Longitudinal observations

Bendix Carstensen

Steno Diabetes Center,
Gentofte, Denmark
& Department of Biostatistics,
University of Copenhagen
bxc@steno.dk
http://BendixCarstensen.com

LEAD symposium, EDEG 2014
31 March 2014

http://BendixCarstensen.com/SDC/LEAD

Two observation points

Bendix Carstensen

LEAD
31 March 2014
LEAD symposium, EDEG 2014
`http://BendixCarstensen.com/SDC/LEAD`

(`twopoints`)

Basic set-up: Two time points

Measurements at two time points

- Randomized study:

Basic set-up: Two time points

Measurements at two time points

- Randomized study:
  - Effect of randomization

Basic set-up: Two time points

Measurements at two time points

- Randomized study:
  - Effect of randomization
  - 1st point special (**pre**-intervention)

Basic set-up: Two time points

Measurements at two time points

- Randomized study:
  - Effect of randomization
  - 1st point special (**pre**-intervention)
- Observational study

Basic set-up: Two time points

Measurements at two time points

- Randomized study:
  - Effect of randomization
  - 1st point special (**pre**-intervention)
- Observational study
  - Describe population processes

Basic set-up: Two time points

Measurements at two time points

- Randomized study:
  - Effect of randomization
  - 1st point special (**pre**-intervention)
- Observational study
  - Describe population processes
  - Nothing special about any one point of observation

Basic set-up: Two time points

Measurements at two time points

- Randomized study:
  - Effect of randomization
  - 1st point special (**pre**-intervention)
- Observational study
  - Describe population processes
  - Nothing special about any one point of observation
  - — except that this was the first measuring occasion.

Two timepoints in randomized study

- Measurements at baseline and follow-up.

Two timepoints in randomized study

- Measurements at baseline and follow-up.
- Two randomized groups

Two timepoints in randomized study

- Measurements at baseline and follow-up.
- Two randomized groups
- Target:

Two timepoints in randomized study

- Measurements at baseline and follow-up.
- Two randomized groups
- Target:
  - What is the change in **each** of the groups,

Two timepoints in randomized study

- Measurements at baseline and follow-up.
- Two randomized groups
- Target:
  - What is the change in **each** of the groups,
  - What is the **difference** in the changes

Two timepoints in randomized study

- Measurements at baseline and follow-up.
- Two randomized groups
- Target:
  - What is the change in **each** of the groups,
  - What is the **difference** in the changes
  - — that is, the intervention effct

Two timepoints in randomized study

- Measurements at baseline and follow-up.
- Two randomized groups
- Target:
  - What is the change in **each** of the groups,
  - What is the **difference** in the changes
  - — that is, the intervention effct
- Thus we know:

Two timepoints in randomized study

- Measurements at baseline and follow-up.
- Two randomized groups
- Target:
  - What is the change in **each** of the groups,
  - What is the **difference** in the changes
  - — that is, the intervention effct
- Thus we know:
  - No difference at baseline (randomization)

Two timepoints in randomized study

- Measurements at baseline and follow-up.
- Two randomized groups
- Target:
  - What is the change in **each** of the groups,
  - What is the **difference** in the changes
  - — that is, the intervention effct
- Thus we know:
  - No difference at baseline (randomization)
  - ny difference at follow-up due to intervention.

Simple approaches

- Compute the change in each group

Simple approaches

- Compute the change in each group
- Compute the differences between changes in the two groups

Simple approaches

- Compute the change in each group
- Compute the differences between changes in the two groups
- — this is the intervention effect

Simple approaches

- Compute the change in each group
- Compute the differences between changes in the two groups
- — this is the intervention effect
- Not quite so: Regression to the mean

Regression to the mean

- The follow up of an exceptional film is rarely as good as the first...

Regression to the mean

- The follow up of an exceptional film is rarely as good as the first...
- Children of tall parents smaller than parents

Regression to the mean

- The follow up of an exceptional film is rarely as good as the first...
- Children of tall parents smaller than parents
- Children of small parents taller than parents

Regression to the mean

- The follow up of an exceptional film is rarely as good as the first...
- Children of tall parents smaller than parents
- Children of small parents taller than parents
- — comes from the make up of measurements:

$$Y_i = \mu_i + e_i$$

Regression to the mean

- The follow up of an exceptional film is rarely as good as the first...
- Children of tall parents smaller than parents
- Children of small parents taller than parents
- — comes from the make up of measurements:

$$Y_i = \mu_i + e_i$$

- The **observed** Y_i is large if μ_i **or** e_i is large

Regression to the mean

- The follow up of an exceptional film is rarely as good as the first...
- Children of tall parents smaller than parents
- Children of small parents taller than parents
- — comes from the make up of measurements:

$$Y_i = \mu_i + e_i$$

- The **observed** Y_i is large if μ_i **or** e_i is large
- Offspring (film no. II) has same μ_i but random e_i!

Regression to the mean

- The follow up of an exceptional film is rarely as good as the first...
- Children of tall parents smaller than parents
- Children of small parents taller than parents
- — comes from the make up of measurements:

$$Y_i = \mu_i + e_i$$

- The **observed** Y_i is large if μ_i **or** e_i is large
- Offspring (film no. II) has same μ_i but random e_i!

Regression to the mean

$$Y_{it} = \mu_i + e_{it}, \quad t = 1, 2$$

Regression to the mean

$$Y_{it} = \mu_i + e_{it}, \quad t = 1, 2$$

- Large measurements at first timepoints Y_{i1} comes around because e_{i1} is large.

