Basics of repeated measurements
Analysis of repeated measurements, 2016

Julie Lyng Forman & Lene Theil Skovgaard
Department of Biostatistics, University of Copenhagen

Outline
About the course
What are repeated measurements?
Basics of longitudinal data (FLW chapters 1 & 2)
The multivariate normal distribution
Analysis of response profiles (FLW chapters 3 & 5)
Baseline adjustment (FLW section 5.6)

Aim of the course
We aim to teach you to:
▶ understand and interpret advanced statistical analyses
▶ perform own analyses using SAS
▶ understand output from a statistical program package - in general, i.e. other than SAS
▶ judge the assumptions behind these analyses and the statistical consequences of your study design.
▶ make suitable presentations of the results from your analyses.

To create a better platform for communication between 'users' of statistics and statisticians, to benefit subsequent collaboration.

Day 1 contents
▶ Basic concepts for correlated and clusted data
▶ Descriptive statistics
▶ The multivariate normal distribution
▶ Analysis of balanced longitudinal data
▶ Baseline-follow up studies
▶ SAS: data handling, descriptive statistics, proc mixed
Prerequisites

You are motivated (from your own research project)

You have an open mind towards mathematical model descriptions and patience to learn statistical programming.

You have a basic knowledge of statistics including:
- the normal distribution, mean and standard deviation/variance
- estimates, standard errors, confidence intervals
- correlation, regression, ANOVA, linear models.
- t-test, $\chi^2$-test, F-test
- generalized linear models (logistic and possion regression)

Topics for the course

Models for dependent data.

Quantitative outcomes (normal distribution):
- Linear mixed models
- Variance component models

Binary and count outcomes:
- Generalized linear mixed models
- Population average models (Generalized estimating equations)

Not covered:
- Censored data (survival analysis)
- Multivariate data (several different outcomes at once)

Course material and practical information

http://staff.pubhealth.ku.dk/~jufo/RepeatedMeasures2016
- Course description and updates
- Lecture notes, demos, and exercises.

As supplementary reading we recommend:

We teach SAS programming.
- Additional R and Stata examples can be found at: www.biostat.harvard.edu/~fitzmaur/ala2e
- We have R solutions to exercises 1–3.

To pass the course

SIGN THE ATTENDANTS LIST which are passed around
- twice a day - one in the morning, one in the afternoon,
- at otherwise randomized times.

80% ATTENDANCE IS REQUIRED TO PASS
- Note: 80% of 12 lists is at least 10 signatures in total.

PLEASE NOTE THAT:
- You cannot sign in advance, for last week, or for your friends!
Outline

About the course

What are repeated measurements?

Basics of longitudinal data (FLW chapters 1 & 2)

The multivariate normal distribution

Analysis of response profiles (FLW chapters 3 & 5)

Baseline adjustment (FLW section 5.6)

What are repeated measurements?

Repeated measurements refer to data where the same outcome has been measured several times, in different situations or at different spots, on the same subjects.

- Special case: longitudinal means repeatedly over time.

Examples of repeated measurements

Subjects should be understood in a wide sense:
- Repeated measurements on a patient or person.
- Or on a mouse, dog, blood sample, or cell line.

Replicates can be made:
- Over time.
- Under different circumstances/treatments.
- With different measurement device.
- On different limbs or locations of the body.

What is clustered data?

Repeated measurements are termed clustered data when the same outcome is measured on groups of subjects that are somehow related.
Examples of clustered data

Clusters could be:
▶ Siblings, families, or school classes.
▶ Clinics, hospitals, or GPs.

But also:
▶ Litters or cages.
▶ Plates (in a laboratory experiment).

Or any kind of clustering due to multiple:
▶ Operators or sessions.

Statistical analysis must account for repetitions

The usual assumption is that observations are independent.

If you have repeated or clustered measurements . . .
▶ the assumption of independence is violated.

Ignoring the repetitions/clustering most often leads to:
▶ p-values that are too small or too large.
▶ confidence intervals that are too wide or too narrow.

It is wrong to analyse repeated measurements data with an ordinary GLM or ANOVA model!!!

