Contemporary Issues in Clinical Trials Methods
Longitudinal Data Analysis Part I

Michael Donohue

Alzheimer's Therapeutic Research Institute
Department of Neurology
University of Southern California

July 14, 2017

Repeated Measures/Longitudinal Data

- In Alzheimer's Disease (AD) clinical trials we typically assess study participants *longitudinally* or *repeatedly* over time
- This gives rise to *serial observations* for each participant at various time points post baseline
- We cannot treat these serial observations as if they came from different people
- Our analysis methods must account for *within-subject correlation*
- There are specialized statistical methods to help accomodate many varieties of *correlated* or *clustered* data

Repeated Measures/Longitudinal Data

- We will explore *longitudinal data analysis* approaches commonly employed in AD clinical trials
- We will demonstrate these methods on a *simulated* clinical trial dataset
- Data is simulated and analyzed using (`cran.r-project.org`; `rstudio.com`)
- All of the code for this session is available from `github.com/atrihub/AAIC2017ClinicalTrialMethods` (see 01LDA.R file)

Let's simulate a hypothetical clinical trial...

- Two groups: placebo vs active (hypothetical)
- $n = 200$ mild to moderate dementia subjects per group
- Alzheimer's Disease Assessment Scale (ADAS-Cog) assessed at 0, 6, 12, 18 months
- Placebo group behaves like ADNI participants
- Weak effects for age and sex (based on ADNI pilot estimates)
- A treatment which slows ADAS-Cog progression by 12.5%
- Typical attrition ($\approx$30%)

Let's simulate a hypothetical clinical trial...

Simulation *reverses* the usual process of statistical modeling/estimation

- Model fitting: Data + Model $\rightarrow$ Parameter Estimates
- Model simulation: Model + Parameter Estimates $\rightarrow$ Pseudo Data
- Given a reasonable model, everything can be simulated: mean, variance, missingness, etc.
- CAUTION: Simulations can only provide information about *models*, but they cannot provide information about *reality*. *Real data* is required for the latter.

Let's simulate a hypothetical clinical trial...

These are all estimates required:

```
# fixed effects parameters estimated from ADNI
Beta <- c(
   '(Intercept)'=19.52, # mean ADAS at baseline
        'female'=-0.29, # better scores for females
         'age_c'= 0.06, # worse change for older at baseline (age mean centered)
         'month'= 0.42, # worse per month post baseline
  'month:active'=-0.05) # improvement per month with treatment

# standard deviations for random effects
sigma_random_intercept <- 6.1
sigma_random_slope <- 0.38
sigma_residual <- 3.3

# other design parameters
months <- c(0, 6, 12, 18)
n <- 200 # per group
attrition_rate <- 0.40/18 # approx per month
```

The pseudo data snapshot

	id	active	female	age	censor	ran.intercept	ran.slope	age_c	month	residual	ADAS11
1	1	0	1	81.0	28.7	-6.68	0.606	5.878	0	3.53	16
2	1	0	1	81.0	28.7	-6.68	0.606	5.878	6	-4.52	15
3	1	0	1	81.0	28.7	-6.68	0.606	5.878	12	4.69	30
4	1	0	1	81.0	28.7	-6.68	0.606	5.878	18	2.78	34
5	2	1	0	75.6	72.1	-10.10	-0.419	0.459	0	2.24	12
6	2	1	0	75.6	72.1	-10.10	-0.419	0.459	6	-3.53	6

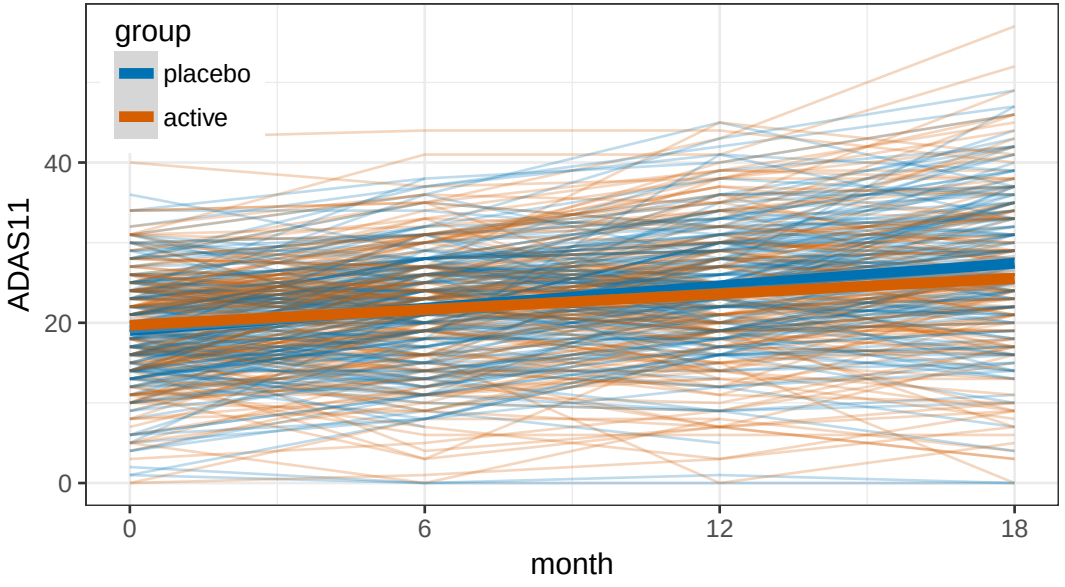
The pseudo data: Baseline characteristics

Table: Descriptive Statistics by group

	placebo $N = 200$	active $N = 200$
female	49% (98)	55% (109)
age	68.36 75.19 79.85 (74.58 ± 8.43)	70.33 75.59 80.80 (75.69 ± 7.78)
ADAS11	15.00 19.00 24.00 (19.11 ± 6.74)	14.00 20.00 24.00 (19.62 ± 7.13)

a b c represent the lower quartile a, the median b, and the upper quartile c for continuous variables. $x \pm s$ represents $\bar{X} \pm 1$ SD.Numbers after percents are frequencies.

