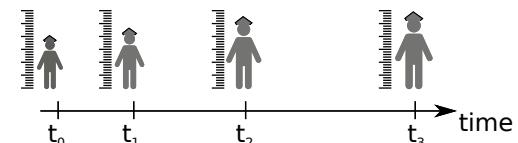


Repeated measurements

Variable(s) measured at **different** occasions
on the **same** experimental unit.

- Longitudinal study: **outcome** measured on the **same patient** at **different timepoints**.



Basic Statistic for health researchers

Lecture 8: repeated measurements

Brice Ozenne - email: broz@sund.ku.dk

¹ Section of Biostatistics, Department of Public Health, University of Copenhagen

² Neurobiology Research Unit, University Hospital of Copenhagen, Rigshospitalet.

28-11-2022

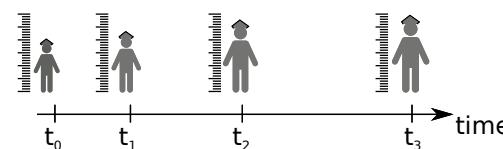
1 / 47

Repeated measurements

Basic Statistics for health researchers - L8: Repeated measurements

Variable(s) measured at **different** occasions
on the **same** experimental unit.

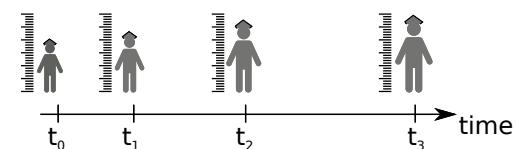
- Longitudinal study: **outcome** measured on the **same patient** at **different timepoints**.



Repeated measurements

Variable(s) measured at **different** occasions
on the **same** experimental unit.

- Longitudinal study: **outcome** measured on the **same patient** at **different timepoints**.



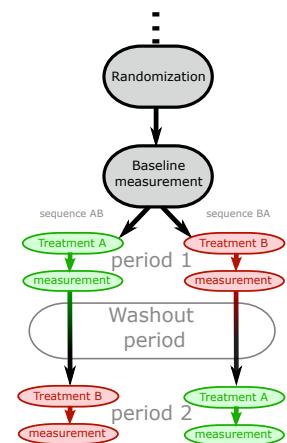
Can you find other examples?

- what motivates collecting repeated measurements?



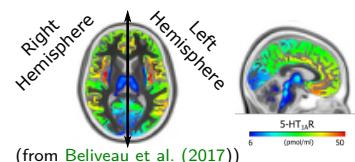
Other designs involving repeated measurements (1/2)

- cross-over: **outcome measured on the same patient under different treatments.**



Other designs involving repeated measurements (2/2)

- the **same type of measurement on the same patient at different locations.**

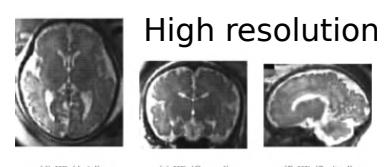


- test re-test study: **different ways of measuring the same quantity on the same patient.**

→ assess the stability of a measurement device

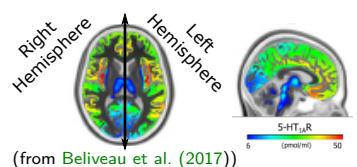


VS.



Other designs involving repeated measurements (2/2)

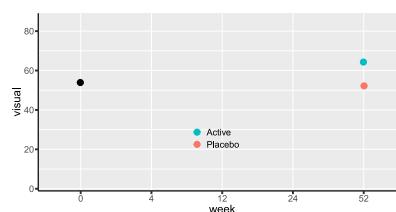
- the **same type of measurement on the same patient at different locations.**



Why using repeated measurements? (1/3)

To better understand the time-dynamic of the **exposure**:

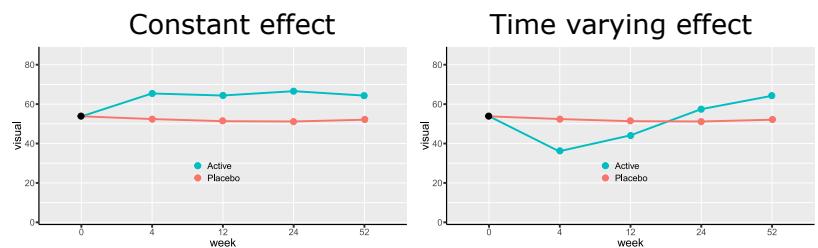
- is there any treatment effect?
- is there a sustained treatment effect?
- is there an immediate treatment effect?
- how do side effects occur after treatment intake?



Why using repeated measurements? (1/3)

To **better understand** the time-dynamic of the **exposure**:

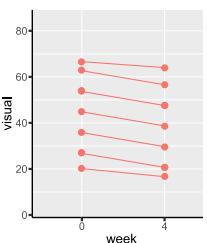
- is there any treatment effect?
- is there a sustained treatment effect?
- is there an immediate treatment effect?
- how do side effects occur after treatment intake?



Why using repeated measurements? (2/3)

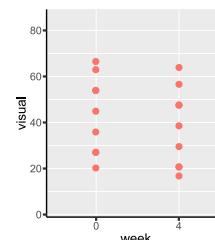
To **improve estimation** of the exposure effect:

- idea: "use each patient as its own control"
- account for some confounders: less bias
- account for some risk factors: more precision



To **improve estimation** of the exposure effect:

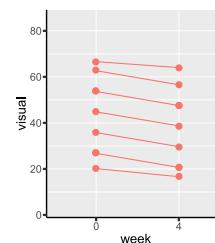
- idea: "use each patient as its own control"
- account for some confounders: less bias
- account for some risk factors: more precision



Why using repeated measurements? (2/3)

To **improve estimation** of the exposure effect:

- idea: "use each patient as its own control"
- account for some confounders: less bias
- account for some risk factors: more precision



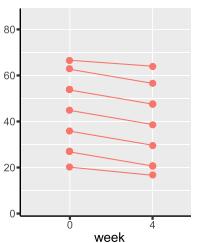
⚠️ Confounders/risk factors changing across repetitions:

- type of device used to make the measurement
- external events, e.g. food intake when monitoring glucose level, require specific modeling

Why using repeated measurements? (2/3)

To **improve estimation** of the exposure effect:

- idea: "use each patient as its own control"
- account for some confounders: less bias
- account for some risk factors: more precision



⚠ Confounders/risk factors changing across repetitions:

- type of device used to make the measurement
- external events, e.g. food intake when monitoring glucose level, require specific modeling

⚠ Does not control for regression to the mean effects induced by inclusion criteria

Why using repeated measurements? (3/3)

To **better handle missing values**:

- as the follow-up time increases, patients are more likely to drop-out
- regular follow-up can help:
 - to understand the reason(s) for drop-out
 - to limit the loss in statistical power due to drop-out
 - to adjust the analysis for informative drop-out

Example of regression to the mean (Kameran and Vollert, 2022)

"It has been recommended that an inclusion threshold of 4 or greater on an 11-point numerical pain rating scale be used when screening for clinical trial participants".