Example: A pre-post study

Average daily dietary intake for 10 women over 10 pre-menstrual and 10 post-menstrual days.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Pre-menstrual</th>
<th>Post-menstrual</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5260</td>
<td>3910</td>
<td>1350</td>
</tr>
<tr>
<td>2</td>
<td>5470</td>
<td>4220</td>
<td>1250</td>
</tr>
<tr>
<td>3</td>
<td>5640</td>
<td>3885</td>
<td>1755</td>
</tr>
<tr>
<td>4</td>
<td>6180</td>
<td>5160</td>
<td>1020</td>
</tr>
<tr>
<td>5</td>
<td>6390</td>
<td>5645</td>
<td>745</td>
</tr>
<tr>
<td>6</td>
<td>6515</td>
<td>4680</td>
<td>1835</td>
</tr>
<tr>
<td>7</td>
<td>6805</td>
<td>5265</td>
<td>1540</td>
</tr>
<tr>
<td>8</td>
<td>7515</td>
<td>5975</td>
<td>1540</td>
</tr>
<tr>
<td>9</td>
<td>7515</td>
<td>6790</td>
<td>725</td>
</tr>
<tr>
<td>10</td>
<td>8230</td>
<td>6900</td>
<td>1330</td>
</tr>
<tr>
<td>11</td>
<td>8770</td>
<td>7335</td>
<td>1435</td>
</tr>
<tr>
<td>Mean</td>
<td>6753.6</td>
<td>5433.2</td>
<td>1320.5</td>
</tr>
<tr>
<td>SD</td>
<td>1142.1</td>
<td>1216.8</td>
<td>366.7</td>
</tr>
</tbody>
</table>

D.G. Altman: Practical Statistics for Medical Research, Section 9.5

Paired data

The most simple example of clustered or repeated measurements.
▶ Two replicates per subject or two subjects per cluster

Examples of paired data:
▶ Same person with treatment and placebo.
▶ A baseline and a follow up measurement.
▶ Twin study.
▶ Comparison of two measurement methods or reliability of a measurement method.

Quantitative outcomes are analysed with the paired t-test.
Example: Paired vs unpaired comparison

To compare pre-menstrual and post-menstrual dietary intake.

- Test $H_0: \mu_1 = \mu_2$.
- Find a confidence interval for $\mu_1 - \mu_2$.

Note the very different results:

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Estimate (95% CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>paired t-test</td>
<td>1320 (1074;1567)</td>
<td>0.0000003</td>
</tr>
<tr>
<td>two-sample t-test</td>
<td>1320 (271; 2370)</td>
<td>P=0.01625</td>
</tr>
</tbody>
</table>

The paired test is much more powerful.

Explanation: it’s all about the correlation

The two-sample t-test assumes the two samples are independent. But there is a strong dependence between pre- and post-intake for the same woman (correlation 0.95, 95% CI: 0.83-0.99).

Outline

- About the course
- What are repeated measurements?
- Basics of longitudinal data (FLW chapters 1 & 2)
- The multivariate normal distribution
- Analysis of response profiles (FLW chapters 3 & 5)
- Baseline adjustment (FLW section 5.6)

Case: A baseline follow-up study

A randomized clinical trial was conducted to compare Eplerenone with standard treatment of patients with chronic kidney disease.

Outcome: Augmentation index (aix), smaller is better.

Repeated measurements at:

- baseline,
- after 12 weeks (safety),
- after 24 weeks (end point).

Note: The study was planned with 37 subjects in each group, but only 25 and 26, resp. could be treated within the time limit.

Typical set-up for longitudinal measurements

Two or more **groups** of subjects
- Often receiving different treatments
- Possibly randomised at baseline.

**Longitudinal measurements**, typically as a function of
- duration (of treatment or disease)
- age

Do the time courses differ between the groups?

Merits of longitudinal studies

In longitudinal studies measurements are taken repeatedly on the same subjects over time.

- This allows us to study changes over time **within subjects** and factors that influence these changes, e.g. treatment.
- By comparing each subject’s responses at two or more occasions we eliminate extraneous but unavoidable sources of variabilitlity among subjects. Thus we obtain more accurate estimates and more certain conclusions about changes over time than in cross-sectional studies.