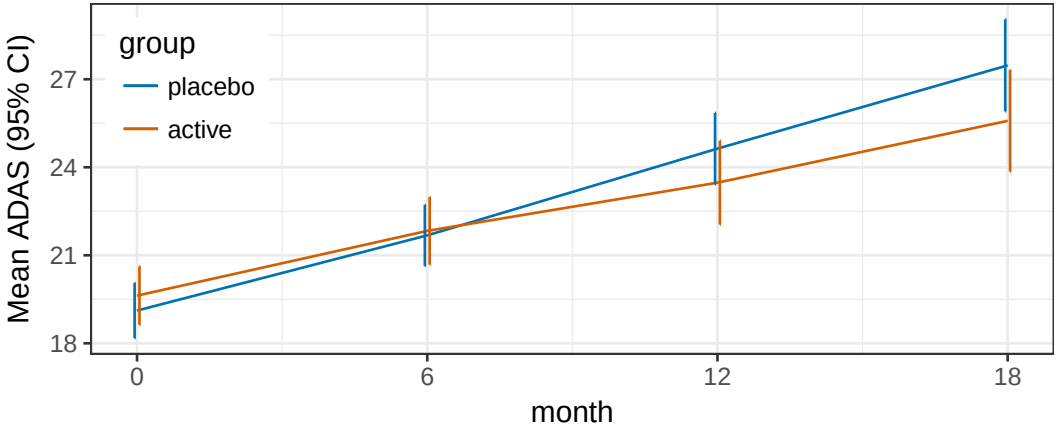
The pseudo data: Spaghetti plot



Basic longitudinal summaries of ADAS11

group	month	n	mean	sd	lower95	upper95	min	max
placebo	0	200	19.1	6.74	18.2	20.1	1	36
placebo	6	181	21.7	7.11	20.6	22.7	0	38
placebo	12	159	24.6	7.77	23.4	25.9	0	45
placebo	18	144	27.5	9.50	25.9	29.0	0	49
active	0	200	19.6	7.13	18.6	20.6	0	43
active	6	179	21.8	7.80	20.7	23.0	0	44
active	12	160	23.5	9.13	22.0	24.9	0	45
active	18	146	25.6	10.57	23.9	27.3	0	57

Mean ADAS

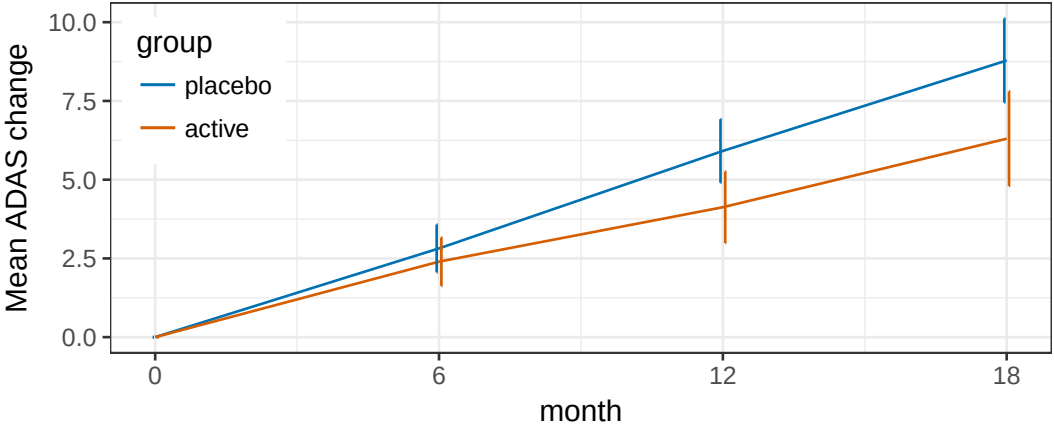


active	200	179	160	146
placebo	200	181	159	144

Basic longitudinal summaries of ADAS11 *change*

group	month	n	mean	sd	lower95	upper95	min	max
placebo	0	200	0.00	0.00	0.00	0.00	0	0
placebo	6	181	2.82	5.17	2.06	3.58	-10	17
placebo	12	159	5.91	6.44	4.90	6.92	-12	28
placebo	18	144	8.78	8.10	7.45	10.12	-12	30
active	0	200	0.00	0.00	0.00	0.00	0	0
active	6	179	2.40	5.21	1.63	3.17	-14	15
active	12	160	4.12	7.27	2.99	5.26	-14	19
active	18	146	6.30	9.20	4.80	7.81	-22	28

Mean ADAS change (95% CI)



active	200	179	160	146
placebo	200	181	159	144

Two sample t-test of mean change at month 18 (completers analysis)

- Difference between group means is 8.785 - 6.301 = 2.483
- (pooled) standard deviation is 8.668
- $t = \frac{2.483}{8.668\sqrt{\frac{1}{144}+\frac{1}{146}}} = 2.439$
- 144 + 146 - 2 = 288 "degrees of freedom"

