Faculty of Health Sciences

# Correlated data

## More about LMMs

Lene Theil Skovgaard

December 4, 2015





## Further topics

- Model check and diagnostics
- Cross-over studies
- Paired T-tests with missing values
- Missing values
- Possible extensions





## Example: Calcium supplements

A total of 112 11-year old girls were randomized to receive either calcium or placebo.

**Outcome**:
BMD=bone mineral density, in $\frac{g}{cm^2}$,
measured every 6 months (5 visits)

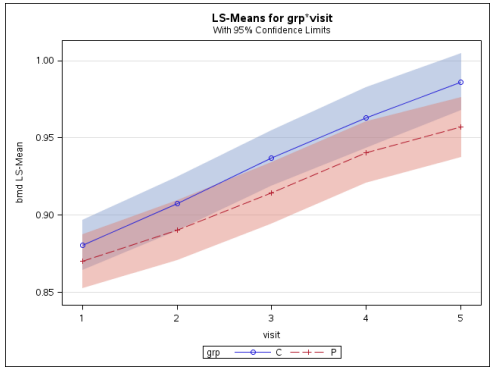
**Scientific question:**
Does calcium improve the rate of bone gain for adolescent women?





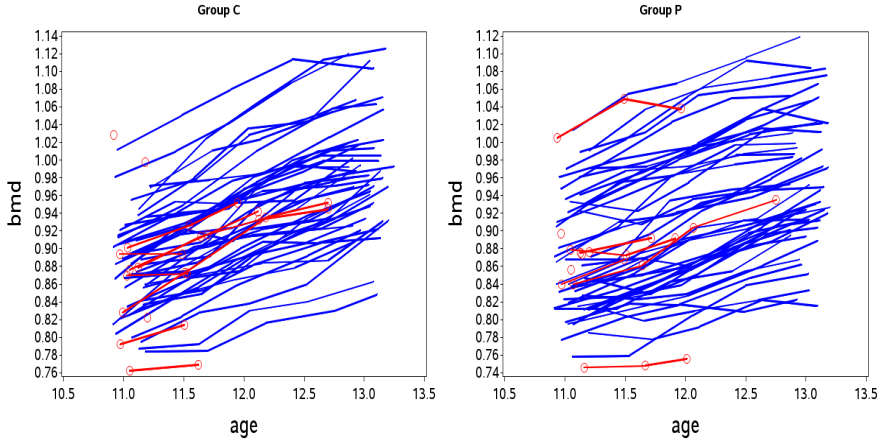
## Average time profiles

with confidence limits
(constructed from `proc glimmix`)




# Individual profiles

## Spaghetti plots






# Plausible models for BMD data

- Response profiles:
  Unstructured mean and unstructured covariance
  (only for balanced data)
- Compound symmetry covariance/correlation
  Synomym for random effect/level for each girl
- Autoregressive covariance/correlation
  or other covariance structures
- Random regression
  Random effects of both intercept and slope for each girl





# How can we choose between models?

- Think....
- Graphical assessment of fit
  e.g. comparison of predicted profiles with average curves
- Inspection of residuals
  Automatic model checks, using `ods graphics`
  More extensice model checks using output data sets
- Tests against more flexible alternatives
  - Fixed effects tested by the *usual* output
  - Covariance patterns evaluated by $\chi^2$-tests on $-2 \log L$



# The mean value structure

Look for:

- Linearity in scatter plot?
- Curves in residual plots?

Alternatives:

- Splines
- More covariates
- Non-linear models

## When the time effect is *difficult* (non-linear)

Penalized spline model:

- Model a flexible time effect, common to the two groups, with several knots in a spline function; e.g. the calcium example with knots at the ages 12 and 12.5 (just for illustration)
- Let the changes in the knots be "*not too big*"....
- Model the difference between the groups as a linear function in time



## ∗Linear splines

- Subdivide the age, using several thresholds, e.g. here: $a_1 = 12, a_2 = 12.5$
- Fit a linear effect in age age group
- Bind the lines together in the threshold values

The result will be a bended (broken) line, but it will still be a linear model). With x=age, we have:

$$\text{Mean} = \alpha + \beta_0(x-11) + \beta_1 I(x > 12)(x-12) + \beta_2 I(x > 12.5)(x-12.5),$$

Splines can also be quadratic, cubic etc.





## ∗Interpretation of parameters

- $\alpha$: "Intercept", expected outcome when age11=0, i.e. at age 11
- $\beta_0$: slope (effect of 1 year increase in age), up to age 12
- $\beta_1$: the snap in the "line" at age 12
  - $\beta_1 = 0$: the "line" continues through 12 without a snap
  - $\beta_1 > 0$: the "line" bends, and becomes steeper after age 12
  - $\beta_1 < 0$: the "line" bends, and becomes less steep after age 12

  The slope in the age interval $(12, 12.5)$ is $\beta_0 + \beta_1$
- $\beta_2$: the snap in the "line" at age 12.5
  Same interpretation as $\beta_1$, only at another threshold
  The slope after age 12.5 is $\beta_0 + \beta_1 + \beta_2$
-

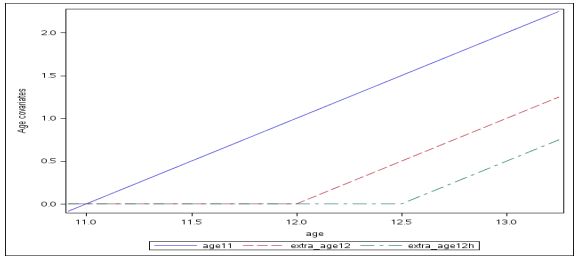


## ∗Coding of penalized spline

Data definitions:

```
age11=age-11;
age12=max(0,age-12);
age12h=max(0,age-12.5);
run;
```

The 3 covariates to describe the age effect:




UNIVERSITY OF COPENHAGEN DEPARTMENT OF BIOSTATISTICS

## ∗Fitting of penalized spline

*Penalized*, because we could (in principle) have many knots and prohibit any single change in slope to be too big.

This is done in the repeated-statement below:

```
title 'penalized spline';
proc mixed data=calcium noclprint;
class grp girl visit;
model bmd=grp age11 extra_age12 extra_age12h grp*age11 /
    outpm=fitpm outp=fitp ddfm=kr s;
random extra_age12 extra_age12h / type=toep(1);
random intercept age11 / subject=girl type=un;
run;
```



## ∗Penalized spline, output

Solution for Fixed Effects

| Effect | grp | Estimate | Standard Error | DF | t Value | Pr > \|t\| |
|---|---|---|---|---|---|---|
| Intercept | | 0.8661 | 0.008754 | 111 | 98.93 | <.0001 |
| grp | C | 0.01113 | 0.01240 | 110 | 0.90 | 0.3715 |
| grp | P | 0 | . | . | . | . |
| age11 | | 0.04483 | 0.006203 | 3.37 | 7.23 | 0.0036 |
| age12 | | 0.000791 | 0.01212 | 1 | 0.07 | 0.9585 |
| age12h | | -0.01440 | 0.009700 | 1 | -1.48 | 0.3773 |
| age11*grp | C | 0.008900 | 0.003087 | 96.2 | 2.88 | 0.0049 |
| age11*grp | P | 0 | . | . | . | . |

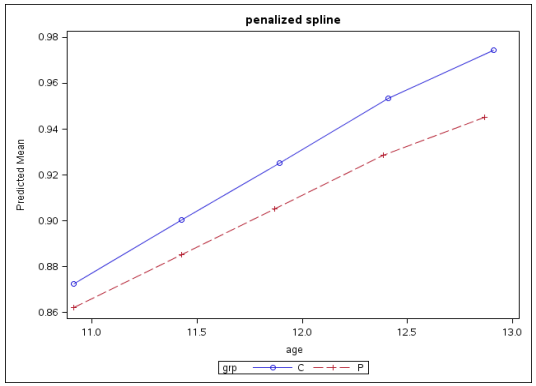
The Calcium group improves an estimated 0.0088 $\frac{g}{cm^2}$ per year more than the Placebo group.