Longitudinal vs cross sectional effect
**Example:** Reading ability, as a function of age and cohort:

Visualizing data: Spaghettiplots
**Good for visual inspection because replicates are connected!**

Note: Missing data due to failed measurements, side effects, relapse or other illness (missing data discussed further in lecture 4).
Why visualization is so important

Graphical description of the data is useful for:

▶ **Exploratory data analysis** and **hypothesis generation**.
▶ **Aiding interpretation** of planned analyses.
▶ **Presentation** or **saying it in figures** rather than in numbers.
▶ **Spotting outliers** that could otherwise spoil your analysis.
▶ **Rough assessment of model assumptions** such as normal distribution or linear trend over time.

**Note:** Having a large dataset is no excuse for omitting graphical description. You can divide your data into subgroups or at least look at a random subsample.

Balanced and complete data

In a planned study the times of measurements will usually be the same for all subjects. We have a **balanced design**

In practice data is most often somewhat unbalanced due to drop-out, missed visits, failed measurements.

▶ In this case we say that **data is incomplete**.
▶ But still the **design is balanced**.

Data from (retrospective) observational studies are most often **unbalanced** both by design and in practice.

Unbalanced designs are treated in lecture 2. Missing data is treated in lecture 4.

Outline

About the course

What are repeated measurements?

Basics of longitudinal data (FLW chapters 1 & 2)

The multivariate normal distribution

Analysis of response profiles (FLW chapters 3 & 5)

Baseline adjustment (FLW section 5.6)

The distribution of repeated outcomes

Repeated measurements are characterized by being

▶ mutually dependent or **correlated**.

We need to characterize their **joint distribution**.

Model for quantitative data: **the multivariate normal distribution**

▶ Location: **mean-vector**
  We have a list of means, one for each occasion.
▶ Variability: **covariance-matrix**
  We have a list of variance, one for each occasion plus a table of cross-correlations.
The multivariate normal distribution

Multivariate normal data

Eplerenone: Scatter plots

Left: Eplerenone. Right: Controls.

Better check of normal distribution: use residual diagnostics (lecture 4).

Eplerenone: Summary statistics

Means, SDs and cross-correlations.

<table>
<thead>
<tr>
<th></th>
<th>Eplerenone group</th>
<th>Control group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week</td>
<td>N</td>
<td>Mean</td>
</tr>
<tr>
<td>0</td>
<td>26</td>
<td>22.29</td>
</tr>
<tr>
<td>12</td>
<td>24</td>
<td>19.94</td>
</tr>
<tr>
<td>24</td>
<td>22</td>
<td>20.39</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Corr</th>
<th>0</th>
<th>12</th>
<th>24</th>
<th>Corr</th>
<th>0</th>
<th>12</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.00</td>
<td>0.68</td>
<td>0.73</td>
<td>0</td>
<td>1.00</td>
<td>0.79</td>
<td>0.76</td>
</tr>
<tr>
<td>12</td>
<td>0.68</td>
<td>1.00</td>
<td>0.82</td>
<td>12</td>
<td>0.79</td>
<td>1.00</td>
<td>0.80</td>
</tr>
<tr>
<td>24</td>
<td>0.73</td>
<td>0.82</td>
<td>1.00</td>
<td>24</td>
<td>0.76</td>
<td>0.80</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Note: May be misleading if data is not normal.
Basic concepts: covariance and correlation

Both are used to describe the linear association between two variables assumed to have a joint normal distribution.

▶ The covariance between two measurements is:

\[ \text{Cov}(Y_1, Y_2) = E\{(Y_1 - \mu_1)(Y_2 - \mu_2)\} \]

...in squared units of the original measurements.

▶ The correlation between two measurements

\[ \text{Cor}(Y_1, Y_2) = \frac{\text{Cov}(Y_1, Y_2)}{\text{SD}(Y_1)\text{SD}(Y_2)} \]

...it has no units - interpretation is free of scale.

Note: SAS PROC MIXED and most other statistical software display the covariances, not correlations, as default output.

Matrix notation

Covariances and correlations of the 3D (normal) distribution:

\[
\text{Cov} = \begin{pmatrix}
\sigma_1^2 & \sigma_{12} & \sigma_{13} \\
\sigma_{21} & \sigma_2^2 & \sigma_{23} \\
\sigma_{31} & \sigma_{32} & \sigma_3^2
\end{pmatrix}, \quad \text{Cor} = \begin{pmatrix}
1 & \rho_{12} & \rho_{13} \\
\rho_{21} & 1 & \rho_{23} \\
\rho_{31} & \rho_{32} & 1
\end{pmatrix}
\]

NOTE:

▶ Variances \( \sigma_1^2, \sigma_2^2, \sigma_3^2 \) along the diagonal in \( \text{Cov} \).
▶ 1’s along the diagonal in \( \text{Cor} \).
▶ Both are symmetric \( \sigma_{ij} = \sigma_{ji} \) and \( \rho_{ij} = \rho_{ji} \).
▶ Note the relation \( \rho_{ij} = \sigma_{ij} / \sqrt{\sigma_i^2 \cdot \sigma_j^2} \).

