.
- Based on likelihood fits, so can conduct likelihood ratio tests to compare two model fits.
- More robust to potential bias due to missing data and dropout.



Consider first longitudinal data



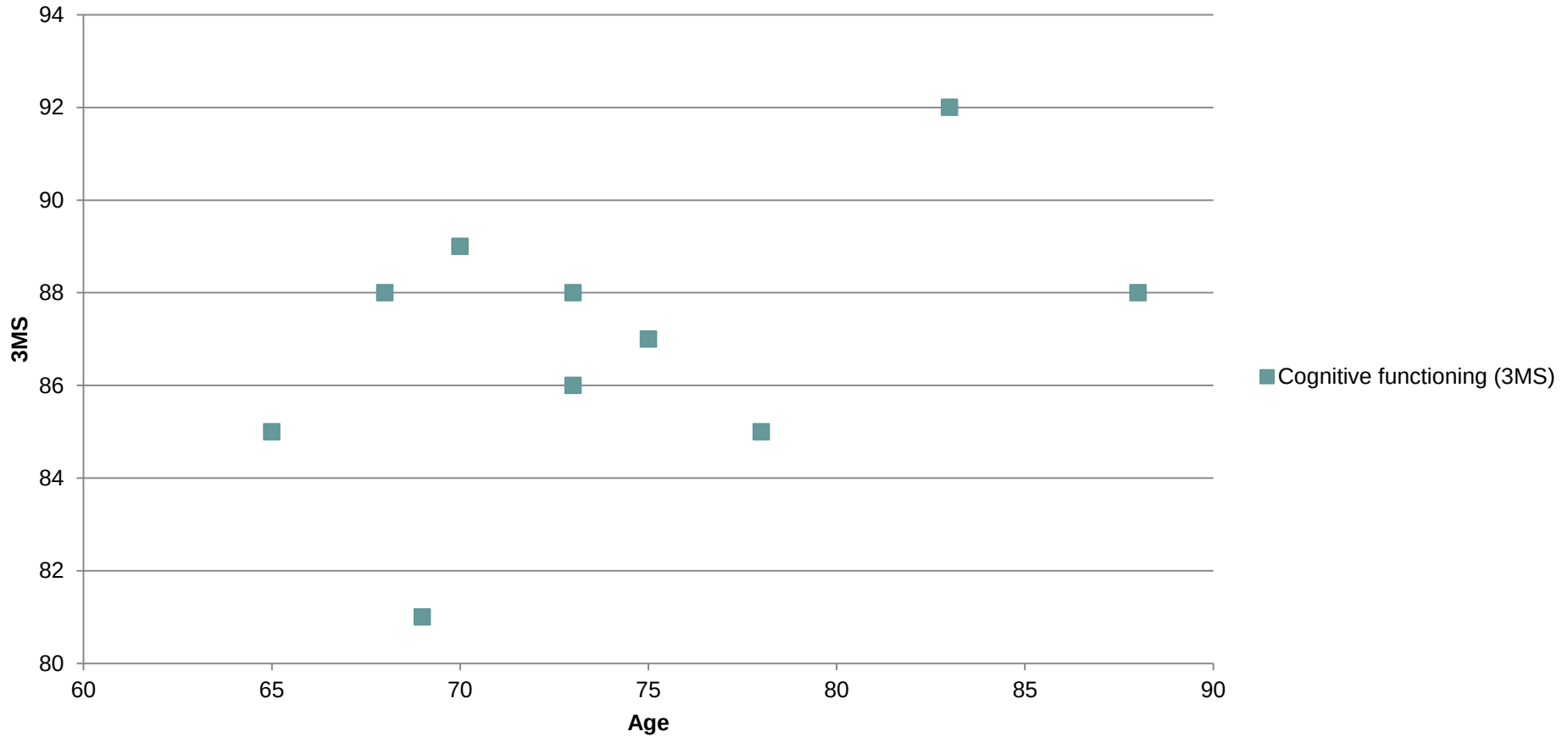
Longitudinal design: Individuals (interpreted broadly) are measured repeatedly over time.

Key features: Can analyze change over time within an individual.

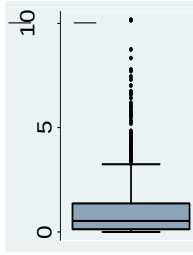
GEE methods are a very common approach for longitudinal data when there are large numbers of individuals and not so many repeated measurements.

Measuring change

Cognitive functioning (3MS)



Modified Mini-mental State Exam



Numerical score from 0 to 100

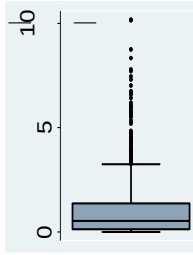
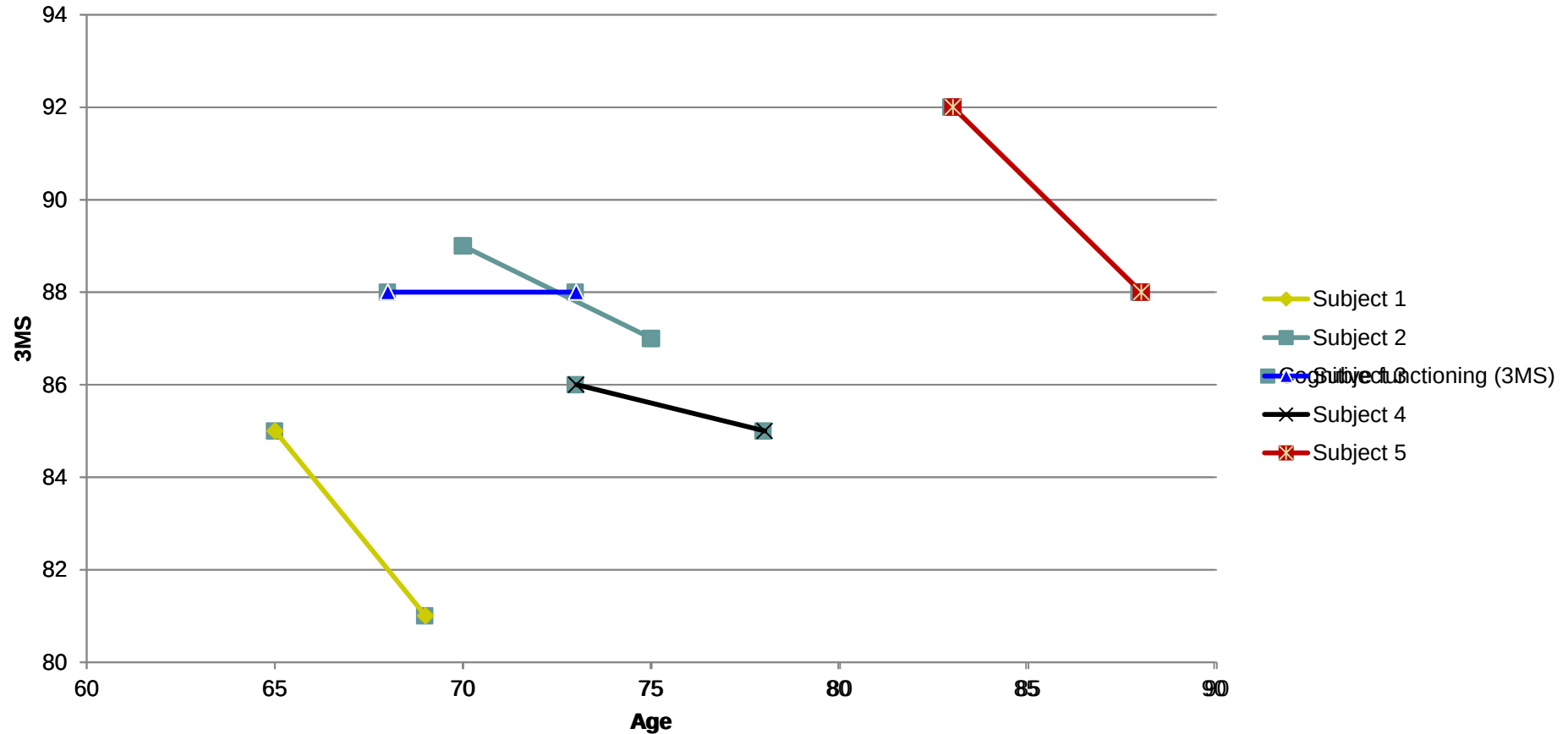
Tell patient “I'd like to test your memory; say these words: boat, cucumber, wire” (up to 3 points for 3 correct answers)