UNIVERSITY OF COPENHAGEN DEPARTMENT OF BIOSTATISTICS

## ∗Penalized spline, fit

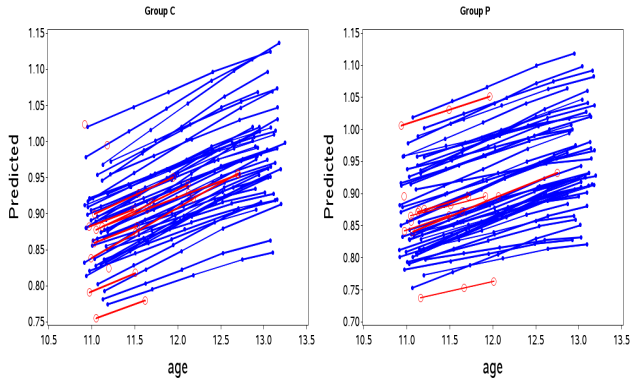
```
proc sgplot data=udpm; where girl in ('101','102');
    series X=age Y=Pred / group=grp;
run;
```





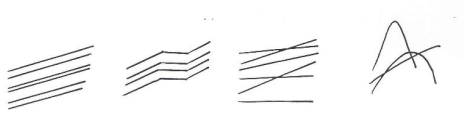

## ∗Penalized spline, individual fit

```
proc sgplot data=udp; by grp;
    series X=age Y=Pred / group=girl markers;
run;
```



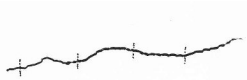

## The covariance/correlation structure

1. Random effects:



2. Serial correlation (*the pattern*)

3. Error of measurement





## The mixed effects model

Individual $i$, "*time*" $j$ (or $t$):
Observations: $Y_{ij}$, fixed effects covariates $X_{ij}$
Additional covariate vector $Z_{ij}$, specifying the random effects.

Mean: $E(Y_{ij}|b_i) = \mu_{ij} = X_{ij}^T\beta + Z_{ij}^T b_i$

Covariance structure:

- $b_i \sim N_p(0, G)$,
  where $G$ is the matrix notation for $\omega_B^2$
- Covariance of residuals:

$$\varepsilon_{i\cdot} \sim N(0, R_i)$$





## ∗Variogram

Variance of difference between time points, $u$ time units apart:

$$\gamma(u) = \frac{1}{2}E(\varepsilon_t - \varepsilon_{t-u})^2$$

If the model has

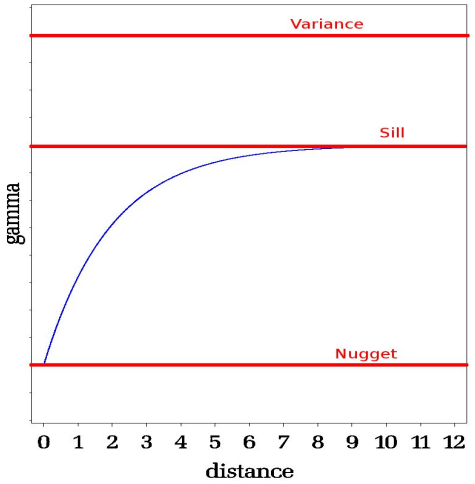
- random level/intercept, $\omega^2$
- serial AR(1)-correlation over time, $\tau^2 \times \rho^u$
- measurement error, $\sigma^2$

we have

$$\gamma(u) = \tau^2(1 - \rho(u)) + \sigma^2$$



## ∗Theoretical variogram




- Nugget: $\sigma^2$
- Sill: $\tau^2 + \sigma^2$
- Variance: $\omega^2 + \tau^2 + \sigma^2$

## Assumptions in a mixed effects model

- Linearity in covariates $X_{ij}$ (including $Z_{ij}$)
- Normality of residuals $\varepsilon_{i\cdot}$
- Normality of random effects $b_i$
- Plausibility of covariance structure
- Independence between individuals
- Independence between $X_{ij}$ and $b_i$
  Example "Reading" in the end of the lecture



## Importance of assumptions

Important:

- Linearity
- Independence between individuals (normally not an issue)
- Independence between $X_{ij}$ and $b_i$
- Appropriateness of the covariance structure:
  (may be circumvented by using the
  empirical sandwich estimator,
  option `empirical` in `proc mixed`)

Less important:
(especially when the number of observations is large)

- Normality of residuals $\varepsilon_{ij}$
- Normality of random effects $b_i$





## Back to "simple" random regression

Random regression, using actual age (age11=age−11):

```
proc mixed covtest data=calcium;
class grp girl;
model bmd=grp age11 grp*age11 / ddfm=kr s cl;
random intercept age11 /
       type=un subject=girl g v vcorr s;
run;
```

Estimated G Matrix

| Row | Effect | grp | girl | Col1 | Col2 |
|---|---|---|---|---|---|
| 1 | Intercept | C | 101 | 0.004215 | 0.000095 |
| 2 | age11 | C | 101 | 0.000095 | 0.000180 |

Estimated V Correlation Matrix for girl(grp) 101 C

| Row | Col1 | Col2 | Col3 | Col4 | Col5 |
|---|---|---|---|---|---|
| 1 | 1.0000 | 0.9664 | 0.9537 | 0.9321 | 0.9056 |
| 2 | 0.9664 | 1.0000 | 0.9687 | 0.9566 | 0.9385 |
| 3 | 0.9537 | 0.9687 | 1.0000 | 0.9697 | 0.9590 |
| 4 | 0.9321 | 0.9566 | 0.9697 | 1.0000 | 0.9723 |
| 5 | 0.9056 | 0.9385 | 0.9590 | 0.9723 | 1.0000 |



## Output, continued

Solution for Fixed Effects

| Effect | grp | Estimate | Standard Error | DF | t Value | Pr > \|t\| | Alpha |
|---|---|---|---|---|---|---|---|
| Intercept | | 0.8667 | 0.008689 | 110 | 99.75 | <.0001 | 0.05 |
| grp | C | 0.01113 | 0.01240 | 110 | 0.90 | 0.3715 | 0.05 |
| grp | P | 0 | . | . | . | . | . |
| age11 | | 0.04529 | 0.002155 | 96 | 21.02 | <.0001 | 0.05 |
| age11*grp | C | 0.008891 | 0.003081 | 96.6 | 2.89 | 0.0048 | 0.05 |
| age11*grp | P | 0 | . | . | . | . | . |

Solution for Fixed Effects

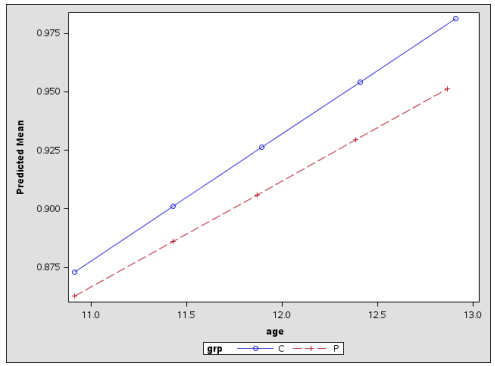
| Effect | grp | Lower | Upper |
|---|---|---|---|
| Intercept | | 0.8495 | 0.8839 |
| grp | C | -0.01345 | 0.03570 |
| grp | P | . | . |
| age11 | | 0.04102 | 0.04957 |
| age11*grp | C | 0.002776 | 0.01501 |
| age11*grp | P | . | . |

In this model, we quantify the effect of a calcium supplement to 0.0089 (0.0031) g per cm$^3$ per **year**.

UNIVERSITY OF COPENHAGEN DEPARTMENT OF BIOSTATISTICS

# Predicted values from random regression

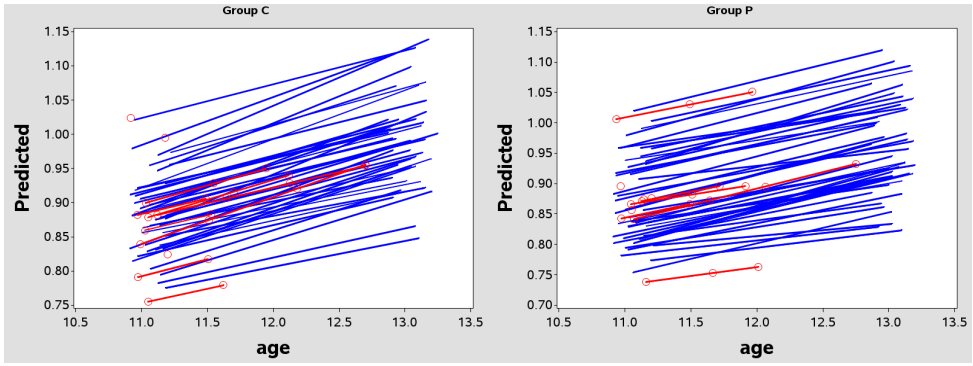
Predicted group means:






# Predicted values from random regression, II

Individual predictions:






UNIVERSITY OF COPENHAGEN DEPARTMENT OF BIOSTATISTICS

# Model checks

Two types of residuals:

**Ordinary** Observed minus predicted group mean (only systematic effects)

$$Y_{ij} - X_{ij}^T \hat{\beta}$$

**Conditional** Observed minus predicted individual mean value (systematic and random effects)

$$\varepsilon_{ij} = Y_{ij} - (X_{ij}^T \hat{\beta} + Z_{ij}^T \hat{b}_i)$$

Conditional residuals are usually *much smaller* than the ordinary, since they describe deviations from subject-specific predictions.



