Faculty of Health Sciences

Introduction and longitudinal data analysis

Analysis of repeated measurements, 9th March 2015

Julie Lyng Forman

Department of Biostatistics, University of Copenhagen

Outline

Plan

Today: Introduction and longitudinal data analysis.

- Basic concepts for correlated and clustered data.
- Analysis of response profiles.
- Baseline adjustment for randomized trials
- Covariance pattern models.

Tomorrow: Linear mixed models in general.

- Random effects and variance components
- Multi-level models for clustered data.
- Cross-over trials.

Practicalities

Lectures in the morning (9.15–12.00)

- Linear mixed models for repeated measurements and clustered data.

Computer labs the afternoon (13.00-15.45)

- Data analysis with `SAS PROC MIXED`

Note: Some of the datasets we use in our case studies are in fact too small to yield interesting conclusion. But due to their small size they are useful for illustrative purposes.

Recommended reading

Lecture notes, exercises etc

- Found at the course webpages

The book:

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

Additional examples in SAS, R and Stata can be found at:

- `www.biostat.harvard.edu/fitzmaur/ala2e`

What are repeated measurements?

Repeated measurements refer to data where the same outcome has been measured in different situations (or at different spots) on the same individuals.

- Special case: longitudinal means repeatedly over time.

What is clustered data?

Repeated measurements are termed clustered data when the same outcome is measured on groups of individuals from the same families/workplaces/school classes/clinics/etc.

Why do we need special models for rep. measurements?

The **general linear model (GLM)** assume that observations are **independent**. If you have clustered or repeated measurements the assumption of independence is violated.

Ignoring the repetitions/clustering would lead to invalid inference:

- p-values that are too small or too large.
- confidence intervals that are too wide or too narrow.

Change in enegy intake (kJ) pre- to post-menstrually in 11 women.

Analysis	Estimate (95% CI)	P-value
paired t-test	1320 (1074;1567)	0.0000003
two-sample t-test	1320 (271; 2370)	0.01625

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

UNIVERSITY OF COPENHAGEN DEPARTMENT OF BIOSTATISTICS

Outline

Typical set-up for longitudinal measurements

Two or more **groups** of subjects

- Often receiving different treatments
- Possibly randomised at baseline.

Longitudinal measurements of the same quantity over time for each subject, typically as a function of

- time (i.e. duration of treatment)
- age
- cumulative dose of drug

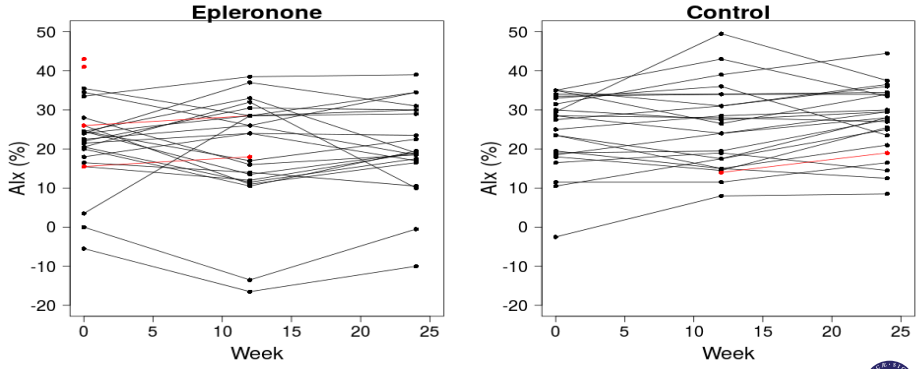
Do the time courses differ between the groups?

UNIVERSITY OF COPENHAGEN DEPARTMENT OF BIOSTATISTICS

Case study: Eplerenone

Response: Augmentation index (aix) in patients with CKD.

- Comparison of novel treatment to standard.
- Follow-up after 12 and 24 weeks.



Reference: Boesby etal, PLOS ONE 8(5), 2013.

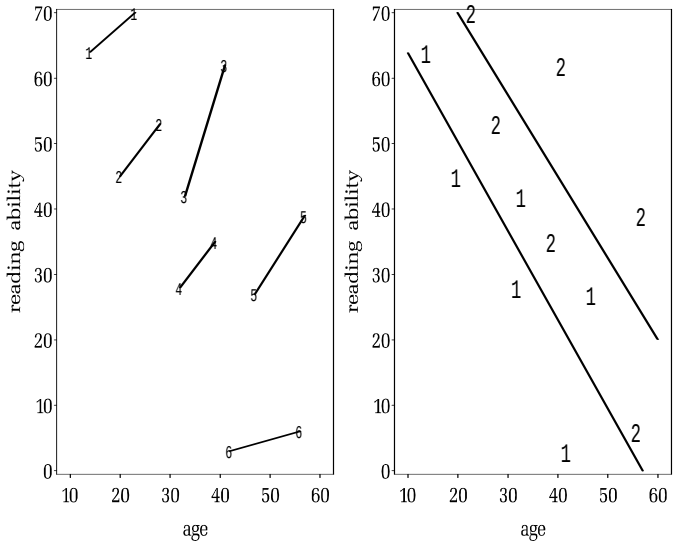
Merits of longitudinal studies

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 individuals responses at two or more occations we eliminate extraneuous but unavoidable sources of variabitlity among individuals. 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:



UNIVERSITY OF COPENHAGEN DEPARTMENT OF BIOSTATISTICS

Unbalanced and incomplete 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 **the design is still balanced**.