Tell patient “Begin with 100 and count backwards by 7” (up to 5 points for five correct answers)

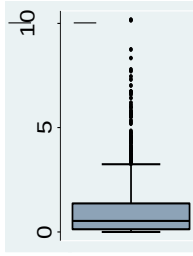
Tell patient “Can you name the three objects I named before?” (up to 3 points for 3 correct answers)

Measuring change

Cognitive functioning (3MS)

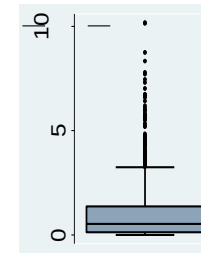


Analyzing change with longitudinal data



Example (SOF): It is well known that women who are overweight and obese have higher bone density. However, is the *change* in bone mineral density related to whether a woman is obese at baseline? We will define obese as $\text{BMI} > 30$.

Example: SOF basics



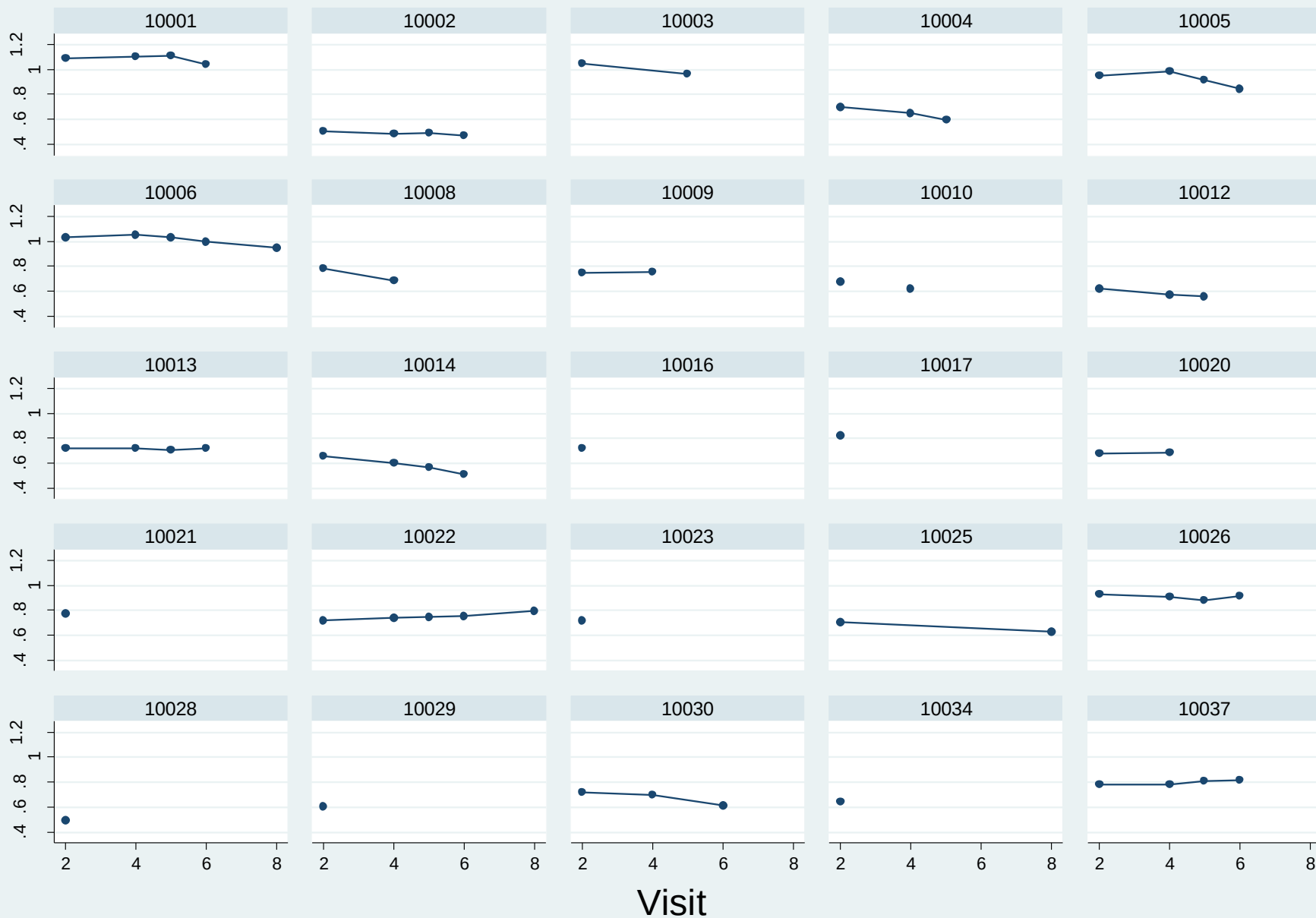
```
. xtset id visit
      panel variable:   id (unbalanced)
      time variable:   visit, 2 to 8, but with gaps
                  delta: 1 unit

. xtodes
      id:  10001, 10002, ..., 42486          n =          8590
      visit:  2, 4, ..., 8                  T =              5
      Delta(visit) = 1 unit      Span(visit)  = 7 periods
      (id*visit uniquely identifies each observation)
```

```
Distribution of T_i:   min   5%   25%   50%   75%   95%   max
                      1    1     2     3     5     5     5
```