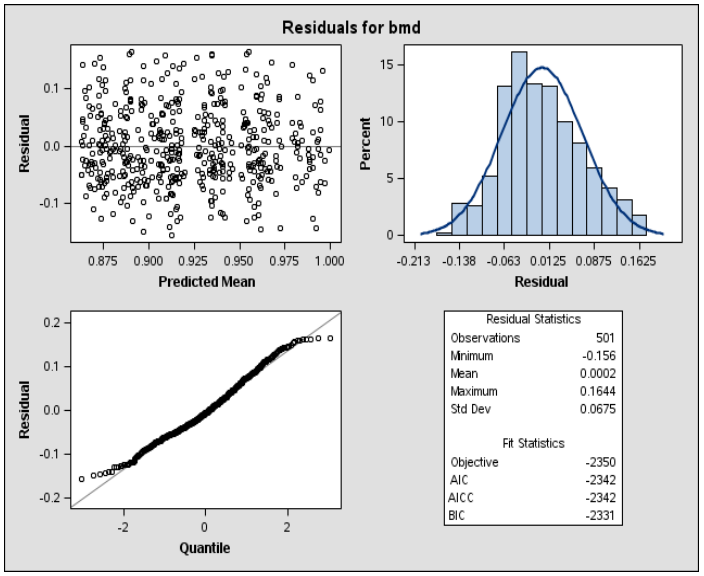
# Coding for model checks

```
proc mixed noclprint data=calcium;
class grp girl;
model bmd=grp age11 grp*age11 / s cl ddfm=kr
          outpm=fitpm outp=fitp residual influence;
random intercept age11 /
       type=un subject=girl g v vcorr;
run;
```

gives us

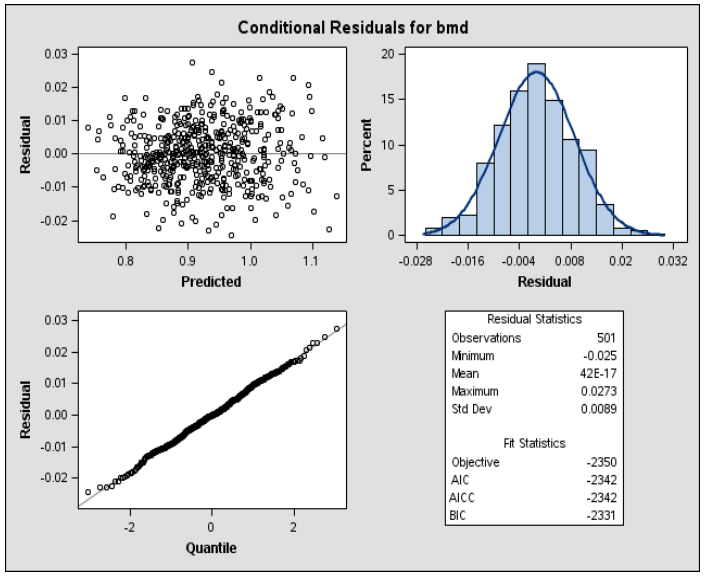
- Panels to check stability of variance and normality of residuals
- creates two output data sets:
  - fitpm: Predicted mean BMD-values, common to girls in the same group
  - fitp: Individually predicted BMD-values, specific for each girl

# Model check, ordinary residuals






# Model check, conditional residuals







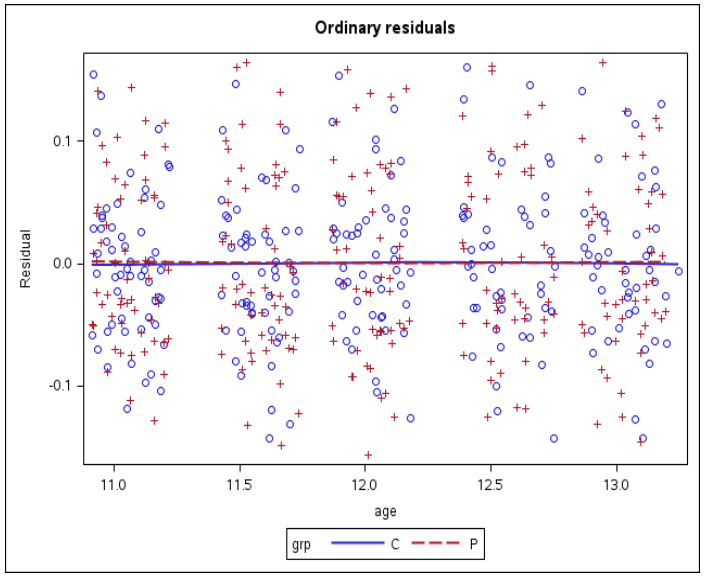

# Additional model checks

Investigating linearity in age:

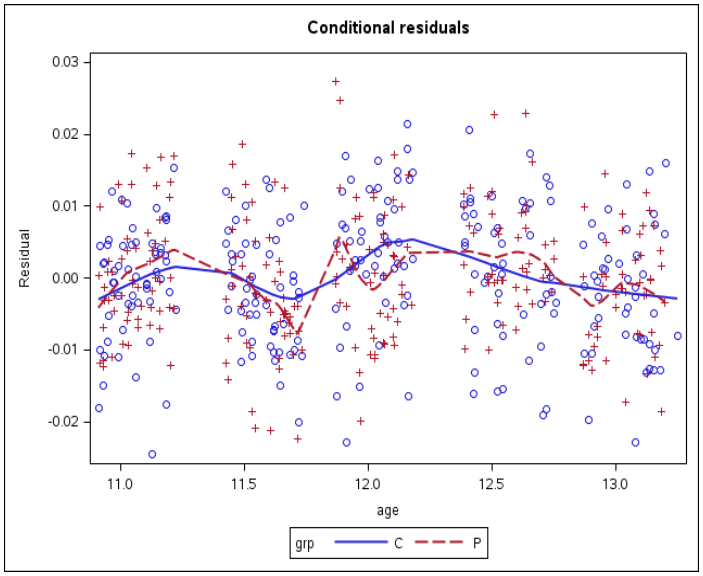
```
proc sort data=fitpm;
  by grp age;
run;

title 'Ordinary residuals';
proc sgplot data=fitpm;
loess Y=Resid X=age / group=grp;
run;

proc sort data=fitp;
  by grp age;
run;

title 'Conditional residuals';
proc sgplot data=fitp;
loess Y=Resid X=age / group=grp;
run;
```



# Check of linearity, ordinary residuals (fitpm)



## Check of linearity, conditional residuals(fitp)






## Comments on model checks

- Ordinary residuals: Slightly skew distribution, almost Normal
- Conditional residuals: Evidently Normal
- Some deviation from linearity seen in conditional residuals, consistently for the two groups
- Deviation from linearity cannot be seen in the ordinary residuals (they "drown" in the between-subject effect)

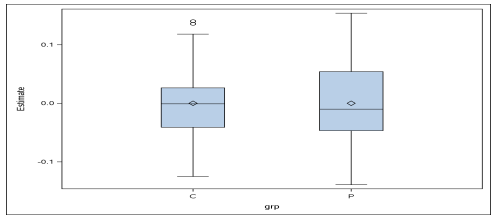




## Normality of random effects?

Histogram or Box plots of estimated $\hat{b}_i$'s from the model is not worth much

```
ods output solutionr=random_effects;
run;

proc sgplot data=random_effects;
  vbox Estimate / category=grp;
run;
```






## Influential observations

i.e. observations with a large influence on the estimates, either on the mean value or on the covariance parameters.

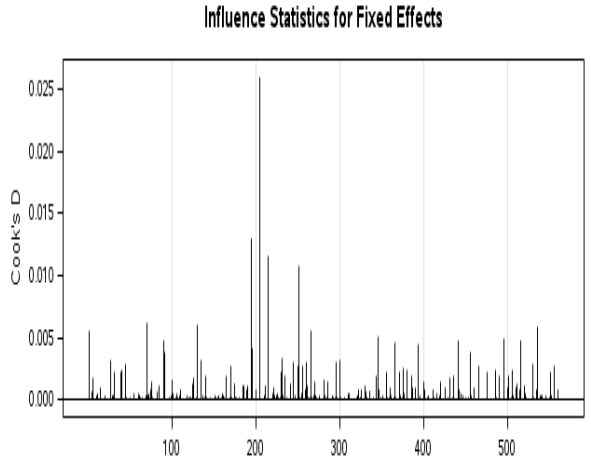
These observations could have

- an unusual combination of covariates $X_i$
- large ordinary residuals (from $X_i\beta$)
- an unusual combination of covariates $Z_i$

or it could be a sign of a

- bad choice of mean value structure
- bad choice of covariance pattern

## Cooks distance




Tentative limit for "influential": $\frac{4}{n} = 0.035$



## Example: Cross-over study of headache

Patients with chronic headache are randomized into two groups:

- Both groups receive LMMA and placebo, on two different days, with a suitable wash-out period in-between
- Group G1 was treated first with placebo (period 1), and then with LNMMA (period 2)
- Group G2 was treated first with LNMMA (period 1), and then with placebo (period 2)

Pain was measured subjectively on a VAS-scale (small is good), at baseline and at 30, 60, 90 and 120 minutes after treatment.