Data from (retro-spective) observational studies are most often unbalanced both by design and in practice.

UNIVERSITY OF COPENHAGEN DEPARTMENT OF BIOSTATISTICS

Outline

UNIVERSITY OF COPENHAGEN DEPARTMENT OF BIOSTATISTICS

Notation

Longitudinal data is described as:

- Subjects $i = 1, \ldots, N$.
- Observations $Y_{i1}, \ldots, Y_{in_i}$ (from subject i).
- Taken at occations $t_{i1}, \ldots, t_{in_i}$ (for subject i).
- Possibly additional covariates $X_{ij2}, \ldots, X_{ijp}$ (for subject i at occation j).

Convention: Subscripts i are dropped when occations $t_1, \ldots, t_n$ are the same for all subjects.

The distribution of repeated outcomes

Repeated measurements $Y_{i1}, \ldots, Y_{in_i}$ are characterized by being

- mutually dependent or **correlated**.

We need to characterize their joint distribution.

Standard model for quantitative data: The multivariate normal

- Location: mean-**vector**
- Variability: covariance-**matrix**

Main interest is in modeling the mean.

- **BUT:** We also need to model the covariance in order to account for it in the analyses.

Notation

Denote mean and variance of the normal distribution by:

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \ldots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \ldots & \sigma_{2n} \\ \vdots & \vdots & & \vdots \\ \sigma_{n1} & \sigma_{n2} & \ldots & \sigma_n^2 \end{pmatrix}$$

The correlation matrix is

$$\begin{pmatrix} 1 & \rho_{12} & \ldots & \rho_{1n} \\ \rho_{21} & 1 & \ldots & \rho_{2n} \\ \vdots & \vdots & & \vdots \\ \rho_{n1} & \rho_{n2} & \ldots & 1 \end{pmatrix}$$

where $\rho_{jk} = \mathrm{Cor}(Y_j, Y_k) = \frac{\sigma_{jk}}{\sigma_j \sigma_k}$.

Modeling the mean

Models for the mean are specified just as in GLM.

- including covariates that can be both cathegorical and continuous.
- E.g. treatment, gender, and age.

The time-effect is always included.

- As a factor
- or as a linear/polynomial trend.

Note that covariates are allowed to change with time.

The linear model for longitudinal data

Assume a linear model for the response-**vectors**:

$$Y_i = X_i \cdot \beta + \varepsilon_i$$

I.e.

$$\begin{pmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{in} \end{pmatrix} = \begin{pmatrix} X_{i11} & X_{i12} & \ldots & X_{i1p} \\ X_{i21} & X_{i22} & \ldots & X_{i2p} \\ \vdots & \vdots & & \vdots \\ X_{in1} & X_{in2} & \ldots & X_{inp} \end{pmatrix} \cdot \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{in} \end{pmatrix}$$

- Where X_i is the $n \times p$ design-matrix for subject i.
- Error terms are multivariate normal $\varepsilon_i \sim \mathcal{N}_n(0, \Sigma)$
  For now we assume the covariance Σ is the same for all subjects, but we could have different Σ's for groups

Modeling the covariance

Several possibilitites.

Unstructured covariance

- 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.

Fully flexible because no assumptions are made about the covariance as a function of time.

Covariance pattern models

- Models borrowed from time series analysis. Make use of the time ordering to describe covariance with fewer parameters.
- Variance component models (tomorrow).

Must be chosen with care due to risk of misspecification.

Likelihood inference

The likelihood function of the longitudinal data (or other repeated measurements) is:

$$\prod_{i=1}^{N} \left(\frac{1}{2\pi|\Sigma(\theta)|} \right)^{\frac{n}{2}} \exp\left\{ -\frac{(y_i - X_i\beta)^T \Sigma(\theta)^{-1} (y_i - X_i\beta)}{2} \right\}$$

Maximize to get estimates of the model parameters:

- β (mean value structure)
- θ (covariance structure)

Generalized least squares (GLS)

Suppose a known covariance Σ_0 is known, then the MLE for β is

- $\hat{\beta} = \left\{ \sum_{i=1}^{N} (X_i^T \Sigma_0^{-1} X_i) \right\}^{-1} \sum_{i=1}^{N} X_i^T \Sigma_0^{-1} Y_i$

This estimate is unbiased even if Σ_0 is not the true covariance.

- We can get an estimate using any *working covariance* (. . . at the price of a possible loss in efficiency).
- E.g. working independence estimator (= OLS).

Restricted likelihood (REML)

The conventional likelihood (ML) yields biased estimates of Σ when $\dim(\beta)$ is large.

- Hence, not often used.

Restricted likelihood yields umbiased estimates of Σ.

1. Use a working covariance to get an initial GLS estimate of $\beta\cdot$,
2. the covariance is estimated from the 'likelihood' of the resulting residuals (non-linear optimization problem),
3. and finally β is re-estimated by GLS weighting by the estimated covariance.

Note: In `proc mixed` REML-estimation is default. To get the ML-estimator you need to use the `METHOD=ML`-option.

When the linear model doesn't seem to fit

Appearant misfis could be due to **time-varying covariates**.

- Model checking should be performed on the residuals.

Some times **transformation** helps.

- E.g. logatrithm if overall change tends to increase with increasing level.

Alternatives to the linear mixed model:

- Non-linear mixed model (difficult).
- Analysis of summary statistics (easy).