Freq.	Percent	Cum.	Pattern
2275	26.48	26.48	1.111.1
1734	20.19	46.67	1.111..
1696	19.74	66.41	1.....
926	10.78	77.19	1.11...
...			
99	1.15	90.94	1.1.1..
89	1.04	91.98	1..11.1
689	8.02	100.00	(other patterns)
8590	100.00		X.XXX.X

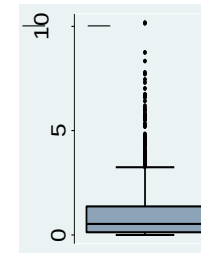
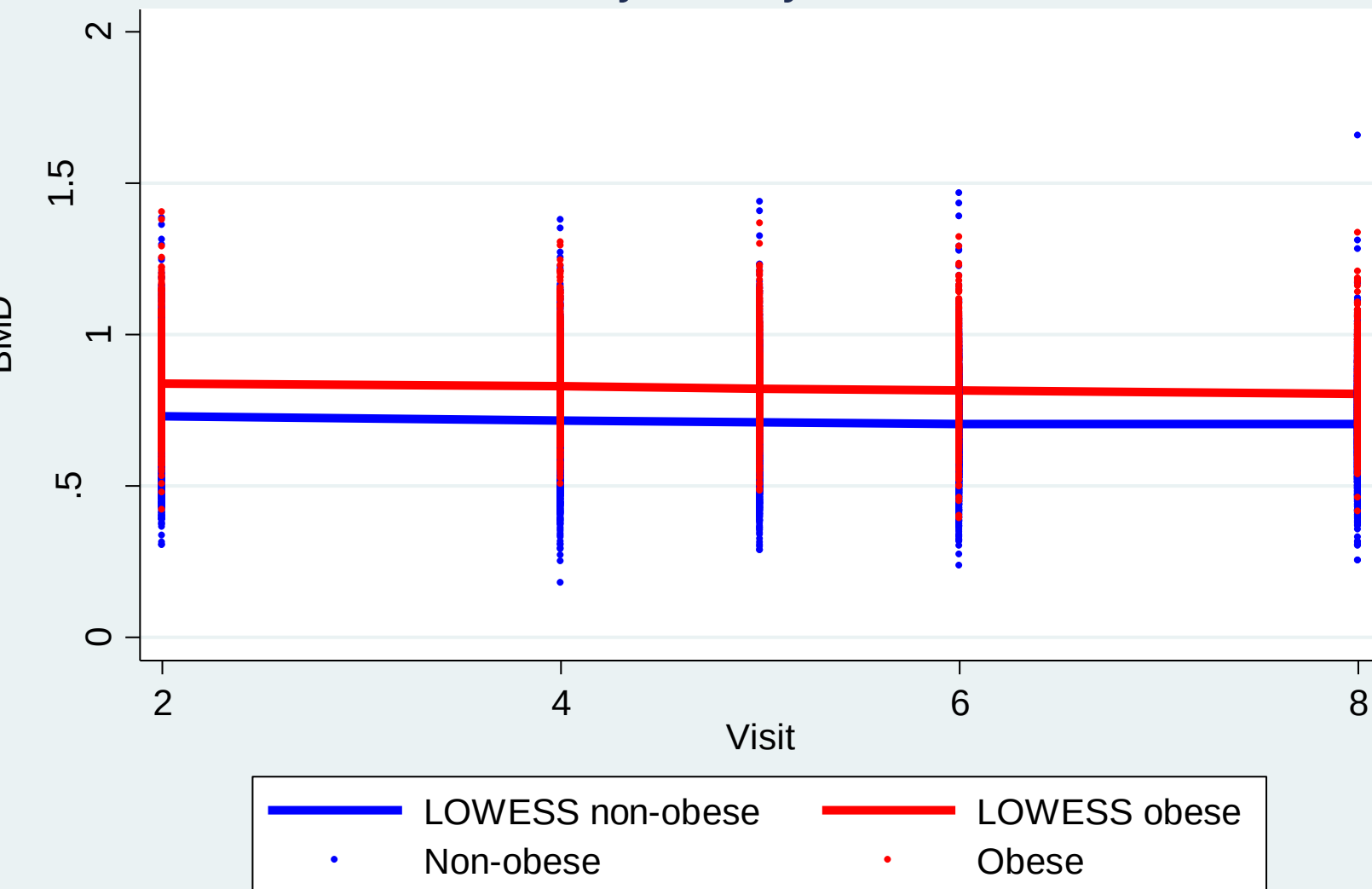
BMD



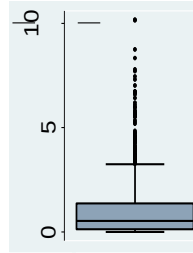
Graphs by ID

Example: BMD/Obesity

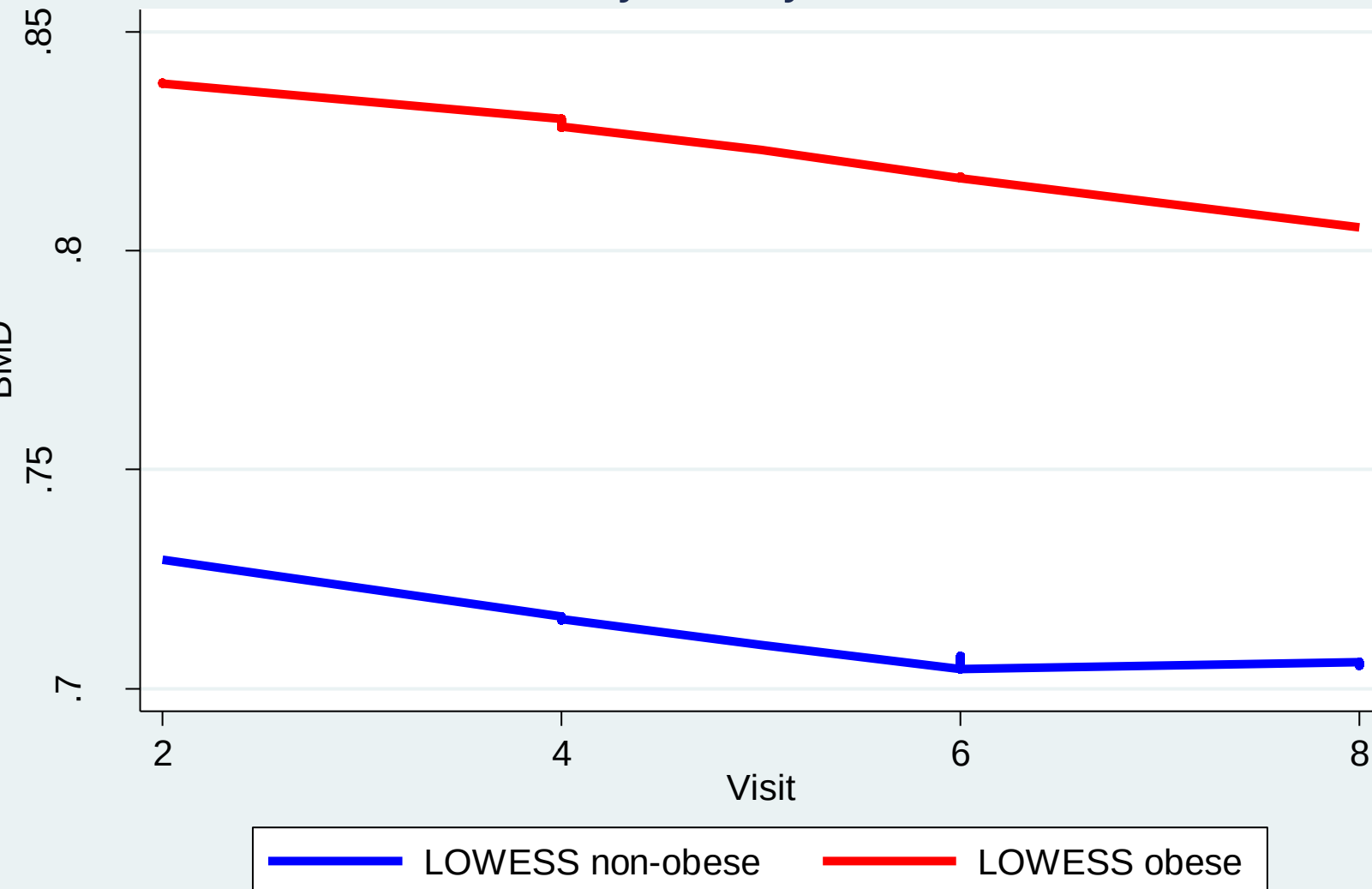
BMD vs visit by obesity status at baseline