What if data is not normally distributed?

The usual assumption is that outcomes from the same subject follow a multivariate normal distribution.

But linear mixed models for repeated outcomes are robust.

▶ If sample size is not too small.
▶ If the distribution of the data is not too skewed.

So your data doesn’t have to be perfectly normal.

Highly skewed data should always be transformed.

Models for counts are treated in lecture 5.
Models for binary outcomes are treated in lecture 6.

Outline

About the course
What are repeated measurements?
Basics of longitudinal data (FLW chapters 1 & 2)
The multivariate normal distribution
Analysis of response profiles (FLW chapters 3 & 5)
Baseline adjustment (FLW section 5.6)
Analysis of response profiles

Comparison of change over \( n \) time points within \( g \) groups of subjects (e.g. different treatments).

- Similar to two-way ANOVA only with correlated data.
- Covariates: group and time (both categorical)
- Balanced design, but possibly incomplete data.
- Do the groups evolve differently with time?

Interest is in the mean parameters (systematic effects)

\[
\begin{array}{ccc}
\text{time}=0 & \mu_{11} & \mu_{21} \\
\text{time}=12 & \mu_{12} & \mu_{22} \\
\text{time}=24 & \mu_{13} & \mu_{23} \\
\end{array}
\]

Main hypothesis

Analysis of response profiles allows for testing a large number of different hypothesis about the mean parameters.

Which hypotheses are relevant of course depend on the subject matter.

Example: The scientific hypothesis was that there would be a positive effect of Eplerenone compared to the standard treatment at final follow up.

The relevant statistical nullhypothesis is:

\[
H_0: \mu_{13} - \mu_{11} = \mu_{23} - \mu_{21},
\]

I.e. same change in means in the two groups at last follow up.

Eplerenone: Averages over time

Seeming improvement over time with Eplerenone.

- But what about statistical uncertainty?
- We also need to consider the (co)variance . . .

Eplerenone: results of analysis of response profiles

<table>
<thead>
<tr>
<th>Week</th>
<th>Control</th>
<th>Eplerenone</th>
<th>Difference</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>1.09 (-2.47;4.65)</td>
<td>-0.86 (-4.38;2.66)</td>
<td>-1.95 (-6.96;3.06)</td>
<td>( P=0.44 )</td>
</tr>
<tr>
<td>24</td>
<td>3.09 (0.07;6.11)</td>
<td>-0.51 (-3.56;2.53)</td>
<td>-3.61 (-7.90;0.68)</td>
<td>( P=0.10 )</td>
</tr>
</tbody>
</table>

There is a seeming improvement at last follow-up with Eplerenone compared to standard treatment with a mean difference in change in AIX of -3.61 (95% CI: -7.90 to 0.68, \( P = 0.10 \)).

But: The difference between the treatments is not significant.
Merits of analysis of response profiles

We can use a linear mixed model (PROC MIXED in SAS) to describe differences between groups at any time point or changes between any two time points (explanation follows).

Computationally this is an advantage compared to making many t-tests. Everything is computed at one go.

Linear mixed models handles data that are missing at random optimally whereas t-tests may be biased (more on this lecture 4).

There is a gain in statistical power when doing baseline adjustment in the analysis of randomized studies (more on this later today).

Linear mixed models (LMMs)

We use linear mixed models to analyze quantitative repeated measurements.

Systematic effects (means) are modeled similar to general linear models (GLM) including relevant explanatory variables such as time, treatment, age, gender, etc.

Additional specification of a model for the covariance is needed due to the repeated measurements. We will consider many such models either given in terms of

- So-called covariance pattern models for the residual covariance
- So-called random effects (e.g. in multi-level models)
- Or a mixture of these for more complex data.

(More about linear mixed models and applications in lectures 2-4).