Nonlinearity

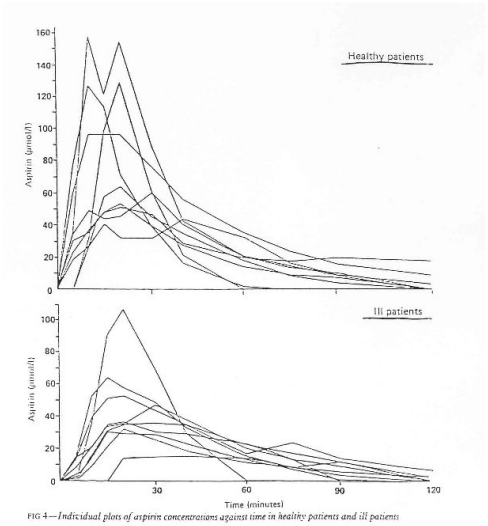


FIG 4—*Individual plots of aspirin concentrations against time in healthy patients and ill patients*

Individual time profiles are not parallel.

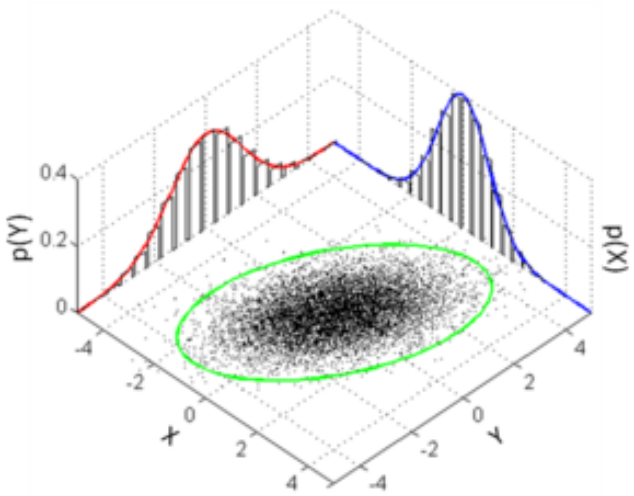
Could analyze:

- AUCs
- times to peak
- peak values
- ...

Or a suitable (which??) non-linear model.

Checking the multivariate normal assumption

... not easy unless $n = 2$.



Source: Wikipedia.

Is normality really needed?

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

But: the linear models for repeated outcomes are robust.

- As long as the linear model for the mean is correct and the covariance is well specified.
- If sample size is not too small.
- If there aren't too many missing observations.
- If the distribution of the data is not too skew.

Highly skew data should always be **transformed**.

Outline

Analysis of response profiles

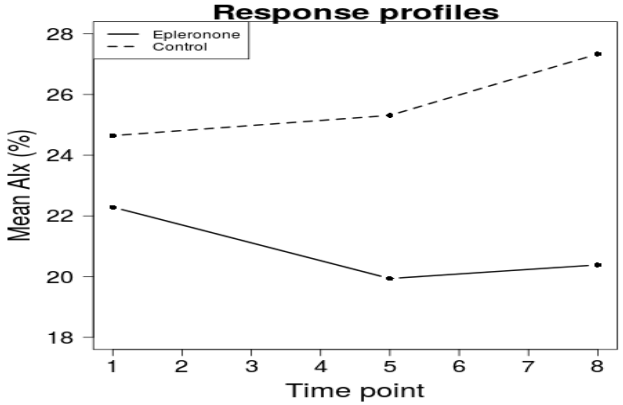
Comparison of change over n time points within g groups of subjects (e.g. different treatments).

- Similar to two-way ANOVA only with correlated data.
- **Covariates:** group and time
- **Balanced design**, but possibly incomplete data.
- An **unstructured covariance** is assumed.
- *Have the groups (treatments) been randomized?*
  - Then do baseline adjustment (later this day)!

Do the groups evolve differently with time?

Case study: Eplerenone

Individual curves are roughly parallel and few data are missing, so we look at averages over time.



Seeming improvement over time with Eplerenone.

Two-way ANOVA model for the means

	Treatment = Control	Treatment = Eplerenone
t=0	β_1	$\beta_1 + \beta_2$
t=12	$\beta_1 + \beta_3$	$\beta_1 + \beta_2 + \beta_3 + \beta_5$
t=24	$\beta_1 + \beta_4$	$\beta_1 + \beta_2 + \beta_4 + \beta_6$

- Standard treatment and baseline is reference (intercept)
- Time effect with standard treatment
- Difference between groups at baseline
  - In fact we know $\beta_2 = 0$ so we ought to do baseline correction.
- Interaction (i.e. difference in time effect)

Eplerenone: Parameter estimates

Effect	time	treat	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept			24.3431	2.0778	49.4	11.72	<.0001
treat		new	-2.0547	2.8988	48.9	-0.71	0.4818
treat		old	0	.	.	.	.
time	after12w		1.0887	1.7658	46.2	0.62	0.5406
time	after24w		3.0895	1.4960	44.5	2.07	0.0448
time	baseline		0	.	.	.	.
time*treat	after12w	new	-1.9493	2.4822	45.8	-0.79	0.4363
time*treat	after12w	old	0	.	.	.	.
time*treat	after24w	new	-3.6078	2.1231	45.3	-1.70	0.0961
time*treat	after24w	old	0	.	.	.	.
time*treat	baseline	new	0	.	.	.	.
time*treat	baseline	old	0	.	.	.	.