"there are numerous studies demonstrating that increased baseline pain score is associated with a greater placebo response in study control arms"

"By including patients only when their pain is high, on average, it becomes likely that a later assessment will be lower because of natural fluctuation, an effect known as regression to the mean."

Outline

✓ Introduction to repeated measurements

- definition and examples of study design
- benefit of having repeated measurements

Example of longitudinal study

Univariate approach

Multivariate approach

Illustration: ARMD trial (int, 1997)

Age-Related Macular Degeneration (ARMD) Trial:

- comparing interferon- α and placebo
- outcome $Y(t)$: change in vision over time

Illustrative example

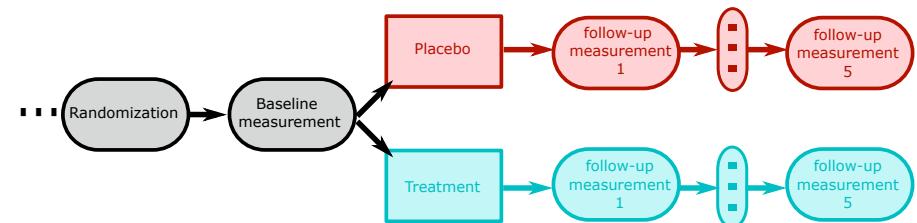
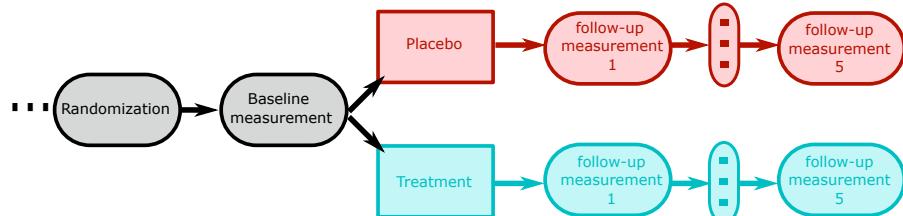


Illustration: ARMD trial (int, 1997)

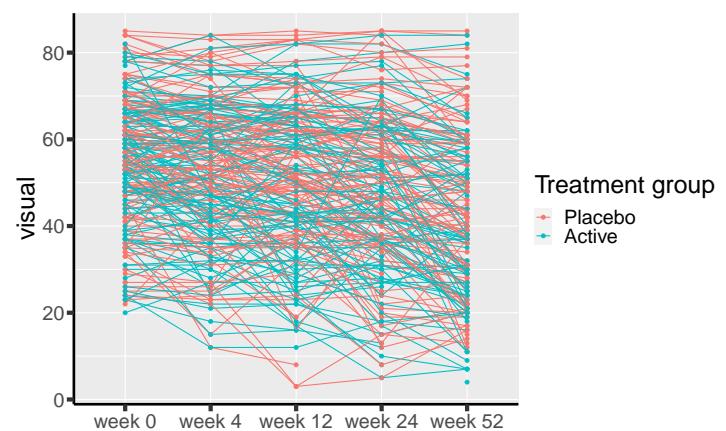
Age-Related Macular Degeneration (ARMD) Trial:

- comparing interferon- α and placebo
- outcome $Y(t)$: change in vision over time

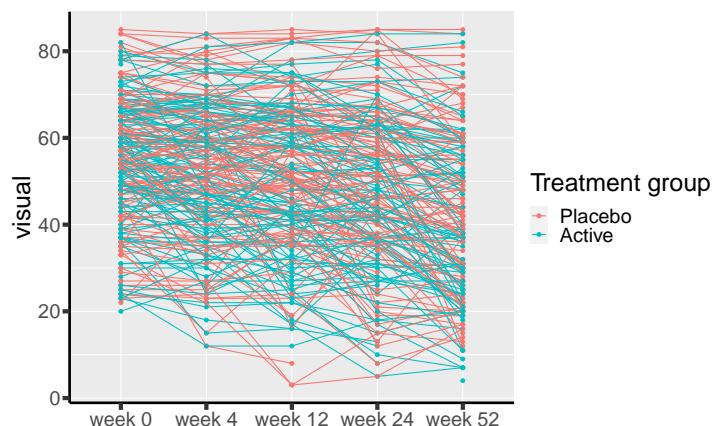


- **cluster** variable: subject (5 observations per cluster)
→ independent outcome replicates at the cluster level
- **repetition** variable: time

Visualizing the data: spaghetti plot



Visualizing the data: spaghetti plot



How would **you** summarize/describe the data ?

Summary statistics (2/3)

- dispersion over time (standard deviation)

week 0 week 4 week 12 week 24 week 52

Placebo	15.33143	15.38915	16.51203	18.61137	18.68844
Active	14.32523	15.99285	17.35207	17.84161	18.36214

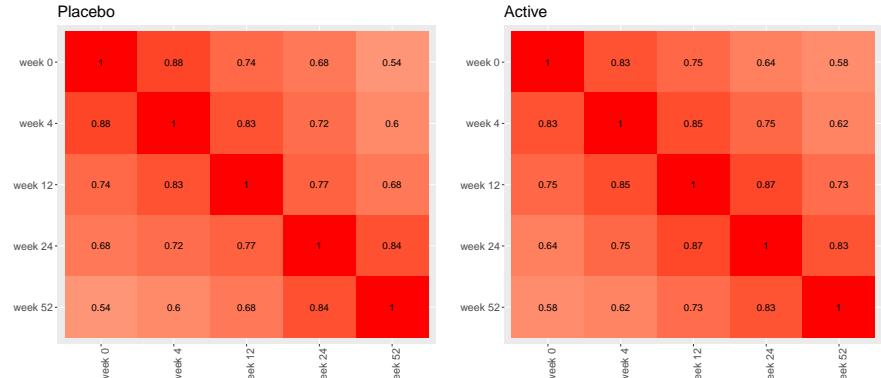
Summary statistics (2/3)