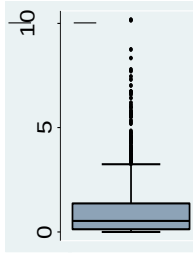
Example: BMD/Obesity



BMD vs visit by obesity status at baseline

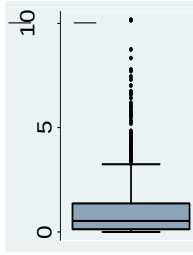


Analyzing change with longitudinal data



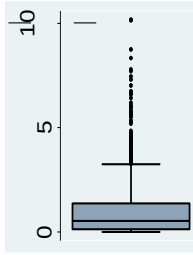
- Including a variable for *time* (or *visit*) describes the change over time.
- Inclusion of *time* (or *visit*) interactions with baseline predictors allows analysis of whether baseline predictors are associated with change over time.
- Inclusion of a time-varying predictor (e.g., BMI at sequential visits) allows analysis of whether change in that predictor is associated with change in the outcome.

Analyzing change with longitudinal data



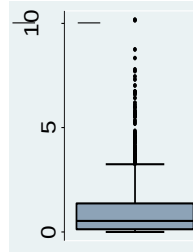
- Analyzing *trajectories* usually implies a functional form over time.
- There is a natural “ladder” of handling a time predictor like visit, moving from simpler (to model *and* interpret) and more restrictive to more flexible: linear, quadratic, spline (flexible smooth fit), categorical.
- Moving up the “ladder” is a simple way to test adequacy of the simpler model.

Example: BMD/Obesity



- Want to characterize the change over time and see if it is the same or different between the obese and non-obese.
- Plot suggests that the trend might not be linear over time.
- So we need to check.

Example: BMD/Obesity



```
. xtgee totbmd visit##obese, i(id) robust
```

Summary information about hierarchy

GEE population-averaged model

Group variable:

id

Link:

identity

Family:

Overall model test (rarely useful)

Gaussian

Correlation:

exchangeable

Scale parameter:

.0169246

Number of obs = 26829

Number of groups = 8468

Obs per group: min = 1

avg = 3.2

max = 5

Wald chi2(9) = 3733.60

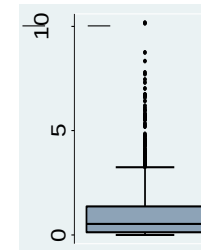
Prob > chi2 = 0.0000

(Std. Err. adjusted for clustering on id)

		Robust				
		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
totbmd						
visit						
4		-.0164933	.0004961	-33.25	0.000	-.0174655 -.015521
5		-.0270842	.0006331	-42.78	0.000	-.028325 -.0258434

etc.

If there are interactions between two variables in the model, the “main effect” of each of those variables is the effect when the other is at its reference value.



		Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
visit							
4		-.0164933	.0004961	-33.25	0.000	-.0174655	-.015521
5		-.0270842	.0006331	-42.78	0.000	-.028325	-.0258434
6		-.0383912	.0007898	-48.			432
8		-.0513558	.001284	-40.			391
1.obese		.033644	.0018641	18.			976
visit#obese							
4 1		.0064715	.0014952	4.			402
5 1		.0044571	.0017778	2.			415
6 1		.0057885	.0021825	2.			61
8 1		.0008468	.0032882	0.2			16
_cons		.7410788	.0013913	532.6			57