Regression to the mean

$$Y_{it} = \mu_i + e_{it}, \quad t = 1, 2$$

- Large measurements at first timepoints Y_{i1} comes around because e_{i1} is large.
- next measurement is with a **random** e_{i2}

Regression to the mean

$$Y_{it} = \mu_i + e_{it}, \quad t = 1, 2$$

- Large measurements at first timepoints Y_{i1} comes around because e_{i1} is large.
- next measurement is with a **random** e_{i2}
- — hence with a random part which on average is smaller.

Regression to the mean

Intervention effect positive:

Regression to the mean

Intervention effect positive:

- Persons who start high likely to have smaller change, their chage is made up of:

Regression to the mean

Intervention effect positive:

- Persons who start high likely to have smaller change, their chage is made up of:
    - the “real” change

Regression to the mean

Intervention effect positive:

- Persons who start high likely to have smaller change, their chage is made up of:
  - the “real” change
  - the differences in random errors:

Regression to the mean

Intervention effect positive:

- Persons who start high likely to have smaller change, their chage is made up of:
  - the “real” change
  - the differences in random errors:
  - first large (high measurement)

Regression to the mean

Intervention effect positive:

- Persons who start high likely to have smaller change, their chage is made up of:
  - the “real” change
  - the differences in random errors:
  - first large (high measurement)
  - second “normal” (presumably smaller)

Regression to the mean

Intervention effect positive:

- Persons who start high likely to have smaller change, their chage is made up of:
  - the “real” change
  - the differences in random errors:
  - first large (high measurement)
  - second “normal” (presumably smaller)
- Persons who start low likely to have larger change

Regression to the mean

Intervention effect positive:

- Persons who start high likely to have smaller change, their chage is made up of:
  - the “real” change
  - the differences in random errors:
  - first large (high measurement)
  - second “normal” (presumably smaller)
- Persons who start low likely to have larger change
  - the “real” change

Regression to the mean

Intervention effect positive:

- Persons who start high likely to have smaller change, their chage is made up of:
  - the “real” change
  - the differences in random errors:
  - first large (high measurement)
  - second “normal” (presumably smaller)
- Persons who start low likely to have larger change
  - the “real” change
  - the differences in random errors:

Regression to the mean

Intervention effect positive:

- Persons who start high likely to have smaller change, their chage is made up of:
  - the "real" change
  - the differences in random errors:
  - first large (high measurement)
  - second "normal" (presumably smaller)
- Persons who start low likely to have larger change
  - the "real" change
  - the differences in random errors:
  - first small (low measurement)

Regression to the mean

Intervention effect positive:

- Persons who start high likely to have smaller change, their chage is made up of:
  - the "real" change
  - the differences in random errors:
  - first large (high measurement)
  - second "normal" (presumably smaller)
- Persons who start low likely to have larger change
  - the "real" change
  - the differences in random errors:
  - first small (low measurement)
  - second "normal" (presumably larger)

How data comes around

Measurement	mean	SD
B_i	μ	σ
F_i	$\mu + \Delta$	σ

How data comes around

Measurement	mean	SD
B_i	μ	σ
F_i	$\mu + \Delta$	σ

F_i & B_i are correlated…

How data comes around

Measurement	mean	SD
B_i	μ	σ
F_i	$\mu + \Delta$	σ

F_i & B_i are correlated...

The **conditional** mean of the difference given the first measurement:

$$\mathrm{E}(F_i - B_i | B_i = x) = \Delta - (x - \mu)(1 - \rho)$$

— ρ is the correlation between F and B.

How data comes around

Measurement	mean	SD
B_i	μ	σ
F_i	$\mu + \Delta$	σ

F_i & B_i are correlated...

The **conditional** mean of the difference given the first measurement:

$$\mathrm{E}(F_i - B_i | B_i = x) = \Delta - (x - \mu)(1 - \rho)$$

— ρ is the correlation between F and B.

So x large (*i.e.* $x > \mu$) means that the conditional mean is **smaller** than Δ - the **true** difference.

Where is the correlation?

Where is the correlation?

The **real** model:

$$y_{it} = \mu + \Delta_2 + a_i + e_{it}$$

Where is the correlation?

The **real** model:

$$y_{it} = \mu + \Delta_2 + a_i + e_{it}$$

with:

- μ — population mean

Where is the correlation?

The **real** model:

$$y_{it} = \mu + \Delta_2 + a_i + e_{it}$$

with:

- μ — population mean
- Δ_2 — change from time 1 to 2

Where is the correlation?

The **real** model:

$$y_{it} = \mu + \Delta_2 + a_i + e_{it}$$

with:

- μ — population mean
- Δ_2 — change from time 1 to 2
- a_i — person i's deviation from population mean:
  Person i has "true" (baseline) mean $\mu + a_i$

Where is the correlation?

The **real** model:

$$y_{it} = \mu + \Delta_2 + a_i + e_{it}$$

with:

- μ — population mean
- Δ_2 — change from time 1 to 2
- a_i — person i's deviation from population mean:
  Person i has "true" (baseline) mean $\mu + a_i$
- $a_i \sim \mathcal{N}, \quad \text{s.d.} = \tau$

Where is the correlation?

The **real** model:

$$y_{it} = \mu + \Delta_2 + a_i + e_{it}$$

with:

- μ — population mean
- Δ_2 — change from time 1 to 2
- a_i — person i's deviation from population mean:
  Person i has "true" (baseline) mean $\mu + a_i$
- $a_i \sim \mathcal{N}, \quad \text{s.d.} = \tau$
- $e_{it} \sim \mathcal{N}, \quad \text{s.d.} = \sigma$

Where is the correlation?

The **real** model:

$$y_{it} = \mu + \Delta_2 + a_i + e_{it}$$

with:

- μ — population mean
- Δ_2 — change from time 1 to 2
- a_i — person i's deviation from population mean:
  Person i has "true" (baseline) mean $\mu + a_i$
- $a_i \sim \mathcal{N}, \quad \text{s.d.} = \tau$
- $e_{it} \sim \mathcal{N}, \quad \text{s.d.} = \sigma$

Where is the correlation?