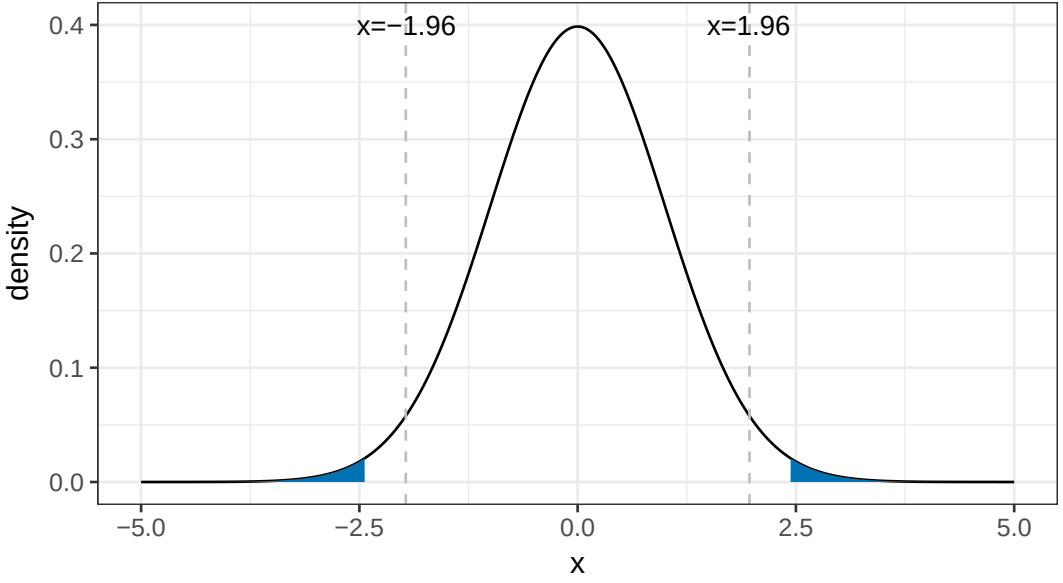
t-test

```
Two Sample t-test

data:  ADAS11.ch by group
t = 2.44, df = 288, p-value = 0.0153
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.48 4.49
sample estimates:
mean in group placebo  mean in group active
              8.78472               6.30137
```

The t_{288}-distribution

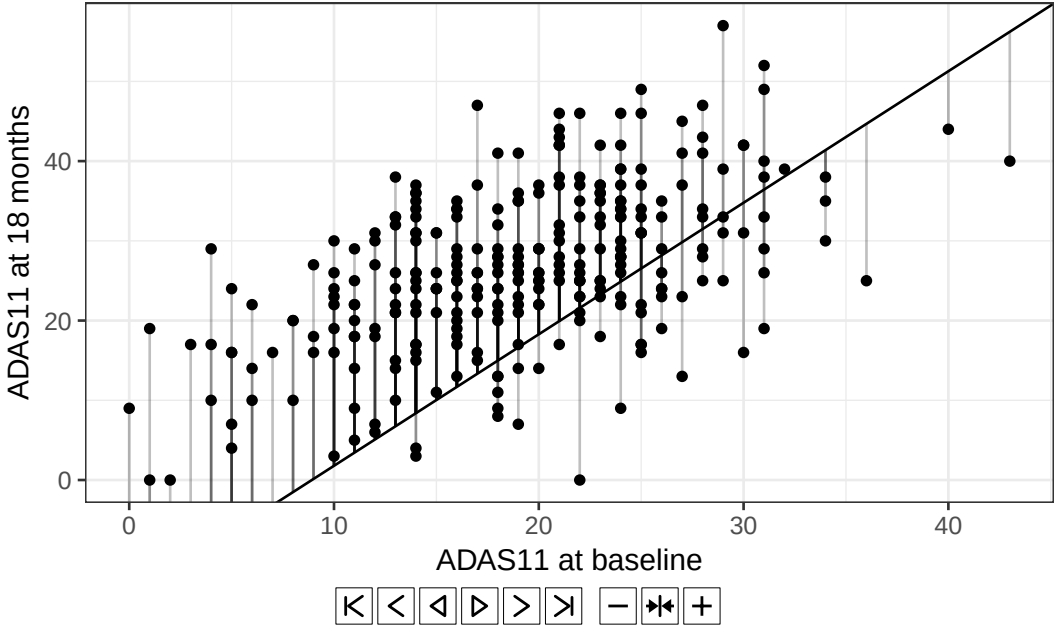
p-value is area under curve for $|x| > 2.439$, the value of the test statistic in this case.