On average, BMD drops by .0165 from

On average, BMD

On average, BMD is higher by 0.033 for the

The difference in the drop for obese v. non-obese between visits 2

On average, BMD drops by .0165 from

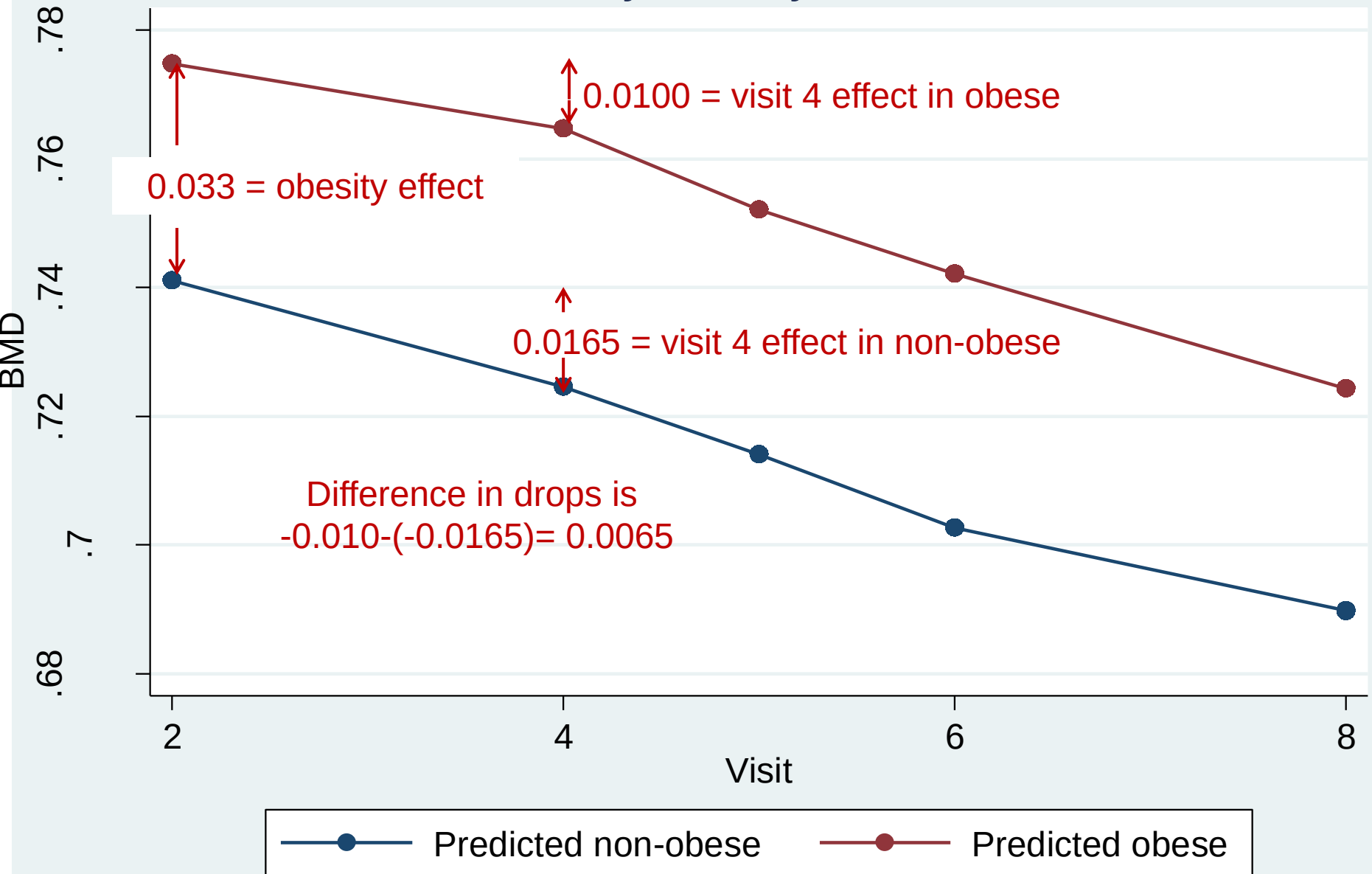
On average, BMD

On average, BMD is

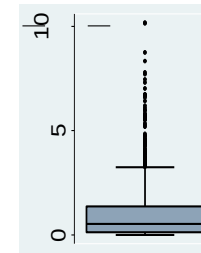
higher by 0.033 for the

The difference in the drop for obese v. non-obese between visits 2 and 4 is 0.0064

BMD vs visit by obesity status at baseline



Back to the science: BMD/Obesity



```
. testparm visit#obese
```

```
( 1)  4.visit#1.obese = 0
```

```
( 2)  5.visit#1.obese = 0
```

```
( 3)  6.visit#1.obese = 0
```

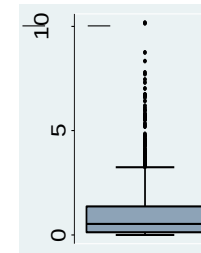
```
( 4)  8.visit#1.obese = 0
```

```
      chi2(  4) =    22.20
```

```
Prob > chi2 =    0.0002
```

The rate of change in BMD during the study is different for the obese and non-obese. The obese change less (since all the interaction coefficients are > 0).

Back to the science: Linear OK?



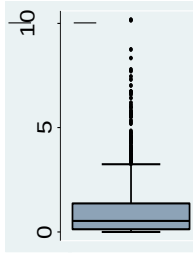
```
. xtgee totbmd c.visit##obese visit##obese, i(id) robust

. testparm visit#obese
( 1)  4.visit#1.obese = 0
( 2)  5.visit#1.obese = 0
( 3)  6.visit#1.obese = 0

      chi2(  3) =    19.02
    Prob > chi2 =    0.0003
```

The categorical interaction is statistically significant after fitting linear, therefore linear inadequate.

Example: BMD/BMI (time varying predictor)



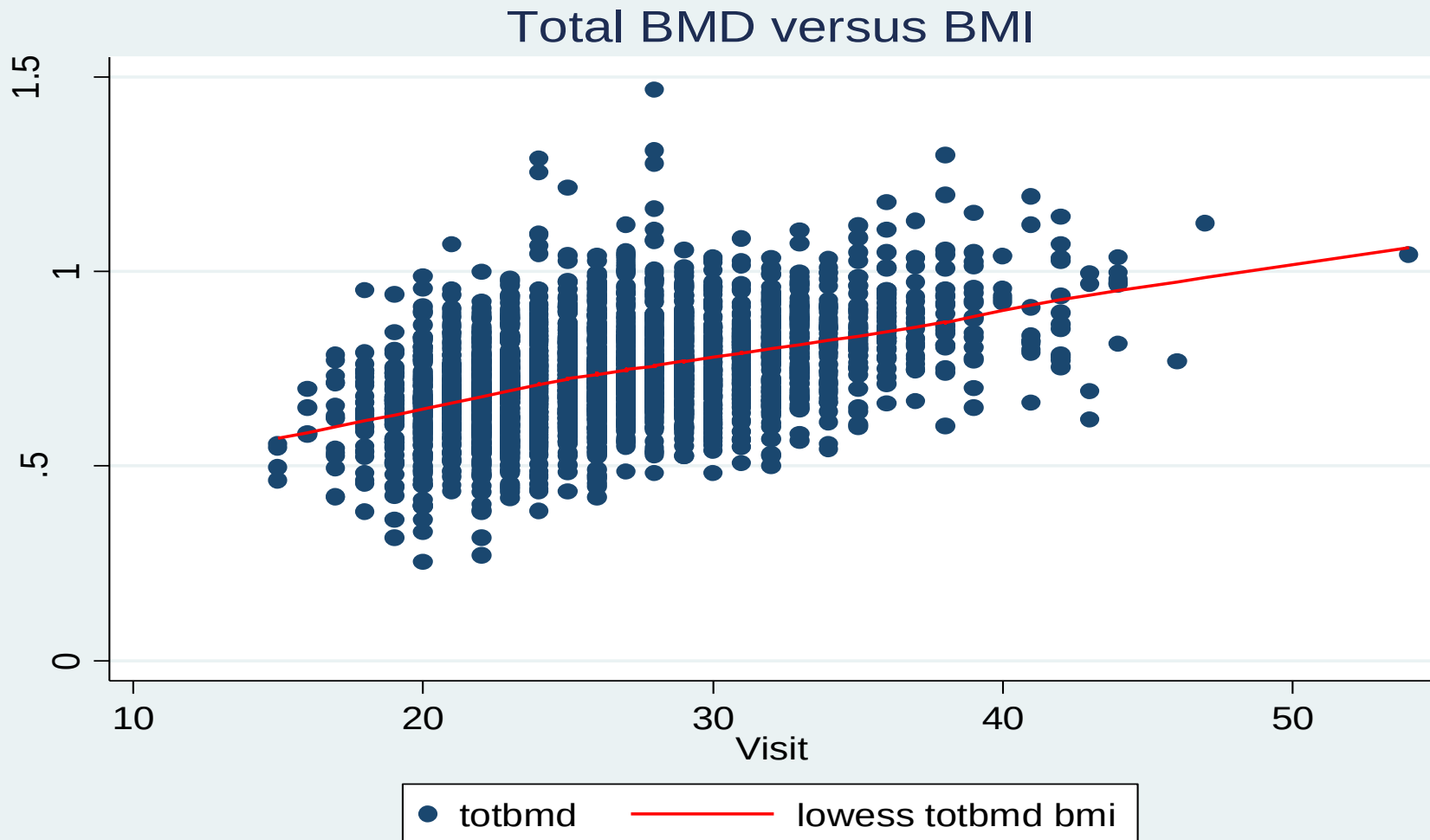
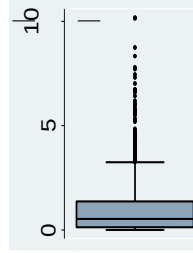
- Including a time-varying predictor automatically models a “trajectory”.
- $BMD_t = b0 + b1 * BMI_t$ implies