The **real** model:

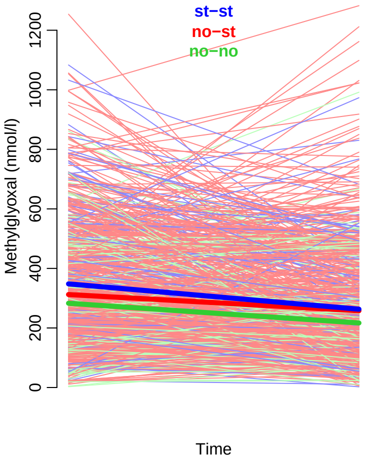
$$y_{it} = \mu + \Delta_2 + a_i + e_{it}$$

with:

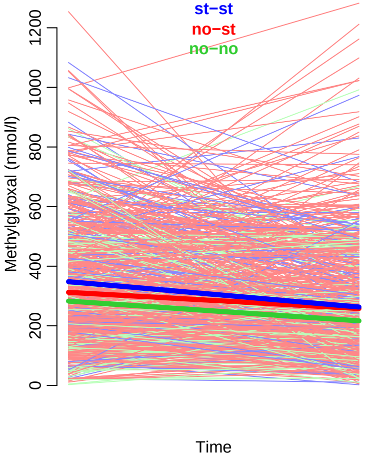
- μ — population mean
- Δ_2 — change from time 1 to 2
- a_i — person i's deviation from population mean:
  Person i has "true" (baseline) mean $\mu + a_i$
- $a_i \sim \mathcal{N}, \quad \text{s.d.} = \tau$
- $e_{it} \sim \mathcal{N}, \quad \text{s.d.} = \sigma$

$$\rho = \text{corr}(F, B) = \text{corr}(y_{t2}, y_{t1}) = \frac{\tau^2}{\tau^2 + \sigma^2}$$

Where is the correlation?

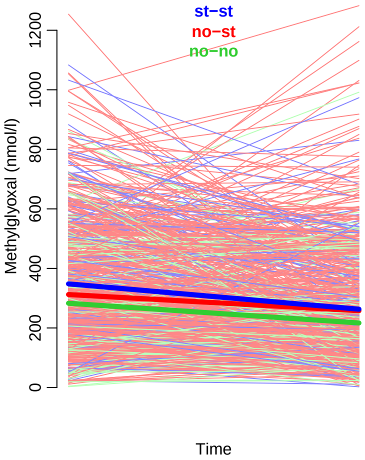


Where is the correlation?



τ is the variation between persons:
Variation between line-midpoints

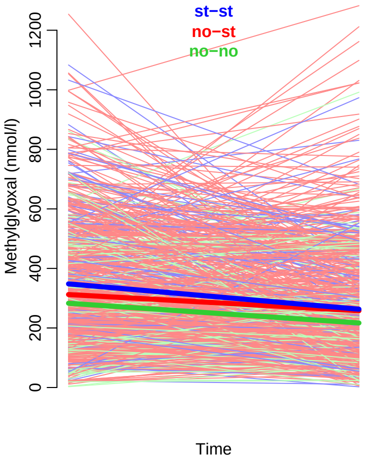
Where is the correlation?



τ is the variation between persons:
Variation between line-midpoints

Δ is the average slope of the lines

Where is the correlation?



τ is the variation between persons:
Variation between line-midpoints

Δ is the average slope of the lines

σ is the variation round these slopes

Two timepoints

- Measurements at baseline and follow-up.

Two timepoints

- Measurements at baseline and follow-up.
- Two randomized groups

Two timepoints

- Measurements at baseline and follow-up.
- Two randomized groups
- Target:

Two timepoints

- Measurements at baseline and follow-up.
- Two randomized groups
- Target:
  - What is the change in each of the groups,

Two timepoints

- Measurements at baseline and follow-up.
- Two randomized groups
- Target:
  - What is the change in each of the groups,
  - What is the difference in the changes

Two timepoints

- Measurements at baseline and follow-up.
- Two randomized groups
- Target:
  - What is the change in each of the groups,
  - What is the difference in the changes
  - — the intervention effct

Simple approach

- Compute the change in each group

Simple approach

- Compute the change in each group
- Compute the differences between groups

Simple approach

- Compute the change in each group
- Compute the differences between groups
- — this is the intervention effect

Simple approach

- Compute the change in each group
- Compute the differences between groups
- — this is the intervention effect
- No so: Regression to the mean

Regression to the mean

- The follow up of an exceptional film is rarely as good as the first...

Regression to the mean

- The follow up of an exceptional film is rarely as good as the first...
- Children of tall parents smaller than parents

Regression to the mean

- The follow up of an exceptional film is rarely as good as the first...
- Children of tall parents smaller than parents
- Children of small parents taller than parents

Regression to the mean

- The follow up of an exceptional film is rarely as good as the first...
- Children of tall parents smaller than parents
- Children of small parents taller than parents
- — comes from the make up of measurements:

$$Y_i = \mu_i + e_i$$

Regression to the mean

- The follow up of an exceptional film is rarely as good as the first...
- Children of tall parents smaller than parents
- Children of small parents taller than parents
- — comes from the make up of measurements:

$$Y_i = \mu_i + e_i$$

- Y_i is large if mu_i **or** e_i is large

Regression to the mean

- The follow up of an exceptional film is rarely as good as the first...
- Children of tall parents smaller than parents
- Children of small parents taller than parents
- — comes from the make up of measurements:

$$Y_i = \mu_i + e_i$$

- Y_i is large if mu_i **or** e_i is large
- Offspring (film no. II) has same μ_i but random e_i!