Regression analysis

- "Regression" generally refers to a relationship between variables that is estimated by data
- "Ordinary Least Squares" regression, for example, describes a linear relationship between two continuous variables that is estimated by the line that minimizes the sum of squared "residuals"
- predictor/covariate $\rightarrow$ response/outcome
- "Residuals" are the differences between observations and values predicted by the regression

Ordinary Least Squares: minimizing the sum of squared residuals



Other types of regression

- "General linear models" can add multiple *covariates/predictors*
- "Generalized linear models" can accomodate other types of *outcome/response* variables (e.g. logistic regression can accomodate binary outcome variables)
- "Mixed-effects models" mix *random effects* with the standard *fixed effects* to account for complex correlation structures

All regression models share the common theme of estimating the best fit relationship between *outcome/response* variables and *covariates/predictors*

ANalysis of COVAriance (ANCOVA) for "*pre-post*" data

- Very common for two groups, and *one* post- assessment
- ADAS_{i1}: baseline or pre- observation for subject i, $i = 1, \ldots, 200$
- ADAS_{i2}: followup or post- observation for subject i, $i = 1, \ldots, 200$
- Active_i: treatment group indicator (e.g. 1 if active, 0 if placebo)
- ANCOVA I: $\text{ADAS}_{i2} = \beta_0 + \text{Active}_i\,\beta_1 + \text{ADAS}_{i1}\,\beta_2 + \varepsilon_i$
  - β is the *intercept*
  - β_1 is the estimate of interest: group difference at 18 months
  - β_2 controls for baseline ADAS
  - ε_i is residual error
- ANCOVA II: $\text{ADAS}_{i2} = \beta_0 + \text{Active}_i\,\beta_1 + \text{ADAS}^*_{i1}\,\beta_2 + \text{Active}_i\,\text{ADAS}^*_{i1}\,\beta_3 + \varepsilon_i$
  - β_3 controls for interaction of treatment assignment and baseline ADAS
  - Need to *mean center* baseline covariates: $\text{ADAS}^*_{i1} = \text{ADAS}_{i1} - \bar{Y}_{\cdot 0}$

Yang & Tsiatis (2001). Efficiency Study of Estimators for a Treatment Effect in a Pretest-Posttest Trial. *The Am. Statistician*, 55(4) 314-321

ANCOVA I for effect of treatment on ADAS11 at 18 months

```
Call:
lm(formula = ADAS11.ch ~ active + ADAS11.m0, data = trial_mmrm)

Residuals:
    Min      1Q  Median      3Q     Max
-25.518  -4.572  -0.027   4.138  26.138

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   10.048      0.680   14.77  < 2e-16 ***
active        -1.326      0.452   -2.94   0.0034 **
ADAS11.m0     -0.236      0.032   -7.38  3.3e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.02 on 966 degrees of freedom
Multiple R-squared:  0.0631,Adjusted R-squared:  0.0612
F-statistic: 32.6 on 2 and 966 DF,  p-value: 2.08e-14
```

ANCOVA II for effect of treatment on ADAS11 at 18 months

```
Call:
lm(formula = ADAS11.ch ~ active * center(ADAS11.m0), data = trial_mmrm)

Residuals:
    Min      1Q  Median      3Q     Max
-25.575  -4.502  -0.025   4.135  25.930

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)                5.5315     0.3196   17.31  < 2e-16 ***
active                    -1.3252     0.4517   -2.93   0.0034 **
center(ADAS11.m0)         -0.2615     0.0470   -5.56  3.5e-08 ***
active:center(ADAS11.m0)   0.0465     0.0642    0.72   0.4693
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.02 on 965 degrees of freedom
Multiple R-squared:  0.0637,Adjusted R-squared:  0.0607
F-statistic: 21.9 on 3 and 965 DF,  p-value: 1.05e-13
```

ANCOVA II with more covariates

```
Call:
lm(formula = ADAS11.ch ~ active * center(ADAS11.m0) + female +
    age_c, data = trial_mmrm)

Residuals:
   Min     1Q Median     3Q    Max
-25.33  -4.58  -0.09   4.21  26.00

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                 5.2568     0.3952   13.30   <2e-16 ***
active                     -1.2836     0.4524   -2.84   0.0046 **
center(ADAS11.m0)          -0.2571     0.0471   -5.46    6e-08 ***
female                      0.4718     0.4547    1.04   0.2997
age_c                      -0.0526     0.0279   -1.89   0.0593 .
active:center(ADAS11.m0)    0.0482     0.0642    0.75   0.4524
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.01 on 963 degrees of freedom
Multiple R-squared:  0.0683,Adjusted R-squared:  0.0634
F-statistic: 14.1 on 5 and 963 DF,  p-value: 2.42e-13
```

ANCOVA & t-test summary

- Ubiquitous, simple, powerful framework
- ANCOVA more powerful/efficient than t-test
- Both are inherently **complete case** analyses!
- With missing data, not intention-to-treat (ITT) analysis
- Do not make use of **incomplete cases!**
- Might be biased and/or inefficient (low power) with missing data

Two-stage models

- *Subject-specific* longitudinal profiles can often be modeled with simple linear regression
- This leads to the 2-stage model:
  - Stage 1: Linear regression model for each subject separately
  - Stage 2: Model subject-specific regression coefficients with covariates of interest
- However, this is **NOT** a recommend analysis approach, but rather a means to introduce mixed-effect models.