$$BMD_t - BMD_{t-1} = b1 * (BMI_t - BMI_{t-1})$$

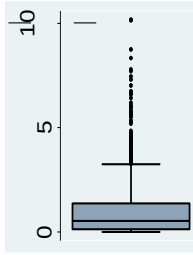
So change in BMD is predicted by change in BMI.

Ex 2: BMD/BMI (time varying predictor)

Does BMI predict total BMD?



Ex 2: BMD/BMI (time varying predictor)

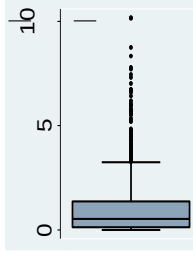


Effect of different analysis methods on the BMI effect

Method	Coef	SE	t-statistic
Regression	0.013	0.00015	82.0
Hierarchical	0.009	0.0002	44.1

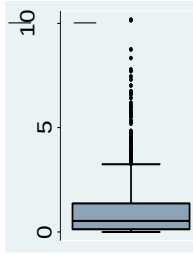
So we see a difference between a naïve and hierarchical analysis.

Mixed effects models



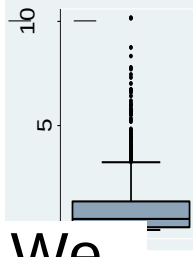
- Models for correlated data are often specified by declaring one or more of the categorical predictors in the model to be *random factors*. (Otherwise they are called *fixed factors*.) Models with both fixed and random factors are called *mixed effects models*.
- Random factors allow for factor-level-specific terms in the model. For example, if subject were specified as a random factor then each subject would be allowed their own intercept. *However*, random factors are different than including subject as a categorical predictor.

Fixed versus Random Factors



- With a random factor we assume that the *effects* associated with the levels of a factor can be regarded as a random sample from some reasonable population of effects. (If yes then random, if no then fixed).
- This is what we ordinarily do for any sample – we ask if we can regard it as a random sample from a larger population. For hierarchical data structures we ask the question over again for each level.

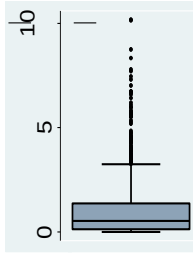
Notes on fixed vs random Factors



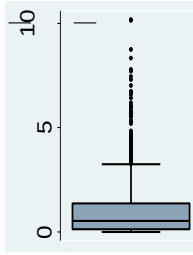
1. Continuous variables are virtually never random effects. We typically treat a variable as continuous because knowing the outcome for one value of the variable tells us something about the nearby values (hence the effects are not random). Do not confuse this with the fact that, if we have a random sample of subjects, their ages are a random sample of ages.
2. Scope of inference: Inferences can be made on a statistical basis to the *population* from which the levels of the random factor have been selected.
3. Incorporation of correlation in the model: Observations that share the same level of the random effect are being modeled as correlated.
4. Accuracy of estimates: Using random factors involves making extra assumptions but gives more accurate estimates.
5. Estimation method: Different estimation methods must be used.

Fixed versus Random Practice

- Fecal fat example. Factors? Fixed? Random?
- Back pain example. Factors? Fixed? Random?
- Study of Osteoporotic Fractures. Factors? Fixed? Random?



MIXED for continuous outcomes



`mixed` is for approximately normally distributed outcomes and is an analog of multiple linear regression. Here is the command syntax:

```
mixed depvar fix_effects || rand  
effects: , cov(corr structure) reml
```

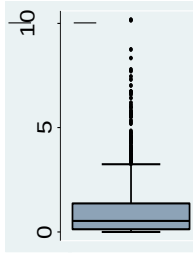
Punctuation important!

Example: Georgia babies

I like to use the
REML option

```
mixed bweight birthord initage ||  
momid: , reml
```

MIXED for continuous outcomes

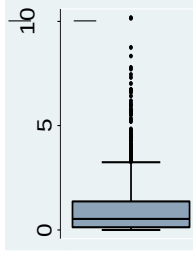


There can be multiple random effects specified by adding additional ||'s and random effects at the end. In this way you can handle multiple hierarchies or levels of clustering. Order is from highest to lowest level of clustering.

Example: backpain data

```
mixed logcost i.pracstyl i.educ ||  
  doctor: || patient:, reml
```

MIXED for continuous outcomes

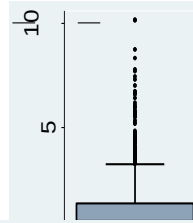


Specifying a random factor with a colon tells MIXED to allow different intercepts for each level of the factor, e.g., different intercepts for each patient.