Regression to the mean

- The follow up of an exceptional film is rarely as good as the first...
- Children of tall parents smaller than parents
- Children of small parents taller than parents
- — comes from the make up of measurements:

$$Y_i = \mu_i + e_i$$

- Y_i is large if mu_i **or** e_i is large
- Offspring (film no. II) has same μ_i but random e_i!

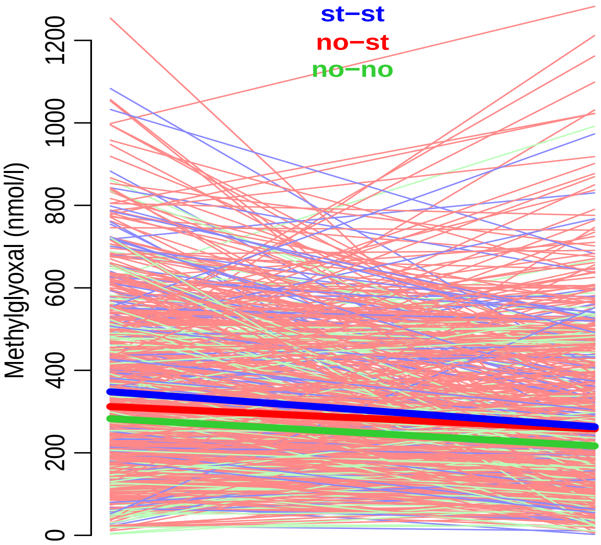
Regression to the mean

- The follow up of an exceptional film is rarely as good as the first...
- Children of tall parents smaller than parents
- Children of small parents taller than parents
- — comes from the make up of measurements:

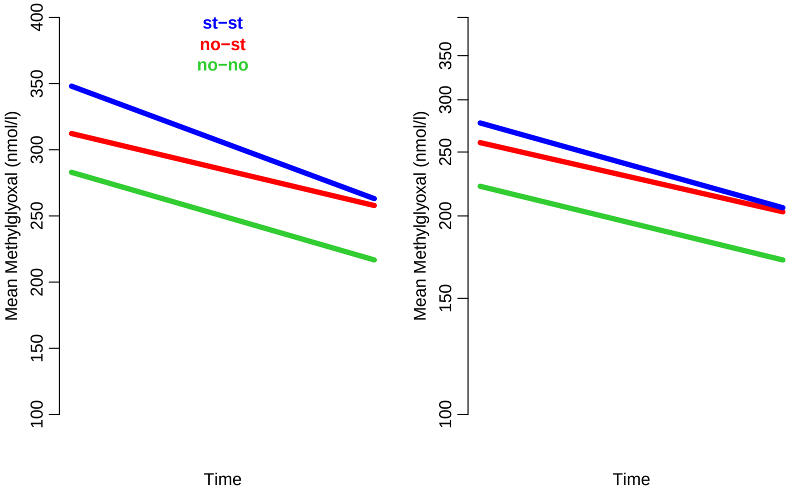
$$Y_i = \mu_i + e_i$$

- Y_i is large if mu_i **or** e_i is large
- Offspring (film no. II) has same μ_i but random e_i!

Methylglyoxal

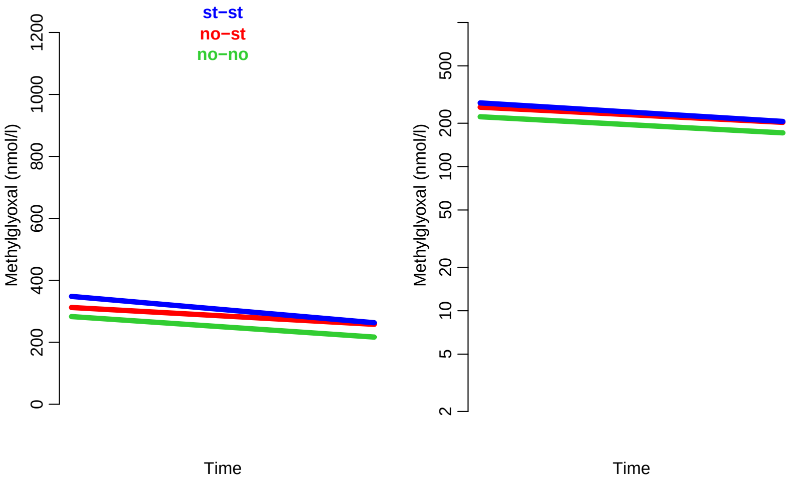


Methylglyoxal



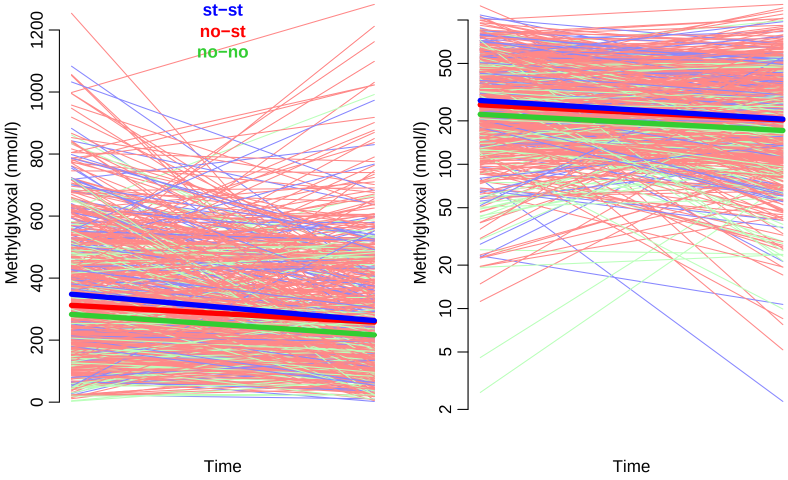
Source: Troels Mygind Jensen & Addition-PRO

