Basic Statistic for health researchers
Lecture 8: repeated measurements

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

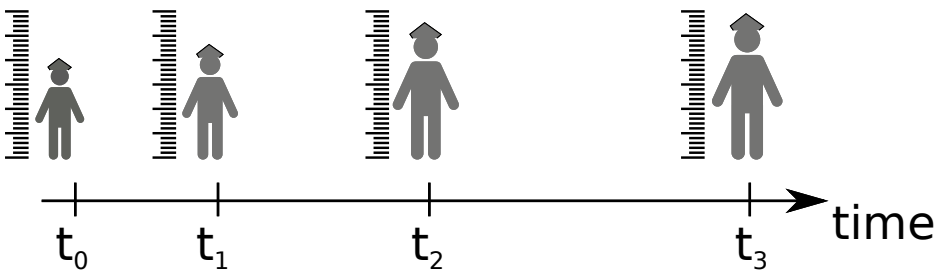
[1] Section of Biostatistics, Department of Public Health, University of Copenhagen
[2] Neurobiology Research Unit, University Hospital of Copenhagen, Rigshospitalet.

20-11-2023

Repeated measurements

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

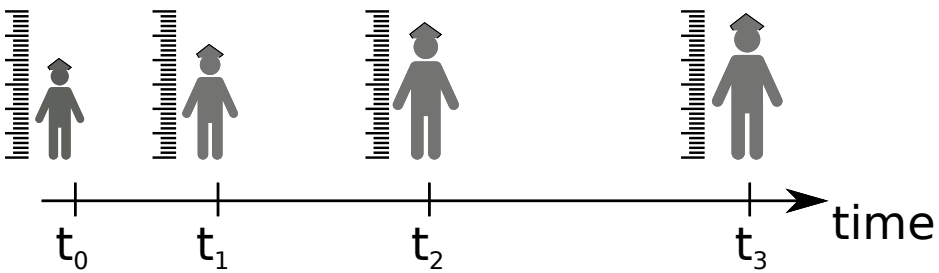
- Longitudinal study: **outcome** measured on the **same patient** at **different timepoints**.



Repeated measurements

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

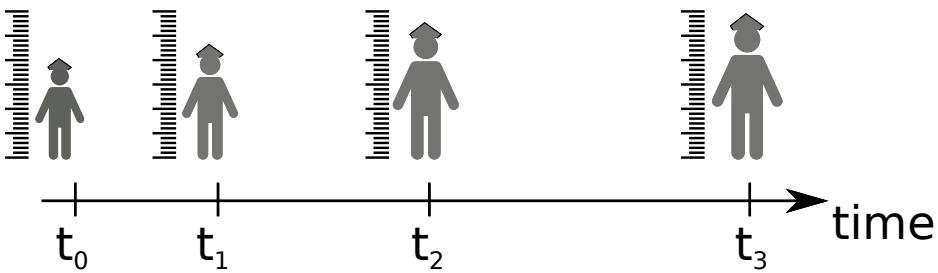
- Longitudinal study: **outcome** measured on the **same patient** at **different timepoints**.



Repeated measurements

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

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

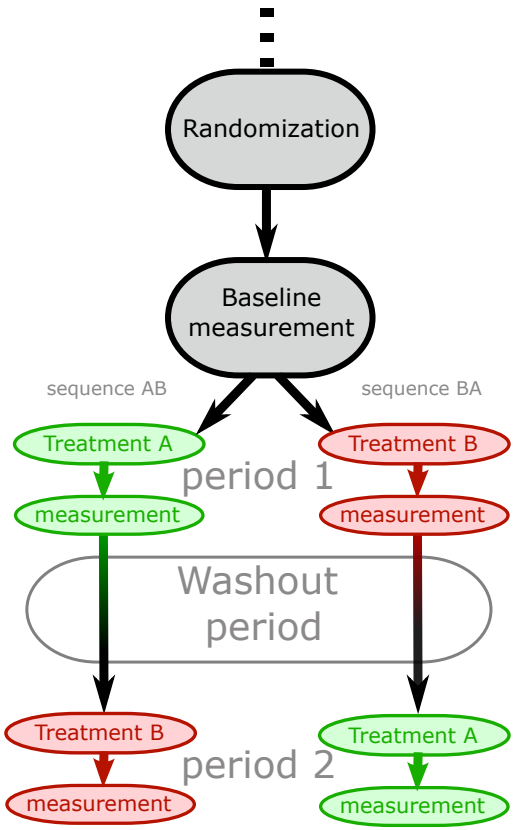


Can you find other examples?

- what motivates collecting repeated measurements?

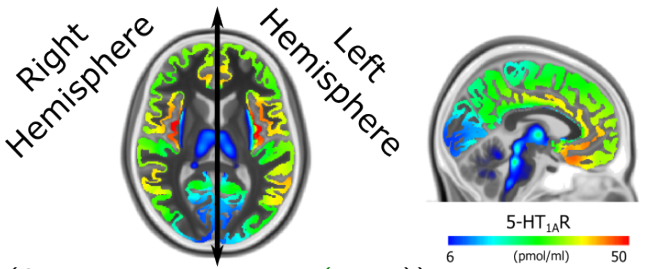
Other designs involving repeated measurements (1/2)

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



Other designs involving repeated measurements (2/2)

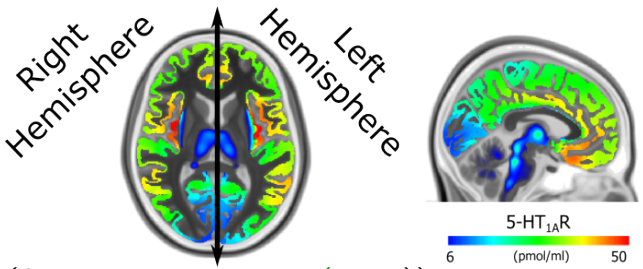
- the **same** type of **measurement** on the **same patient** at **different locations**.



(from Beliveau et al. (2017))

Other designs involving repeated measurements (2/2)

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

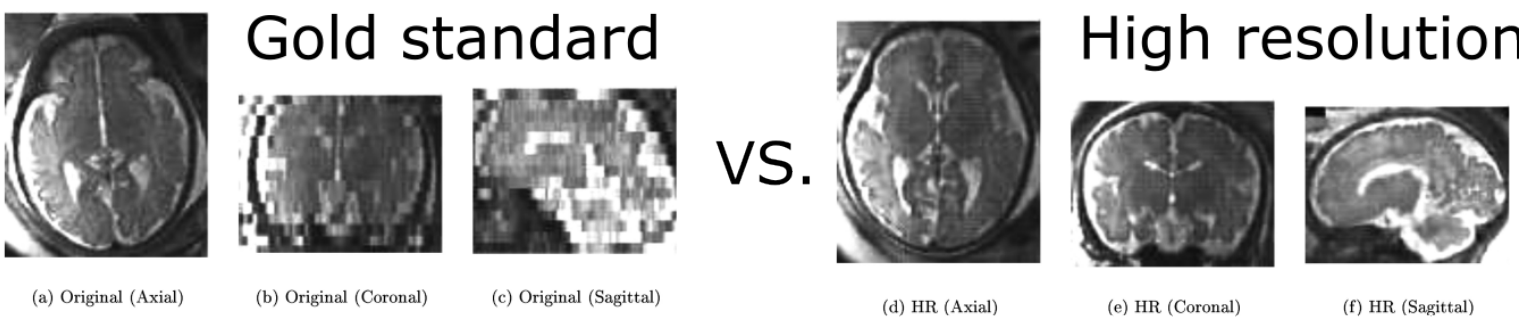


(from Beliveau et al. (2017))

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

$\rightarrow$ assess the stability of a measurement device

$\rightarrow$ comparison of diagnostic tests (Mc Nemar test in lecture 5)

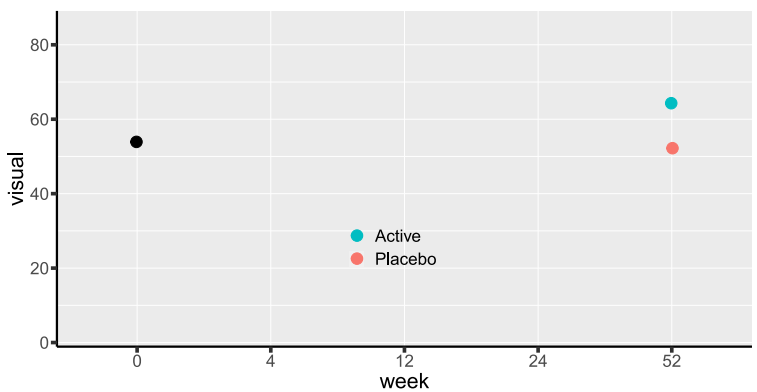


(from Van Reeth et al. (2012))

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?

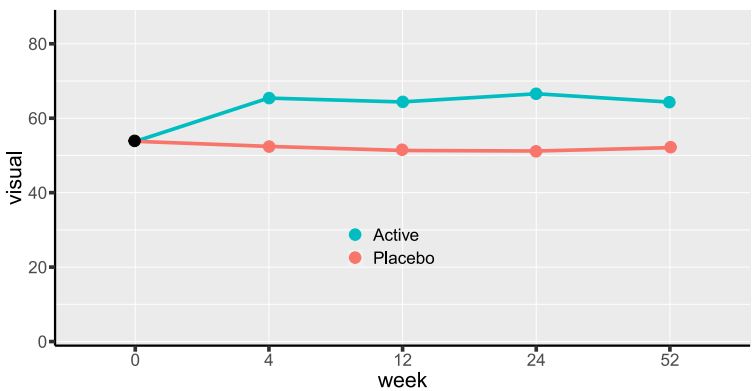


Why using repeated measurements? (1/3)

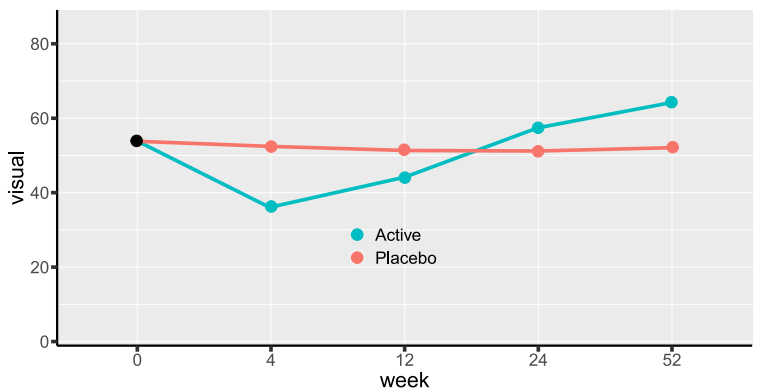
To **better understand** the time-dynamic of the **exposure**:

- 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?

Constant effect



Time varying effect



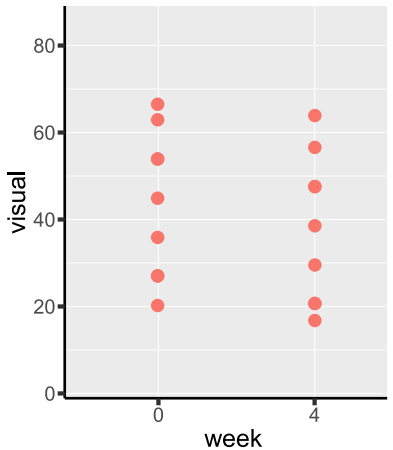
Why using repeated measurements? (2/3)

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



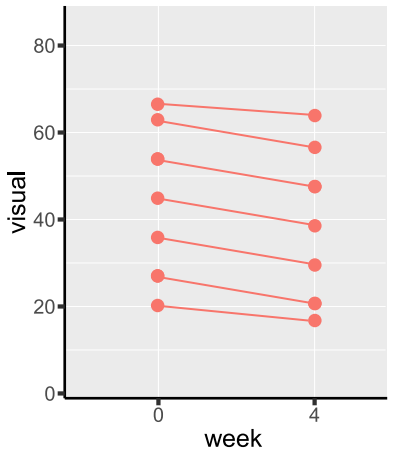
Why using repeated measurements? (2/3)

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



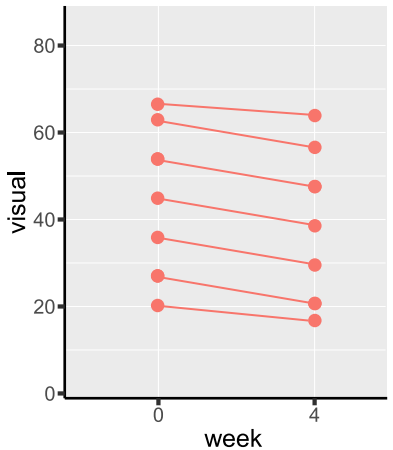
Why using repeated measurements? (2/3)

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. time trend, regression to the mean

require specific modeling

Example of regression to the mean (Kamerman 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."