Ashina, Lassen, Bendtsen, Jensen og Olesen (1999), Lancet, pp.287-289





## Picture ignoring period effect and pairing

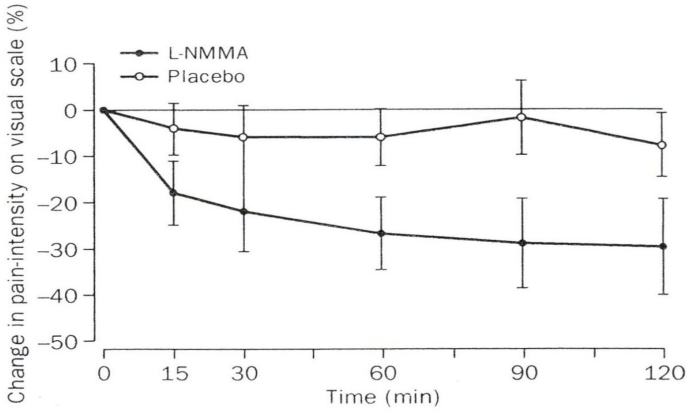



Figure 2: **Mean percentage change from baseline in pain intensity on 100 mm visual analogue scale**
Bars=SE.



## Model building for cross over study

Fixed effect:

- `time`, maybe linear on a logarithmic scale
- `treat`: `lnmma` or `placebo`
- interaction `treat*time`
- period: 1 or 2
- carry-over effect: `treat*period`(`*time`?)

## Model building for cross over study, II

Covariance structure:

- All observations from the same patient should be correlated
  `random patient;`
- Observations from the same period (and same patient) are more closely correlated, depending upon the time difference, e.g.

```
repeated time / subject=patient*period
                type=sp(pow)(time) local;
```

- An additional measurement error may be added, using option `local` in the statement above



## But but but

(Un)fortunately, we do not have access to the repeated measurements over time

Here, we shall reduce the complexity by simply looking at difference between baseline and follow up
(remember that for this to be a good idea,
the correlation must be strong)

Outcome:
Difference between follow-up measurements and baseline, i.e.
$Y_{30} + Y_{60} + Y_{120} - 3\,Y_0$

Large values are good





## Overview of data

| Obs | group | treat | period | effect |
|---|---|---|---|---|
| 1 | G1 | placebo | 1 | -102 |
| 2 | G1 | lnmma | 2 | -167 |
| 3 | G1 | placebo | 1 | -39 |
| 4 | G1 | lnmma | 2 | -127 |
| ..... | | | | many lines here |
| 29 | G2 | placebo | 2 | 3 |
| 30 | G2 | lnmma | 1 | -74 |
| 31 | G2 | placebo | 2 | -36 |
| 32 | G2 | lnmma | 1 | -72 |

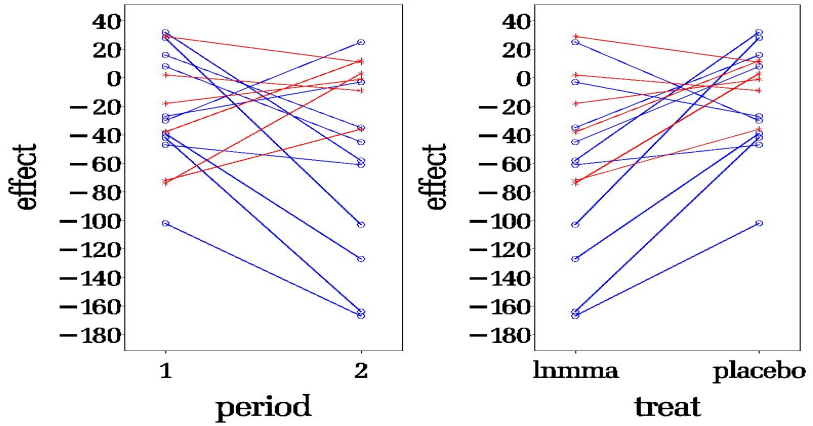
Analysis Variable : effect

| treat | period | N Obs | N | Mean | Std Dev |
|---|---|---|---|---|---|
| lnmma | 1 | 6 | 6 | -28.5000000 | 40.9865832 |
| | 2 | 10 | 10 | -73.8000000 | 65.0022222 |
| placebo | 1 | 10 | 10 | -20.3000000 | 41.5452899 |
| | 2 | 6 | 6 | -3.3333333 | 17.8063659 |

Note the large effect for group G1 in period 2 (active treatment)

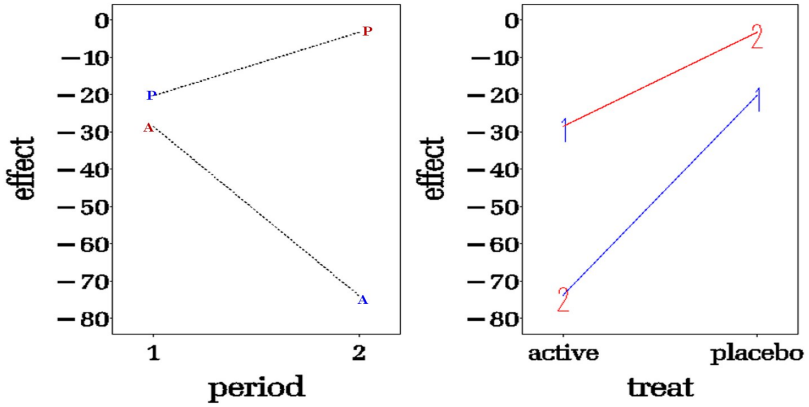


## Observations, vs. period and treatment




Legend: Group G1 (P+A), Group G2 (A+P)
Correlation looks reasonably strong

## Average over patients




A, P denote the treatments, A is best
1 and 2 denote the periods, perhaps, 2 is best



## ∗Ignoring periods: Paired T-test for treatment effect

```
The TTEST Procedure
 Difference:  lnmma - placebo

  N        Mean      Std Dev      Std Err      Minimum      Maximum
 16    -42.8750      53.1462      13.2865       -131.0      55.0000

    Mean        95% CL Mean         Std Dev       95% CL Std Dev
-42.8750    -71.1946 -14.5554      53.1462      39.2593  82.2539

    DF     t Value     Pr > |t|
    15       -3.23       0.0056
```

Remember: Small values are good

LNMMA is more effective than placebo: 42.9 (14.6, 71.2)
but we have ignored a possible effect of period.....





## ∗What about simple ANOVA?

Two-way anova in treat and period

| Parameter | | Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|
| Intercept | | -22.79166667 B | 16.72738044 | -1.36 | 0.1835 |
| treat | lnmma | -39.33333333 B | 18.04415614 | -2.18 | 0.0375 |
| treat | placebo | 0.00000000 B | . | . | . |
| period | 1 | 14.16666667 B | 18.04415614 | 0.79 | 0.4388 |
| period | 2 | 0.00000000 B | . | . | . |

| Parameter | | 95% Confidence Limits | |
|---|---|---|---|
| Intercept | | -57.00300097 | 11.41966763 |
| treat | lnmma | -76.23777634 | -2.42889032 |
| treat | placebo | . | . |
| period | 1 | -22.73777634 | 51.07110968 |
| period | 2 | . | . |

LNMMA is more effective than placebo:
Almost identical estimate: 39.33, but wide confidence interval (-76.24, -2.43), because we have ignored the pairing....



## Traditional approach to analyses

Test equality of treatments assuming no carry-over effect:
Unpaired T-test (compare groups G1 and G2) for period differences (Period2-Period1), since these are

- G1: Active - Placebo
- G2: Placebo - Active

Test of no carry-over effect : usually taken to mean an extra effect of placebo when given after the active treatment:
Unpaired T-test (compare groups G1 and G2) for the sum of the two periods, since these are

- G1: Placebo + Active
- G2: Active + Placebo + Carry-over

## Mixed model for cross-over study

$t =$ active, placebo, $\quad p = 1, 2$ (periods), $\quad i = 1, 2, \ldots$ (individuals)

With no carry-over effect:

$$Y_{tpi} = \alpha_t + \beta_p + b_i + \varepsilon_{tpi}$$

where

$$b_i \sim N(0, \omega_B^2)$$
$$\varepsilon_{tpi} \sim N(0, \sigma_W^2)$$

With carry-over effect:

$$Y_{tpi} = \alpha_t + \beta_p + \gamma_{tp} + b_i + \varepsilon_{tpi}$$



## Coded as a mixed effects model

Without carry-over effect:

```
proc mixed data=ashina;
  class patient group treat period;
  model effect=treat period /
        s cl ddfm=kr;
  random intercept / subject=patient;
run;
```





## Mixed, no carry-over effect

Covariance Parameter Estimates

| Cov Parm | Subject | Estimate |
|---|---|---|
| Intercept | patient(group) | 1001.77 |
| Residual | | 1405.62 |