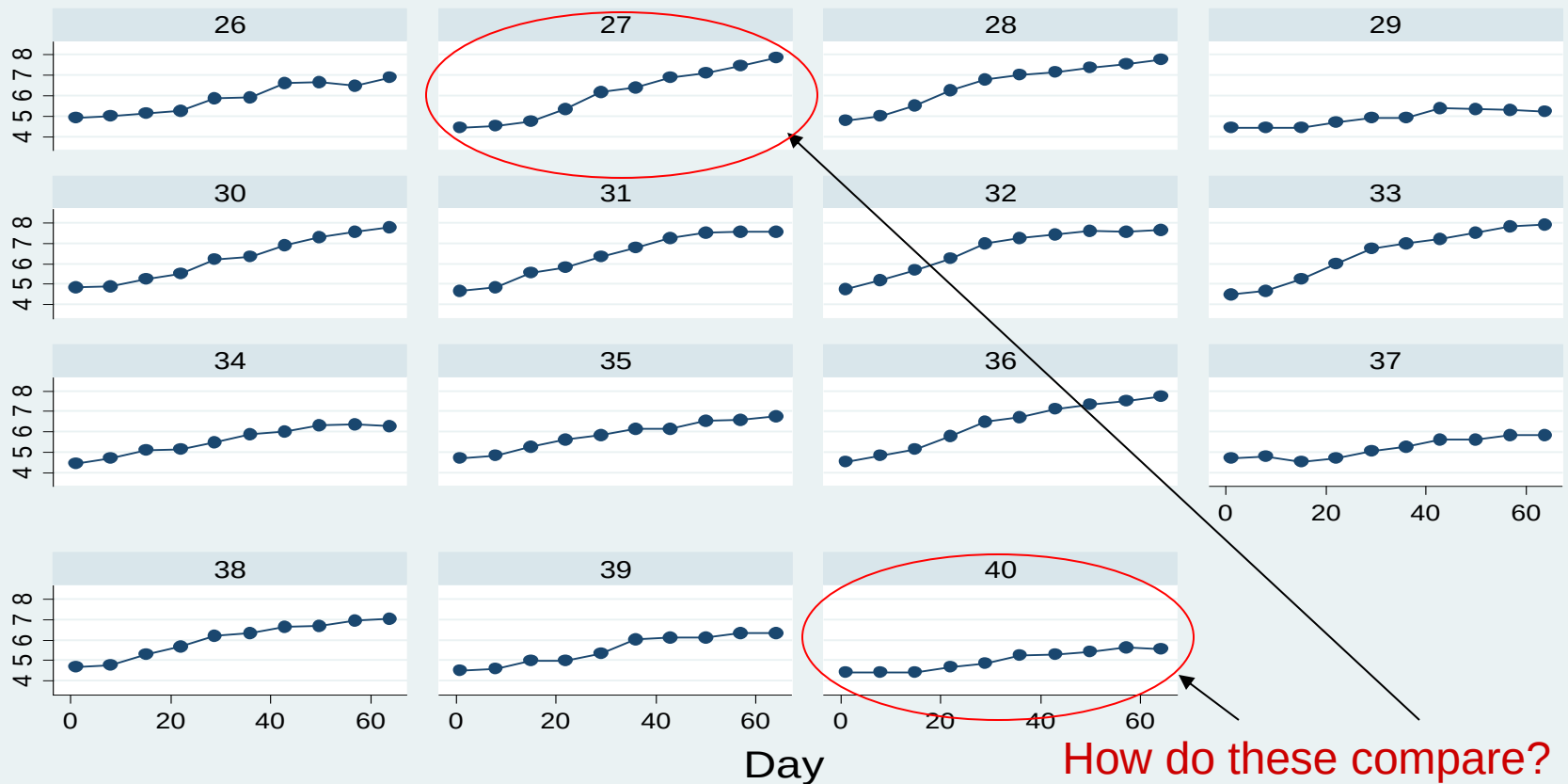
This is often sufficient, but there are cases in which more features of the model need to be included, for example, patient, animal or physician specific terms.

If so, these are added after the colon.

Mouse tumor/weight data



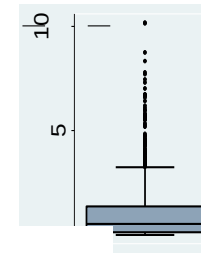
Mouse-specific trajectories of $\log(\text{weight})$
For a single treatment group



Graphs by mouse

How do these compare?

Mouse tumor/weight data



- So - an adequate model would need animal specific slopes and intercepts.

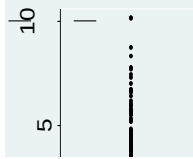
Stata code

- ```
mixed logw i.group day
i.group*day || mouse: day,
cov(un) reml
```

Mouse specific  
intercepts

Mouse specific day  
coefficients (slopes  
with day).

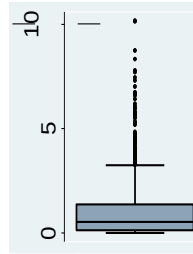
## Mouse tumor/weight data



More details in lab, but the “|| mouse:” portion specifies that mouse is a random effect (i.e., allows a mouse specific intercept or constant term) and the “day” term specifies that the slopes over days that are associated with a mouse are also random effects. That is, it allows each mouse to have its own growth trend.

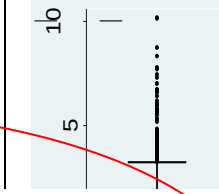
The “cov(un)” specifies that the variance and covariances of the random intercepts and slopes are unstructured. That is, the variances of the random intercepts and slopes are not assumed the same and they may be correlated.

## Recall: Fecal fat example



| PatID<br>/ Sex | Fecal Fat (g/day) |        |         |                | Avg    |
|----------------|-------------------|--------|---------|----------------|--------|
|                | Pill type         |        |         |                |        |
|                | None              | Tablet | Capsule | Coated Capsule |        |
| 1 – M          | 44.5              | 7.3    | 3.4     | 12.4           | 16.900 |
| 2 – M          | 33.0              | 21.0   | 23.1    | 25.4           | 25.625 |
| 3 – M          | 19.1              | 5.0    | 11.8    | 22.0           | 14.475 |
| 4 – F          | 9.4               | 4.6    | 4.6     | 5.8            | 6.100  |
| 5 – F          | 71.3              | 23.3   | 25.6    | 68.2           | 47.100 |
| 6 – F          | 51.2              | 38.0   | 36.0    | 52.6           | 44.450 |
| Avg            | 38.08             | 16.53  | 17.42   | 31.07          | 25.775 |

# mixed for the fecal fat example



## Summary information about hierarchy

Mixed-effects REML regression  
Group variable: patid

Number of obs = 24  
Number of groups = 6  
Obs per group: min = 4  
                  avg = 4.0  
                  max = 4  
Wald chi2(3) = 18.77  
Prob > chi2 = 0.0003

## The usual coef and SE

Log restricted-likelihood = -84.555945

| fecfat       | Coef.     | Std. Err. | z     | P> z  | [95% Conf. Interval] |           |
|--------------|-----------|-----------|-------|-------|----------------------|-----------|
| _Ipilltype_2 | -21.55    | 5.972127  | -3.61 | 0.000 | -33.25515            | -9.844848 |
| _Ipilltype_3 | -20.66667 | 5.972127  | -3.46 | 0.001 | -32.37182            | -8.961514 |
| _Ipilltype_4 | -7.016668 | 5.972127  | -1.17 | 0.240 | -18.72182            | 4.688485  |
| _cons        | 22.90852  | 5.972127  | 3.84  | 0.000 | 22.90852             | 53.25815  |

Summary of variation in random intercepts ( sd(\_cons) )  
and residuals ( sd(Residual) )

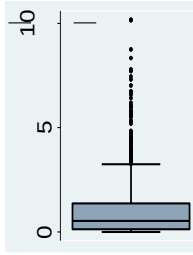
| RANDOM-EFFECTS PARAMETERS |              | Estimate | Std. Err. | [95% Conf. Interval] |          |
|---------------------------|--------------|----------|-----------|----------------------|----------|
| patid: Identity           | sd(_cons)    | 15.89557 | 5.567268  | 8.00113              | 31.5792  |
|                           | sd(Residual) | 10.34403 | 1.888552  | 7.232396             | 14.79439 |

Test of H0: no clustering

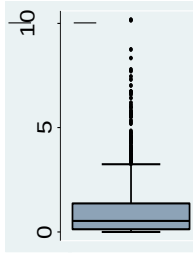
LR test vs. linear regression: chibar2(01) = 12.52 Prob >= chibar2 = 0.0002

# Mixed model analyses

- You get estimates of the regression coefficients (with the same interpretation as in regular linear regression), but accounting for the correlated data.
- Also get an understanding of how the variability in the data is attributable to various levels in the hierarchy.
- `sd(_cons)` is the standard deviation of the person-specific intercepts. So 15.89557 is the standard deviation, across people, of their true baseline values.