- dispersion over time (standard deviation)

week 0 week 4 week 12 week 24 week 52

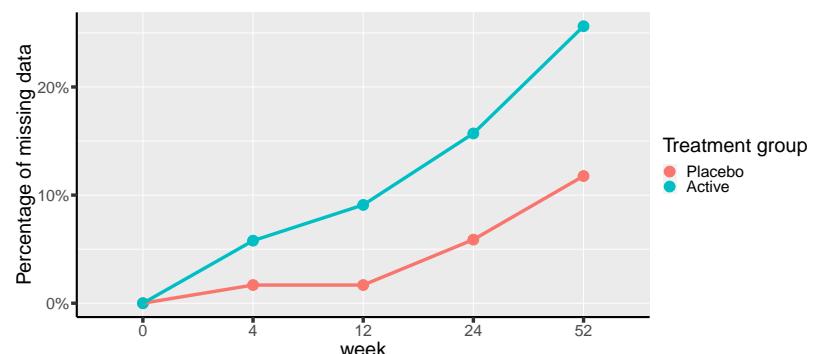
Placebo	15.33143	15.38915	16.51203	18.61137	18.68844
Active	14.32523	15.99285	17.35207	17.84161	18.36214

- dependency in visual acuity over time (Pearson correlation)



Summary statistics (3/3)

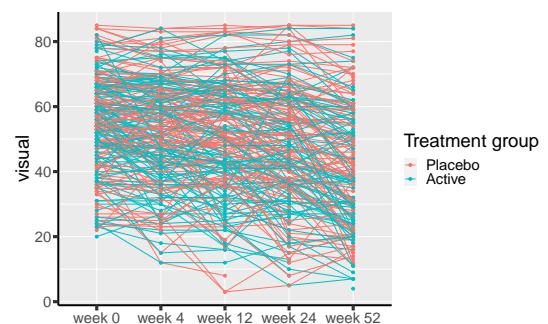
- what about missing values?



Planning the statistical analysis

How would **you** approach the analysis?

- challenges you anticipate
- specificities of working with repeated measurements



Concerns:

- treatment side effect(s) not measured by the outcome
- missing not at random may bias the estimated mean (upward bias if patients with weak vision are more likely to drop)

Challenge I - non independence

The usual **assumption of independent observations** is **violated**

- required when using `t.test`, `wilcox.test`, `lm`, `glm`, ...

If ignored, this can lead to:

- incorrect p-values/confidence intervals (almost always)
- biased estimates (unless certain assumptions are met)

Challenge I - non independence

The usual **assumption of independent observations** is violated

- required when using `t.test`, `wilcox.test`, `lm`, `glm`, ...

If ignored, this can lead to:

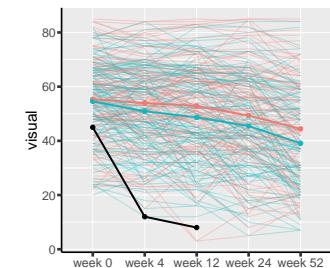
- ✗ incorrect p-values/confidence intervals (almost always)
- ✗ biased estimates (unless certain assumptions are met)

Possible solutions:

- **summary-statistic**: summarize repetitions into one number (e.g. average, area under the curve, peak value)
- **univariate**: perform separate analyses at each timepoint.
- **multivariate**: simultaneously analyze all timepoints

Challenge III - missing data

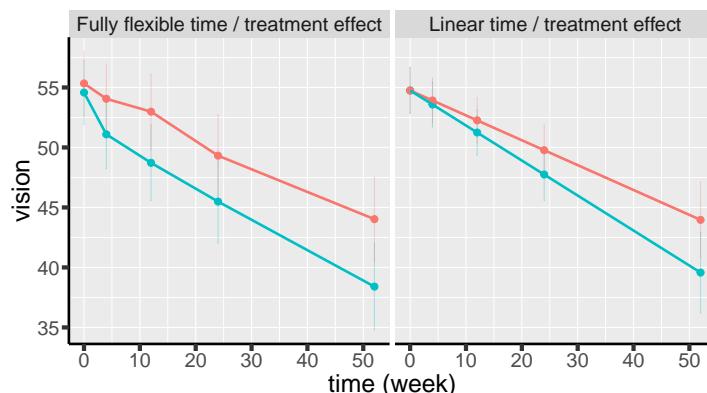
Different types of missing data:



Challenge II - multiple testing

Multiple testing:

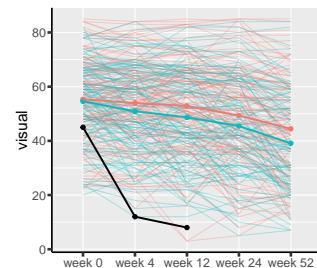
- several ways to test the treatment effect at any time, short term, long term, ...
- several possible statistical approaches t-test, mixed model with linear / non-linear treatment effect ...



Challenge III - missing data

Different types of missing data:

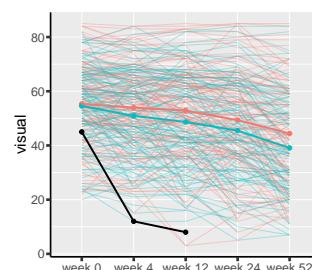
- **drop-out (patients leaving the study)**
Informative censoring
vs. censoring completely at random



Challenge III - missing data

Different types of missing data:

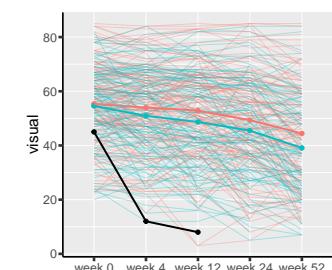
- drop-out (patients leaving the study)
Informative censoring
vs. censoring completely at random
- competing risks (e.g. death)
Complete case analysis usually wrong



Challenge III - missing data

Different types of missing data:

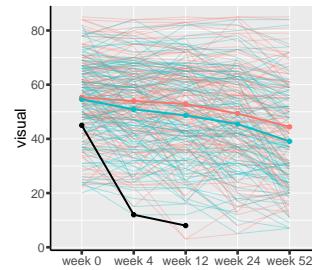
- drop-out (patients leaving the study)
Informative censoring
vs. censoring completely at random
- competing risks (e.g. death)
Complete case analysis usually wrong
- unbalanced data: measurement times differ between patients
Selection bias when sick patients have earlier or more frequent visits



Challenge III - missing data

Different types of missing data:

- drop-out (patients leaving the study)
Informative censoring
vs. censoring completely at random
- competing risks (e.g. death)
Complete case analysis usually wrong
- unbalanced data: measurement times differ between patients
Selection bias when sick patients have earlier or more frequent visits



⚠ Serious issues: remedies are beyond the scope of this lecture:

- reach out to a statistician!