Solution for Fixed Effects

| Effect | treat | period | Estimate | Standard Error | DF | t Value | Pr > \|t\| |
|---|---|---|---|---|---|---|---|
| Intercept | | | -22.7917 | 14.9556 | 15 | -1.52 | 0.1483 |
| treat | lmmma | | -39.3333 | 13.6900 | 14 | -2.87 | 0.0123 |
| treat | placebo | | 0 | . | . | . | . |
| period | | 1 | 14.1667 | 13.6900 | 14 | 1.03 | 0.3183 |
| period | | 2 | 0 | . | . | . | . |

| Effect | treat | period | Alpha | Lower | Upper |
|---|---|---|---|---|---|
| Intercept | | | 0.05 | -54.6689 | 9.0855 |
| treat | lnmma | | 0.05 | -68.6954 | -9.9712 |
| treat | placebo | | . | . | . |
| period | | 1 | 0.05 | -15.1954 | 43.5288 |
| period | | 2 | . | . | . |

Intra-individual correlation: $\frac{1001.77}{1001.77+1405.62} = 0.42$



## Conclusion on treatment effect

| Method | Effect | Confidence Interval | P-value |
|---|---|---|---|
| Period 1 | 8.20 | (-37.59, 53.99) | 0.71 |
| Period 2 | 70.47 | (11.55, 129.40) | 0.022 |
| Paired T-test | 42.88 | (14.55, 71.19) | 0.0056 |
| 2-way Anova | 39.33 | (2.43, 76.24) | 0.0375 |
| T-test, period diff. | 39.33 | (9.97, 68.70) | 0.012 |
| Mixed model | 39.33 | (9.97, 68.70) | 0.012 |

Colour codes: wrong, perhaps feasible, correct

## Not much effect of period adjustment

Period 1 is an estimated 14.1667 above period 2,
CI: (-15.1954, 43.5288), $P = 0.32$

but the pictures suggest
that there may be an interaction between treatment and period.

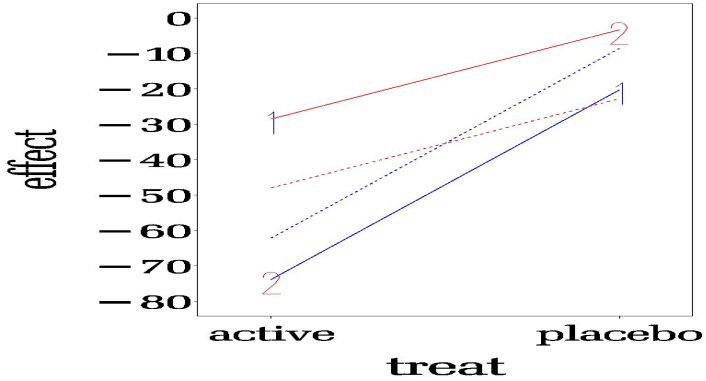
Maybe we have a carry-over effect

Include a carry-over effect as the interaction `treat*period`



## Predictions in contrast to observations

assuming no carry-over effect:




Coloured lines denote the groups. Numbers denote the period
Some deviations from the model is obvious





## Output, with carry-over effect

Traditional parametrization, with interaction term `treat*period`:

Solution for Fixed Effects

| Effect | treat | period | Estimate | Standard Error | DF | t Value | Pr > \|t\| |
|---|---|---|---|---|---|---|---|
| Intercept | | | -3.3333 | 19.4487 | 14 | -0.17 | 0.8664 |
| treat | lmmma | | -70.4667 | 24.6009 | 14 | -2.86 | 0.0125 |
| treat | placebo | | 0 | . | . | . | . |
| period | | 1 | -16.9667 | 24.6009 | 14 | -0.69 | 0.5017 |
| period | | 2 | 0 | . | . | . | . |
| treat*period | lmmma | 1 | 62.2667 | 40.8798 | 14 | 1.52 | 0.1500 |
| treat*period | lmmma | 2 | 0 | . | . | . | . |
| treat*period | placebo | 1 | 0 | . | . | . | . |
| treat*period | placebo | 2 | 0 | . | . | . | . |

Interaction is not significant,
but how do we interpret the estimate 62.2667?



## Carry-over effect, intuitive parametrization

Parameters:

- Intercept: Placebo, period 1, $\beta_1$
- Active vs. Placebo, in period 1, $\beta_2$
- Period 2 vs. period 1, for Active, $\beta_3$
- Extra placebo effect in period 2 (carry-over), $\beta_4$

| Treatment | Period | Mean (on log scale) |
|---|---|---|
| Placebo | 1 | $\beta_1$ |
| Placebo | 2 | $\beta_1 + \beta_3 + \beta_4$ |
| Active | 1 | $\beta_1 + \beta_2$ |
| Active | 2 | $\beta_1 + \beta_2 + \beta_3$ |

# Coding of intuitive parametrization

```
data ashina;
set ashina;

active=(treat="LNMMA");
period2=(period=2);
carry_over=(treat="placebo")*(period=2);
run;

proc mixed data=ashina;
  class patient group;
  model effect=active period2 carry_over /
               ddfm=kr s cl;
  random intercept / subject=patient;
run;
```



# Output, intuitive parametrization

Class Level Information

| Class | Levels | Values |
|---|---|---|
| patient | 16 | 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 |
| group | 2 | A B |

Covariance Parameter Estimates

| Cov Parm | Subject | Estimate |
|---|---|---|
| Intercept | patient(group) | 863.90 |
| Residual | | 1405.62 |

Solution for Fixed Effects

| Effect | Estimate | Standard Error | DF | t Value | Pr > \|t\| | Alpha |
|---|---|---|---|---|---|---|
| Intercept | -20.3000 | 15.0649 | 14 | -1.35 | 0.1992 | 0.05 |
| active | -8.2000 | 24.6009 | 14 | -0.33 | 0.7438 | 0.05 |
| period2 | -45.3000 | 24.6009 | 14 | -1.84 | 0.0868 | 0.05 |
| carry_over | 62.2667 | 40.8798 | 14 | 1.52 | 0.1500 | 0.05 |

Extra effect of Placebo in Period 2. Why?





# Output, intuitive parametrization II

Solution for Fixed Effects

| Effect | Lower | Upper |
|---|---|---|
| Intercept | -52.6110 | 12.0110 |
| active | -60.9637 | 44.5637 |
| period2 | -98.0637 | 7.4637 |
| carry_over | -25.4118 | 149.95 |

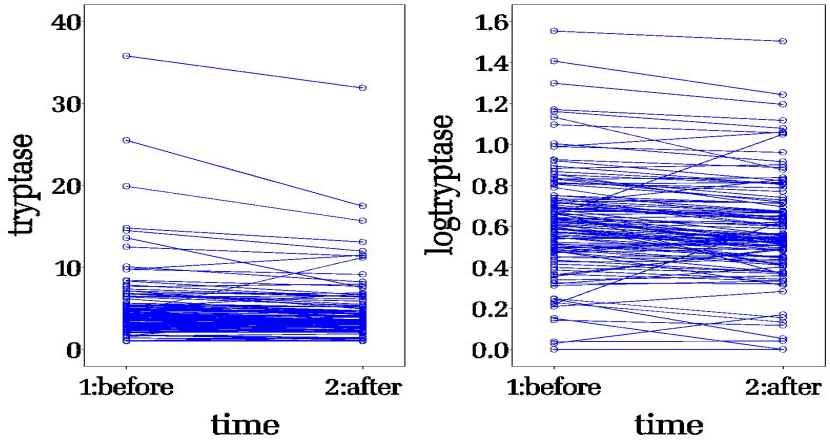
The carry-over effect is estimated to be an extra effect of placebo in period 2, estimated to 62.3, with confidence interval $(-25.4, 149.9)$

The carry-over effect (placebo following active) has a positive value, corresponding to a worsening of the headache.
This could be explained as a psychological effect, in the sense that you expect something better (namely what you experienced in the previous period)



# Paired T-test, revisited (from lecture 1)

Tryptase level before and after operation, for 120 patients




Garvey et al. (2010a,b)

## Paired T-tests, on a $\log_{10}$-scale

```
Difference:  logafter - logbefore

   N        Mean     Std Dev      Std Err     Minimum     Maximum
 120     -0.0414      0.0929      0.00848     -0.3020      0.4062

    Mean        95% CL Mean         Std Dev      95% CL Std Dev
 -0.0414     -0.0582  -0.0246       0.0929      0.0825   0.1064

    DF     t Value    Pr > |t|
   119       -4.88       <.0001
```

Tryptase levels decrease with a factor $10^{-0.0414} = 0.909$,
i.e. approximately 9%

(CI: 0.875-0.945, i.e. a 5.5%-12.5% decrease)

Note: Two-way ANOVA in `patient` and `time` will give exactly the same.