Methylglyoxal



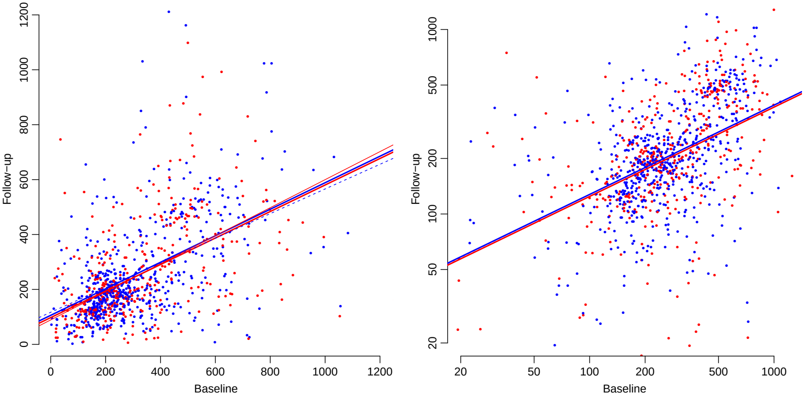
Source: Troels Mygind Jensen & Addition-PRO

Methylglyoxal



Source: Troels Mygind Jensen & Addition-PRO

Methylglyoxal



Source: Troels Mygind Jensen & Addition-PRO

Analysis by lm I

```
cf <- coef( m0 <- lm( log10(mf) ~ log10(mb) + factor(gr), data=
round( ci.lin( m0 ), 2 )
```

```
              Estimate StdErr      z    P  2.5% 97.5%
(Intercept)       1.14   0.07 15.50 0.00  0.99  1.28
log10(mb)         0.48   0.03 16.26 0.00  0.43  0.54
factor(gr)1      -0.01   0.02 -0.59 0.56 -0.05  0.03
```

Multiple measurements

Bendix Carstensen

LEAD
31 March 2014
LEAD symposium, EDEG 2014
`http://BendixCarstensen.com/SDC/LEAD`

(`multpt`)

More than two timepoints

- Identical time points:

More than two timepoints

- Identical time points:
  - Slightly simpler analysis:

More than two timepoints

- Identical time points:
  - Slightly simpler analysis:
  - time effects can be specified arbitrarily (not neccessarily sensible)

More than two timepoints

- Identical time points:
    - Slightly simpler analysis:
    - time effects can be specified arbitrarily (not neccessarily sensible)
    - resembles 2-way analysis of variance

More than two timepoints

- Identical time points:
  - Slightly simpler analysis:
  - time effects can be specified arbitrarily (not neccessarily sensible)
  - resembles 2-way analysis of variance
  - essentially fitting data(structure) to available methodology

More than two timepoints

- Identical time points:
  - Slightly simpler analysis:
  - time effects can be specified arbitrarily (not neccessarily sensible)
  - resembles 2-way analysis of variance
  - essentially fitting data(structure) to available methodology
- Time points different between persons:

More than two timepoints

- Identical time points:
  - Slightly simpler analysis:
  - time effects can be specified arbitrarily (not neccessarily sensible)
  - resembles 2-way analysis of variance
  - essentially fitting data(structure) to available methodology
- Time points different between persons:
  - time effects must be specified as functions of time

More than two timepoints

- Identical time points:
  - Slightly simpler analysis:
  - time effects can be specified arbitrarily (not neccessarily sensible)
  - resembles 2-way analysis of variance
  - essentially fitting data(structure) to available methodology
- Time points different between persons:
  - time effects must be specified as functions of time
  - — to be estimated. . .

More than two timepoints

- Identical time points:
  - Slightly simpler analysis:
  - time effects can be specified arbitrarily (not neccessarily sensible)
  - resembles 2-way analysis of variance
  - essentially fitting data(structure) to available methodology
- Time points different between persons:
  - time effects must be specified as functions of time
  - — to be estimated...
- Model data by random effects models for **mean** and **between person variation**

More than two timepoints

- Identical time points:
  - Slightly simpler analysis:
  - time effects can be specified arbitrarily (not neccessarily sensible)
  - resembles 2-way analysis of variance
  - essentially fitting data(structure) to available methodology
- Time points different between persons:
  - time effects must be specified as functions of time
  - — to be estimated...
- Model data by random effects models for **mean** and **between person variation**
- Limited amount of information per person.

Random effects — error structure

- Because of limited information per person, we model the distribution of person-level measuremnst by a normal distribution. (could be another type of dist'n)

Random effects — error structure

- Because of limited information per person, we model the distribution of person-level measuremnst by a normal distribution. (could be another type of dist'n)
- A single random person-effect is hardy ever sufficient with several time points

Random effects — error structure

- Because of limited information per person, we model the distribution of person-level measuremnst by a normal distribution. (could be another type of dist'n)
- A single random person-effect is hardy ever sufficient with several time points
- Random slopes, random higher-order terms can be added

Random effects — error structure

- Because of limited information per person, we model the distribution of person-level measuremnst by a normal distribution. (could be another type of dist'n)
- A single random person-effect is hardy ever sufficient with several time points
- Random slopes, random higher-order terms can be added
- Neither approach requires the same number of timepoints (let alone identical timepoints) between persons' measurements.

Random effects — error structure

- Because of limited information per person, we model the distribution of person-level measuremnst by a normal distribution. (could be another type of dist'n)
- A single random person-effect is hardy ever sufficient with several time points
- Random slopes, random higher-order terms can be added
- Neither approach requires the same number of timepoints (let alone identical timepoints) between persons' measurements.
- This is how the world usually looks.