Two-stage model example

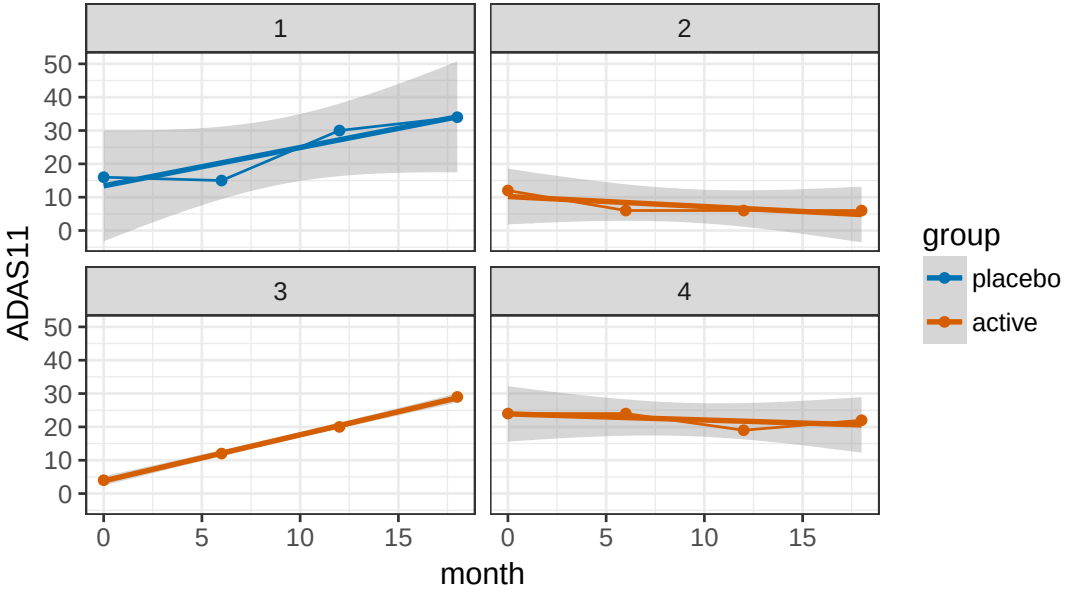
- Stage 1:

$$\text{ADAS}_{ij} = \beta_{0i} + t_{ij}\beta_{1i} + \varepsilon_i \tag{1}$$

  for subject i at time t_{ij}
- Provides estimates of subject-specific intercepts, $\hat{\beta}_{0i}$ and slopes $\hat{\beta}_{0i}$
- $\varepsilon_i \sim \mathcal{N}(0, \sigma_i^2 I_{n_i})$ estimates *within*-subject variability
- *Between*-subject variability can now be modeled by treating $\hat{\beta}_i$ as "response variables"
- Stage 2:

$$\hat{\beta}_{1i} = X_i\beta + \varepsilon_i' \tag{2}$$

Stage 1 models of simulated trial



Stage 1 model of simulated trial

	id	beta.(Intercept)	beta.month	sigma	active	group	age_c	female
1	1	13.4	1.150	4.621	0	placebo	5.878	1
2	2	10.2	-0.300	2.324	1	active	0.459	0
3	3	3.8	1.383	0.387	1	active	-5.163	0
4	4	23.9	-0.183	2.313	1	active	-3.609	0
5	5	27.5	0.750	1.225	1	active	8.972	1
6	6	25.0	0.833	NaN	1	active	1.064	1

Stage 1 model of simulated trial

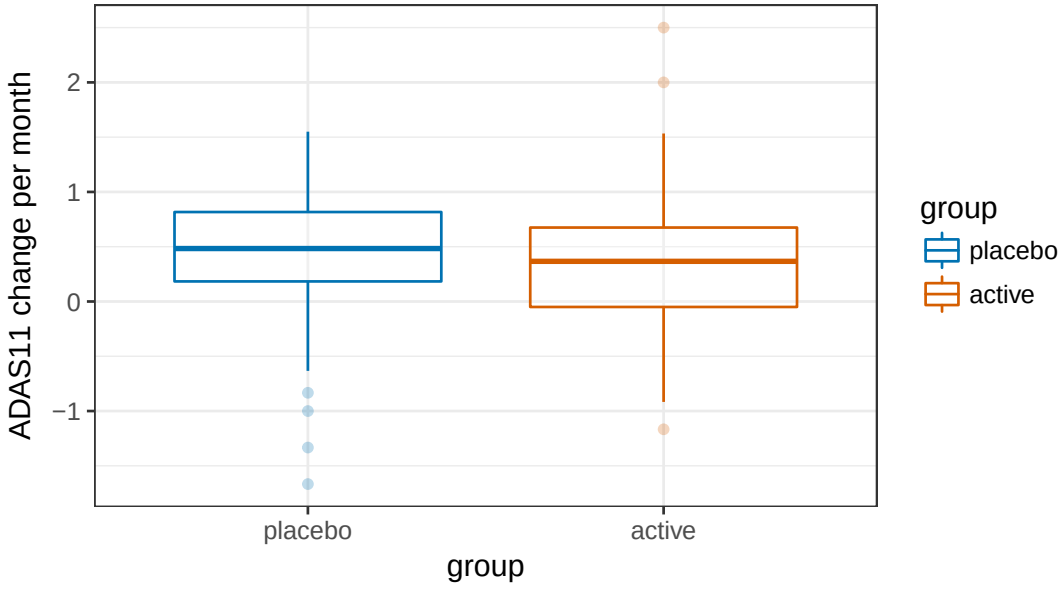
```
Call:
lm(formula = ADAS11 ~ month, data = trial_obs, subset = id ==
    1)

Residuals:
   1    2    3    4
 2.6 -5.3  2.8 -0.1

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   13.400      3.866    3.47    0.074 .
month          1.150      0.344    3.34    0.079 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.62 on 2 degrees of freedom
Multiple R-squared:  0.848,Adjusted R-squared:  0.772
F-statistic: 11.1 on 1 and 2 DF,  p-value: 0.0792
```