The unstructured covariance

With a balanced design and few different time points we don’t have to make any specific assumptions about the covariance:

An unstructured covariance pattern model is assumed.

- One variance parameter for each time point
- One correlation parameter for each pair of time points
- \( n + \frac{n(n-1)}{2} \) parameters in total with \( n \) time points.

Usually all groups are assumed to have the same covariance, but this assumption can be relaxed.

Note: Not feasible with many time points or groups.

Two-way ANOVA type model for the means

Describe means for the six time-treatment combinations as:

<table>
<thead>
<tr>
<th>Time</th>
<th>group = Control</th>
<th>group = Eplerenone</th>
</tr>
</thead>
<tbody>
<tr>
<td>time=0</td>
<td>( \beta_1 )</td>
<td>( \beta_1 + \beta_4 )</td>
</tr>
<tr>
<td>time=12</td>
<td>( \beta_1 + \beta_2 )</td>
<td>( \beta_1 + \beta_2 + \beta_4 + \beta_5 )</td>
</tr>
<tr>
<td>time=24</td>
<td>( \beta_1 + \beta_3 )</td>
<td>( \beta_1 + \beta_3 + \beta_4 + \beta_6 )</td>
</tr>
</tbody>
</table>

- Mean of standard treatment at baseline is reference (intercept)
- Change over time with standard treatment (time estimates)
- Difference between groups at baseline (group estimate)
- Differences in time effects between the groups (interaction or time\(*\)group estimates)
SAS output (program on SAS slides)

```
MODEL aix = week group week*group / SOLUTION;
```

<table>
<thead>
<tr>
<th>Effect</th>
<th>week</th>
<th>treat</th>
<th>Estimate</th>
<th>StdError</th>
<th>DF</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>t</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td></td>
<td></td>
<td>24.3431</td>
<td>2.0793</td>
<td>49.4</td>
<td>11.71</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>week</td>
<td>12</td>
<td>0</td>
<td>1.0887</td>
<td>1.7694</td>
<td>46.2</td>
<td>0.62</td>
<td>0.5414</td>
<td></td>
<td></td>
</tr>
<tr>
<td>week</td>
<td>24</td>
<td>0</td>
<td>3.0895</td>
<td>1.4995</td>
<td>44.5</td>
<td>2.06</td>
<td>0.0452</td>
<td></td>
<td></td>
</tr>
<tr>
<td>group</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>group</td>
<td>1</td>
<td>0</td>
<td>-2.0547</td>
<td>2.8999</td>
<td>48.9</td>
<td>-0.71</td>
<td>0.4820</td>
<td></td>
<td></td>
</tr>
<tr>
<td>week*group</td>
<td>12</td>
<td>1</td>
<td>-1.9493</td>
<td>2.4871</td>
<td>45.6</td>
<td>-0.78</td>
<td>0.4372</td>
<td></td>
<td></td>
</tr>
<tr>
<td>week*group</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>week*group</td>
<td>24</td>
<td>1</td>
<td>-3.6078</td>
<td>2.1298</td>
<td>45.3</td>
<td>-1.69</td>
<td>0.0971</td>
<td></td>
<td></td>
</tr>
<tr>
<td>week*group</td>
<td>24</td>
<td>0</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>week*group</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>week*group</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Confidence intervals omitted due to lack of space)

Note: Here baseline (week=0) and the control group (group=0) is the reference point (intercept).

Estimated response profiles

![Estimated response profiles graph]

Type 3 test of the time*group-interaction

<table>
<thead>
<tr>
<th>Effect</th>
<th>Num</th>
<th>Den</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>week</td>
<td>2</td>
<td>44.5</td>
<td>0.99</td>
<td>0.3794</td>
</tr>
<tr>
<td>group</td>
<td>1</td>
<td>47</td>
<td>1.84</td>
<td>0.1817</td>
</tr>
<tr>
<td>week*group</td>
<td>2</td>
<td>44.5</td>
<td>1.43</td>
<td>0.2490</td>
</tr>
</tbody>
</table>

By testing $H_0$: No group*time-interaction we test that

- mean change over time is identical in all groups...
- ...at all follow-up times.

If we aim to show that there is a treatment effect we will get more power by focusing on a specific time point;

- Usually the time point where the difference is largest.

This is a so-called one degree of freedom test (FLW section 5.5).