Why using repeated measurements? (3/3)

To **better handle missing values**:

- as the follow-up time increases, patient 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

Outline

✔ Introduction to repeated measurements

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

Example of longitudinal study

Univariate approach

Multivariate approach

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

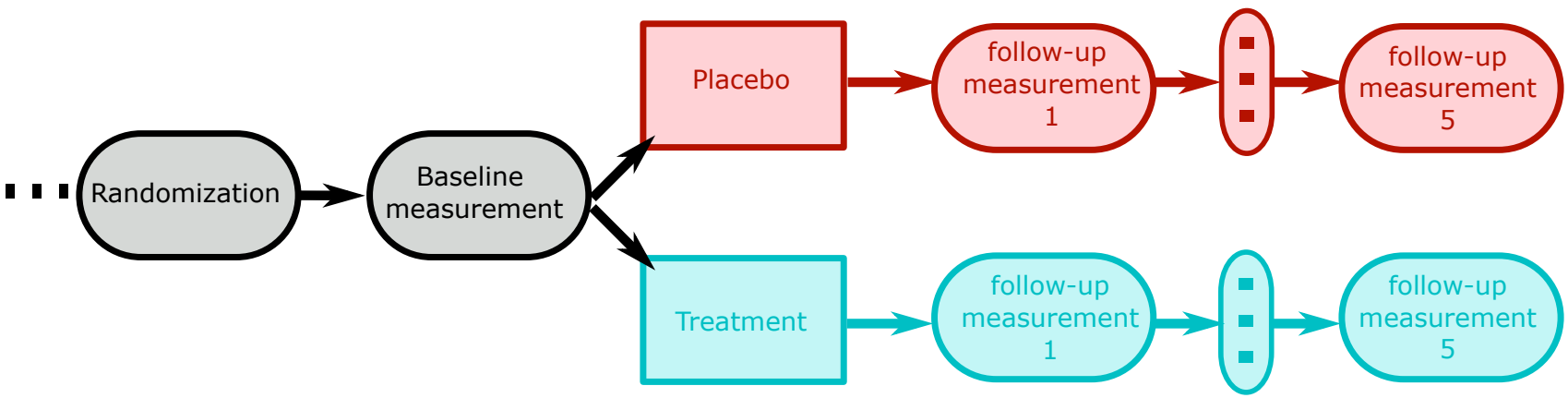
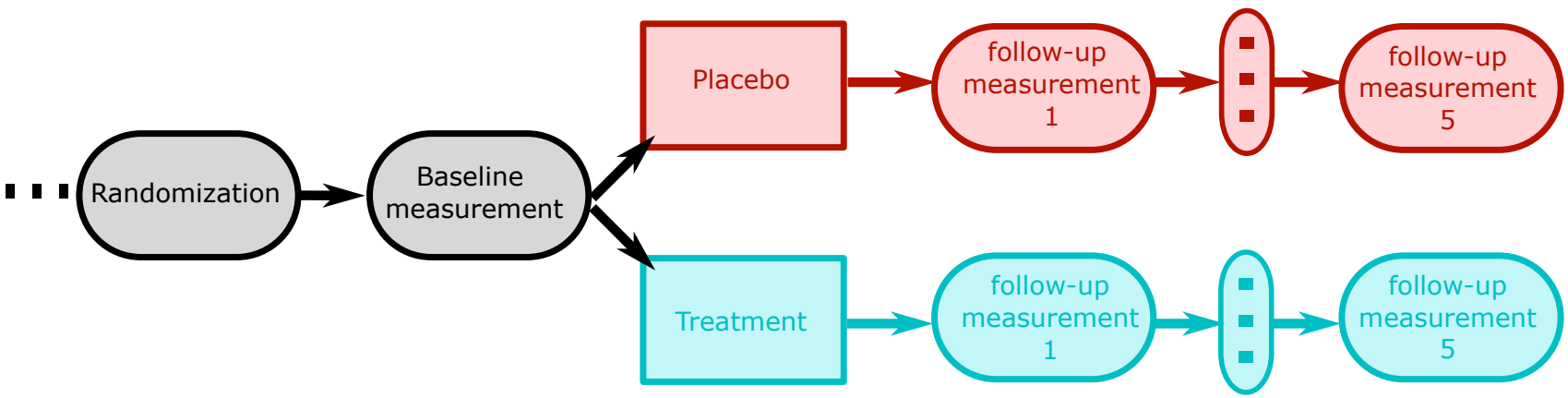


Illustration: ARMD trial (int, 1997)

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)
  $\rightarrow$ independent outcome replicates at the cluster level
- **repetition** variable: time

Wide format

Data in the wide format (`dfW`):

- 1 row = 1 subject ("level 1 data")