Stage 2 model of simulated trial



Stage 2 model of simulated trial

```
Call:
lm(formula = beta.month ~ female + age_c + active, data = trial_stage1)

Residuals:
    Min      1Q  Median      3Q     Max
-2.1450 -0.3195  0.0068  0.3313  2.1516

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.42662    0.04988    8.55  3.6e-16 ***
female       0.04370    0.05762    0.76    0.449
age_c       -0.00464    0.00352   -1.32    0.188
active      -0.09680    0.05769   -1.68    0.094 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.545 on 356 degrees of freedom
  (40 observations deleted due to missingness)
Multiple R-squared:  0.0148,Adjusted R-squared:  0.00654
F-statistic: 1.79 on 3 and 356 DF,  p-value: 0.149
```

Two-stage models

- In contrast to ANCOVA and t-test, two-stage models allow all randomized subject with at least one followup to be included into analysis ("modified intention-to-treat")
- However, second stage models ignore the variability/uncertainty of the slope estimates from the first stage
- This means that p-values from second stage might be smaller than they should be and Type I error could be inflated

Linear mixed-effects model (LME)

Linear mixed-effects models provide a cleaner, more efficient, and more accurate one-step alternative to two-stage models

$$\left.\begin{array}{ll} \text{Stage 1: } \text{ADAS}_{ij} & = \beta_{0i} + t_{ij}\beta_{1i} + \varepsilon_{ij} \\ \text{Stage 2: } \hat{\beta}_{1i} & = X_i\beta + \varepsilon'_{ij} \end{array}\right\} \rightarrow \text{ADAS}_{ij} = X_i\beta + b_{0i} + t_{ij}b_{1i} + \varepsilon_{ij}$$

- X_i: covariates for subject $i = 1, \ldots, 400$
- β: population level "fixed effects"
- $b_i \sim \mathcal{N}(0, D)$: subject-specific "random effects" for subject $i = 1, \ldots, 400$
- $(\varepsilon_{i1}, \ldots, \varepsilon_{i4}) \sim \mathcal{N}(0, \Sigma)$: vector of "residuals" for subject $i = 1, \ldots, 400$
- D, Σ: "variance components"

$b_1, \ldots, b_N, \varepsilon_1, \ldots, \varepsilon_N$ are assumed independent

Linear mixed-effects models of simulated trial

```
Linear mixed-effects model fit by REML
 Data: trial_obs
   AIC  BIC logLik
  8623 8659  -4304

Random effects:
 Formula: ~month | id
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev Corr
(Intercept) 6.175  (Intr)
month       0.406  -0.055
Residual    3.327

Fixed effects: ADAS11 ~ month + month:active
             Value Std.Error  DF t-value p-value
(Intercept)  19.43     0.341 967    57.0  0.0000
month         0.46     0.037 967    12.5  0.0000
month:active -0.11     0.052 967    -2.2  0.0263
 Correlation:
             (Intr) month
month        -0.153
month:active  0.001 -0.699

Standardized Within-Group Residuals:
     Min       Q1      Med       Q3      Max
-2.36800 -0.50399 -0.00885  0.51994  3.38014

Number of Observations: 1369
Number of Groups: 400
```

LME model with additional covariates

```
Linear mixed-effects model fit by REML
 Data: trial_obs
   AIC  BIC logLik
  8626 8673  -4304

Random effects:
 Formula: ~month | id
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev Corr
(Intercept) 6.138  (Intr)
month       0.406  -0.044
Residual    3.327

Fixed effects: ADAS11 ~ age_c + female + month + month:active
             Value Std.Error  DF t-value p-value
(Intercept)  20.17     0.483 967    41.8  0.0000
age_c         0.00     0.041 397     0.0  0.9742
female       -1.44     0.664 397    -2.2  0.0309
month         0.46     0.037 967    12.4  0.0000
month:active -0.11     0.052 967    -2.2  0.0275
 Correlation:
             (Intr) age_c  female month
age_c        -0.010
female       -0.712  0.015
month        -0.107  0.013  0.004
month:active  0.007 -0.015 -0.009 -0.700

Standardized Within-Group Residuals:
     Min       Q1      Med       Q3      Max
-2.36080 -0.50520 -0.00914  0.52270  3.37069

Number of Observations: 1369
Number of Groups: 400
```

Linear mixed-effects models (R code)

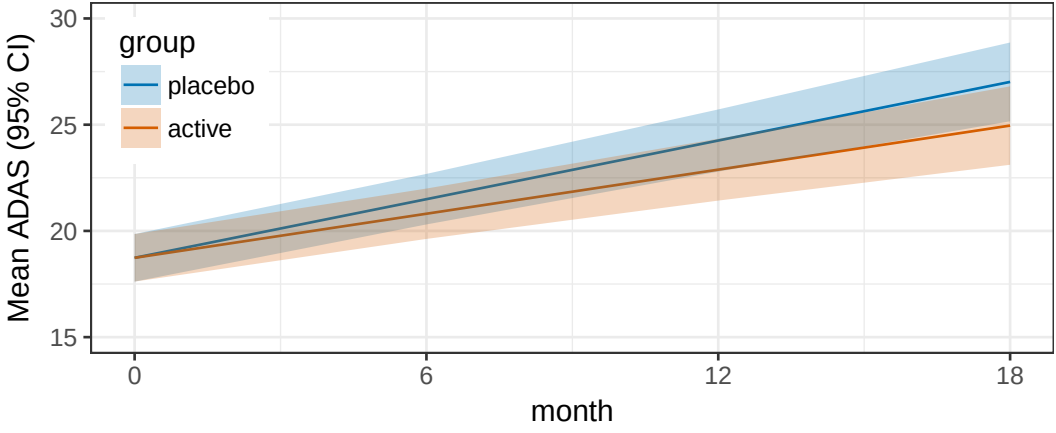
```
lme(ADAS11 ~ month + month:active,
  data = trial_obs, random = ~month|id)

lme(ADAS11 ~ age_c + female + month + month:active,
  data = trial_obs, random = ~month|id)
```