# Mixed model analyses



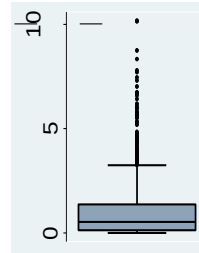
- Total variance = sum of all estimated variances.

$$\begin{aligned}\text{Total variance} &= (\text{patient SD})^2 + (\text{residual SD})^2 \\ &= (15.89557)^2 + (10.34403)^2 \\ &= 252.6691 + 106.9990 = 359.6681\end{aligned}$$

- Intraclass correlation (ICC)

$$\begin{aligned}\text{ICC} &= (\text{patient SD})^2 / (\text{total variance}) \\ &= 252.6691 / 359.6681 = 0.703\end{aligned}$$

# xtgee: Fecal fat example



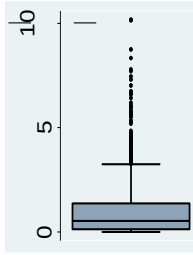
```
. xtgee fecfat i.pilltype i.sex, i(patid)
```

|                               |              |                    |   |        |
|-------------------------------|--------------|--------------------|---|--------|
| GEE population-averaged model |              | Number of obs      | = | 24     |
| Group variable:               | patid        | Number of groups   | = | 6      |
| Link:                         | identity     | Obs per group: min | = | 4      |
| Family:                       | Gaussian     | avg                | = | 4.0    |
| Correlation:                  | exchangeable | max                | = | 4      |
|                               |              | Wald chi2(4)       | = | 24.00  |
| Scale parameter:              | 253.8228     | Prob > chi2        | = | 0.0001 |

| -----    |  |           |           |       |       |                      |           |
|----------|--|-----------|-----------|-------|-------|----------------------|-----------|
| fecfat   |  | Coef.     | Std. Err. | z     | P> z  | [95% Conf. Interval] |           |
| -----    |  |           |           |       |       |                      |           |
| pilltype |  |           |           |       |       |                      |           |
| 2        |  | -21.55    | 5.451781  | -3.95 | 0.000 | -32.23529            | -10.86471 |
| 3        |  | -20.66667 | 5.451781  | -3.79 | 0.000 | -31.35196            | -9.981373 |
| 4        |  | -7.016668 | 5.451781  | -1.29 | 0.198 | -17.70196            | 3.668626  |
|          |  |           |           |       |       |                      |           |
| 1.sex    |  | 13.55     | 11.16389  | 1.21  | 0.225 | -8.330816            | 35.43082  |
| _cons    |  | 31.30833  | 8.570992  | 3.65  | 0.000 | 14.5095              | 48.10717  |
| -----    |  |           |           |       |       |                      |           |



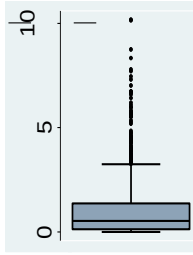
# Study of cognitive decline



Outcome is the mini-mental state exam, a numerical score from 0 to 30.

Question: Do participants with better physical functioning at baseline (measured in quartiles of a physical function exam) show lower rates of cognitive decline?

# Study of cognitive decline

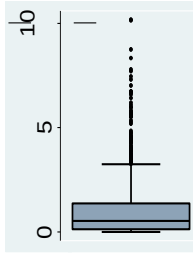


Outcome is the mini-mental state exam, a numerical score from 0 to 30 (original version of the 3MS).

Question: Do participants with better physical functioning at baseline (measured in quartiles of a physical function exam) show lower rates of cognitive decline?

MIXED model syntax:

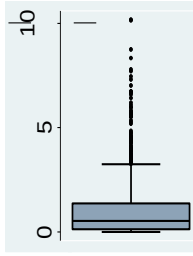
# Model diagnostics: predictors



Nothing new in these models on the predictor side of the equation. For checking linearity do the usual:

Plot residuals versus predictors (RVP), transform predictors (e.g., try quadratic), try splines, categorize predictors.

# Model diagnostics: normality/outliers



Calculate residuals

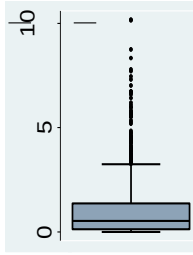
```
mixed: predict resids, residuals
```

```
xtgee: predict preds
```

```
gen resids=outcome-preds
```

Plot residuals versus predicted values and look for outliers, unequal variances (but some mixed models allow unequal variances as does the robust option in `xtgee`), histogram of residuals.

# Model diagnostics: normality/outliers



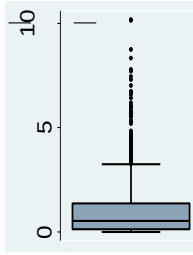
If you find issues – do the “usual”:

1. Try removing outliers to assess their influence.
2. Try transformations to alleviate non-normality, unequal variances.
3. Use bootstrap. But - only works for one level of hierarchical data, you need to use the `cluster()` option, and requires a fair number of clusters.

or

4. Use `xtgee` but specify a different distribution.

# Terminology



GEE type models are sometimes called “population averaged” or “marginal” models because they hypothesize a relationship (e.g., logistic regression) that holds averaged over all subjects in a population. Random effects models (like those fit by MIXED) are sometimes called “subject specific” or “conditional” because they are built using random effects that are specific to a subject (e.g., a doctor or mouse effect).

# Summary

- Approximately normally distributed outcomes can be handled with mixed models (`mixed`) or generalized estimating equations (`xtgee`).
- Mixed models have the advantage of handling multiple levels of clustering and more explicit modeling of sources of variability and correlation.
- Generalized estimating equations (when using the robust option) makes fewer assumptions.
- Model checking is similar to regression models for non-hierarchical data with some exceptions:
  - With a bootstrap need to cluster resample.
  - `mixed` can model certain forms of unequal variance.
  - `xtgee` can directly model non-normally distributed outcomes
  - `xtgee` can accommodate unequal variances with the robust option. More on these latter two next lecture.