$\rightarrow$ 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 visual4 "..." 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
```

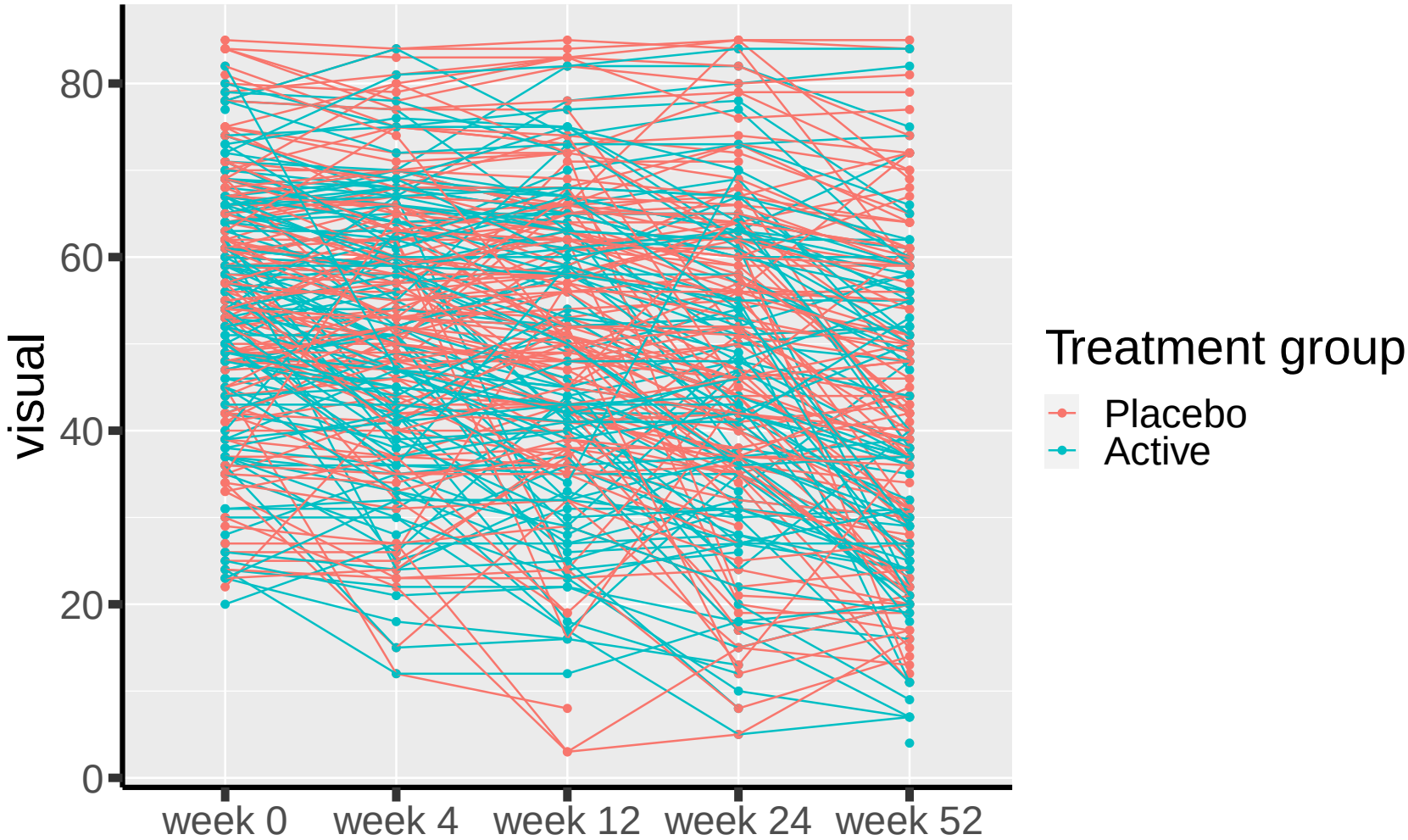
Long format

Data in the long format (`dfL`):

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

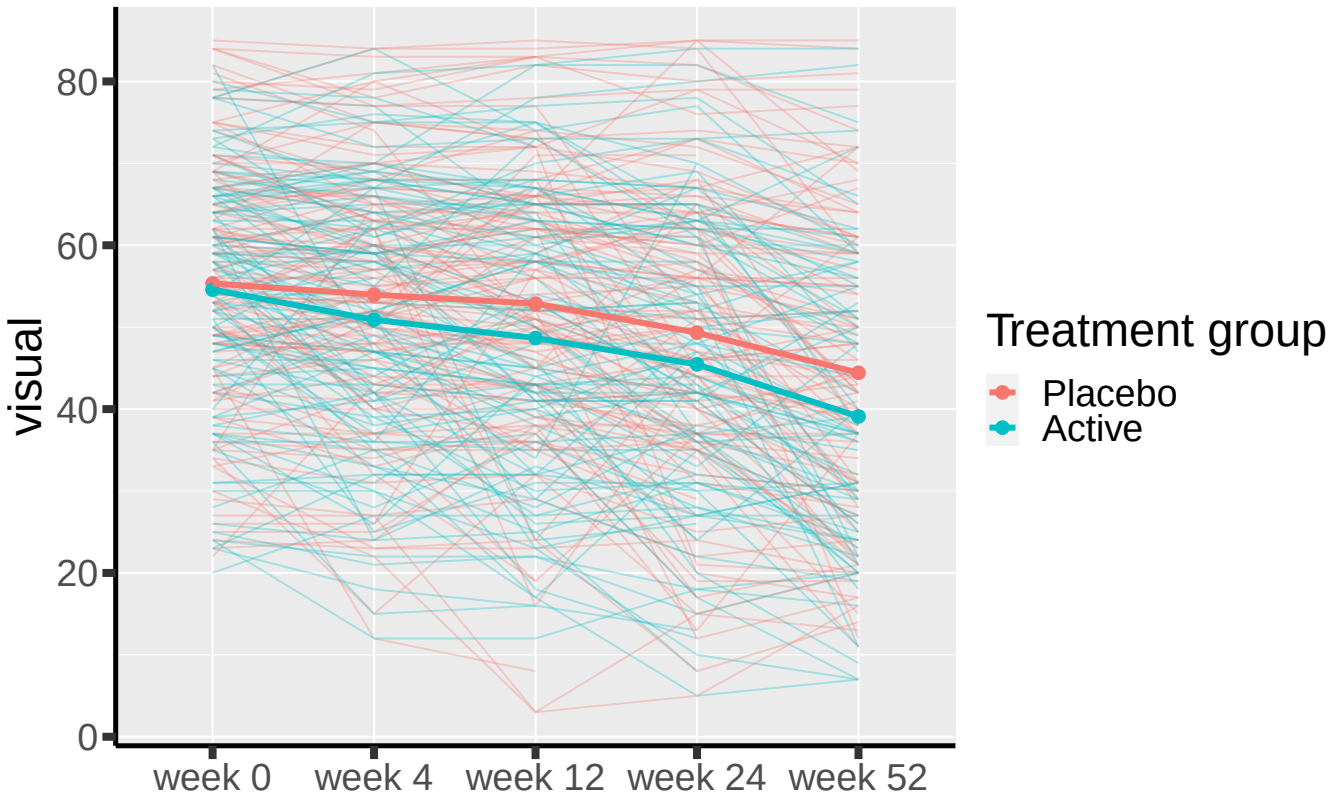
	subject	treat.f	week	visual
1	1	Active	0	59
2	1	Active	4	55
3	1	Active	12	45
4	1	Active	24	NA
5	1	Active	52	NA
6	2	Active	0	65
7	2	Active	4	70
8	2	Active	12	65
9	2	Active	24	65
10	2	Active	52	55

Visualizing the data: spaghetti plot



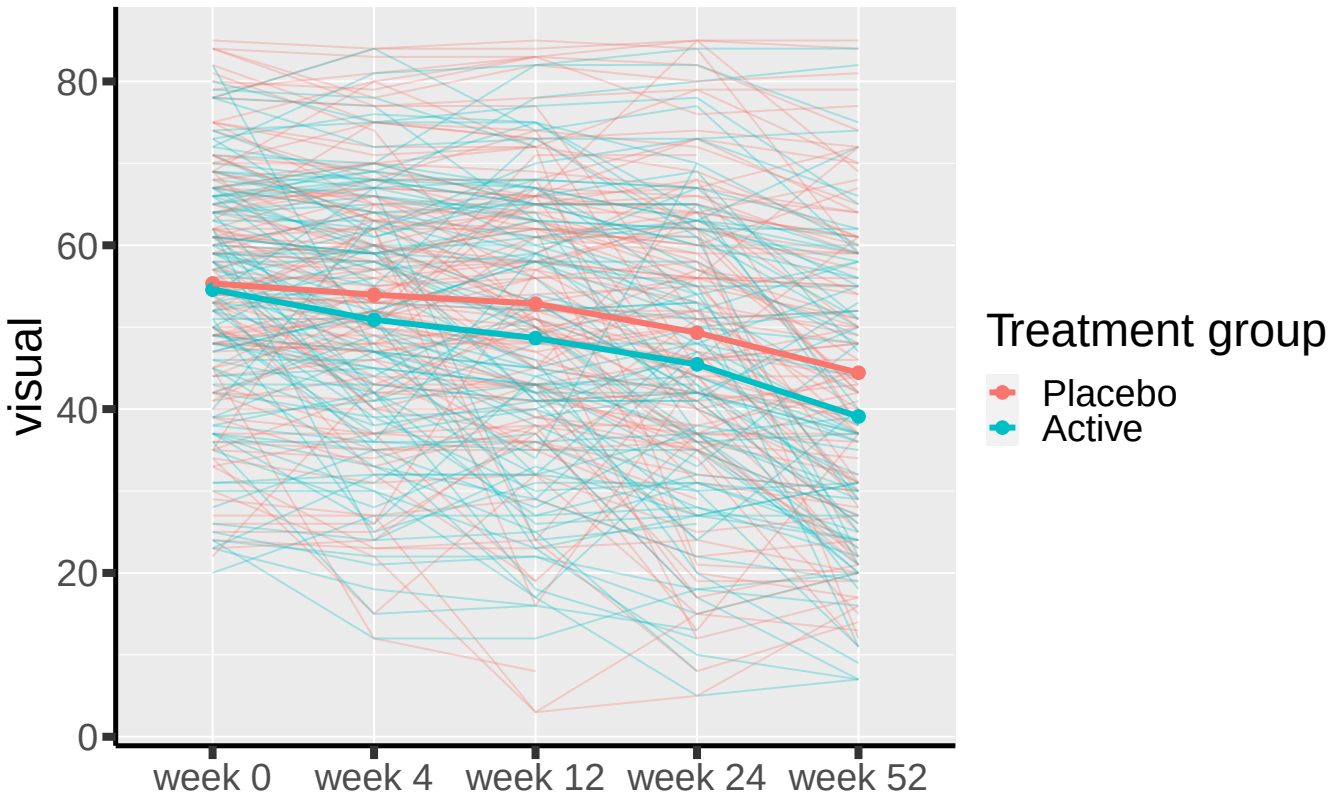
Summary statistics (1/3)

- using the mean by group and timepoint:



Summary statistics (1/3)

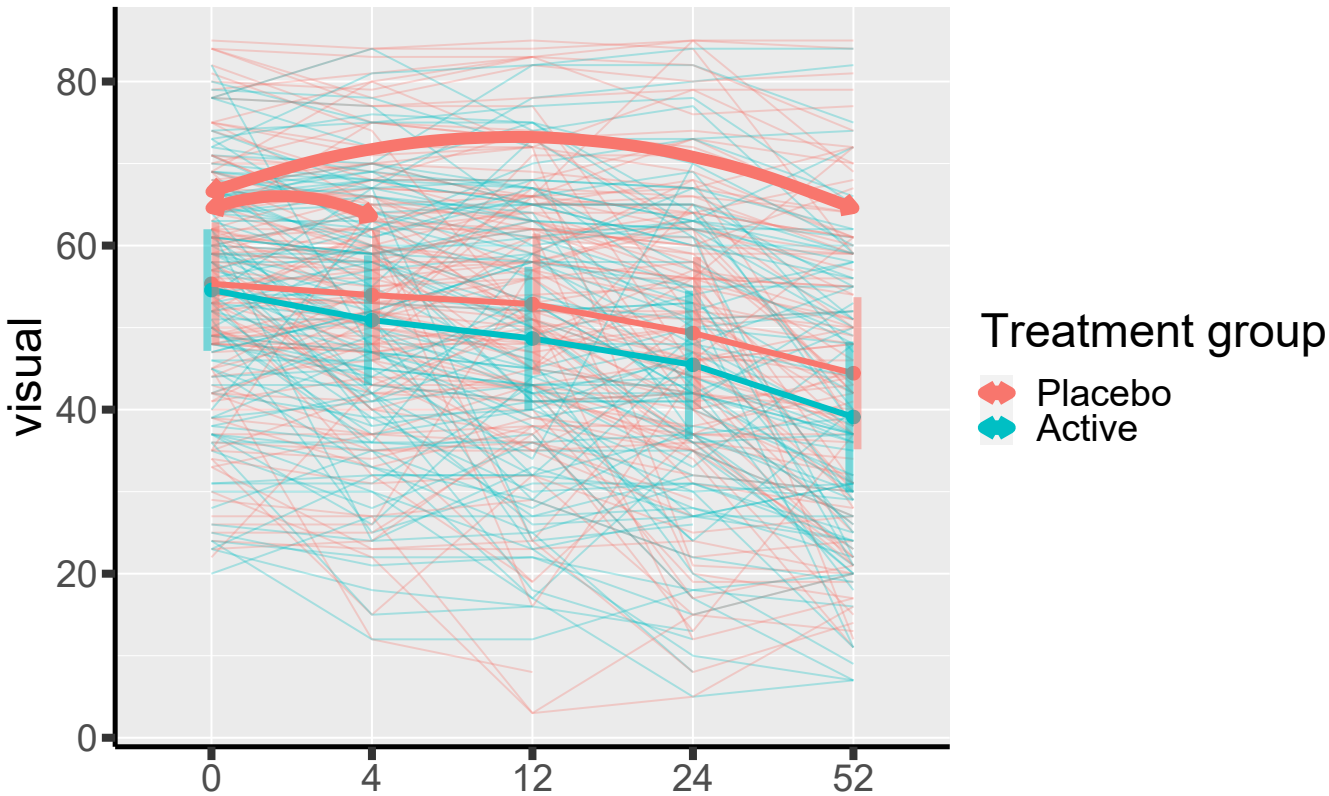
- using the mean by group and timepoint:



Other statistics you would use to summarize the data

Summary statistics (1/3)

- using the mean by group and timepoint:



Other statistics you would use to summarize the data

Summary statistics (2/3)

- dispertion over time (standard deviation)

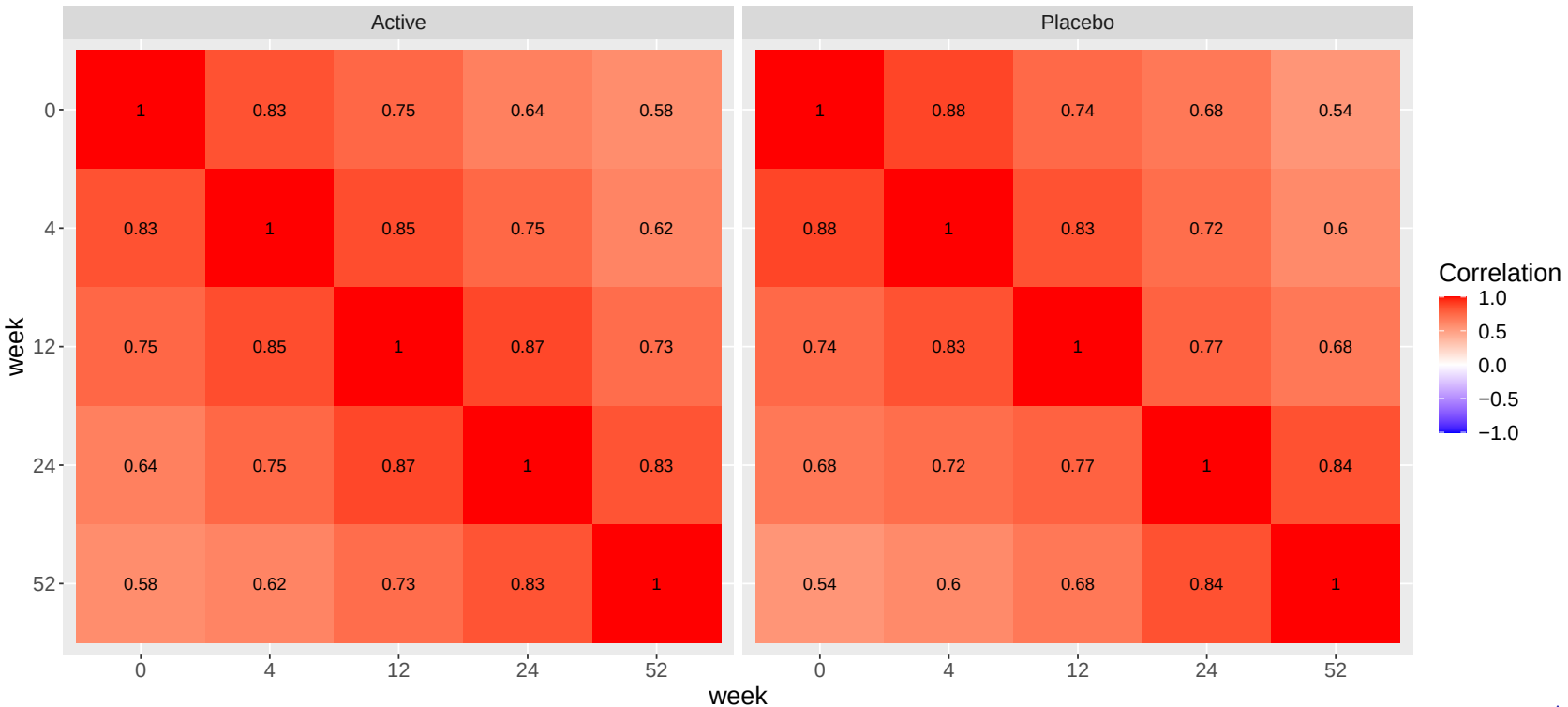
	week 0	week 4	week 12	week 24	week 52
Placebo	15.33143	15.38915	16.51203	18.61137	18.68844
Active	14.32523	15.99285	17.35207	17.84161	18.36214

Summary statistics (2/3)

- dispertion over time (standard deviation)

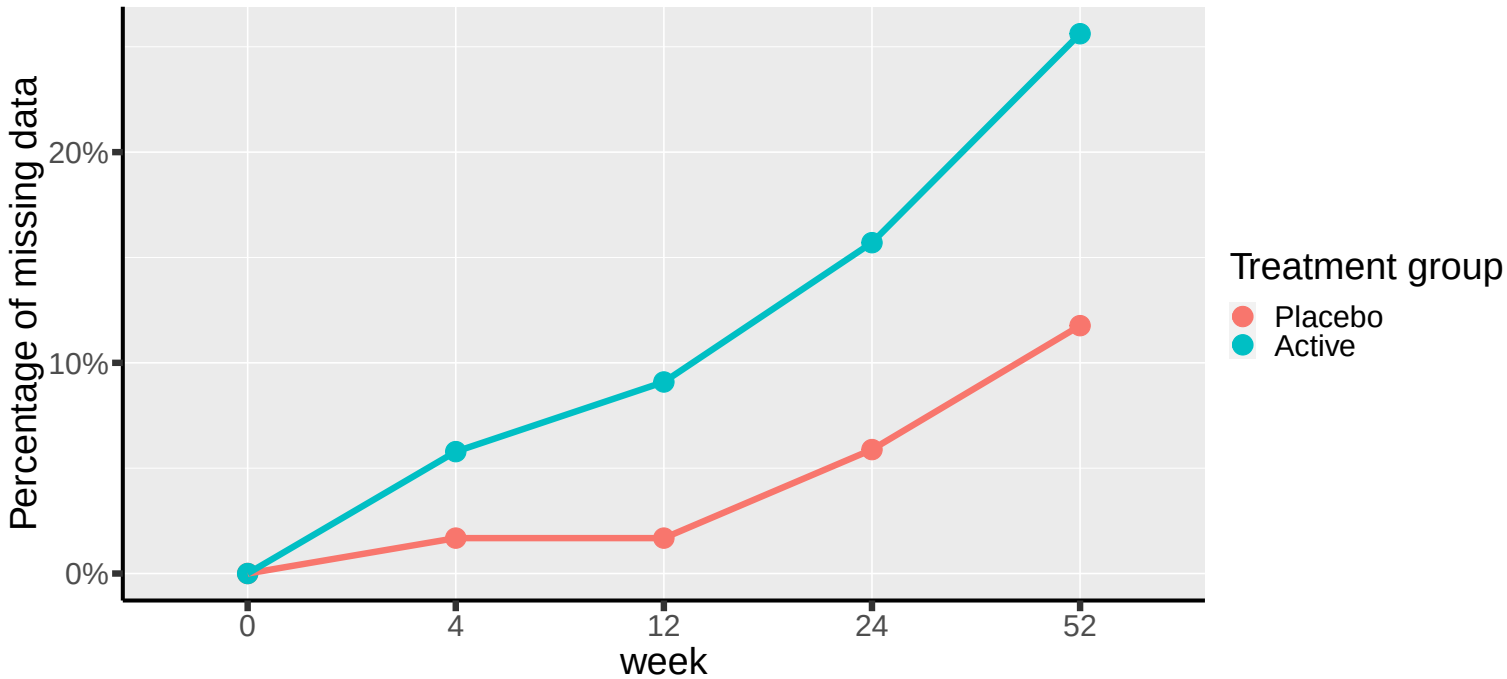
	week 0	week 4	week 12	week 24	week 52
Placebo	15.33143	15.38915	16.51203	18.61137	18.68844
Active	14.32523	15.99285	17.35207	17.84161	18.36214

- dependency in visual acuity over time (Pearson correlation)



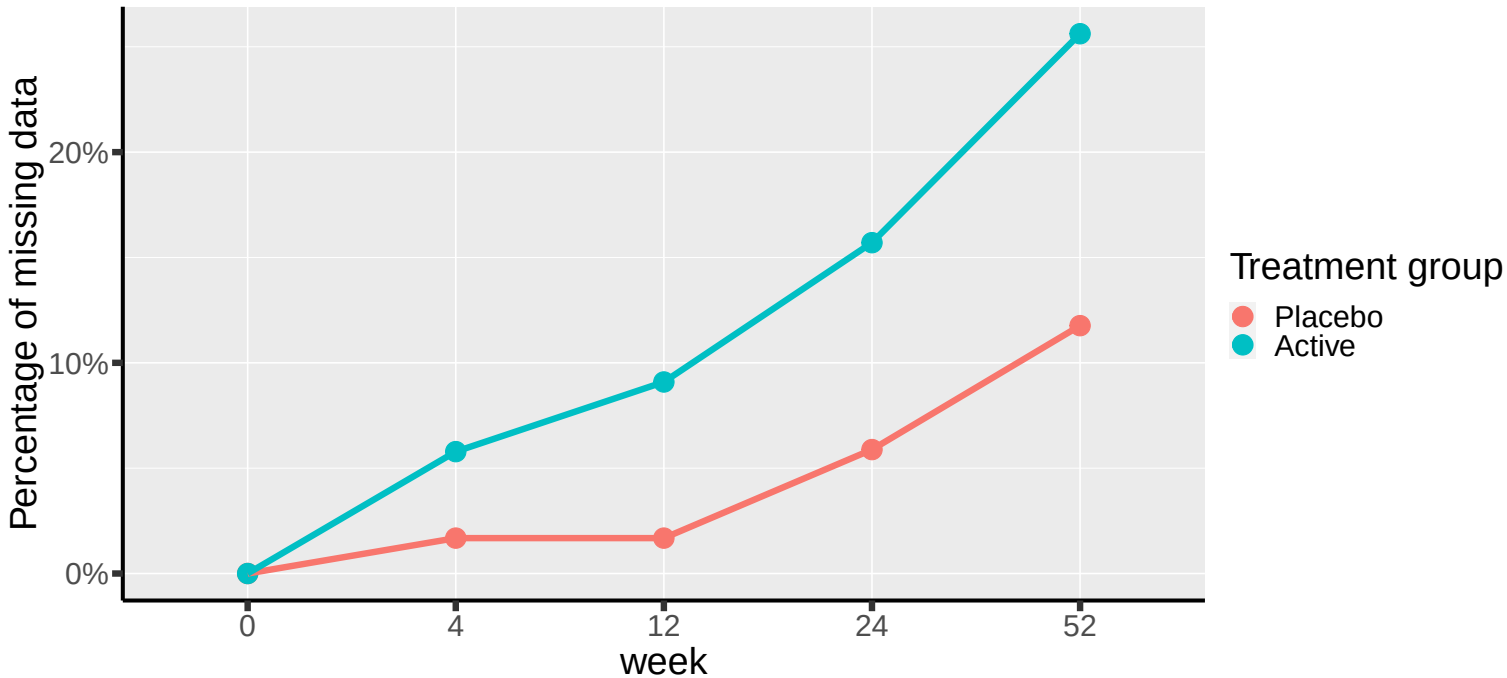
Summary statistics (3/4)

- what about missing values?



Summary statistics (3/4)

- what about missing values?



Concerns:

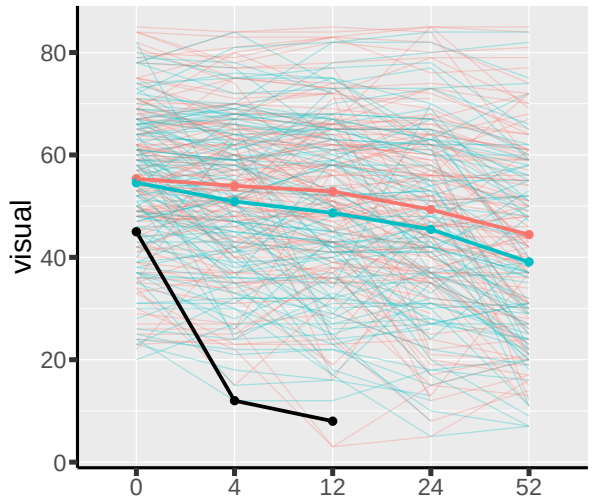
- treatment side effect(s) not measured by the outcome
- missing not at random may bias the estimated mean