Mean profiles

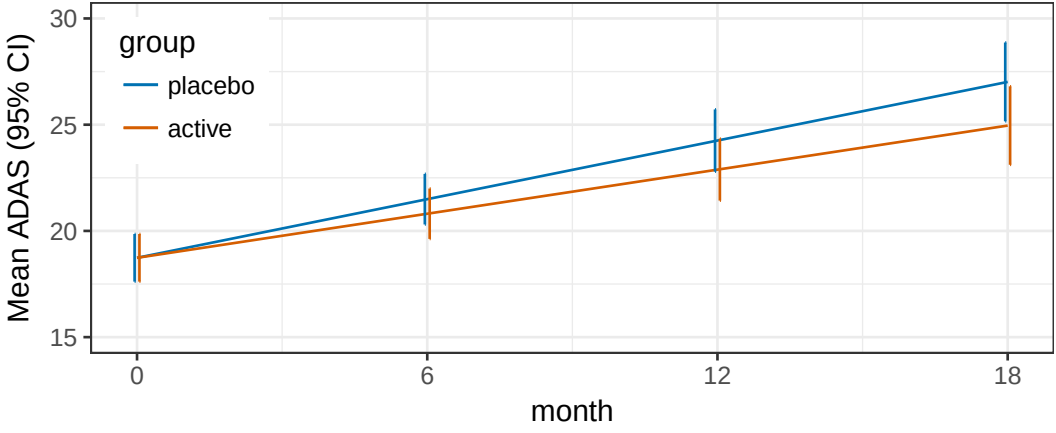
group	month	Estimate	lwr	upr
placebo	0	18.7	17.6	19.9
placebo	6	21.5	20.3	22.7
placebo	12	24.3	22.8	25.7
placebo	18	27.0	25.2	28.9
active	0	18.7	17.6	19.9
active	6	20.8	19.6	22.0
active	12	22.9	21.4	24.3
active	18	25.0	23.1	26.8

Modeled mean profiles (shaded CIs)



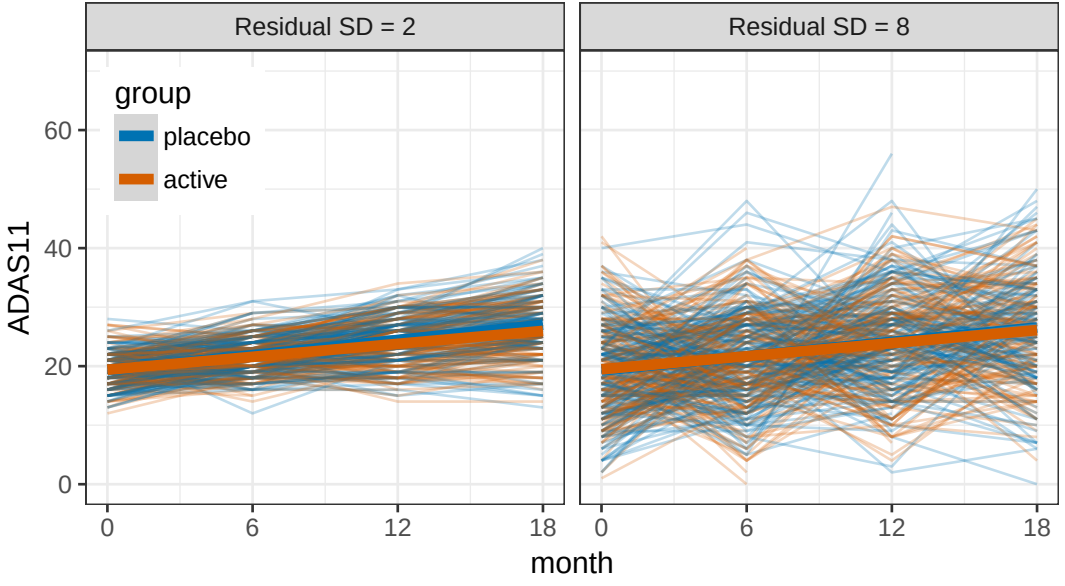
active	200	179	160	146
placebo	200	181	159	144

Plotting profiles (error bar CIs)