## Same analysis, written as a mixed model

```
proc mixed data=tryptase;
    class patient time;
    model logtryptase=time / s cl ddfm=kr;
    random patient;
run;
```

```
The Mixed Procedure

Covariance Parameter Estimates
Cov Parm     Estimate
patient       0.07328
Residual     0.004318

                    Solution for Fixed Effects

                                 Standard
Effect      time      Estimate      Error     DF  t Value  Pr > |t|   Alpha
Intercept               0.5573    0.02543    119    21.92    <.0001    0.05
time        1:before    0.04139   0.008483   119     4.88    <.0001    0.05
time        2:after           0          .     .        .         .       .

Effect      time          Lower       Upper
Intercept                0.5070      0.6077
time        1:before    0.02459     0.05819
time        2:after           .           .
```





## The balanced situation

as here,
where we have both observations for all patients:

All three types of analyses give the same result:

- The paired T-test
- The two-way Anova
- The mixed model

But only as long as we have all observations



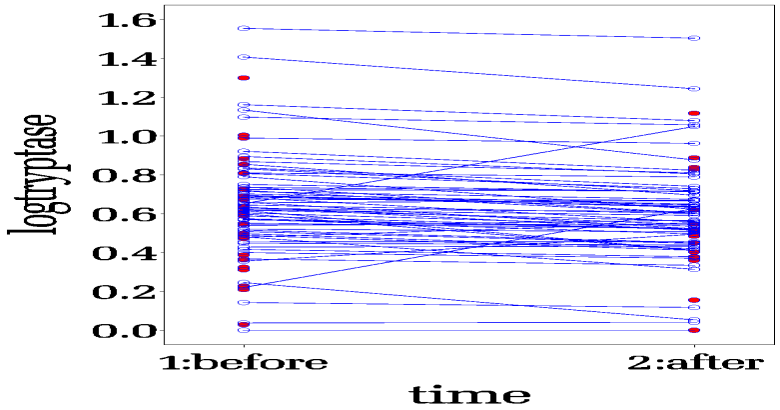
## Missing values

What would happen if some patients were only measured once?
We randomly leave out a number of observations:

```
Analysis Variable : tryptase

                N               N
time         Obs       N     Miss
-----------------------------------
1:before     120      96       24

2:after      120      89       31
-----------------------------------
Total        240     185       55
```

UNIVERSITY OF COPENHAGEN DEPARTMENT OF BIOSTATISTICS

# Missing values, completely at random




Red observations are the singletons:



# Paired T-test (ignoring singletons)

In order to respect the pairing,
only 70 patients have observations both before and after operation,
so only 140 out of the available 185 are used.

```
The TTEST Procedure
Difference:  logafter - logbefore

  N        Mean      Std Dev      Std Err      Minimum      Maximum
 70     -0.0458       0.1015       0.0121      -0.2567       0.4062

    Mean        95% CL Mean         Std Dev       95% CL Std Dev
 -0.0458     -0.0700  -0.0216        0.1015       0.0871    0.1218

    DF     t Value     Pr > |t|
    69       -3.77       0.0003
```

Two-way ANOVA would give the exact same result



UNIVERSITY OF COPENHAGEN DEPARTMENT OF BIOSTATISTICS

# Unpaired T-test (ignoring the pairing)

in order to use all available observations

```
The GLM Procedure
Dependent Variable: logtryptase

                                 Sum of
Source                     DF       Squares   Mean Square  F Value  Pr > F
Model                       1    0.13126389    0.13126389     1.58  0.2104
Error                     183   15.20446687    0.08308452
Corrected Total           184   15.33573075

R-Square     Coeff Var      Root MSE    logtryptase Mean
0.008559      49.77365      0.288244            0.579109

                                          Standard
Parameter                 Estimate           Error   t Value   Pr > |t|
Intercept            0.5514445435 B     0.03055379     18.05     <.0001
time      1:before   0.0533123559 B     0.04241459      1.26     0.2104
time      2:after    0.0000000000 B      .               .        .

Parameter              95% Confidence Limits
Intercept            0.4911615573 0.6117275298
time      1:before   -.0303721402 0.1369968521
time      2:after     .            .
```



# Mixed effects model

In order to use all available observations
and to account for correlations/pairing

```
The Mixed Procedure
Dependent Variable           logtryptase

          Number of Observations
Number of Observations Read             240
Number of Observations Used             185
Number of Observations Not Used          55

Covariance Parameter Estimates
Cov Parm      Estimate
patient        0.07677     ----------------Between
Residual      0.005146     ----------------Within

                    Solution for Fixed Effects

                                 Standard
Effect      time      Estimate      Error    DF  t Value  Pr > |t|   Alpha
Intercept               0.5501    0.02718   127    20.24     <.0001    0.05
time        1:before    0.04640   0.01201  71.9     3.86     0.0002    0.05
time        2:after           0         .     .        .          .       .

Effect      time          Lower       Upper
Intercept                0.4963      0.6039
time        1:before     0.02247     0.07034
time        2:after           .           .
```

## Estimated decrease after tryptase, log-scale

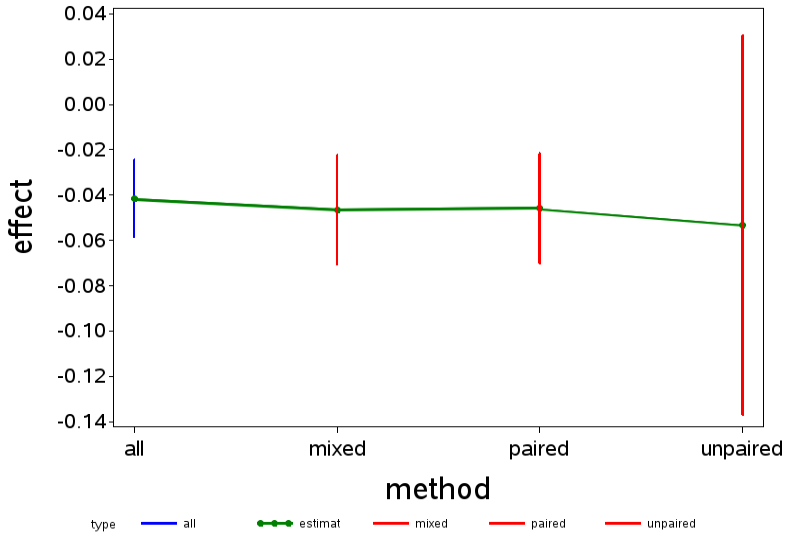
on incomplete data, to sum up:

| Method | Effect | N | Confidence Interval | P-value |
|---|---|---|---|---|
| **Paired T-test** | -0.0458 | 140 | (-0.0700, -0.0216) | 0.0003 |
| **Unpaired T-test** | -0.0533 | 185 | (-0.1369, 0.0304) | 0.21 |
| **Mixed model** | -0.0464 | 185 | (-0.0703, -0.0225) | 0.0002 |
| **All data** | -0.0414 | 240 | (-0.0582, -0.0246) | $< 0.0001$ |

Back-transformed:
Ratio `after/before` estimated to $10^{-0.0464} = 0.899$, i.e. approximately a 10% decrease, with confidence interval $(10^{-0.0703}, 10^{-0.0225}) = (0.851, 0.950)$, i.e. from 5-15%



## Confidence interval comparison








## Within patient correlation

$$\rho = \frac{\omega_B^2}{\omega_B^2 + \sigma_W^2}$$

- From full data set:
  $\frac{0.07328}{0.07328+0.004318} = 0.944$
- From data set with missing values:
  $\frac{0.07677}{0.07677+0.005146} = 0.937$

We expect the pairing to be important here,
as indeed we have seen



## When should we use what approach?

Paired T-test: When the correlation is strong, and only few observations are missing

Unpaired T-test: When the correlation is weak, and many observations are missing

Mixed effects model: Always possible

But note: The missingness has to be random!
Otherwise, bias is expected

## Missing values

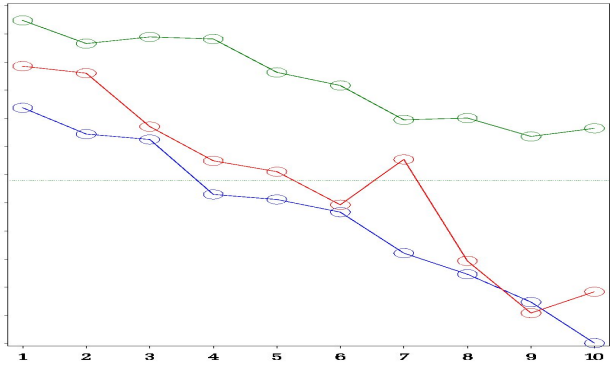
Most investigations are planned to be balanced
but almost inevitable turn out to have missing values,
or patients who drop-out for some reason.