Wide format

Data in the wide format:

- 1 row = 1 subject ("level 1 data")
→ independent replicate of ($Y(0)$, $Y(4)$, $Y(12)$, $Y(24)$, $Y(52)$)
- convenient when working with one or two timepoints

subject	treat.f	visual0	visual14	"..."	visual52
1	1 Active	59	55	...	NA
2	2 Active	65	70	...	55
3	3 Placebo	40	40	...	NA
4	4 Placebo	67	64	...	68
5	5 Active	70	NA	...	NA
6	6 Active	59	53	...	42

Long format

Data in the long format:

- 1 row = 1 measurement of 1 subject ("level 0 data")
- convenient when performing operations over all timepoints

	subject	treat.f	week	visual
1	1	Active	0	59
2	1	Active	4	55
3	1	Active	12	45
4	1	Active	24	NA
5	1	Active	52	NA
6	2	Active	0	65
7	2	Active	4	70
8	2	Active	12	65
9	2	Active	24	65
10	2	Active	52	55

Long			
subject	group	time	outcome
1	A	0	59
1	A	4	55
2	A	0	65
2	A	4	70
3	P	0	40
3	P	4	40
..

Technical note: summary statistic in the long format

Example: average outcome value **at each timepoint**

Sort

Long			
subject	group	time	outcome
1	A	0	59
1	A	4	55
2	A	0	65
2	A	4	70
3	P	0	40
3	P	4	40
..

Long			
subject	group	time	outcome
1	A	0	59
2	A	0	65
3	P	0	40
..	..	0
1	A	4	55
2	A	4	70
3	P	4	40
..	..	4

Technical note: summary statistic in the long format

Example: average outcome value **at each timepoint**

Split

Long			
subject	group	time	outcome
1	A	0	59
2	A	0	65
3	P	0	40
..	..	0
1	A	4	55
2	A	4	70
3	P	4	40
..	..	4

Split 1

subject	group	time	outcome
1	A	0	59
2	A	0	65
3	P	0	40
..	..	0

Split 2

subject	group	time	outcome
1	A	4	55
2	A	4	70
3	P	4	40
..	..	4

Technical note: summary statistic in the long format

Example: average outcome value **at each timepoint**

Split 1				Summarize
subject	group	time	outcome	
1	A	0	59	
2	A	0	65	
3	P	0	40	
..	..	0	

Split 2				
subject	group	time	outcome	
1	A	4	55	
2	A	4	70	
3	P	4	40	
..	..	4	

mean

59
65
40
....

mean

55
70
40
....

Univariate approach

A simple approach

What would **you** do if you were asked to assess the long term treatment effect?



To simplify, consider the case of:

- complete data (no missing data)
- no covariate/confounder

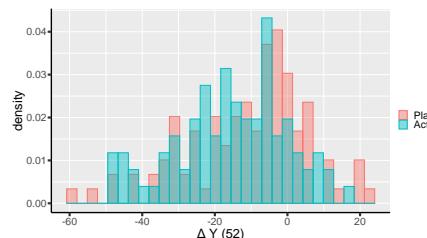
```
dfW.CC <- dfW[rowSums(is.na(dfW))==0,]
head(dfW.CC, 4)
```

subject	treat.f	visual0	visual4	visual12	visual24	visual52
2	2 Active	65	70	65	65	55
4	4 Placebo	67	64	64	64	68
6	6 Active	59	53	52	53	42
7	7 Placebo	64	68	74	72	65

1. Compute the difference in outcome between baseline and the timepoint of interest

```
dfW.CC$change <- dfW.CC$visual52 - dfW.CC$visual0
```

2. Visualize the change per group



General idea (2/2)

3. Compare the change between groups using a statistical test

3. Compare the change between groups using a statistical test

- two sample t-test: optimal for normally distributed data

General idea (2/2)

3. Compare the change between groups using a statistical test

3. Compare the change between groups using a statistical test

- two sample t-test: optimal for normally distributed data

- two sample t-test: optimal for normally distributed data

Formally:

- $Y(t)$ outcome at time t
- $\Delta Y = Y(52) - Y(0)$ change in outcome
- $\mathbb{E}[Y(t)|G = g] = \mu_g(t)$ expected outcome in group g at time t
- $\mathbb{E}[\Delta Y|G = g] = \Delta\mu_g = \mu_g(52) - \mu_g(0)$ expected change

Formally:

- $Y(t)$ outcome at time t
- $\Delta Y = Y(52) - Y(0)$ change in outcome
- $\mathbb{E}[Y(t)|G = g] = \mu_g(t)$ expected outcome in group g at time t
- $\mathbb{E}[\Delta Y|G = g] = \Delta\mu_g = \mu_g(52) - \mu_g(0)$ expected change

The underlying statistical model is

$$\begin{aligned}\Delta Y|G = 0 &\sim \mathcal{N}(\Delta\mu_0, \tau_0) \\ \Delta Y|G = 1 &\sim \mathcal{N}(\Delta\mu_1, \tau_1)\end{aligned}$$

and we are testing whether $\Delta\mu_0 = \Delta\mu_1$

Why working on the change?

Consider a simple model for individual i from the placebo group:

$$Y_i(t) = \mu_0(t) + \beta X_i + \gamma Z_i(t) + \varepsilon_i(t)$$

- X_i : traits of the individual (e.g. gender)
- Z_i : experimental setting that may change over time (e.g. distance between eyes and eye chart)
- unknown factors $\varepsilon_i(t)$ with variance σ^2

The change in outcome between baseline and week 52 is:

Why working on the change?

Consider a simple model for individual i from the placebo group:

$$Y_i(t) = \mu_0(t) + \beta X_i + \gamma Z_i(t) + \varepsilon_i(t)$$

- X_i : traits of the individual (e.g. gender)
- Z_i : experimental setting that may change over time (e.g. distance between eyes and eye chart)
- unknown factors $\varepsilon_i(t)$ with variance σ^2

The change in outcome between baseline and week 52 is:

$$Y_i(52) - Y_i(0) = \Delta\mu_0 + \gamma(Z_i(52) - Z_i(0)) + \varepsilon_i(52) - \varepsilon_i(0)$$

Why working on the change?

Consider a simple model for individual i from the placebo group:

$$Y_i(t) = \mu_0(t) + \beta X_i + \gamma Z_i(t) + \varepsilon_i(t)$$

- X_i : traits of the individual (e.g. gender)
- Z_i : experimental setting that may change over time (e.g. distance between eyes and eye chart)
- unknown factors $\varepsilon_i(t)$ with variance σ^2

The change in outcome between baseline and week 52 is:

$$Y_i(52) - Y_i(0) = \Delta\mu_0 + \gamma(Z_i(52) - Z_i(0)) + \varepsilon_i(52) - \varepsilon_i(0)$$

- we **only need to adjust for the change** in Z
- when $\rho > 1/2$, lower residual variance with ΔY vs. $Y(52)$
→ **gain in statistical power!**

Is it a good idea? (1/2)

- ✓ makes no assumption about the treatment effect over time
- ✓ simple to carry out
- ✓ naturally accounts for some covariates, even when unobserved.