Post hoc testing

That the group*time interaction is significant indicate that there is a difference in changes over time between the groups, but

- not between which time points.
- not between which groups, if there are more than two.

To find out where differences occur we have to look at estimated differences between specific groups at specific time points.

- The total number of comparisons may become large in particular if there are many time points (or several groups).
- Shouldn't we adjust for multiple testing?

LSMEANS

SAS proc mixed allows you to estimate the means for all time-group-combinations and all possible differences between them.

| Effect      | week | group | _week | _group | Estimate | Error | DF   | t Value | Pr > |t| |
|-------------|------|-------|-------|--------|----------|-------|------|---------|-------|---|
| week*group  | 12   | 1     | 12    | 0      | -4.0040  | 3.5909 | 45.5 | -1.12   | 0.2707 |
| week*group  | 12   | 1     | 24    | 1      | -0.3423  | 1.5282 | 46.8 | -0.22   | 0.8237 |
| week*group  | 12   | 1     | 24    | 0      | -6.0048  | 3.2739 | 54.8 | -1.83   | 0.0721 |
| week*group  | 12   | 0     | 24    | 1      | 3.6617   | 3.2988 | 55.4 | 1.11    | 0.2718 |
| week*group  | 12   | 0     | 24    | 0      | -1.0098  | 1.4868 | 44.9 | -0.64   | 0.5248 |
| week*group  | 12   | 0     | 0     | 1      | 1.0887   | 1.7694 | 46.2 | 0.62    | 0.5414 |
| week*group  | 24   | 1     | 24    | 1      | -5.6625  | 2.9506 | 46.4 | -1.92   | 0.0562 |
| week*group  | 24   | 1     | 0     | 1      | -0.5183  | 1.5125 | 46.0 | -0.34   | 0.7334 |
| week*group  | 24   | 0     | 0     | 0      | -2.5730  | 2.9485 | 63.7 | -0.87   | 0.3861 |
| week*group  | 24   | 0     | 0     | 0      | 5.1442   | 2.9019 | 60.7 | 1.77    | 0.0813 |
| week*group  | 0    | 1     | 0     | 0      | -2.0547  | 2.8999 | 48.9 | -0.71   | 0.4820 |

Not all comparisons are interesting, though.

Baseline in randomized studies

Example: CKD patients were randomized to Eplerenone or standard treatment.

- We know that the two treatment groups cannot differ systematically at baseline since they represent two random samples from the same population.
- We are wasting statistical power when estimating the difference between the baseline means.

So should we leave out the baseline measurement?

- No, then we loose information about changes over time and again the power of the test of treatment effect is reduced.

Solution: Constrained model (cLMM)

<table>
<thead>
<tr>
<th>time</th>
<th>group = Control</th>
<th>group = Eplerenone</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$\beta_1$</td>
<td>$\beta_1 + 0$</td>
</tr>
<tr>
<td>12</td>
<td>$\beta_1 + \beta_2$</td>
<td>$\beta_1 + \beta_2 + 0 + \beta_4$</td>
</tr>
<tr>
<td>24</td>
<td>$\beta_1 + \beta_3$</td>
<td>$\beta_1 + \beta_3 + 0 + \beta_5$</td>
</tr>
</tbody>
</table>

- Intercept.
- Time effect with standard treatment
- Difference between groups at baseline = 0!
- Differences in time-effects (interaction)
Eplerenone: cLMM results

Changes in mean AIX (%) since baseline (with adjustment).

<table>
<thead>
<tr>
<th>Week</th>
<th>Control</th>
<th>Eplerenone</th>
<th>Difference</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>1.20 (-2.38;4.78)</td>
<td>-0.95 (-4.50;2.59)</td>
<td>-2.16 (-7.24;2.92)</td>
<td>P=0.40</td>
</tr>
<tr>
<td>24</td>
<td>3.36 (0.42;6.30)</td>
<td>-0.76 (-3.75;2.22)</td>
<td>-4.12 (-8.24;-0.01)</td>
<td>P=0.049</td>
</tr>
</tbody>
</table>