  (upward bias if patients with weak vision are more likely to drop)

Summary statistics (4/4)

- missing data patterns:

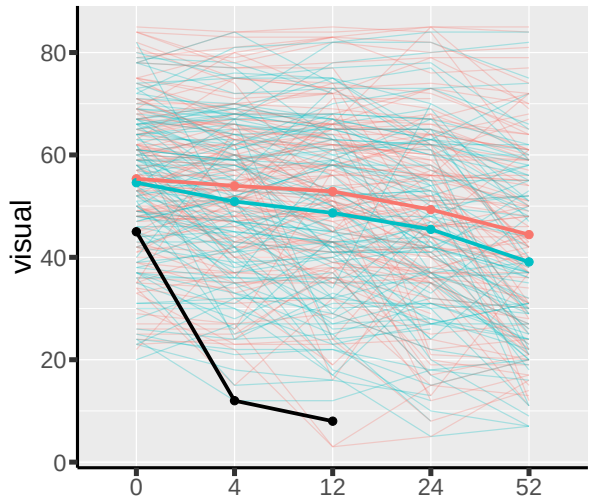
frequency	pattern	visual0	visual4	visual12	visual24	visual52
188	00000	0	0	0	0	0
24	00001	0	0	0	0	1
4	00010	0	0	0	1	0
8	00011	0	0	0	1	1
1	00110	0	0	1	1	0
6	00111	0	0	1	1	1
2	01000	0	1	0	0	0
1	01011	0	1	0	1	1
6	01111	0	1	1	1	1

Different types of missing data



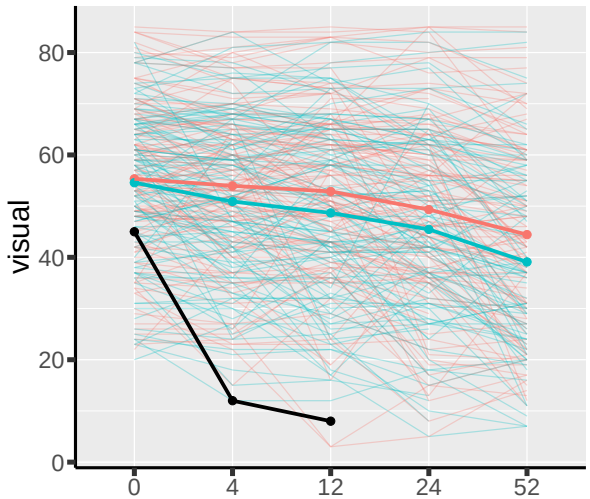
Different types of missing data

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



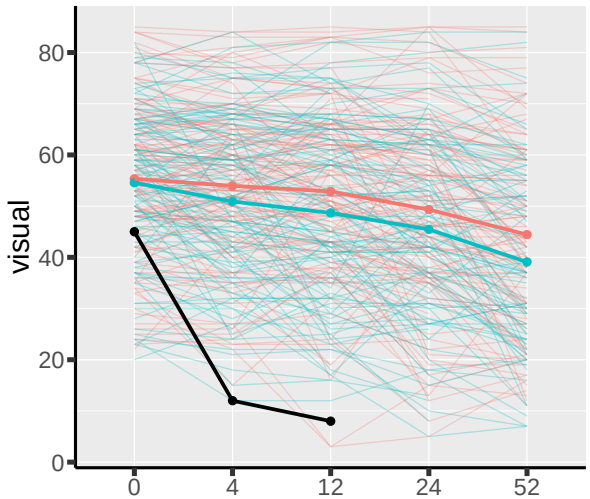
Different types of missing data

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



Different types of missing data

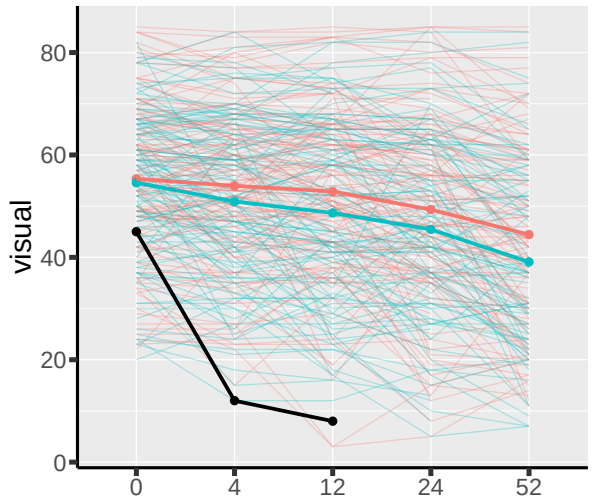
- drop-out (patients leaving the study)
  Informative censoring
  vs. censoring completely at random
- competing risks (e.g. death)
  Complete case analysis usually wrong



- unbalanced data: measurement times differ between patients
  Selection bias when sick patients have earlier or more frequent visits

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!

Software considerations

Data management is more complex with repeated measurements:

- conversion from wide to long format
- evaluating statistics (e.g. mean) per group of rows (e.g. per time)
- connect points on a graph belonging to the same subject

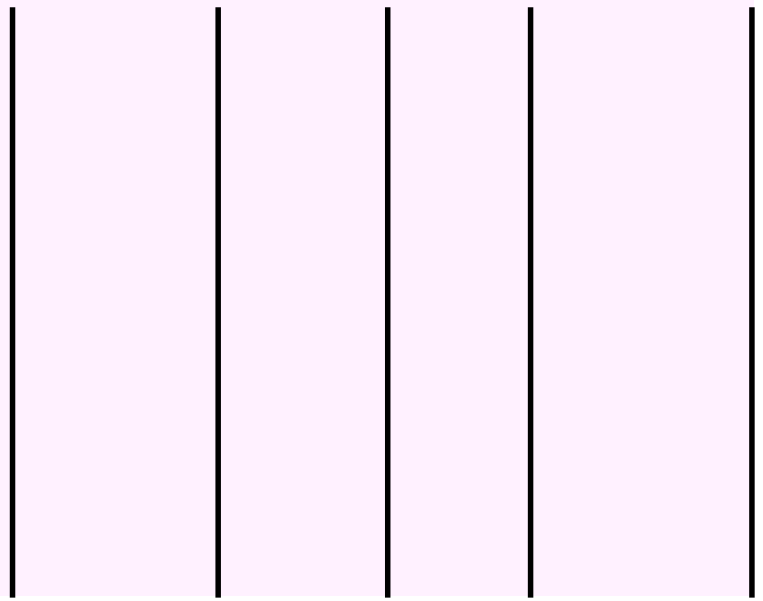
$\rightarrow$ 'new'  functions can be helpful

Wide to long format

Wide

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

Long



Wide to long format

Wide

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

Long

subject	group	time	outcome
		0	
		4	
		0	
		4	
		0	
		4	
		..	

Wide to long format

Wide

subject	group	outcome (time = 0)	outcome (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	
		..	

R code: from wide to long format