Data structure: "long" format

- Always advisable to have data in the long form:

```
head( gluc )
     id  fpg    ds       time gruppe   end   tfe
1  4521 5.35 13895 -10.512011      0 17724 -3829
2  4521 5.30 15890  -5.035003      0 17724 -1834
3  4521 5.90 17724   0.000000      0 17724     0
4 10613 5.00 12116   0.000000      0 12116     0
5 11934 5.30 11849  -2.954015      0 11849     0
6 16753 5.06 13919  -8.312972      0 15865 -1946
```

Data structure: "long" format

- Always advisable to have data in the long form:

```
head( gluc )
     id  fpg    ds       time gruppe   end   tfe
1  4521 5.35 13895 -10.512011      0 17724 -3829
2  4521 5.30 15890  -5.035003      0 17724 -1834
3  4521 5.90 17724   0.000000      0 17724     0
4 10613 5.00 12116   0.000000      0 12116     0
5 11934 5.30 11849  -2.954015      0 11849     0
6 16753 5.06 13919  -8.312972      0 15865 -1946
```

  - each record in data represents **one** measurement

Data structure: "long" format

- Always advisable to have data in the long form:

```
head( gluc )
     id  fpg    ds       time gruppe   end   tfe
1  4521 5.35 13895 -10.512011      0 17724 -3829
2  4521 5.30 15890  -5.035003      0 17724 -1834
3  4521 5.90 17724   0.000000      0 17724     0
4 10613 5.00 12116   0.000000      0 12116     0
5 11934 5.30 11849  -2.954015      0 11849     0
6 16753 5.06 13919  -8.312972      0 15865 -1946
```

  - each record in data represents **one** measurement
  - and the corresponding covariate values

Data structure: “long” format

- Always advisable to have data in the long form:

```
head( gluc )
     id  fpg    ds       time gruppe   end   tfe
1  4521 5.35 13895 -10.512011      0 17724 -3829
2  4521 5.30 15890  -5.035003      0 17724 -1834
3  4521 5.90 17724   0.000000      0 17724     0
4 10613 5.00 12116   0.000000      0 12116     0
5 11934 5.30 11849  -2.954015      0 11849     0
6 16753 5.06 13919  -8.312972      0 15865 -1946
```

  - each record in data represents **one** measurement
  - and the corresponding covariate values
- Most programs use this format, and it imposes fewer restrictions on your data

Data structure: "long" format

- Always advisable to have data in the long form:

```
head( gluc )
     id  fpg    ds       time gruppe   end   tfe
1  4521 5.35 13895 -10.512011      0 17724 -3829
2  4521 5.30 15890  -5.035003      0 17724 -1834
3  4521 5.90 17724   0.000000      0 17724     0
4 10613 5.00 12116   0.000000      0 12116     0
5 11934 5.30 11849  -2.954015      0 11849     0
6 16753 5.06 13919  -8.312972      0 15865 -1946
```

  - each record in data represents **one** measurement
  - and the corresponding covariate values
- Most programs use this format, and it imposes fewer restrictions on your data
- A bad idea to taylor your data to fit a given computer representation,

Data structure: "long" format

- Always advisable to have data in the long form:

```
head( gluc )
     id  fpg    ds       time gruppe   end   tfe
1  4521 5.35 13895 -10.512011      0 17724 -3829
2  4521 5.30 15890  -5.035003      0 17724 -1834
3  4521 5.90 17724   0.000000      0 17724     0
4 10613 5.00 12116   0.000000      0 12116     0
5 11934 5.30 11849  -2.954015      0 11849     0
6 16753 5.06 13919  -8.312972      0 15865 -1946
```

  - each record in data represents **one** measurement
  - and the corresponding covariate values
- Most programs use this format, and it imposes fewer restrictions on your data
- A bad idea to taylor your data to fit a given computer representation,

Data structure: "long" format

- Always advisable to have data in the long form:

```
head( gluc )
     id  fpg    ds       time gruppe   end   tfe
1  4521 5.35 13895 -10.512011      0 17724 -3829
2  4521 5.30 15890  -5.035003      0 17724 -1834
3  4521 5.90 17724   0.000000      0 17724     0
4 10613 5.00 12116   0.000000      0 12116     0
5 11934 5.30 11849  -2.954015      0 11849     0
6 16753 5.06 13919  -8.312972      0 15865 -1946
```

  - each record in data represents **one** measurement
  - and the corresponding covariate values
- Most programs use this format, and it imposes fewer restrictions on your data
- A bad idea to taylor your data to fit a given computer representation, vice versa is better.

Simple model for repeated measures

Measurement on individual i at timepoint t

$$y_{ti} = \mu + [\mathsf{cov}] + a_i + e_{it}$$

Simple model for repeated measures

Measurement on individual i at timepoint t

$$y_{ti} = \mu + [\mathsf{cov}] + a_i + e_{it}$$

a_i is a random effect for person i: represents the (random) **deviation** of the person-mean from the population mean

Simple model for repeated measures

Measurement on individual i at timepoint t

$$y_{ti} = \mu + [\mathsf{cov}] + a_i + e_{it}$$

a_i is a random effect for person i: represents the (random) **deviation** of the person-mean from the population mean — that is the predicted population mean for persons with **similar** values of the covariates,

Simple model for repeated measures

Measurement on individual i at timepoint t

$$y_{ti} = \mu + [\text{cov}] + a_i + e_{it}$$

a_i is a random effect for person i: represents the (random) **deviation** of the person-mean from the population mean — that is the predicted population mean for persons with **similar** values of the covariates, $\mu + [\text{cov}]$

Simple model for repeated measures

Measurement on individual i at timepoint t

$$y_{ti} = \mu + [\text{cov}] + a_i + e_{it}$$

a_i is a random effect for person i: represents the (random) **deviation** of the person-mean from the population mean — that is the predicted population mean for persons with **similar** values of the covariates, $\mu + [\text{cov}]$

e_{it} is a random effect representing the measurement error on any measurement

Simple model for repeated measures

Measurement on individual i at timepoint t

$$y_{ti} = \mu + [\textsf{cov}] + a_i + e_{it}$$

The variation in a_i is the **between** person variation.