Estimated R Matrix for id 1

Row	Col1	Col2	Col3
1	106.23	96.3802	80.1893
2	96.3802	159.64	106.48
3	80.1893	106.48	106.38

Estimated R Correlation Matrix for id 1

Row	Col1	Col2	Col3
1	1.0000	0.7401	0.7544
2	0.7401	1.0000	0.8171
3	0.7544	0.8171	1.0000

Hypothesis testing I: Omnibus test

H_0: No `group*time`-interaction.

- i.e. mean changes over time are identical in all groups.

```
* MODEL aix = treat time treat*time ;
```

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
treat	1	47	1.84	0.1815
time	2	45.3	1.02	0.3690
time*treat	2	45.3	1.48	0.2394

- No overall significant difference in mean changes over time between the two treatments.

Hypothesis testing I: Post hoc tests

Suppose we had found a significant `group*time`-interaction.

```
* MODEL aix = treat*time / NOINT;
* LSMEANS treat*time / DIFF SLICE=time;
```

Differences of Least Squares Means

Effect	time	treat	_time	_treat	Estimate	Standard Error	DF	t Value	Pr >
time*treat	after12w	new	after12w	old	-4.0040	3.5885	45.5	-1.12	0.2
time*treat	after12w	new	after24w	new	-0.3423	1.5221	46.8	-0.22	0.8
time*treat	after12w	new	after24w	old	-6.0048	3.2715	54.8	-1.84	0.0
time*treat	after12w	new	baseline	new	-0.8606	1.7445	45.4	-0.49	0.6
time*treat	after12w	new	baseline	old	-2.9153	3.2694	62	-0.89	0.3
time*treat	after12w	old	after24w	new	3.6617	3.2952	55.4	1.11	0.2
time*treat	after12w	old	after24w	old	-2.0008	1.4854	44.9	-1.35	0.1
time*treat	after12w	old	baseline	new	3.1434	3.2545	60.1	0.97	0.3
time*treat	after12w	old	baseline	old	1.0887	1.7658	46.2	0.62	0.5
time*treat	after24w	new	after24w	old	-5.6625	2.9468	46	-1.92	0.0
time*treat	after24w	new	baseline	new	-0.5183	1.5065	46	-0.34	0.7
time*treat	after24w	new	baseline	old	-2.5730	2.9444	63.7	-0.87	0.3
time*treat	after24w	old	baseline	new	5.1442	2.9012	60.7	1.77	0.0
time*treat	after24w	old	baseline	old	3.0895	1.4960	44.5	2.07	0.0
time*treat	baseline	new	baseline	old	-2.0547	2.8988	48.9	-0.71	0.4

Tests of Effect Slices

Effect	time	Num DF	Den DF	F Value	Pr > F
time*treat	after12w	1	45.5	1.24	0.2704
time*treat	after24w	1	46	3.69	0.0609
time*treat	baseline	1	48.9	0.50	0.4818

Hypothesis testing II: treatment contrast

H_0: "No difference for a specific treatment contrast", e.g.

- Change at last follow-up is the same in all groups.
- Average change over time is the same in both groups.

```
* MODEL aix = treat*time / NOINT;
* ESTIMATE 'diff in ch last follow-up' treat*time 1 0 -1 -1 0 1;
* ESTIMATE 'diff in average change'    treat*time 1 -0.5 -0.5 -1 0.5 0.5;
```

Label	Estimate	Standard Error	DF	t Value	Pr > |t|
diff in ch last follow-up	3.6078	2.1231	45.3	1.70	0.0961
diff in average change	2.7786	2.0503	45.4	1.36	0.1821

- Seemingly improvement after eplerenone therapy, but non-significant difference at final evaluation.
- But we ought to make baseline adjustment.

Outline

Syntax: Analysis of response profiles

```
PROC MIXED DATA=kidney PLOTS=all;
  CLASS id time treat timepoint;
  MODEL aix = treat time treat*time / SOLUTION CL DDFM=SATTERTHW
  REPEATED timepoint / subject=id TYPE=UN R RCORR;
RUN;
```

- Syntax is similar to PROC GLM with a MODEL-statement specifying the (linear) relationship between outcome and covariates.
- Cathegorical variable must be declared with CLASS.
- The model for the covariance (UN=ustructured) is specified in a separate REPEATED-statement.
- Diagnostic plots with PLOTS-option.

Why do we need two version of the `time`-variable?

Only one version is needed, but I used ...

`time` (cathegorical variable)

- Labeled *baseline*, *after12w*, *after24w*.
- Baseline is default reference being last in alphabetic order.

for labeling the estimates for the mean.

`timepoint` (numerical variable)

- Labeled *1*, *2*, *3*.
- Point 3 is default reference being last in numerical order.

that matches the labels of the estimates for the covariance.

Simple alternative (in more recent versions of SAS):

```
CLASS id treat timepoint (REF=1);
```

The option `DDFM=SATTERTHWAITE`

(or `DDFM=KENWARDROGERS`).

A technical option intended to improve the statistical performance of the F-tests.

- It has no effect on balanced data.
- In unbalanced situations (i.e for almost all observational designs and in case of missing observations) degrees of freedom are computed by a more complicated formulae.
- The computations may require a little more time, but in most cases this will not be noticable.

When in doubt, use it!