```
col.visual <- paste0("visual",c(0,4,12,24,52))
```

```
[1] "visual0"  "visual4"  "visual12" "visual24" "visual52"
```

```
dfL <- reshape(dfW, direction = "long",
   ## information to retrieve in dfW
   varying = col.visual, idvar = "subject",
   ## column names & values in dfL
   timevar = "week.num", times = c(0,4,12,24,52),
   v.names = "visual")
```

	subject	lesion	line0	treat.f	miss.pat	week.num	visual
1	1	3	12	Active	--XX	0	59
241	1	3	12	Active	--XX	4	55
481	1	3	12	Active	--XX	12	45
721	1	3	12	Active	--XX	24	NA
961	1	3	12	Active	--XX	52	NA

code: summary statistics with LMMstar

```
summarize(visual ~ week.num, data = dfL, na.rm = TRUE)
```

```
week.num observed missing pc.missing mean   sd
       0      240       0          0   55 14.9 ...
       4      231       9     0.0375 52.5 15.9 ...
     ...      ...     ...        ...  ...  ... ...
```

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
..	..	..	....

R code: summary statistics with LMMstar

```
summarize(visual ~ week.num, data = dfL, na.rm = TRUE)
```

```
week.num observed missing pc.missing mean   sd
       0      240       0          0   55 14.9 ...
       4      231       9     0.0375 52.5 15.9 ...
     ...      ...     ...       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	....

code: summary statistics with LMMstar

```
summarize(visual ~ week.num, data = dfL, na.rm = TRUE)
```

```
week.num observed missing pc.missing mean   sd
       0      240       0          0   55 14.9 ...
       4      231       9     0.0375 52.5 15.9
     ...      ...     ...        ...  ...  ...
```

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	....

code: summary statistics with LMMstar

```
summarize(visual ~ week.num, data = dfL, na.rm = TRUE)
```

```
week.num observed missing pc.missing mean   sd
       0      240       0          0   55 14.9 ...
       4      231       9     0.0375 52.5 15.9 ...
                                      ... ... ...
```

Split 1

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

Summarize

mean

59
65
40
....

Split 2

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

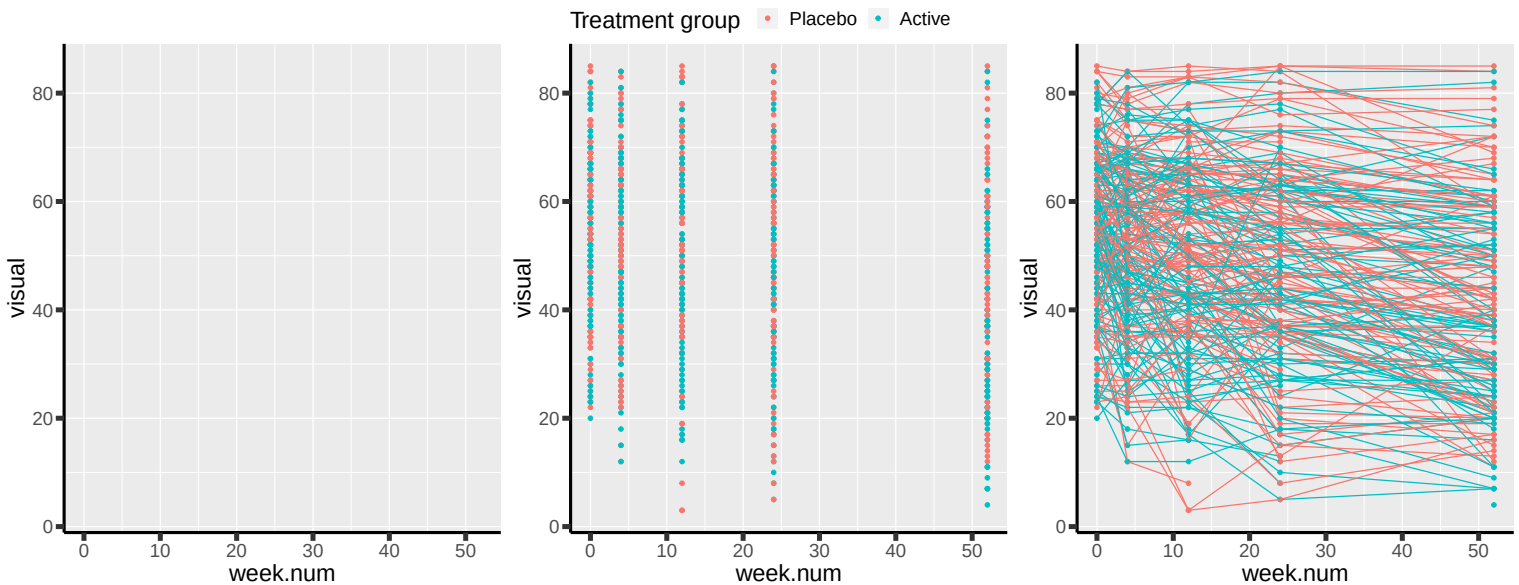
mean

55
70
40
....

code: spaghetti plot

```
library(ggplot2)

gg.spa <- ggplot(dfL, aes(x = week.num, y = visual,
                          group = subject, color = treat.f))
gg.spa <- gg.spa + labs(color = "Treatment group")
gg.spa <- gg.spa + geom_point() + geom_line()
gg.spa
```



Univariate approach

Data analysis

What would you do if you we were asked to assess the long term treatment effect?
(to simplify assume no missing data and no covariate)

- what should you not do? Why?

```
dfW.CC <- dfW[rowSums(is.na(dfW))==0,
              c("subject","treat.f",col.visual)]
```

	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

Challenge: 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: 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

Data analysis - possibilities

- linear regression/t-test on the final value

```
lm(visual52 ~ treat.f, data = armd.wide)
t.test(visual52 ~ treat.f, data = armd.wide)
```

- linear regression/t-test on the change from baseline

```
lm(visual52-visual0 ~ treat.f, data = armd.wide)
t.test(visual52-visual0 ~ treat.f, data = armd.wide)
```

- linear regression on the final value adjusted for baseline

```
lm(visual52 ~ visual0 + treat.f, data = armd.wide)
```

- ~~linear regression of vision as a function of time and group~~

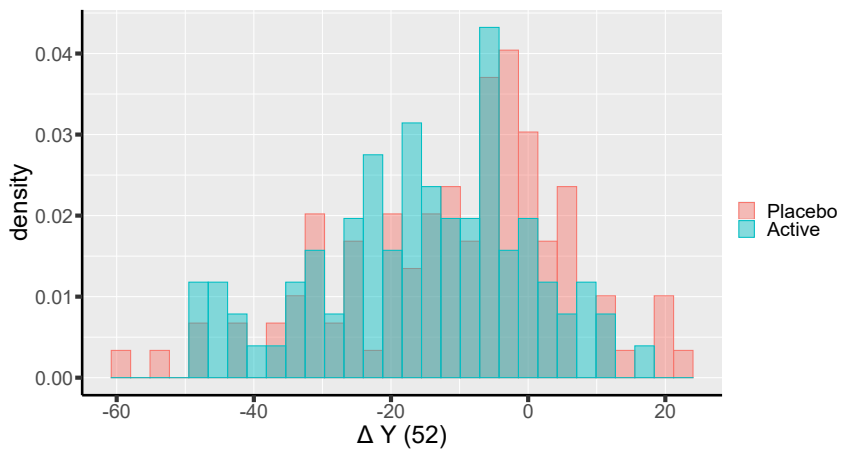
```
lm(visual ~ week * treat.f, data = armd.long)
```

t-test on the change (1/2)

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



t-test on the change (2/2)

3. Compare the change between groups using a statistical test

t-test on the change (2/2)

3. Compare the change between groups using a statistical test

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

t-test on the change (2/2)

3. Compare the change between groups using a statistical test

- 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

t-test on the change (2/2)

3. Compare the change between groups using a statistical test

- 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

The underlying statistical model is

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

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. $Y(52)$ $\rightarrow$ **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).