- just by coincidence (blood sample lost or ruined)
- we lost track of the patient (may be worrysome)
- because the patient has recovered
  (worrying, i.e. carrying information)
- the patient is too ill to show up
  (very serious, i.e. carrying unretriavable information)





## Hypothetical time courses

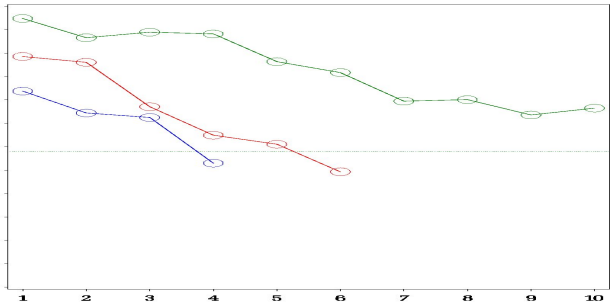







## Possible missing mechanism, I

Low values are good (e.g. blood pressure):
When the patient is well treated, he is dismissed
(missing at random, MAR)



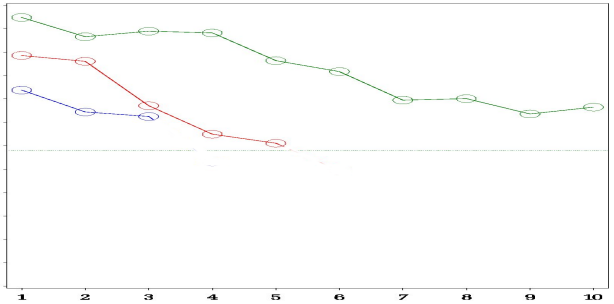

Average values stay high, i.e.
The treatment does not look good





## Possible missing mechanism, II

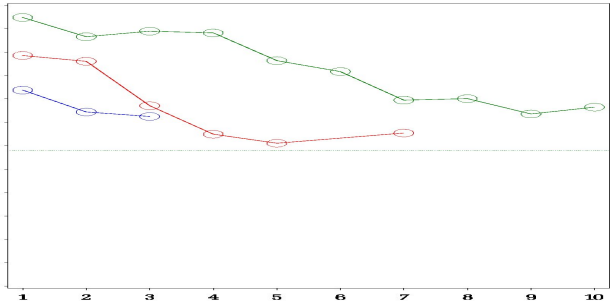
Low values are bad (e.g. lung function):
When the patient is sufficiently ill, he drops out of the labour market.
(informative missing, NI)




Average values stay high, i.e. Healthy worker effect

## Possible missing mechanism, III

Low values are bad (e.g. lung function):
Below some threshold, the patient is too ill to show up (informative missing, NI)




Only high values are observed, so the treatment looks good





## Types of missingness

MCAR Missing completely at random
pure coincidence

MAR Missing at random
Missingness may depend on covariate values ($x$) and possible also on previously observed outcome values ($y$)

NI Non-ignorable: (Informative missing)
Missingness depends upon the values of the unobserved outcome values

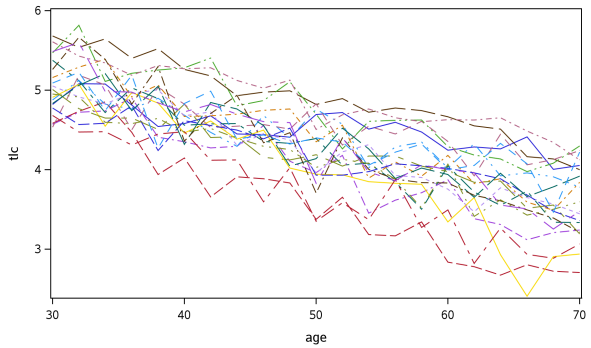




## Hypothetical TLC-observations

Lung capacity measured at regular time intervals for two groups, that we want to compare

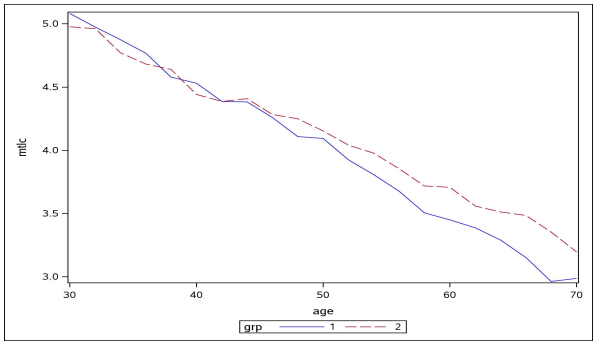
For a single group, we might have







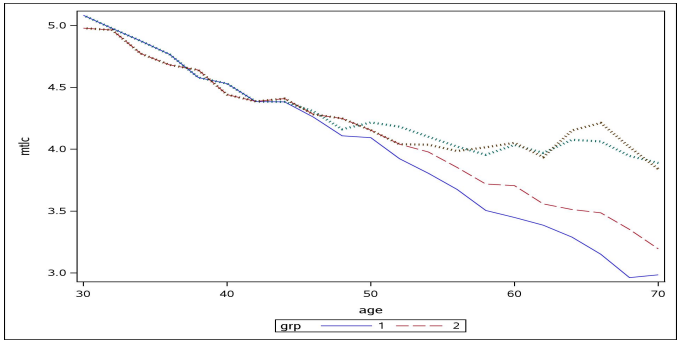

## Average for the two groups




The groups differ for high ages

## Hypothetical example: Informative missing

Patients who would have been below 3.5 are not seen




Averages change, the groups look more identical
and the linearity disappears





## Traditional handling of missing data

- Complete case analysis
- LOCF (or LVCF): Last observation (value) carried forward

More up-to-date methods

- Likelihood methods
- Inverse probablility weighting (IPW)
- Imputation





## Complete case analysis

Make an analysis including only those individuals
who are observed at all available time points

- Default choice for oldfashioned software
- Information loss
- Potential bias, if there is a specific reason for the missingness
- Requires MCAR

not recommended





## LOCF: Last observation carried forward

If an individual has no observed value at time $t$, replace the missing value by the previously observed value

For drop-outs, all subsequent values will equal this same value

Typical consequences:

- The time effect will be less pronounced
- Large residuals, i.e. overestimation of residual variation

definitely not recommended

## Likelihood methods

Requires MAR, but not (necessarily) MCAR

Proper modeling of the observed outcomes, including a model for the covariance: model, i.e:

- Mixed models for normally distributed outcomes
- SS-specific models for other types of outcome (*later*)
- not marginal (PA) models for other types of outcome (this requires the stronger MCAR, *later*)

Important in case of not-MCAR:

- The model has to fit the data properly, *including the model for the covariance*

recommended whenever possible





## Likelihood methods, II

Covariate dependent missingness is included in the MAR definition, and likelihood methods apply
if the covariates in question are actually included in the model.

Hypothetical example: Compare two treatments:

- Outcome is e.g. "general health"
- The treatments have different probability of side-effects
- Side effects may affect the risk of missing values and must therefore be included in the model

If the presence of side effects has an effect on general health, it may be impossible to answer the original question.





## Modeling the missing mechanism

- when likelihood methods are unavailable, or unsatisfactory (e.g. for non-normal outcome, *to follow in the subsequent lectures*),

MAR: Missing at random, not *completely* at random:
The probability of being observed may depend upon

- Covariates (possibly not in the model): group, time, age
- Previously observed outcomes:
  If treatment has worked sufficiently well, it may be stopped

We must model this missing mechanism (typically logistic regression), so that each (patient,time)-combination gets an estimated probability of being observed, $\hat{\pi}_{ij}$.





## Inverse probability modeling (IPW)

The analysis is then performed, with each observation weighted according to the inverse of this probability of being observed, i.e.

$$w_{ij} = \frac{1}{\hat{\pi}_{ij}}$$

Rationale:
The values actually observed should account for those that are not observed. So if the probability of being missing for *this kind of observation* is large, the ones that *are* in fact observed should have a large weight.

If the probability of being observed is low, an actual observation should carry a lot of weight.

## Imputation of missing values

Replace the missing values with some *clever guesses* of what they might have been:

- Single imputation
  - Time average imputation:
    Replace missing value with average over remaining subjects
    - Subject effect will be underestimated
    - Large residuals, i.e. overestimation of residual variation
  - Model prediction imputation:
    Replace missing value with predicted value from some clever model for Y
    - Too small residuals, i.e. downwards bias of of SD
- Multiple imputation is recommended





## MCAR: Missing completely at random

- Average curves are OK
- Complete case analysis OK, but inefficient
- LOCF is unreasonable
- If only few observations are missing, single imputation could work
  but will affect the standard deviations
- Multiple imputation OK, but unneccessary
- Mixed models work OK and
  uses all available information





## MAR: Missing at random

- Average curves are biased
- Complete case analysis is biased
  We downweight subjects with special characteristics
- If only few observations are missing, imputations according to some plausible model could work, but
  multiple imputation is better
- IPW works OK, depending upon the *validity* of the modeling of the missing mechanism
- Likelihood approaches are OK,
  provided that they are specified *correctly*.





## Non-ignorable

Nothing works!