Alternative syntax: Treatment contrasts

```
PROC MIXED DATA=kidney PLOTS=all;
  CLASS id time treat timepoint;
  MODEL aix = treat*timepoint / NOINT DDFM=SATTERTH;
  ESTIMATE 'diff in ch.last follow-up'
           treat*timepoint 1 0 -1 -1 0 1;
  ESTIMATE 'diff in average change'
           treat*timepoint 1 -0.5 -0.5 -1 0.5 0.5;
  REPEATED timepoint / subject=id type=un R RCORR;
RUN; QUIT;
```

Or simply:

```
  LSMEANS treat*timepoint / DIFF SLICE=timepoint;
```

SAS: proc mixed output

```
The Mixed Procedure

                Model Information

Data Set                     WORK.KIDNEY
Dependent Variable           aix
Covariance Structure         Unstructured
Subject Effect               id
Estimation Method            REML
Residual Variance Method     None
Fixed Effects SE Method      Model-Based
Degrees of Freedom Method    Satterthwaite


             Class Level Information

Class        Levels    Values

id               51    1 2 3 4 5 6 7 8 9 10 11 12 13
                       14 15 16 17 18 19 20 21 22 23
                       24 25 26 28 29 30 31 32 33 34
                       35 36 37 38 39 40 41 42 43 45
                       46 47 48 49 51 52 53 54
time              3    after12w after24w baseline
treat             2    new old
timepoint         3    1 2 3
```


SAS: proc mixed output

```
           Dimensions

Covariance Parameters            6
Columns in X                    12
Columns in Z                     0
Subjects                        51
Max Obs Per Subject              3

         Number of Observations

Number of Observations Read           153
Number of Observations Used           144
Number of Observations Not Used         9


                  Iteration History

Iteration    Evaluations    -2 Res Log Like       Criterion

        0              1      1070.85454941
        1              2       982.86560047      0.00144735
        2              1       982.26253864      0.00009905
        3              1       982.22468047      0.00000061
        4              1       982.22445749      0.00000000

               Convergence criteria met.
```

Always check that the numerical optimisation has converged.

SAS: proc mixed output

Options R and RCORR asks that estimated residual covariance and correlation matrices be printed.

```
The Mixed Procedure

      Estimated R Matrix for id 1

 Row        Col1        Col2        Col3

   1      106.23     96.3802     80.1893
   2     96.3802      159.64      106.48
   3     80.1893      106.48      106.38


 Estimated R Correlation Matrix for id 1

 Row        Col1        Col2        Col3

   1      1.0000      0.7401      0.7544
   2      0.7401      1.0000      0.8171
   3      0.7544      0.8171      1.0000
```

SAS: proc mixed output

Fit Statistics

-2 Res Log Likelihood	982.2
AIC (smaller is better)	994.2
AICC (smaller is better)	994.9
BIC (smaller is better)	1005.8

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
5	88.63	<.0001

Used for comparison of different models.

- `METHOD=REML` is default.
- But `METHOD=ML` should be used if you want to test a nested submodel for the mean-structure by the likelihood ratio test.

BUT: Often a `CONTRAST`-statement could be used instead.

SAS: proc mixed output

At last: Parameter estimates and tests.

Solution for Fixed Effects

Effect	time	treat	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept			24.3431	2.0778	49.4	11.72	<.0001
treat		new	-2.0547	2.8988	48.9	-0.71	0.4818
treat		old	0	.	.	.	.
time	after12w		1.0887	1.7658	46.2	0.62	0.5406
time	after24w		3.0895	1.4960	44.5	2.07	0.0448
time	baseline		0	.	.	.	.
time*treat	after12w	new	-1.9493	2.4822	45.8	-0.79	0.4363
time*treat	after12w	old	0	.	.	.	.
time*treat	after24w	new	-3.6078	2.1231	45.3	-1.70	0.0961
time*treat	after24w	old	0	.	.	.	.
time*treat	baseline	new	0	.	.	.	.
time*treat	baseline	old	0	.	.	.	.

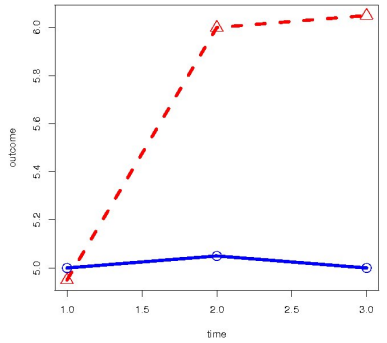
Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
treat	1	47	1.84	0.1815
time	2	45.3	1.02	0.3690
time*treat	2	45.3	1.48	0.2394

Outline

Baseline measurements

In randomized clinical trials, the first measurement is often a baseline measurement.



- The group means **must** be equal at baseline.

Hypothetical comparison of two treatment groups

What happens if we ignore the baseline problem?

- The non-existing difference at baseline makes the overall treatment effect appear smaller. Thus, the power of the test is reduced.

So should we leave out the baseline measurement?

- We loose information about change over time and again the power of the test of treatment effect is reduced.

Classical approaches for handling baseline

Vickers & Altman, *Analysing controlled clinical trials with baseline follow-up measurements*, BMJ **323**, 1123–1124.

Three possibilities: **1. End point, 2. Change, 3. ANCOVA**

1. Discard baseline, ok if correlation is small
2. Subtract baseline, ok if correlation is large
3. Condition on baseline, using it as covariate, always ok.

Conclusion: ANCOVA is most efficient.

Why ANCOVA is superior

For simplicity assume treatment ($x = 1$) vs placebo ($x = 0$) with only one time of follow-up ($t_1 = 0,\ t_2 = 1$),

$$Y_{ij} = \beta_1 + \beta_2 \cdot t_j + \beta_3 \cdot x \cdot t_j + \varepsilon_{ij}$$

where

$$\begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma^2\rho \\ \sigma^2\rho & \sigma^2 \end{pmatrix} \right)$$

Implied residual variances for the three models.