$\rightarrow$ 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

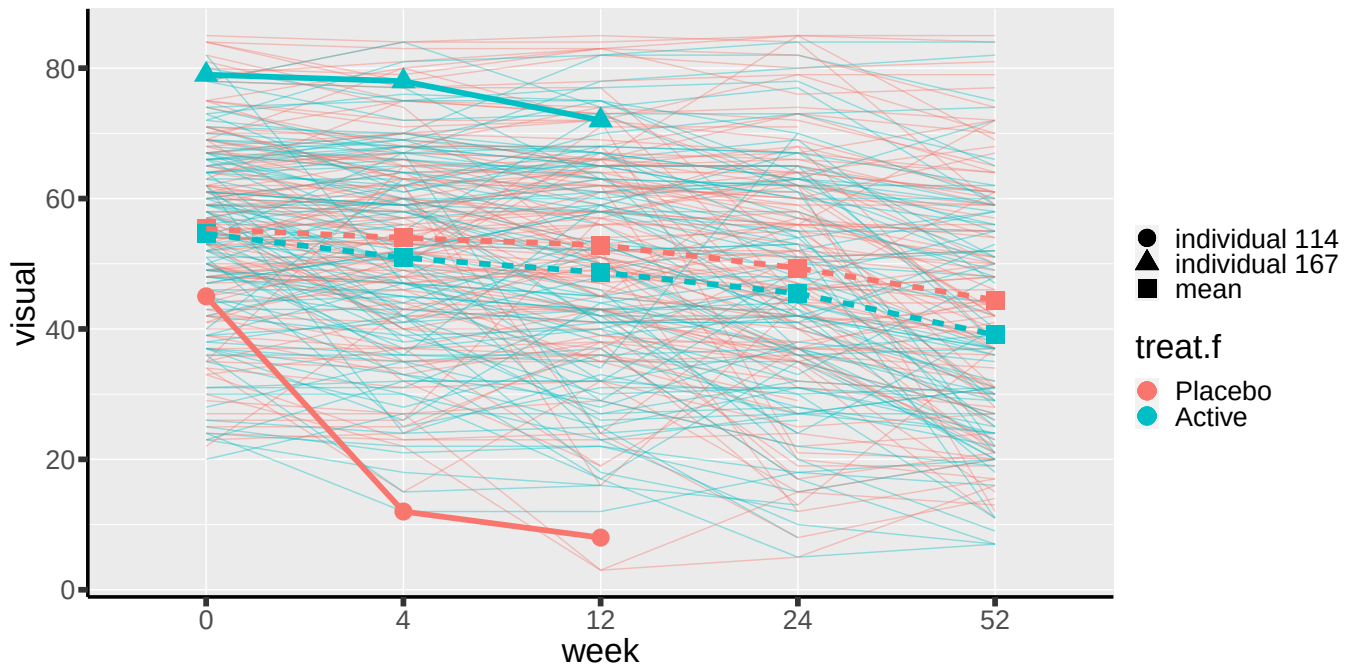
⚠ multiple testing issue

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

Multivariate approach

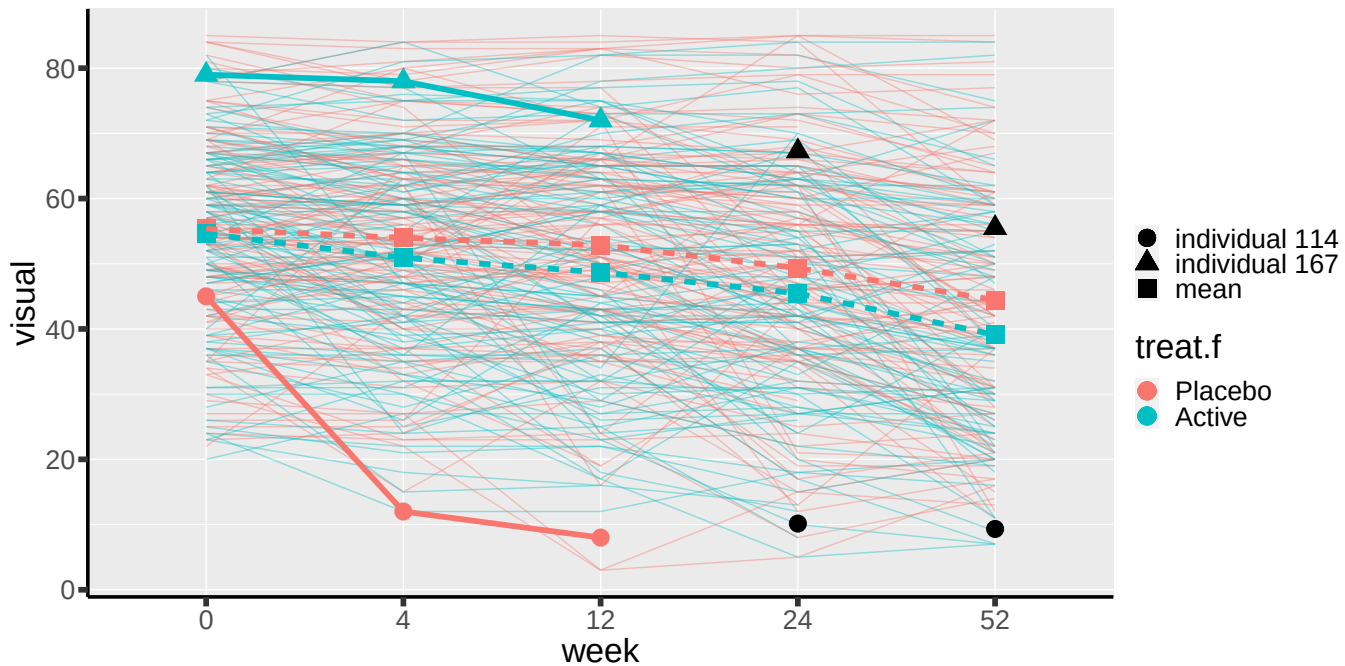
Better handling missing values

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?



Yes! Using the observed outcomes and fitted mean & covariance.

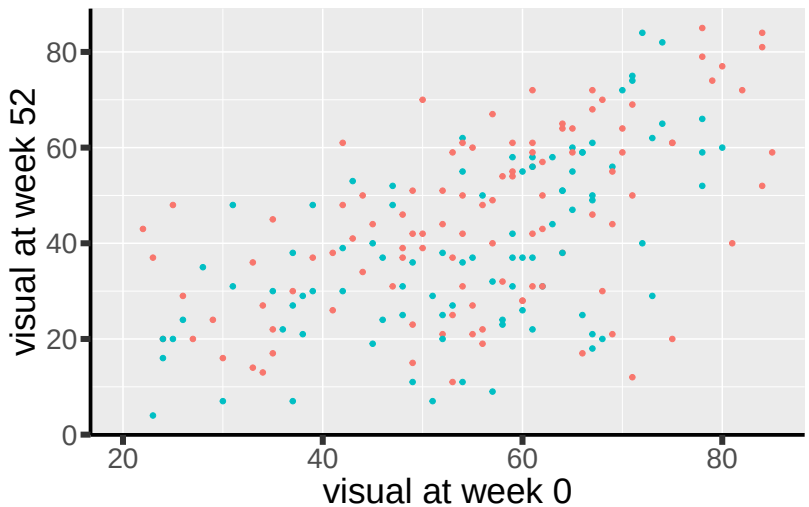
How can we do that? (intuition)

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

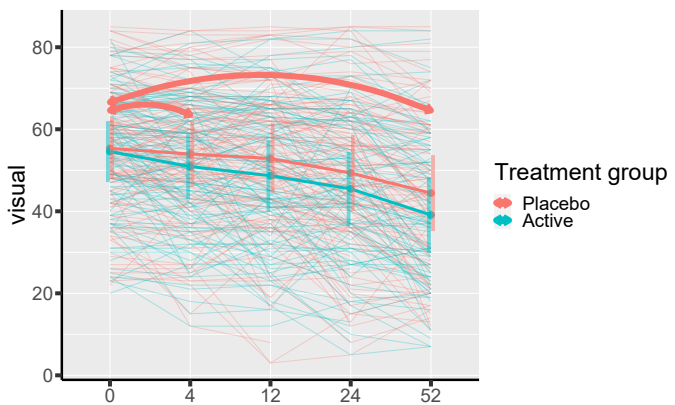


How can we do that? (formula)

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

$$\widehat{Y}_i(52) = \mu(52) + \rho(0, 52)\frac{\sigma(52)}{\sigma(0)}\left(Y_i(0) - \mu(0)\right)$$

- $\mu(t)$, $\sigma(t)$: mean and variance of the outcome at time t
- $\rho(t_1, t_2)$: correlation between the outcome at time t_1 and t_2



How can we do that? (formula)

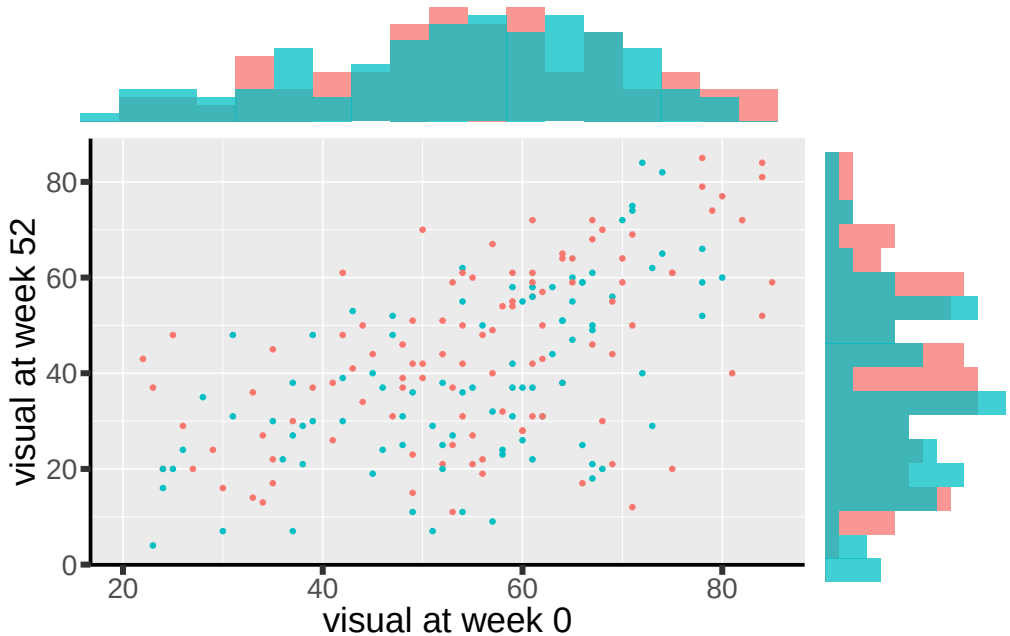
Formally, the expected value at one timepoint given the observed value at another is:

$$\widehat{Y}_i(52) = \mu(52) + \rho(0, 52)\frac{\sigma(52)}{\sigma(0)}\left(Y_i(0) - \mu(0)\right)$$

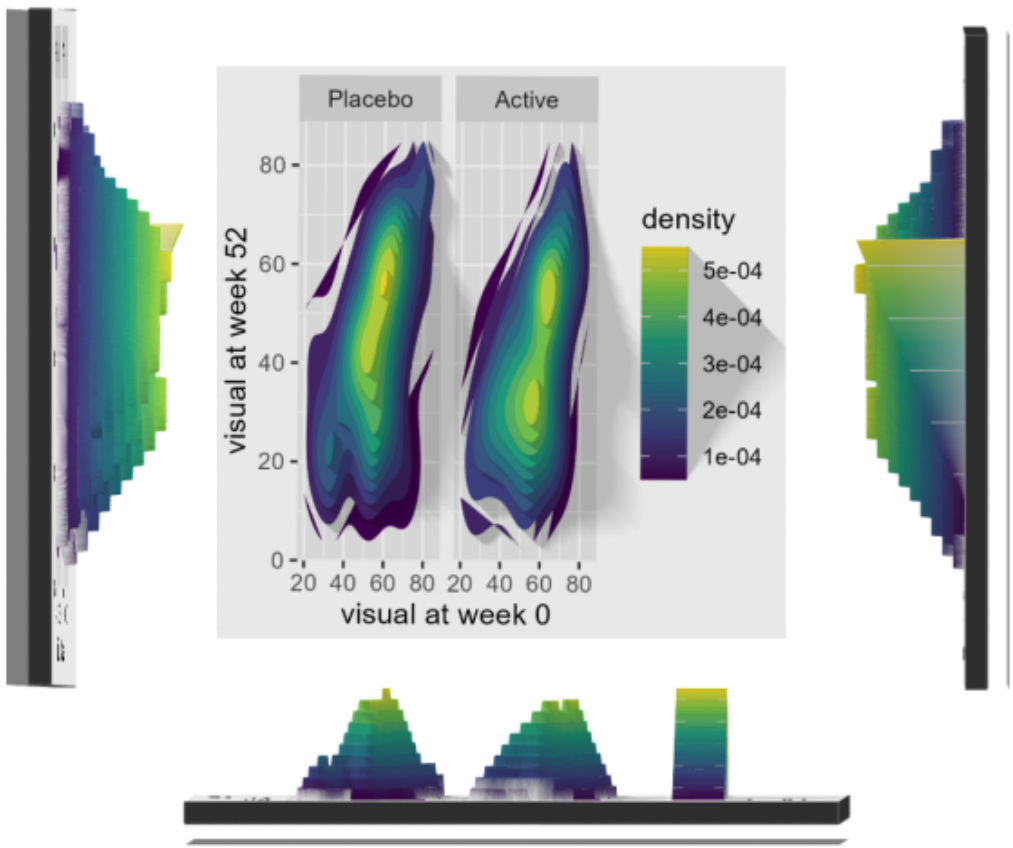
- $\mu(t)$, $\sigma(t)$: mean and variance of the outcome at time t
- $\rho(t_1, t_2)$: correlation between the outcome at time t_1 and t_2

$\rightarrow$ avoids the need for multiple linear regression
(one for each combination of timepoints)
$\rightarrow$ we need not only to model the **mean**
but also the **variance** and **correlation** over time!
$\rightarrow$ we assume a joint normal distribution over time