Is it a good idea? (1/2)

- ✓ makes no assumption about the treatment effect over time
- ✓ simple to carry out
- ⚠ ... except in presence of missing values!
in the previous slides, complete case analysis was performed which is biased in presence of informative censoring.
- ✓ naturally accounts for some covariates, even when unobserved.

Is it a good idea? (1/2)

- ✓ makes no assumption about the treatment effect over time
- ✓ simple to carry out
- ⚠ ... except in presence of missing values!
in the previous slides, complete case analysis was performed which is biased in presence of informative censoring.
- ✓ naturally accounts for some covariates, even when unobserved.
- ⚠ does not account for unbalanced in baseline score which can lead to bias if baseline score is correlated to change (Vickers and Altman, 2001).
→ use a linear model instead $Y_i(52) = \alpha + \beta X_i + \gamma Y_i(0) + \varepsilon_i(52)$

Is it a good idea? (1/2)

- ✓ makes no assumption about the treatment effect over time
 - ✓ simple to carry out
 - ⚠ ... except in presence of missing values!
in the previous slides, complete case analysis was performed which is biased in presence of informative censoring.
 - ✓ naturally accounts for some covariates, even when unobserved.
 - ⚠ does not account for unbalanced in baseline score which can lead to bias if baseline score is correlated to change (Vickers and Altman, 2001).
- use a linear model instead $Y_i(52) = \alpha + \beta X_i + \gamma Y_i(0) + \varepsilon_i(52)$
- ⚠ by default, assume no treatment effect on the outcome variability

Is it a good idea? (2/2)

When looking at several timepoints:

	dmean in Placebo	dmean in Active	difference	p.value
week 4	-1.30	-3.51	-2.21	0.04
week 12	-2.27	-5.88	-3.61	0.02
week 24	-5.71	-9.07	-3.36	0.08
week 52	-11.18	-15.48	-4.30	0.06

Is it a good idea? (2/2)

When looking at several timepoints:

	dmean in Placebo	dmean in Active	difference	p.value
week 4	-1.30	-3.51	-2.21	0.04
week 12	-2.27	-5.88	-3.61	0.02
week 24	-5.71	-9.07	-3.36	0.08
week 52	-11.18	-15.48	-4.30	0.06

⚠ multiple testing issue

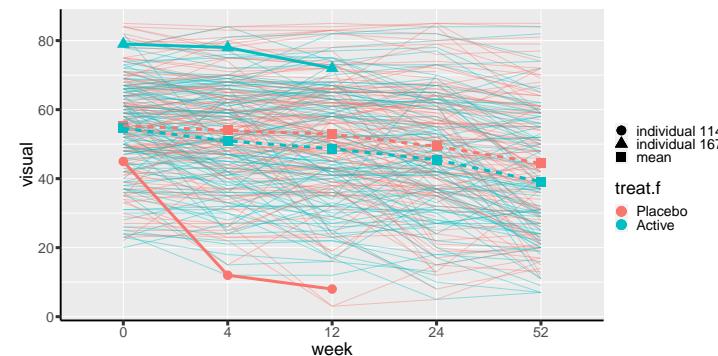
⚠ estimates are timepoint-specific: what about week 30?

Multivariate approach

Better handling missing values

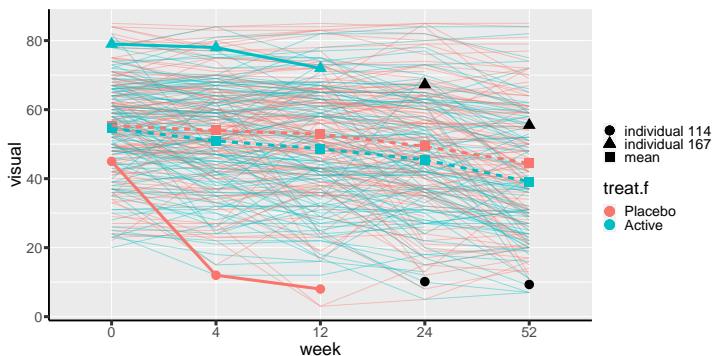
Previously, individuals 114 and 167 were removed from the analysis because of missing outcome at week 52.

Can we do better?



Better handling missing values

Previously, individuals 114 and 167 were removed from the analysis because of missing outcome at week 52. Can we do better?



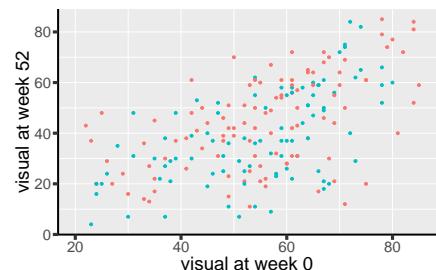
How can we do that? (intuition)

Using a linear model relating the outcome at timepoint(s):

- where the subject has data (e.g. week 0)
- where the subject has no data (e.g. week 52)

The relationship is estimated using data from the other subjects.

We then predict the missing value(s) based on the observed one(s)
using the fitted linear model

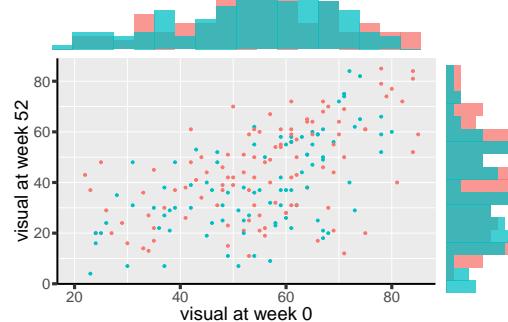


Basic Statistics for health researchers - L8: Repeated measurements

31 / 47

How can we do that? (model)

We can see linear regression as a bivariate normal distribution:

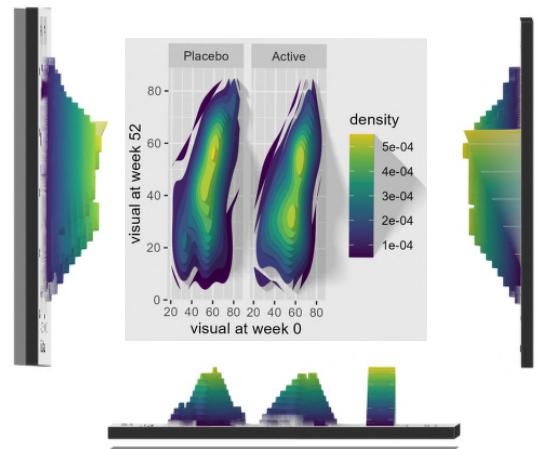


Basic Statistics for health researchers - L8: Repeated measurements

32 / 47

How can we do that? (model)

We can see linear regression as a bivariate normal distribution:

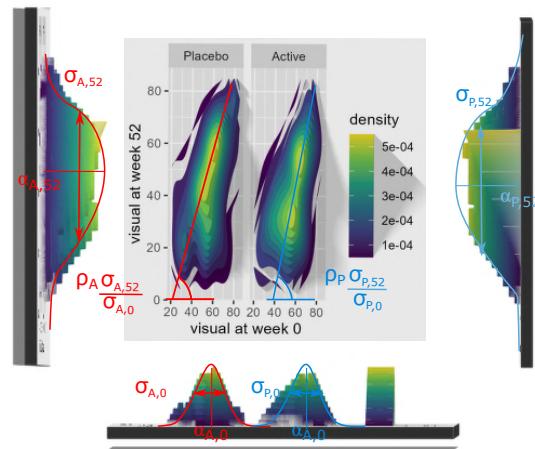


Basic Statistics for health researchers - L8: Repeated measurements

32 / 47

How can we do that? (model)

We can see linear regression as a bivariate normal distribution:



Basic Statistics for health researchers - L8: Repeated measurements

32 / 47

How can we do that? (formula)

Formally, the expected value at one timepoint given the observed value at another is:

$$\hat{Y}_i(52) = \alpha(52) + \rho(0, 52) \frac{\sigma(52)}{\sigma(0)} (Y_i(0) - \alpha(0))$$

→ we need not only to model the **mean** but also the **variance** and **correlation** over time!

→ we assume a joint normal distribution over time

Is it a good idea?

- ✓ the mean will be robust to drop-out depending on past observed outcome values.
(not the case when using complete case analysis)
- ✓ the estimation of the mean will be more precise.

However it requires a more complex model:

- more difficult analysis to carry-out

With complete data, estimates from an adequately parametrized mixed model will match the results from a t-test.

Implementation

In practice we will use what is called a **mixed model**:

- generalization of the univariate linear model (`lm` in R)
- need more inputs: variance and correlation structure
- format of these "new" inputs is software dependent

There are several R package implementing mixed models:

- `nlme` and `lme4`: recommended ones
- `LMMstar`: narrower scope but should be more user-friendly

Example in R with 2 timepoints

```
library(LMMstar)

e.lmm <- lmm(visual ~ treat.f*time, ## mean structure
               repetition = ~ time | subject,
               structure = "UN", ## variance/correlation structure
               data = dfL52)

model.tables(e.lmm)
```

	estimate	se	df	lower	upper	p.value
(Intercept)	55.336	1.37	238	52.64	58.029	0.00e+00
treat.fActive	-0.758	1.93	238	-4.55	3.035	6.94e-01
time52	-11.095	1.55	196	-14.15	-8.038	1.61e-11
treat.fActive:time52	-4.383	2.27	198	-8.87	0.103	5.54e-02

Example in R with 2 timepoints

```
library(LMMstar)

e.lmm <- lmm(visual ~ treat.f*time, ## mean structure
               repetition = ~ time | subject,
               structure = "UN", ## variance/correlation structure
               data = dfL52)

model.tables(e.lmm)
```

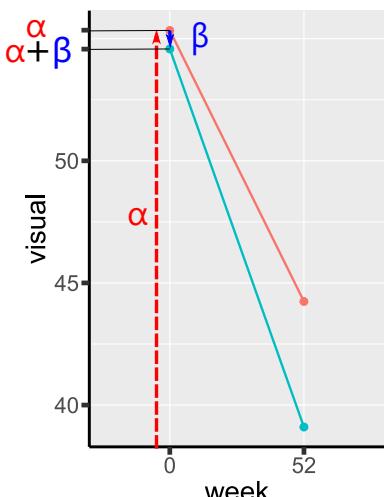
	estimate	se	df	lower	upper	p.value
(Intercept)	55.336	1.37	238	52.64	58.029	0.00e+00
treat.fActive	-0.758	1.93	238	-4.55	3.035	6.94e-01
time52	-11.095	1.55	196	-14.15	-8.038	1.61e-11
treat.fActive:time52	-4.383	2.27	198	-8.87	0.103	5.54e-02



What are those coefficients?

Let's look at the fitted values (1/2)

```
plot(e.lmm, ci = FALSE)
```

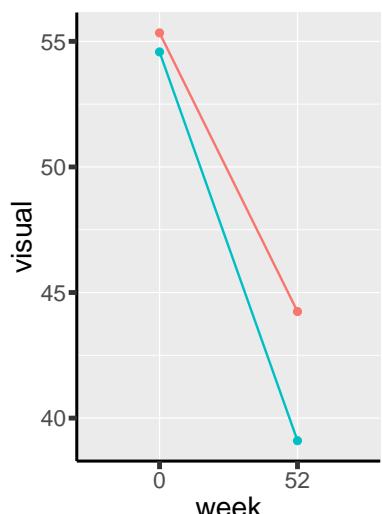


coef(e.lmm)

treat.f	Placebo	Active
(Intercept)	55.3361345	
treat.fActive	-0.7576221	
time52	-11.0948777	
treat.fActive:time52	-4.3831236	

Let's look at the fitted values (1/2)

```
plot(e.lmm, ci = FALSE)
```

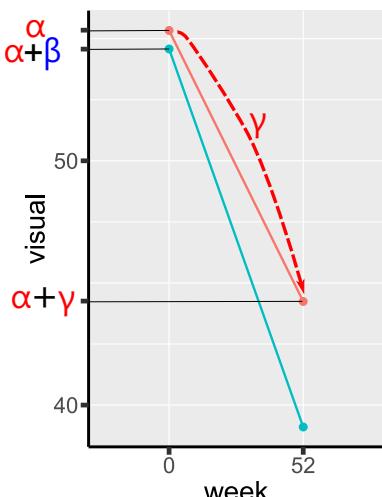


coef(e.lmm)

treat.f	Placebo	Active
(Intercept)	55.3361345	
treat.fActive	-0.7576221	
time52	-11.0948777	
treat.fActive:time52	-4.3831236	