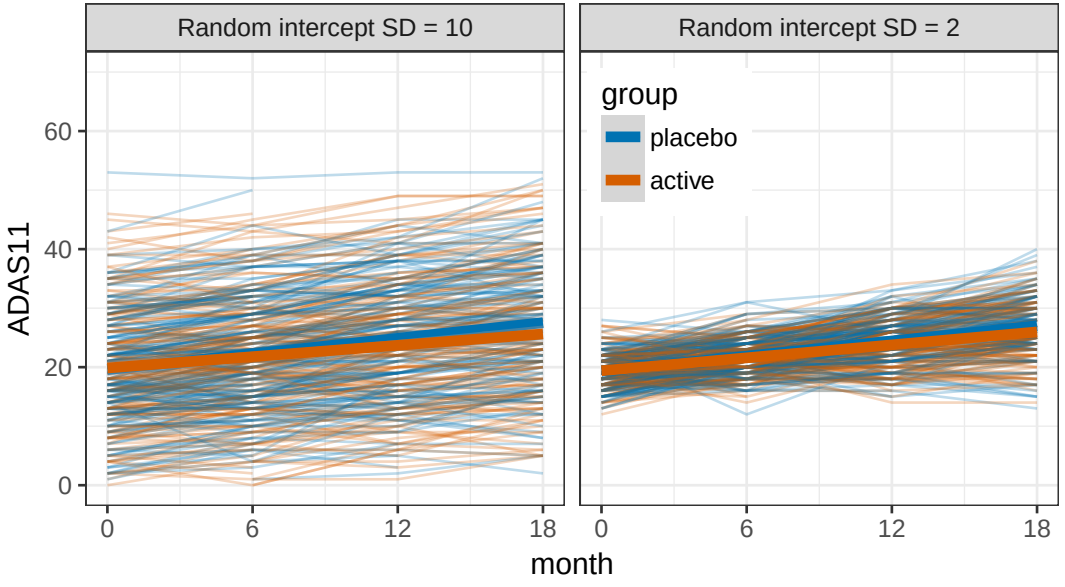


active	200	179	160	146
placebo	200	181	159	144

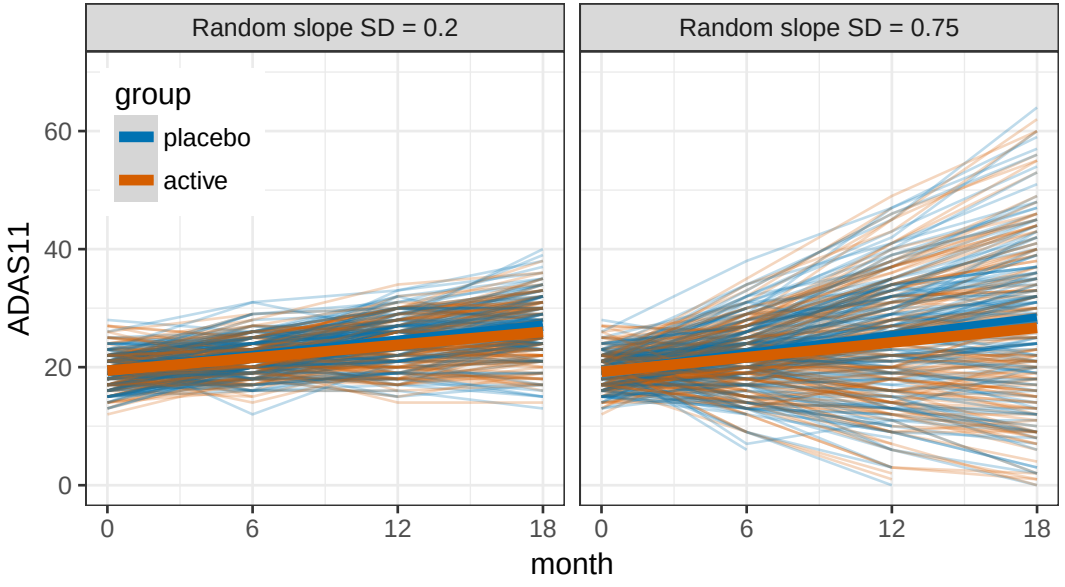
Mixed effect models: standard deviation of residuals



Mixed effect models: standard deviation of random intercepts



Mixed effect models: standard deviation of random slopes



Random intercepts model

- NOTE: With only two timepoints, it is impossible to fit a model with random slopes
- If we drop the *random slope* term, $t_{ij}b_{1i}$, what remains is called a *random intercepts* model:

$$\text{ADAS}_{ij} = X_i\beta + b_{0i} + \varepsilon_i$$

Random intercepts model

```
Linear mixed-effects model fit by REML
 Data: trial_obs
   AIC  BIC logLik
  8859 8886  -4425

Random effects:
 Formula: ~1 | id
        (Intercept) Residual
StdDev:        6.66     4.51

Fixed effects: ADAS11 ~ month + month:active
             Value Std.Error  DF t-value p-value
(Intercept)  19.40     0.385 967    50.4   0e+00
month         0.47     0.026 967    18.1   0e+00
month:active -0.12     0.035 967    -3.5   5e-04
 Correlation:
             (Intr) month
month        -0.271
month:active  0.001 -0.682

Standardized Within-Group Residuals:
     Min       Q1      Med       Q3      Max
-3.98906 -0.55214 -0.00459  0.55101  3.28226

Number of Observations: 1369
Number of Groups: 400
```

Random intercepts model vs model with random slopes

```
anova(fit_lme_int, fit_lme)

            Model df  AIC  BIC logLik   Test L.Ratio p-value
fit_lme_int     1  5 8859 8886  -4425
fit_lme         2  7 8623 8659  -4304 1 vs 2     240  <.0001
```

The model with random slopes is preferred (smaller AIC is better)

Further reading

- Fitzmaurice, G. M., Laird, N. M., Ware, J. H. (2012). *Applied Longitudinal Analysis*. Hoboken: Wiley.
- Diggle, P., Heagerty, P., Liang, K.-Y., Zeger, S. (2002). *Analysis of Longitudinal Data*. Oxford University Press
- Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer Series in Statistics. New-York: Springer.
- Molenberghs, G. and Kenward, M.G. (2007). *Missing Data in Clinical Studies*. Wiley.