Simple model for repeated measures

Measurement on individual i at timepoint t

$$y_{ti} = \mu + [\textsf{cov}] + a_i + e_{it}$$

The variation in a_i is the **between** person variation.

Standard deviation of the a_is is τ, say;
you get an estimate of this from statistics programmes.

Interpretation of btw. person s.d.

- Select two persons at random with the same covariate values ([cov]).

Interpretation of btw. person $\mathrm{s.d.}$

- Select two persons at random with the same covariate values ([cov]).
- The $\mathrm{s.d.}$ of the difference of their measurements is $\sqrt{2}\tau$; the absolute difference follow a half-normal distribution with this $\mathrm{s.d.}$,

Interpretation of btw. person s.d.

- Select two persons at random with the same covariate values ([cov]).
- The s.d. of the difference of their measurements is $\sqrt{2}\tau$; the absolute difference follow a half-normal distribution with this s.d.,
- The median of this corresponds to the 75^{th} percentile of a normal with this scale, that is $0.953 \times \tau$.

Interpretation of btw. person s.d.

- Select two persons at random with the same covariate values ([cov]).
- The s.d. of the difference of their measurements is $\sqrt{2}\tau$; the absolute difference follow a half-normal distribution with this s.d.,
- The median of this corresponds to the 75^{th} percentile of a normal with this scale, that is $0.953 \times \tau$.
- Thus the median absolute difference between measuremnts on two identical persons (in terms of covariates) is $0.953 \times \tau$.

Interpretation of btw. person $\mathrm{s.d.}$

- Select two persons at random with the same covariate values ([cov]).
- The $\mathrm{s.d.}$ of the difference of their measurements is $\sqrt{2}\tau$; the absolute difference follow a half-normal distribution with this $\mathrm{s.d.}$,
- The median of this corresponds to the 75^{th} percentile of a normal with this scale, that is $0.953 \times \tau$.
- Thus the median absolute difference between measuremnts on two identical persons (in terms of covariates) is $0.953 \times \tau$.
- This is the way to report between person variation [**?**]

Extended model: Random slopes

Measurement on individual i at timepoint t

$$y_{ti} = \mu + [\textsf{cov}] + a_i + b_i t + e_{it}$$

The variation in $a_i + b_i t$ is now the **between** person variation; depending on t.

Extended model: Random slopes

Measurement on individual i at timepoint t

$$y_{ti} = \mu + [\text{cov}] + a_i + b_i t + e_{it}$$

The variation in $a_i + b_i t$ is now the **between** person variation; depending on t.

Note: The distribution of (a_i, b_i) must be specified as a bivariate normal, with arbitrary correlation.

Extended model: Random slopes

Measurement on individual i at timepoint t

$$y_{ti} = \mu + [\text{cov}] + a_i + b_i t + e_{it}$$

The variation in $a_i + b_i t$ is now the **between** person variation; depending on t.

Note: The distribution of (a_i, b_i) must be specified as a bivariate normal, with arbitrary correlation.

Otherwise the model is dependent on the scaling and origin of t

Extended model: Random slopes

Measurement on individual i at timepoint t

$$y_{ti} = \mu + [\textsf{cov}] + a_i + b_i t + e_{it}$$

The variation in $a_i + b_i t$ is now the **between** person variation; depending on t.

Note: The distribution of (a_i, b_i) must be specified as a bivariate normal, with arbitrary correlation.

Otherwise the model is dependent on the scaling and origin of t

The $\mathrm{s.d.}$ of a_i normally meaningless, but the $\mathrm{s.d.}$ of the b_is is interpretable

Extended model: Random slopes

Measurement on individual i at timepoint t

$$y_{ti} = \mu + [\text{cov}] + a_i + b_i t + e_{it}$$

The variation in $a_i + b_i t$ is now the **between** person variation; depending on t.

Note: The distribution of (a_i, b_i) must be specified as a bivariate normal, with arbitrary correlation.

Otherwise the model is dependent on the scaling and origin of t

The $\mathrm{s.d.}$ of a_i normally meaningless, but the $\mathrm{s.d.}$ of the b_is is interpretable (principle of marginality).

Changing the times individually

Bendix Carstensen

LEAD
31 March 2014
LEAD symposium, EDEG 2014
`http://BendixCarstensen.com/SDC/LEAD`

(`reshuf`)

Relative changes of times

- Time is usually an explanatory variable

Relative changes of times

- Time is usually an explanatory variable
- used in modelling the outcome

Relative changes of times

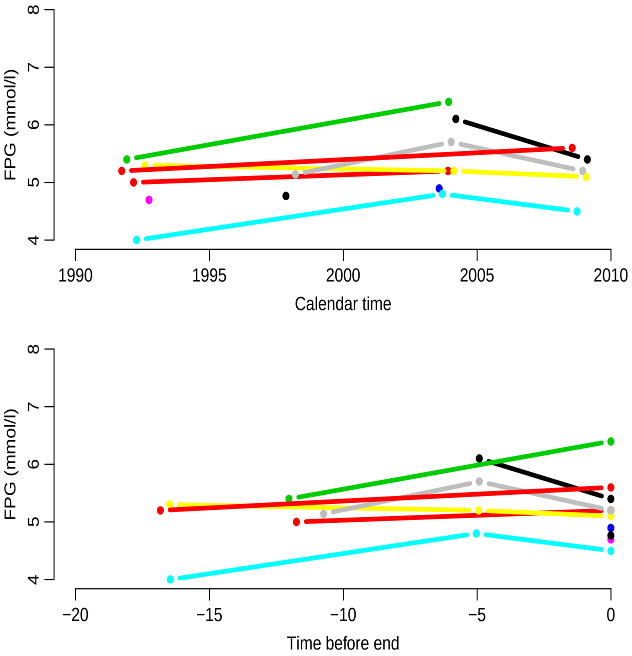
- Time is usually an explanatory variable
- used in modelling the outcome
- Meaningless to change the **relative** position of times **within** a person.