1. $\mathrm{Var}(Y_2) = \sigma^2$
2. $\mathrm{Var}(Y_2 - Y_1) = 2\sigma^2(1-\rho)$
3. $\mathrm{Var}(Y_2 | Y_1) = \sigma^2(1-\rho^2)$

Note: The assumption that the variance is constant over time could be dropped.

ANCOVA with multiple times of follow-up

Different effects of baseline at different time points due to stronger correlation between baseline and early follow up.

- The model should include a `baseline*time` interaction.

Example:

```
PROC MIXED DATA=kidney0;
  CLASS id time treat timepoint;
  MODEL aix = aix0*time treat*time / NOINT DDFM=SATTERTHWAITE;
  LSMEANS treat*time / DIFF SLICE=time;
  REPEATED timepoint / SUBJECT=id TYPE=UN R RCORR;
RUN;
```

Eplerenone: ANCOVA

Differences of Least Squares Means

Effect	time	treat	_time	_treat	Estimate	Standard Error	DF	t Value	Pr > |t|
time*treat	after12w	new	after12w	old	-2.3494	2.5643	44	-0.92	0.3646
time*treat	after12w	new	after24w	new	-0.2999	1.5388	44.8	-0.19	0.8464
time*treat	after12w	new	after24w	old	-4.5328	2.3048	63.6	-1.97	0.0536
time*treat	after12w	old	after24w	new	2.0495	2.3410	64.8	0.88	0.3845
time*treat	after12w	old	after24w	old	-2.1834	1.5312	43.1	-1.43	0.1611
time*treat	after24w	new	after24w	old	-4.2329	2.0762	43.7	-2.04	0.0475

Tests of Effect Slices

Effect	time	Num DF	Den DF	F Value	Pr > F
time*treat	after12we	1	44	0.84	0.3646
time*treat	after24we	1	43.7	4.16	0.0475

Conclusion: Significant difference at last follow-up. Estimated difference in change over time -4.23% (95% CI: -8.42% to -0.05%, P=0.0475) in favor of Eplerenone. So hopefully the protocol dictated ANCOVA with this particular parameter of interest.

Eplerenone: ANCOVA

Note: Covariance estimates change substantially when baseline is included as a covariate (it explains a lot of variation in the data).

The Mixed Procedure

Estimated R Matrix for id 1

Row	Col1	Col2
1	74.1578	34.2828
2	34.2828	47.2484

Estimated R Correlation Matrix for id 1

Row	Col1	Col2
1	1.0000	0.5792
2	0.5792	1.0000

A fourth option

Constrained linear mixed model (cLMM):

- Analysis of response profiles.
- include baseline as a response, but redefine covariates so that **identical group means at baseline** are modeled.

Liu et al: *Should baseline be a covariate or dependent variable in analyses of change from baseline in clinical trials?*, Statist. Med. **28**, 2509–2530, (2009):

Conclusions:

- similar power to ANCOVA (with no missing data).
- by default handles missing data (MAR) optimally.

But: ANCOVA has a computational advantage, so . . .

cLMM parametrisation

	Treatment = Standard	Treatment = Eplerenone
t=0	β_1	$\beta_1 + 0$
t=12	$\beta_1 + \beta_2$	$\beta_1 + 0 + \beta_2 + \beta_4$
t=24	$\beta_1 + \beta_3$	$\beta_1 + 0 + \beta_3 + \beta_5$

- Intercept.
- Time effect with standard treatment
- difference between groups at baseline = **0!**
- Interactions (differences in time-effects)

Eplerenone: cLMM

```
DATA kidney;
SET kidney;
timenew='zero0';
IF week EQ 5 AND treat EQ 'new' THEN timenew ='week12';
IF week EQ 8 AND treat EQ 'new' THEN timenew='week24';
RUN;

PROC MIXED DATA=kidney;
  CLASS id time timenew timepoint;
  MODEL aix = time timenew / SOLUTION CL DDFM=SATTERTHWAITE;
  REPEATED timepoint / SUBJECT=id TYPE=UN R RCORR;
RUN;
```

Eplerenone: cLMM output

Solution for Fixed Effects

Effect	time	timenew	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept			23.2879	1.4425	50	16.14	<.0001
time	after12we		1.2017	1.7575	46.8	0.68	0.4975
time	after24we		3.3608	1.4450	48.3	2.33	0.0243
time	baseline		0	.	.	.	.
timenew		week12	-2.1552	2.4637	46	-0.87	0.3862
timenew		week24	-4.1247	1.9931	45.9	-2.07	0.0442
timenew		zero0	0	.	.	.	.

Conclusion: Significant difference at last follow-up. Estimated difference in change over time -4.12% (95% CI: -8.14% to -0.11%, P=0.0442) in favor of Eplerenone.

Eplerenone: cLMM output

Note: Covariance estimates hardly change when the constraint is put on the mean parameters at baseline.

Estimated R Matrix for id 1

Row	Col1	Col2	Col3
1	105.29	95.5196	79.4628
2	95.5196	158.77	105.76
3	79.4628	105.76	105.77

Estimated R Correlation Matrix for id 1

Row	Col1	Col2	Col3
1	1.0000	0.7388	0.7530
2	0.7388	1.0000	0.8161
3	0.7530	0.8161	1.0000

ANCOVA and cLMM interpretation

ANCOVA and cLMMs estimate the treatment effect with similar accuracy.