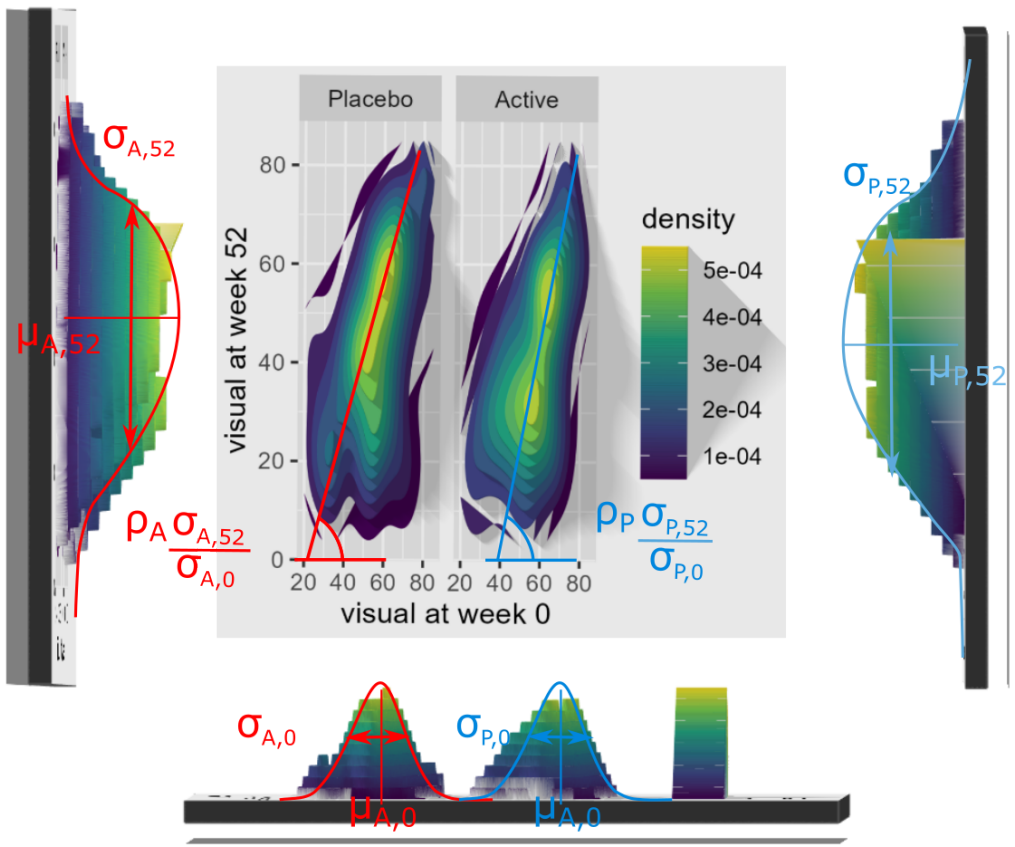
From linear regression to multivariate normal distribution



From linear regression to multivariate normal distribution



From linear regression to multivariate normal distribution



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.

✘ requires a more complex model

With complete data, estimates from an adequately parametrized multivariate 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

```
dfL52 <- dfL[dfL$week %in% c(0,52),]
dfL52$week <- factor(dfL52$week, levels = c(0,52))

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

model.tables(e.lmm)
```

	estimate	se	df	lower	upper	p.value
(Intercept)	55.34	1.4	238	52.6	58.0	0.0e+00
treat.fActive	-0.76	1.9	238	-4.6	3.0	6.9e-01
week52	-11.09	1.6	196	-14.2	-8.0	1.6e-11
treat.fActive:week52	-4.38	2.3	198	-8.9	0.1	5.5e-02

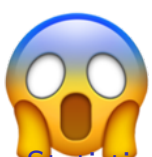
Example in R with 2 timepoints

```
dfL52 <- dfL[dfL$week %in% c(0,52),]
dfL52$week <- factor(dfL52$week, levels = c(0,52))

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

model.tables(e.lmm)
```

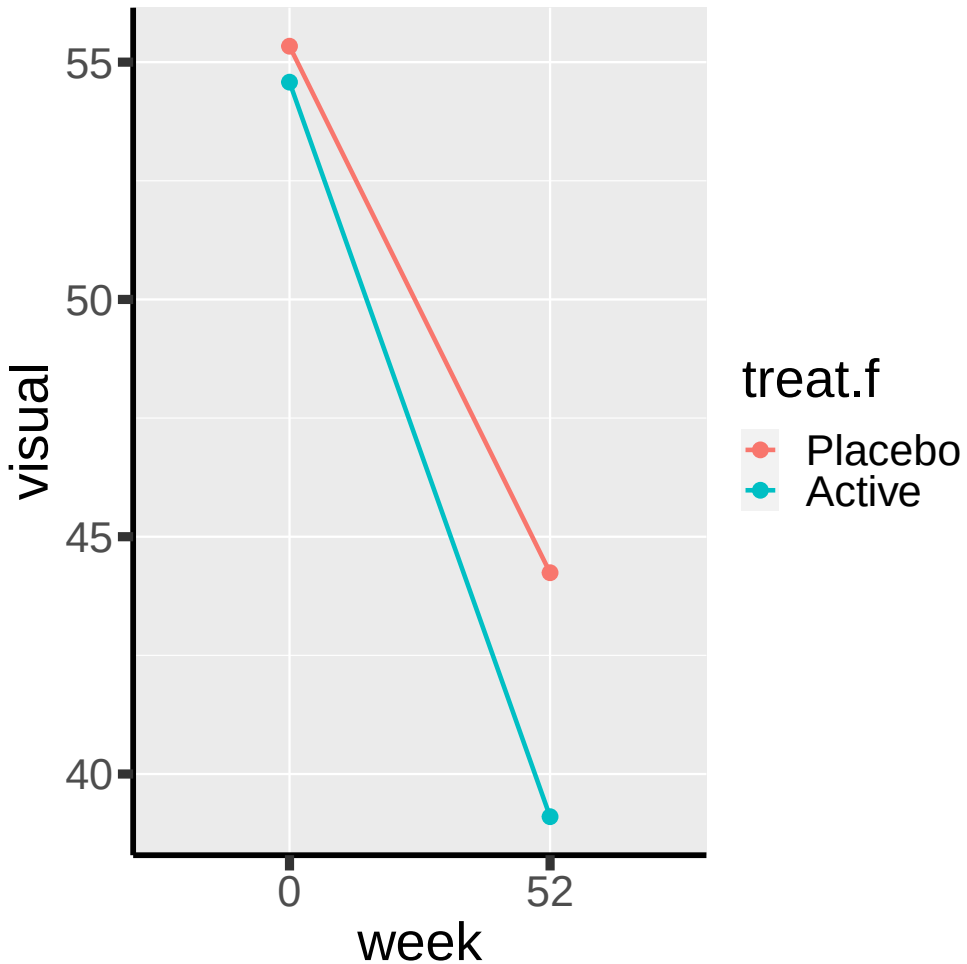
	estimate	se	df	lower	upper	p.value
(Intercept)	55.34	1.4	238	52.6	58.0	0.0e+00
treat.fActive	-0.76	1.9	238	-4.6	3.0	6.9e-01
week52	-11.09	1.6	196	-14.2	-8.0	1.6e-11
treat.fActive:week52	-4.38	2.3	198	-8.9	0.1	5.5e-02



What are those coefficients?

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

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

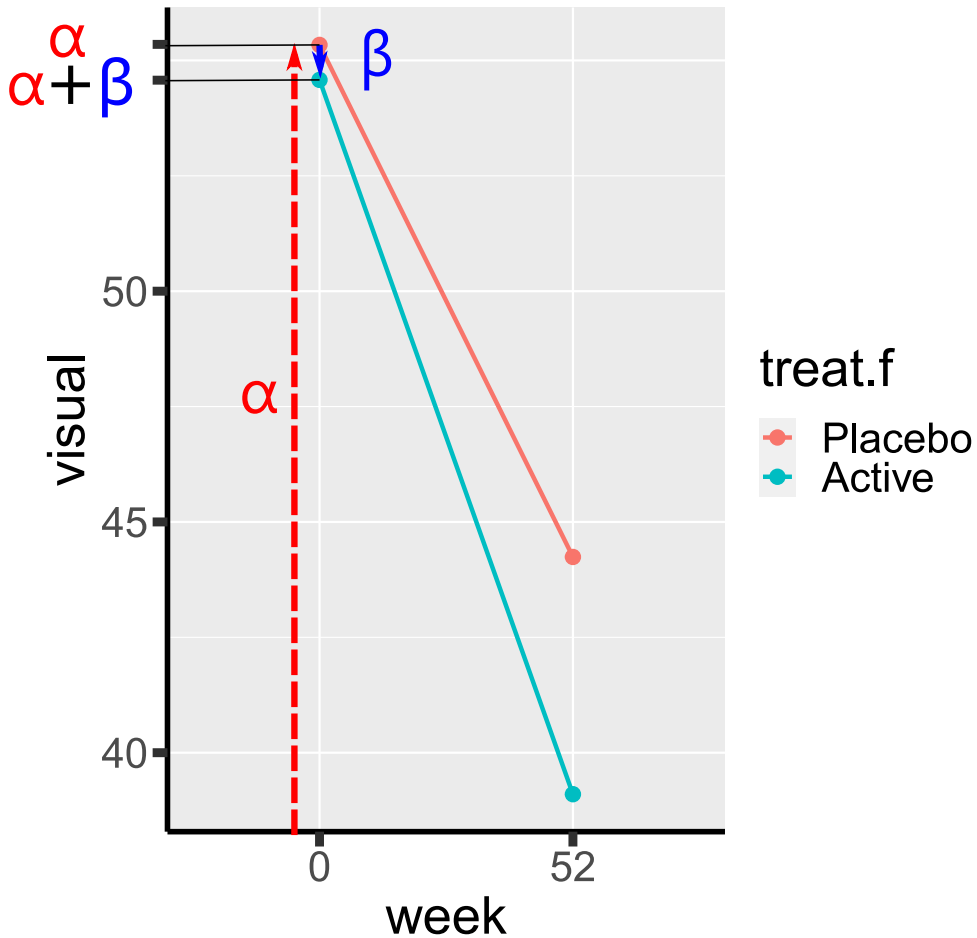



```
coef(e.lmm)
```

```
(Intercept)
55.3361345
treat.fActive
-0.7576221
week.f52
-11.0948777
treat.fActive:week.f52
-4.3831236
```

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

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

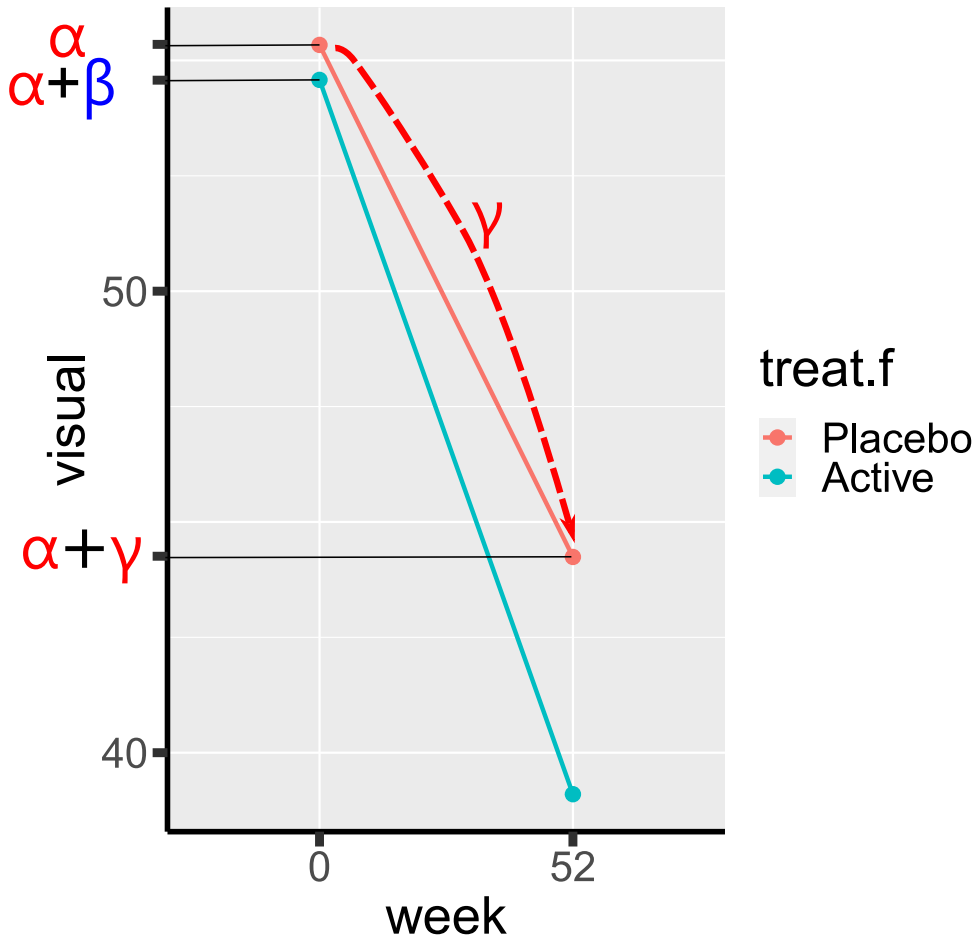



```
coef(e.lmm)
```

```
(Intercept)
55.3361345
treat.fActive
-0.7576221
week.f52
-11.0948777
treat.fActive:week.f52
-4.3831236
```