**Note:** Significant difference at last follow-up in favour of Eplerenone.

SAS-output:

```
Effect       week  treat  Estimate  StdError  DF  t Value  Pr > |t|  Alpha
Intercept    22.2879  1.4430  50  16.14  <.0001  0.05
week 12      1.2017  1.7816  46.8  0.67  0.5033  0.05
week 24      3.3608  1.4643  48.3  2.30  0.0261  0.05
week*group   12  1  -2.1552  2.5240  46  -0.85  0.3976  0.05
week*group   12  0  0  .  .  .  .  .  .
week*group   24  1  -4.1247  2.0436  45.9 -2.02  0.0494  0.05
week*group   24  0  0  .  .  .  .  .  .
week*group   0  0  0  .  .  .  .  .  .
(Confidence intervals omitted)
```

Classical approaches for handling baseline

1. **Two-sample t-test on the end point measurements.**
   - Has less power than the others if the correlation is strong.

2. **Two-sample t-test on the changes.**
   - Has less power than the others if the correlation is weak.

3. **ANCOVA model including baseline as a covariate.**
   - Has optimal power (when there are no missing data).


Eplerenone: ANCOVA results

Expected change in AIX (%) since baseline ...

<table>
<thead>
<tr>
<th>Week</th>
<th>Control</th>
<th>Eplerenone</th>
<th>Difference</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>1.40 (-2.24;5.03)</td>
<td>-0.95 (-4.56;2.66)</td>
<td>-2.35 (-7.51;2.82)</td>
<td>P=0.36</td>
</tr>
<tr>
<td>24</td>
<td>3.47 (0.57;6.38)</td>
<td>-0.76 (-3.73;2.21)</td>
<td>-4.23 (-8.42;-0.05)</td>
<td>P=0.048</td>
</tr>
</tbody>
</table>

...for subjects with average baseline value - !!!!!

SAS-output:

```
Effect       week  group  Estimate  StdError  DF  t Value  Pr > |t|  Alpha
Intercept    1.3967  1.8030  44  0.77  0.4427  0.05
week 12      1.3738  1.4392  43  2.41  0.0201  0.05
week*group   12  1  -2.3494  2.5643  44 -0.92  0.3646  0.05
week*group   12  0  0  .  .  .  .  .  .
week*group   24  1  -4.2329  2.0776  43 -2.04  0.0477  0.05
week*group   24  0  0  .  .  .  .  .  .
baseline*week 12  -0.09247  0.1320  44 -0.70  0.4872  0.05
baseline*week 24  -0.2430  0.1058  43.2 -2.30  0.0266  0.05
(Confidence intervals omitted)
```
ANCOVA: predicted changes over time

The two models have somewhat different interpretations.

- cLMM estimates the difference in population mean response.
- ANCOVA estimates the expected difference in response between subjects having the same baseline response.
- only ANCOVA is valid if randomization is *conditional on baseline values*.

The two models estimate the same treatment effect with similar accuracy/power

- Estimates and p-values are usually very similar.
- cLMM can better handle missing data. ANCOVA deletes subjects with a missing baseline or entirely missing follow-up.

**Note:** In models for non-normal data results differ between the constrained approach and when using baseline as a covariate (lectures 5+6).

Baseline in observational studies

Compare the outcomes for individuals from different groups (e.g. gender or illness groups):

- The groups are likely to differ in many respects ... *including the baseline outcome value*!
- Differences in response profiles may be due to many factors, and quantifications will depend on which of these are factors are included in the model.
- Adjust for the covariates that are sensible in the context.

Is the baseline measurement a *sensible* covariate?

Baseline in observational studies

Fitzmaurice et al. (2011)[Section 5.6]:

For example, in an observational study examining gender differences in weight gain of infants between 12 months (baseline) and 24 months (...) At baseline boys are on average 1 1/2 pounds heavier than girls, but there is no evidence of a gender effect on the 12 month change in body weight, with boys and girls both gaining approximately 5 1/4 pound. In contrast the analysis of covariance of the same data reveals a discernible gender effect with boys showing more weight gain than girls.

(...) the analysis of covariance is directed at the conditional question of whether boys are expected to gain more weight than girls given that they have the same initial weight at 12 months. (...) The reasoning is that if a boy and girl have the same initial weight at 12 months, then there are two possibilities: (1) the girl is initially overweight and is expected to gain less weight or (2) the boy is initially underweight and is expected to gain more weight over the 12 months.

We advise readers to employ the analysis of covariance approach in longitudinal settings only if the approach and its implications are fully understood.