Many attempts have been tried to model the missing mechanisms, but they all rely on assumptions that cannot be checked.

UNIVERSITY OF COPENHAGEN DEPARTMENT OF BIOSTATISTICS

## Missing values in the BMD-example

First BMD-measurement, and slope,
subdivided according to group and dropout:

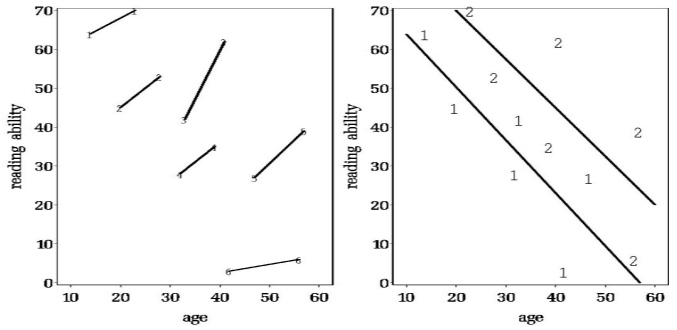
| Treatment | Dropout | N | Mean BMD at first visit | P for dropout |
|---|---|---|---|---|
| Placebo | No | 47 | 0.8703 | 0.95 |
| Placebo | Yes | 10 | 0.8688 | |
| Calcium | No | 44 | 0.8813 | 0.88 |
| Calcium | Yes | 11 | 0.8773 | |

| Treatment | Dropout | N | Slope | P for dropout |
|---|---|---|---|---|
| Placebo | No | 47 | 0.0459 | 0.54 |
| Placebo | Yes | 4 | 0.0295 | |
| Calcium | No | 44 | 0.0547 | 0.87 |
| Calcium | Yes | 4 | 0.0585 | |



## Example: Reading ability

An example of covariates that are correlated to random effects:
Longitudinal (*within-individual*, $\beta_W$) effect of age/training, vs.
Cross-sectional (*between-individual*, $\beta_B$) effect of age/cohort:




Might also be income vs. time



UNIVERSITY OF COPENHAGEN DEPARTMENT OF BIOSTATISTICS

## Data

- Two observations for each individual
- Covariates
  - age1: Age at the first measurement
  - difage: The number of years between first and second measurement

| Obs | id | age | read | age1 | difage |
|---|---|---|---|---|---|
| 1 | 1 | 14 | 64 | 14 | 0 |
| 2 | 1 | 23 | 70 | 14 | 9 |
| 3 | 2 | 20 | 45 | 20 | 0 |
| 4 | 2 | 28 | 53 | 20 | 8 |
| 5 | 3 | 33 | 42 | 33 | 0 |
| 6 | 3 | 41 | 62 | 33 | 8 |
| 7 | 4 | 32 | 28 | 32 | 0 |
| 8 | 4 | 39 | 35 | 32 | 7 |
| 9 | 5 | 47 | 27 | 47 | 0 |
| 10 | 5 | 57 | 39 | 47 | 10 |
| 11 | 6 | 42 | 3 | 42 | 0 |
| 12 | 6 | 56 | 6 | 42 | 14 |



## *Model (technicalities)

- Baseline level: $a_{p1}$ at the age $x_{p1}$:
  $a_{p1} = \alpha + \beta_B x_{p1} + A_p$, $\beta_B$ negative
- Baseline measurement: $y_{p1} = a_{p1} + \varepsilon_{p1}$
- Follow-up level: $a_{p2} = a_{p1} + \beta_W(x_{p2} - x_{p1})$ at the age $x_{p2}$:
- Follow-up measurement: $y_{p2} = a_{p2} + \varepsilon_{p2}$
- Difference: $y_{p2} - y_{p1} = \beta_W(x_{p2} - x_{p1}) - A_p + (\varepsilon_{p2} - \varepsilon_{p1})$

Model for all $y$-observations:

$$y_{pj} = \alpha + \beta_B x_{p1} + \beta_W(x_{pj} - x_{p1}) + A_p + \varepsilon_{pj}$$

## Actual analyses

Regression with inter- as well as intra-individual effect of age/time:

```
proc mixed data=reading;
  class id;
  model read=age1 difage / s;
random id;
run;
```

```
Covariance Parameter Estimates
 Cov Parm     Estimate
id             245.35
Residual      27.0449
```

| Effect | Estimate | Standard Error | DF | t Value | Pr > \|t\| |
|---|---|---|---|---|---|
| Intercept | 78.1267 | 19.1124 | 4 | 4.09 | 0.0150 |
| age1 | -1.3615 | 0.5722 | 5 | -2.38 | 0.0632 |
| difage | 0.8646 | 0.3121 | 5 | 2.77 | 0.0394 |





## Estimation results

| Method | cross sectional ($\beta_B$) Cohort effect | longitudinal ($\beta_W$) Age effect |
|---|---|---|
| $y_{i1}$ vs. $x_{i1}$ | -1.359 (0.458) | – |
| $y_{i2}$ vs. $x_{i2}$ | -1.245 (0.534) | – |
| $y_{ij}$ vs. $x_{ij}$ no individual effect | -1.000 (0.384) | – |
| $y_{i2} - y_{i1}$ vs. $x_{i2} - x_{i1}$ no intercept | – | 0.883 (0.211) |
| $y_{ij}$ vs. $x_{ij}$ random individual level | – | 0.676 (0.307) |
| $y_{ij}$ vs. $x_{i1}$ and $(x_{i2} - x_{i1})$ | -1.362 (0.572) | 0.865 (0.312) |

Colour codes: wrong, feasible, correct





## Explained variation in percent, $R^2$

In the simplest case, the two-level model,
we have two (or more) different variances to explain!

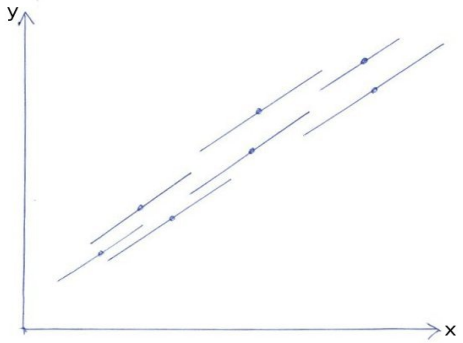
- Residual variation (variation *within* individuals, $\sigma^2_W$)
  - decreases (as usual)
    when we include an important $x$ covariate (level 1)
  - may decrease
    when we include an important $z$ covariate (level 2)
- Variation between individuals , $\omega^2_B$
  - decreases
    when we include an important $z$ covariate (level 2)
  - *may increase or decrease*,
    when we include an important $x$ covariate (level 1)
    which differ between individuals (confounding)





## Hypothetical example I, covariate $x$ included

The $x$'s vary between individuals, and the average outcomes ($\bar{y}$)
are mostly due to this variation:

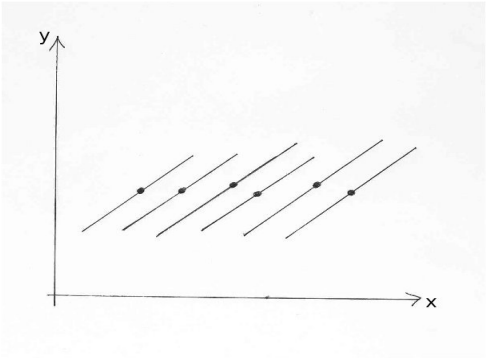



Levels of $y$, for fixed $x$ are quite alike!
$\omega^2$ **decreases**

UNIVERSITY OF COPENHAGEN DEPARTMENT OF BIOSTATISTICS

## Another hypothetical example II, covariate $x$ included

The $x$'s vary between individuals, but the average outcomes ($\bar{y}$) are almost identical:




Levels of $y$, for fixed $x$ are very different!
$\omega^2$ **increases**



## Advice

1. Be sure to formulate in advance the questions you want to answer, as precisely as possible, preferably including handmade sketches of the desired kinds of illustrations.
2. Do not design unneccessarily complicated investigations, i.e. do not try to answer all questions simultaneously.
3. Do your very best in collecting all of the planned information on all individuals.
4. Make illustrations - lots of them!



UNIVERSITY OF COPENHAGEN DEPARTMENT OF BIOSTATISTICS

## Advice, II

5. Try to see whether a conclusion can be reached in a simple fashion - at least a temporary, or approximate conclusion.
6. Try to be calm and systematic when you build your model. Don't just put it into your favourite computer program and run (away).
7. When interpreting your output, check whether the results look plausible - and whether they in fact answer the original questions asked.



## Take care

You may encounter trobles, if

- your outcome is binary, with only a few observations for each subject
- the random effects $b_i$ depend on the covariates, e.g. if the variation is much larger in one group than in the other
- the covariates depend on previous outcome values (e.g. if dose is adjusted according to need)
- Consult our hotline and get some help. We are there for you, as long as you are Ph.D.students - and possibly also afterwards (if you are affiliated with Univ. Cop. - or if your problem is *cute*).