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

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

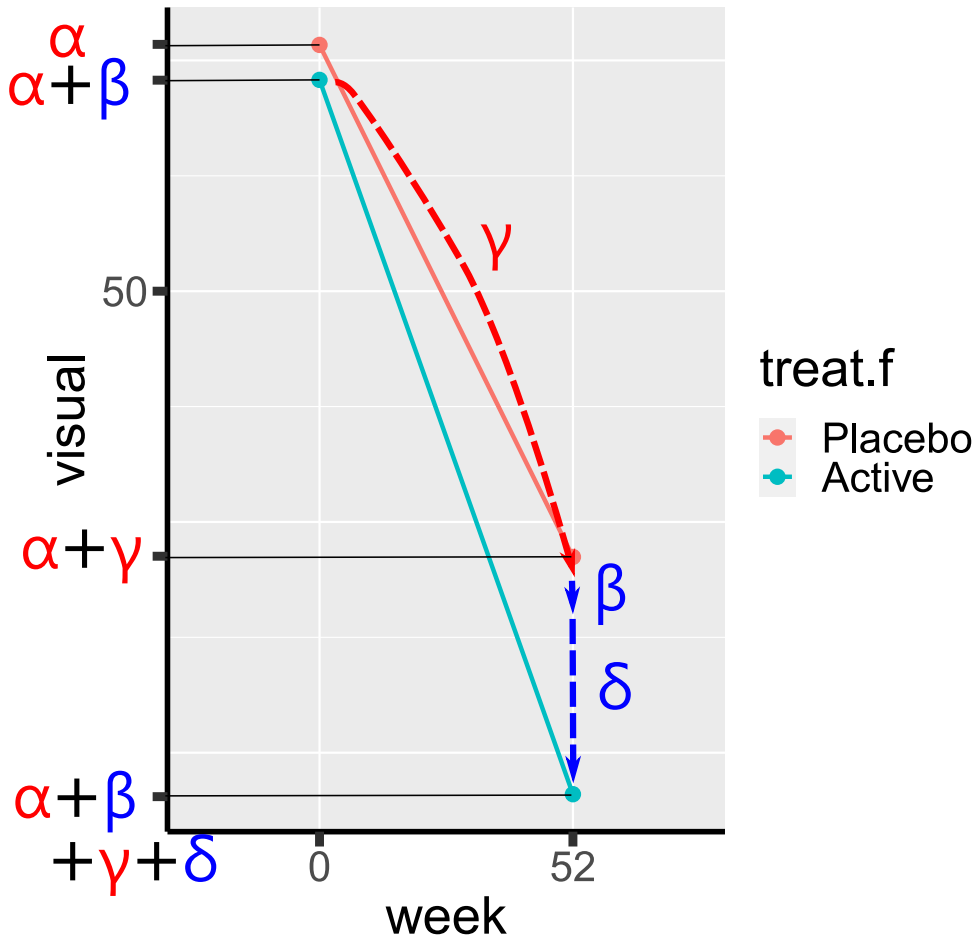



```
coef(e.lmm)
```

```
(Intercept)
55.3361345
treat.fActive
-0.7576221
week.f52
-11.0948777
treat.fActive:week.f52
-4.3831236
```

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

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

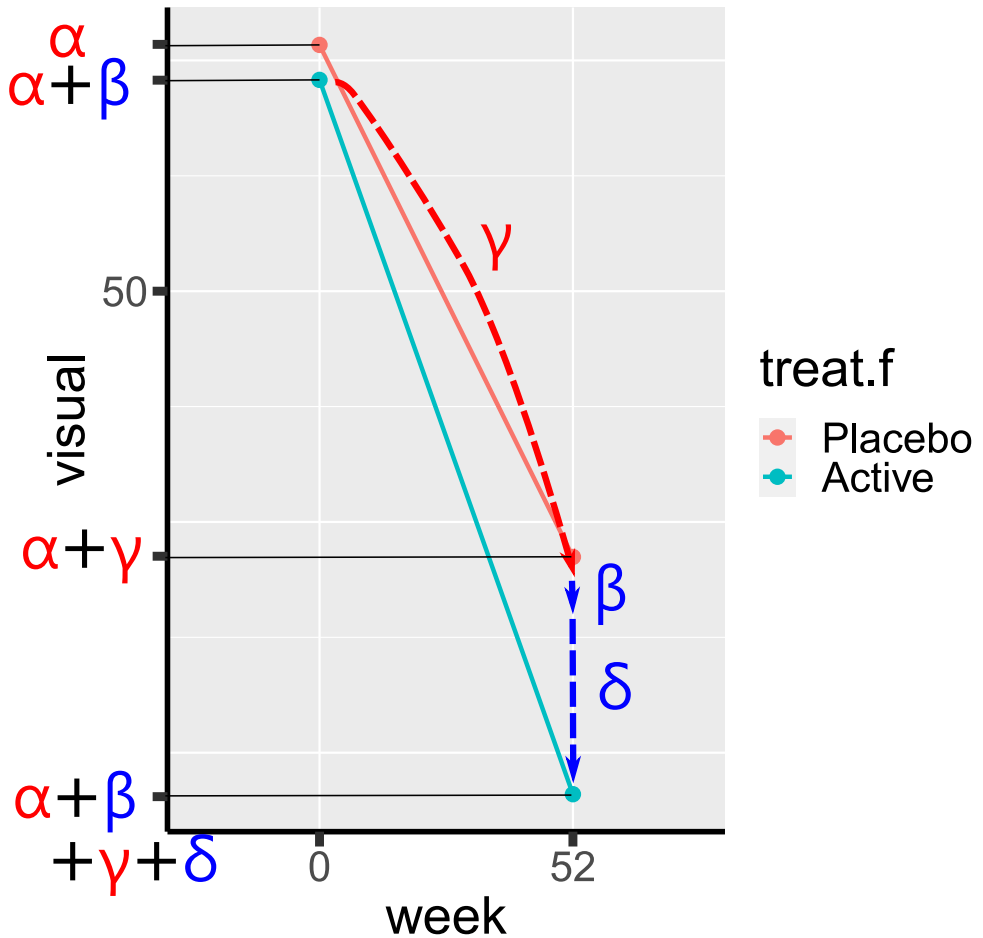



```
coef(e.lmm)
```

```
(Intercept)
55.3361345
treat.fActive
-0.7576221
week.f52
-11.0948777
treat.fActive:week.f52
-4.3831236
```

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

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




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

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

Underlying Gaussian model

Unstructured variance/correlation:

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

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

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

```
     sigma        k.52  rho(0,52)
14.9115118  1.2397277  0.5612167
```

Underlying Gaussian model

Unstructured variance/correlation:

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

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

⚠ we assume no treatment effect on the variance/correlation

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

```
     sigma        k.52  rho(0,52)
14.9115118  1.2397277  0.5612167
```

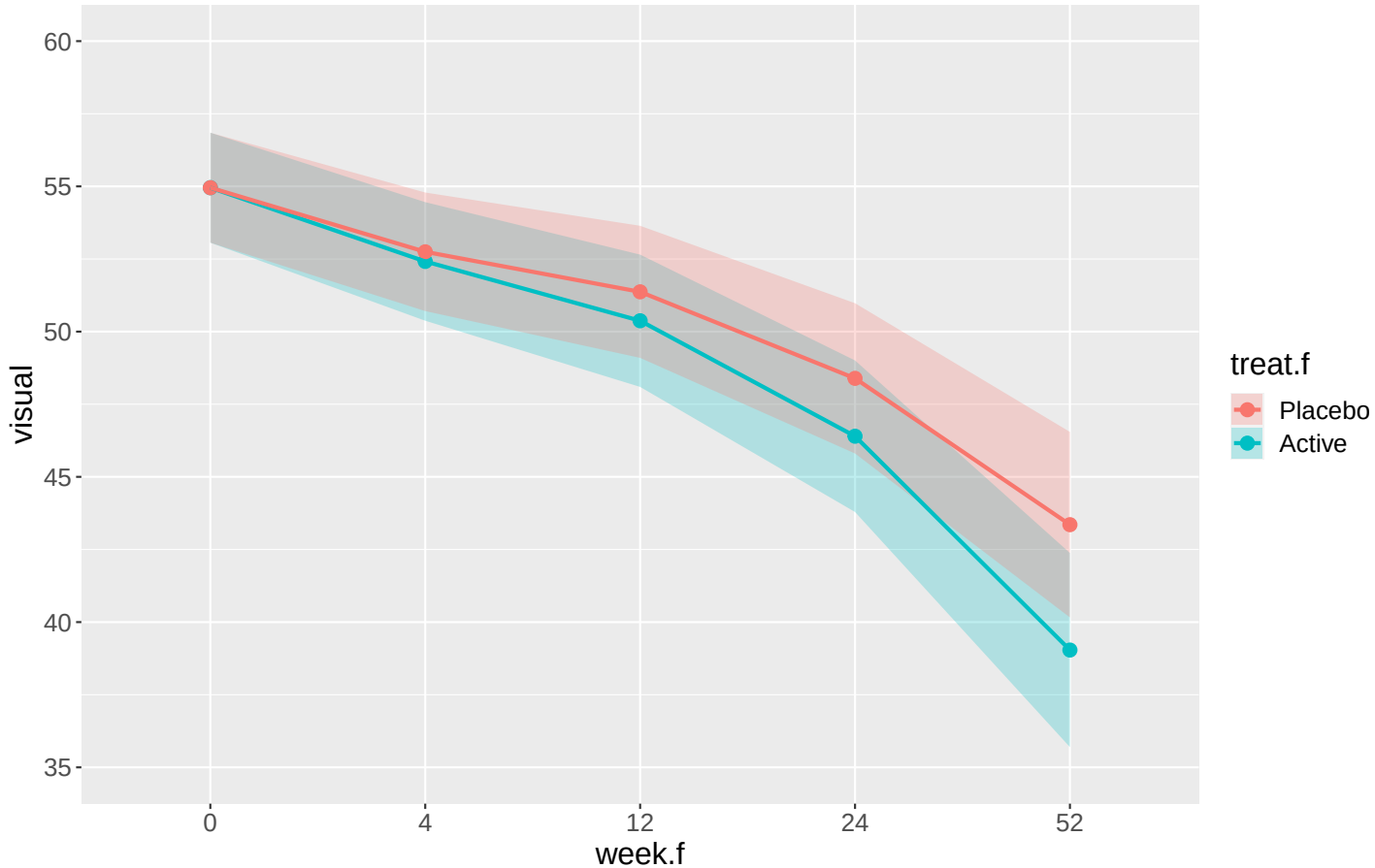
Treatment effect proportional to duration

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

model.tables(eLin.lmm)
```

	estimate	se	df	lower	upper	p.value
week.f0	54.954	0.961	239	53.06	56.8469	0.000
week.f4	52.748	1.036	240	50.71	54.7882	0.000
week.f12	51.369	1.154	257	49.10	53.6426	0.000
week.f24	48.391	1.314	281	45.80	50.9776	0.000
week.f52	43.354	1.621	232	40.16	46.5471	0.000
week.num:treat.fActive	-0.083	0.041	187	-0.16	-0.0023	0.044

Visualisation



Warp-up

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/ANCOVA 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 . . .

Conclusion

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 obsertions 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.

Kamerman, 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.

Reference II

Van Reeth, E., Tham, I. W., Tan, C. H., and Poh, C. L. (2012). Super-resolution in magnetic resonance imaging: a review. *Concepts in Magnetic Resonance Part A*, 40(6):306–325.

Vickers, A. J. and Altman, D. G. (2001). Analysing controlled trials with baseline and follow up measurements. *Bmj*, 323(7321):1123–1124.

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 (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 (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    Active
-11.18095 -15.47778
```