Let's look at the fitted values (1/2)

```
plot(e.lmm, ci = FALSE)
```

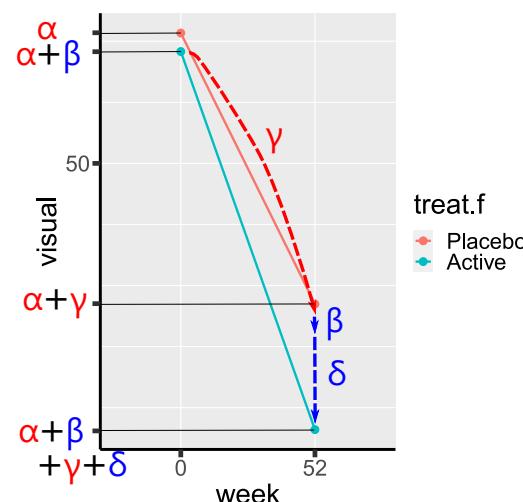


coef(e.lmm)

treat.f	Placebo	Active
(Intercept)	55.3361345	
treat.fActive	-0.7576221	
time52	-11.0948777	
treat.fActive:time52	-4.3831236	

Let's look at the fitted values (1/2)

```
plot(e.lmm, ci = FALSE)
```

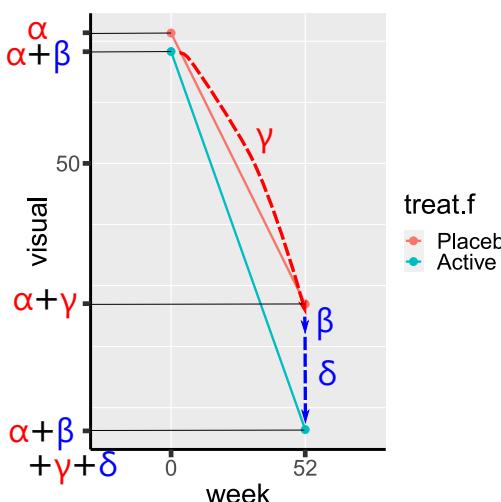


```
coef(e.lmm)
```

	estimate
(Intercept)	55.3361345
treat.fActive	-0.7576221
time52	-11.0948777
treat.fActive:time52	-4.3831236

37 / 47

```
plot(e.lmm, type = "fit", ci = FALSE)
```



```
dummy.coef(e.lmm)[,1:3]
```

treat.f	time	estimate
1 Placebo	0	55.33613
2 Active	0	54.57851
3 Placebo	52	44.24126
4 Active	52	39.10051

38 / 47

Model for the covariance

Unstructured variance/correlation:

$$\text{placebo} \begin{bmatrix} Y_0 \\ Y_{52} \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \alpha \\ \alpha + \gamma \end{bmatrix}, \sigma^2 \begin{bmatrix} 1 & \rho(0, 52)k_{52} \\ \rho(0, 52)k_{52} & k_{52}^2 \end{bmatrix} \right)$$

$$\text{active} \begin{bmatrix} Y_0 \\ Y_{52} \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \alpha + \beta \\ \alpha + \beta + \gamma + \delta \end{bmatrix}, \sigma^2 \begin{bmatrix} 1 & \rho(0, 52)k_{52} \\ \rho(0, 52)k_{52} & k_{52}^2 \end{bmatrix} \right)$$

```
coef(e.lmm, effects = c("variance", "correlation"))
```

sigma	k.52	rho(0,52)
14.9115119	1.2397277	0.5612167

39 / 47

Model for the covariance

Unstructured variance/correlation:

$$\text{placebo} \begin{bmatrix} Y_0 \\ Y_{52} \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \alpha \\ \alpha + \gamma \end{bmatrix}, \sigma^2 \begin{bmatrix} 1 & \rho(0, 52)k_{52} \\ \rho(0, 52)k_{52} & k_{52}^2 \end{bmatrix} \right)$$

$$\text{active} \begin{bmatrix} Y_0 \\ Y_{52} \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \alpha + \beta \\ \alpha + \beta + \gamma + \delta \end{bmatrix}, \sigma^2 \begin{bmatrix} 1 & \rho(0, 52)k_{52} \\ \rho(0, 52)k_{52} & k_{52}^2 \end{bmatrix} \right)$$

⚠ we assume no treatment effect on the variance/correlation

```
coef(e.lmm, effects = c("variance", "correlation"))
```

sigma	k.52	rho(0,52)
14.9115119	1.2397277	0.5612167

39 / 47

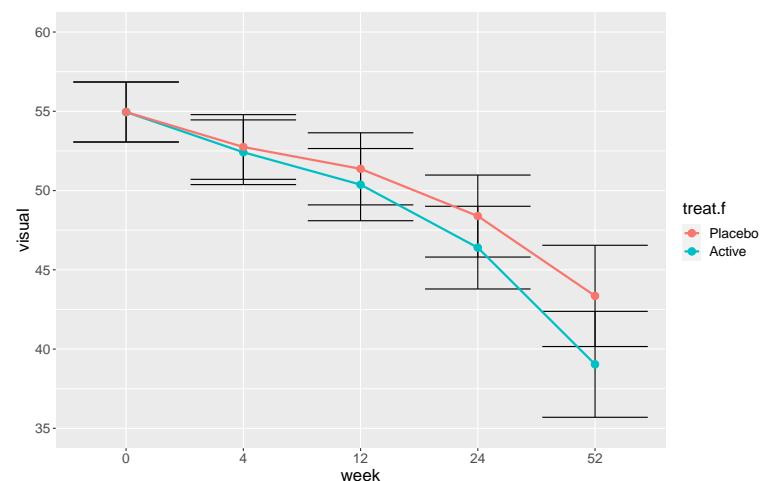
Treatment effect proportional to duration

```
## time: categorical variable ("0", "4", ...)
## week: numeric variable (0, 4, ...)
eLin.lmm <- lmm(visual ~ 0 + time + week:treat.f,
  repetition = ~ week | subject,
  structure = "UN",
  data = dfL)

model.tables(eLin.lmm)
```

	estimate	se	df	lower	upper	p.value
time0	54.954	0.9608	239	53.061	56.84693	0.0000
time4	52.748	1.0359	240	50.707	54.78820	0.0000
time12	51.369	1.1544	257	49.096	53.64263	0.0000
time24	48.391	1.3141	281	45.804	50.97755	0.0000
time52	43.354	1.6209	232	40.160	46.54706	0.0000
week:treat.fActive	-0.083	0.0409	187	-0.164	-0.00231	0.0439

Visualisation



Why using mixed models?