- P-values and parameter estimates are very similar
- but not identical.

The models have different interpretations.

- cLMM describes the joint distribution of the response over time including baseline.
- ANCOVA describes the response at follow-up conditional on the baseline response.

cLMM implies ANCOVA

Assume a cLMM for $(Y_{i1}, \ldots, Y_{in})$. That is, means described as

$$E(Y_{i1}) = \beta_1, \quad E(Y_{ij}) = \beta_1 + \beta_j + X_i\beta_{n+j-1}$$

for $j = 2, \ldots, n$, and with X denoting the indicator of treatment. A similar **unstructured covariance** Σ is assumed for **both groups**.

Then the conditional distribution of $(Y_{i2}, \ldots, Y_{in})$ given Y_{i1} is again a normal distribution with

$$E(Y_{ij}|Y_{i1}) = \beta_1 + \beta_j + X_i\beta_{n+j-1} + \sigma_{ij}/\sigma_1^2(Y_{i1} - \beta_1)$$

$$\mathrm{Cov}(Y_{ij}, Y_{ik}|Y_{i1}) = \sigma_{jk} - \frac{\sigma_{1k}\sigma_{1j}}{\sigma_1^2}$$

Hence the conditional model is the ANCOVA.

- **Note:** the treatment effects are the same in both models.

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 intial 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.**

Outline

The unstructured covariance

Advantages

- We make no wrong assumptions about the covariance of our observations.
- We gain insight in the actual structure of the covariance.

Drawbacks

- We use quite a lot of parameters to describe the covariance structure. Thus our analysis becomes less efficient.
- No good with small data sets/many time points; The results may be unstable.
- It can only be used in case of balanced design, i.e. all subjects have to be measured at identical times.

Choosing a model for the covariance

Explorative data analysis suggested by FLW (2011):

1. Put up a plausible (e.g. saturated) model for the mean
2. Fit the data so far ignoring correlation (GLM).
3. Check the residuals for assessing the adequacy of the model for the mean and in order to get an impression of the error covariance.
4. Pick a reasonable model for the covariance (if possible test against the unstructured model).
5. Re-check the model fit.
6. Compute estimates, confidence intervals, and p-values.

Example: Calcium supplements

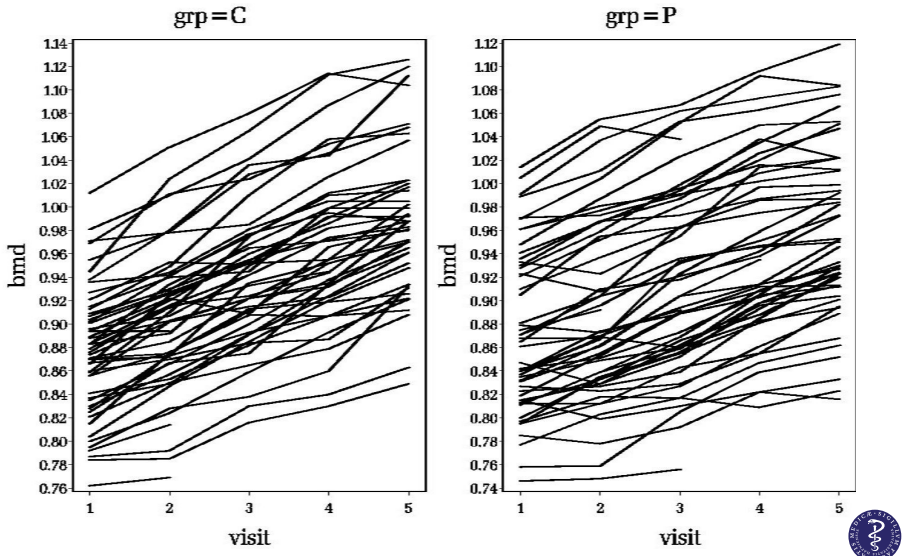
A total of 112 11-year old girls were randomized to receive either calcium or placebo.

Outcome: BMD=bone mineral density, in $\frac{g}{cm^2}$

Follow-up: every 6 months, 5 visits in total including baseline

Does calcium improve the bone gain for adolescent women?

Individual profiles



Analysis of response profiles

At first do not assume any specific structure
neither for the mean nor for the covariances.

- Use the saturated model as reference point.

Note: We ought to do baseline adjustment.

```
PROC MIXED DATA=calcium;
CLASS grp girl visit;
MODEL bmd=grp visit grp*visit / DDFM=SATTERTHWAITE;
REPEATED visit / TYPE=UN SUBJECT=girl(grp) R RCORR;
RUN;
```

Output

Estimated R Correlation Matrix for girl(grp) 101 C

Row	Col1	Col2	Col3	Col4	Col5
1	1.0000	0.9699	0.9414	0.9250	0.8987
2	0.9699	1.0000	0.9727	0.9585	0.9399
3	0.9414	0.9727	1.0000	0.9809	0.9592
4	0.9250	0.9585	0.9809	1.0000	0.9755
5	0.8987	0.9399	0.9592	0.9755	1.0000

Fit Statistics

-2 Res Log Likelihood	-2346.3 <----------used later

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
grp	1	109	2.55	0.1129
visit	4	97.1	258.08	<.0001
grp*visit	4	97.1	2.79	0.0303

Stationary covariance patterns

Most available models for are for equidistant observations,
assuming both variances and correlations are stationary, i.e.