Generalize t-test on the change:

- equivalent with 2 endpoints and no missing data

Better handling of missing values:

- full information instead of complete case analysis



- ✓ no need to model the cause of censoring



- ⚠ require valid model for the mean/covariance structure

Can ease interpretability:

- imposing constant or linear treatment effect over time

When not to use mixed models?

No missing data and only two timepoints

- a univariate analysis on the change from baseline is often enough

Very small sample size:

- model parameters can be difficult to estimate
- possible inflation of type 1 error
(can be solved with specialized tests)

In presence of competing risks (e.g. death)

- mixed model are not a "magic" solution for missing values ...

Collecting several measurements per subject is a good idea:

- give more insight into the treatment effect
- better handling of missing data
- reduce uncertainty/confounding
(each subject is its own control)

Scheduled measurement times is recommended.

But is also challenging:

- more demanding for the patient (drop-out!)
- more complex to organize
(e.g. ensure subjects follow the schedule)
- often require dedicated/advanced statistical tools

What we have seen today

✓ Introduction to repeated measurements

- definition and examples of study design
- benefit of having repeated measurements
- challenges for the statistical analysis

✓ Example of longitudinal study

- descriptive statistics and plots for repeated measurements
- concerns due to the presence of missing values
- what is a long and wide format

✓ Univariate approach

- adjustment resulting from working on change from baseline
- treatment effect assessment using a two sample t-test on the change
- pros and cons

✓ Multivariate approach

- intuition behind handling missing values using a multivariate model
- parametrization of a linear mixed model (mean and covariance)
- pros and cons

Want to know more?

Ph.D. course:

- Statistical analysis of correlated and repeated measurements
(course director: Julie Forman)

Contents

This course is concerned with analysis of correlated quantitative data arising e.g. when taking observations from clusters of subjects, repeatedly over time on the same subjects, or by applying different treatment to different parts of the body. Pitfalls of traditional statistical analyses will be discussed and appropriate models for the analysis of e.g. baseline follow-up studies, cross-over studies, and cluster randomized trials will be exemplified.

For supplementary reading we recommend:

- FLW: G.M. Fitzmaurice, N.M. Laird and J.H. Ware, Applied Longitudinal Analysis (2nd edition), John Wiley & sons, 2011.

Please note that the book is available as e-book on KB (free download for KU students).

Day	Topics	Suggested reading*
1	Introduction to repeated measurements and clustered data. Basic theory of linear mixed models. Analysis of single group studies. Handling repeated measurements in SAS/R.	FLW 1-3. Tutorial 1.
2	Longitudinal data analysis. Models for balanced and unbalanced designs. Analysis of randomized baseline follow-up studies.	FLW 5-7. Tutorial 2.
3	Analysis of clustered data. Variance components. Multi-level models. The linear growth model.	FLW 8, 21 & 22.
4	Select topics in linear mixed models. Cross-over studies. Repeatability and reproducibility of measurement methods.	Lecture notes only.
5	Models for binary and count data. Generalized linear mixed models. Marginal models and generalized estimating equations.	FLW 10-16
6	Missing data. Consequences and statistical handling.	FLW 17-18

Reference I

(1997). Interferon alfa-2a is ineffective for patients with choroidal neovascularization secondary to age-related macular degeneration: Results of a prospective randomized placebo-controlled clinical trial. *Archives of Ophthalmology*, 115(7):865–872.

Beliveau, V., Ganz, M., Feng, L., Ozenne, B., Højgaard, L., Fisher, P. M., Svarer, C., Greve, D. N., and Knudsen, G. M. (2017). A high-resolution *in vivo* atlas of the human brain's serotonin system. *Journal of Neuroscience*, 37(1):120–128.

Kameran, P. R. and Vollert, J. (2022). Greater baseline pain inclusion criteria in clinical trials increase regression to the mean effect: a modelling study. *Pain*, 163(6):e748–e758.

Wide to long format

subject	group	outcome	
		(time = 0)	(time = 4)
1	A	59	55
2	A	65	70
3	P	40	40
..

Long

Wide to long format

subject	group	outcome	
		(time = 0)	(time = 4)
1	A	59	55
2	A	65	70
3	P	40	40
..

subject	group	time	outcome	
			0	4
1	A	0	59	55
2	A	4	65	70
3	P	0	40	40
..

Wide to long format

		outcome	
subject	group	(time = 0)	(time = 4)
1	A	59	55
2	A	65	70
3	P	40	40
..

Long			
subject	group	time	outcome
1	A	0	59
1	A	4	55
?	?	0	?
?	?	4	?
		0	
		4	
		..	

Equivalence t-test and mixed model (2/3)

Mixed model on the change (complete case week 0 and 52):

```
armd.wideCC$change <- armd.wideCC$visual52-armd.wideCC$visual0
e2CC.lmm <- lmm(change ~ treat.f,
  repetition = ~ treat.f | subject, structure = "UN",
  data = armd.wideCC)
model.tables(e2CC.lmm)
```

	estimate	se	df	lower	upper	p.value
(Intercept)	-11.2	1.60	104	-14.36	-8.002	2.94e-10
treat.fActive	-4.3	2.28	192	-8.79	0.201	6.11e-02

Equivalence t-test and mixed model (1/3)

t-test (complete case week 0 and 52):

```
armd.wideCC <- na.omit(armd.wide[,c("subject","treat.f",
  "visual0","visual52")])
t.test(visual52-visual0 ~ treat.f,
  data = armd.wideCC)
```

Welch Two Sample t-test

```
data: visual52 - visual0 by treat.f
t = 1.8842, df = 191.47, p-value = 0.06106
alternative hypothesis: true difference in means between group P
95 percent confidence interval:
-0.2013017 8.7949525
sample estimates:
mean in group Placebo mean in group Active
-11.18095 -15.47778
```

Equivalence t-test and mixed model (3/3)

Mixed model on time-specific data (complete case week 0 and 52):

```
test.CC <- armd.long$week %in% c("0","52") & armd.long$subject %in% subjCC
e2CC.lmm <- lmm(visual ~ week*treat.f,
  repetition = ~ week | subject, structure = "UN",
  data = armd.long[testCC,])
model.tables(e2CC.lmm)[ "week52:treat.fActive", , drop=FALSE]
c("Placebo" = as.double(coef(e2CC.lmm)[ "week52"]),
  "Active" = sum(coef(e2CC.lmm)[ c("week52", "week52:treat.fActive")]))
```

	estimate	se	df	lower	upper	p.value
week52:treat.fActive	-4.3	2.29	193	-8.82	0.224	0.0624
Placebo	-11.18095	-15.47778				