- The variances are **all the same**
- Correlation **depend only on the time-distance** between the observations.

proc mixed type=	$\mathrm{Cov}(Y_{ij}, Y_{ik})$	no. par
CS	$\sigma^2[I\{j=k\} + \rho \cdot I\{j \neq k\}]$	2
AR(1)	$\sigma^2 \rho^{\lvert k-j \rvert}$	2
ARMA(1,1)	$\sigma^2[I\{j=k\} + \gamma \cdot \rho^{\lvert k-j \rvert - 1} I\{j \neq k\}]$	3
TOEP	$\sigma^2[I\{j=k\} + \rho_{\lvert k-j \rvert} \cdot I\{j \neq k\}]$	n

aka the *compound symmetry*, *autoregressive*, *autoregressive moving average*, and the *Toeplitz* models.

Heterogeneous covariance patterns

The assumption that the variances are stationary can be dropped
in which case we have a heterogeneous model for the variances.

- **No restrictions** on the variances
- Correlation **depend only on the time-distance** between the observations.

proc mixed type=	$\mathrm{Cov}(Y_{ij}, Y_{ik})$	no. par
CSH	$\sigma_j \sigma_k [I\{j=k\} + \rho \cdot I\{j \neq k\}]$	$n+1$
ARH(1)	$\sigma_j \sigma_k \rho^{\lvert k-j \rvert}$	$n+1$
TOEPH	$\sigma_j \sigma_k [I\{j=k\} + \rho_{\lvert k-j \rvert} \cdot I\{j \neq k\}]$	$2n-1$
ANTE(1)	$\sigma_j \sigma_k \prod_{l=j}^{k-1} \rho_l$	$2n-1$

aka the *heterogeneous compound symmetry*, *heterogeneous autoregressive*, the *heterogeneous Toeplitz*, and the *antedependence* covariance sturctures.

Autoregressive covariance structure in SAS

Fit the calcium data with:

```
PROC MIXED DATA=calcium;
CLASS grp girl visit;
MODEL bmd=grp visit grp*visit / DDFM=SATTERTHWAITE;
REPEATED visit / TYPE=AR(1) SUBJECT=girl(grp) R RCORR;
RUN;
```

Note: Similar syntax is valid for the other types of covariance patterns in the table above.

Output from TYPE=AR(1) structure

Estimated R Correlation Matrix for girl(grp) 101 C

Row	Col1	Col2	Col3	Col4	Col5
1	1.0000	0.9708	0.9425	0.9150	0.8883
2	0.9708	1.0000	0.9708	0.9425	0.9150
3	0.9425	0.9708	1.0000	0.9708	0.9425
4	0.9150	0.9425	0.9708	1.0000	0.9708
5	0.8883	0.9150	0.9425	0.9708	1.0000

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
AR(1)	girl(grp)	0.9708
Residual		0.004412

Fit Statistics

-2 Res Log Likelihood -2318.6 <----------used later

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
grp	1	113	2.74	0.1005
visit	4	382	233.91	<.0001
grp*visit	4	382	2.86	0.0232

Comparison of covariance structures

Use the likelihood (either of REML or ML):

- Better fitting models have large values of likelihood L and therefore small values of deviance: $-2\log L$
- Compute differences in deviances ($\Delta = -2\log Q$) and compare to a χ^2-distribution with df$=\Delta$no. params.

Note: Only *nested* models can be compared.

- We can use the unstructured covariance as reference point since it contains all other models as submodels.

Example: AR vs UN

$$\begin{aligned} -2\log Q &= 2346.3 - 2318.6 = 27.7 \\ &\sim \chi^2(15-2) = \chi^2(13) \Rightarrow P = 0.01 \end{aligned}$$

Comparison of covariance structures

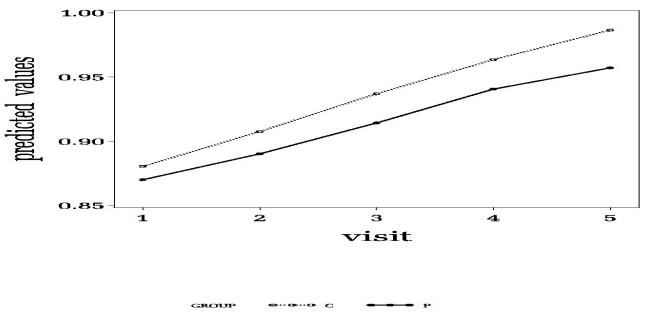
Model	-2 log L	par.	$-2\log Q$	*df*	P
UN	-2346.3	15			
ARMA(1,1)	-2318.6	3	27.7	12	0.006
AR(1)	-2318.6	2	27.7	13	0.010
CS	-2188.8	2	129.8	13	< 0.0001

Better stick to the unstructured covariance.

(Or try ARH(1) since the variances seem to increase with time).

UNIVERSITY OF COPENHAGEN DEPARTMENT OF BIOSTATISTICS

Predicted mean time profiles



Note: Estimated profiles are almost identical for all choices of covariance structures. In fact, for balanced data, they agree completely (since they are equal to the `group*time`-averages).

Tests of treatment effect

BUT: We cannot make inference from profiles alone

Tests of the interaction term `group*visit`:.

Covariance structure	Test statistic $\sim$ distribution	P value*
Independence	$0.35 \sim F(4,491)$	0.84
Compound symmetry	$5.30 \sim F(4,382)$	0.0004
Autoregressive	$2.86 \sim F(4,382)$	0.023
Unstructured	$2.72 \sim F(4,107)$	0.034

- Confidence intervals and tests depend on the covariance.