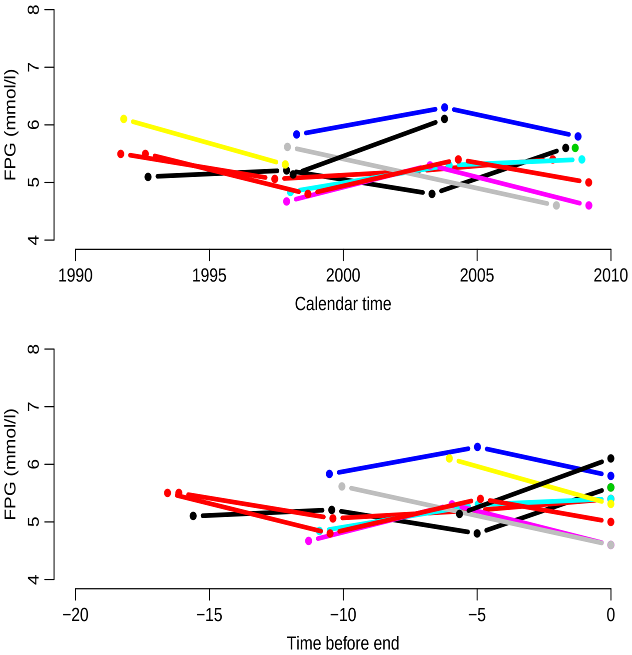
Relative changes of times

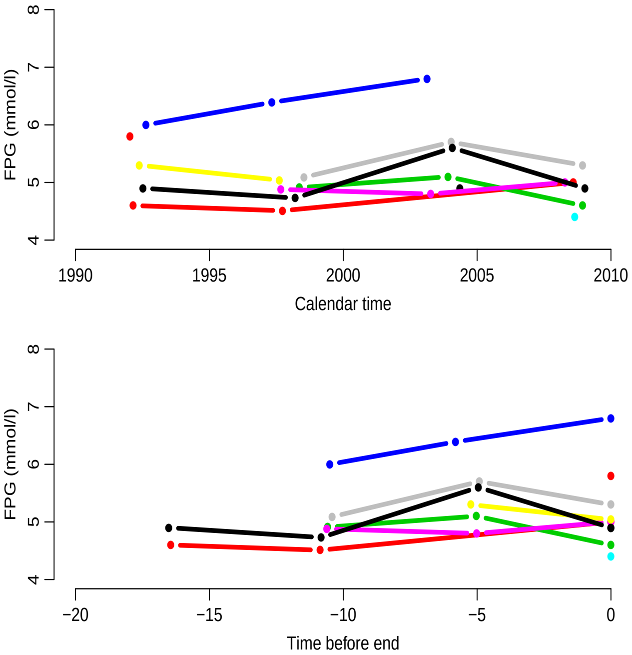
- Time is usually an explanatory variable
- used in modelling the outcome
- Meaningless to change the **relative** position of times **within** a person.
- Changing times between persons just amounts to using a different timescale. Age instead of time since diagnosis...

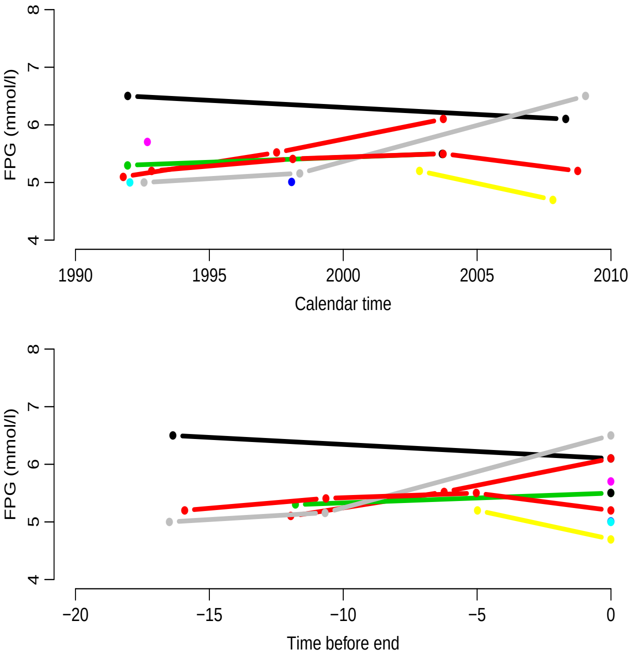
Relative changes of times

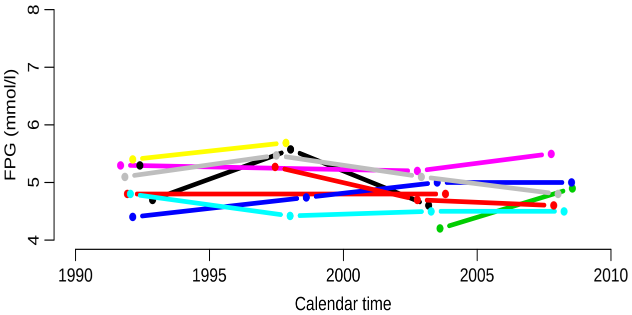
- Time is usually an explanatory variable
- used in modelling the outcome
- Meaningless to change the **relative** position of times **within** a person.
- Changing times between persons just amounts to using a different timescale. Age instead of time since diagnosis...
- Change of the statistical model in terms of interpretation

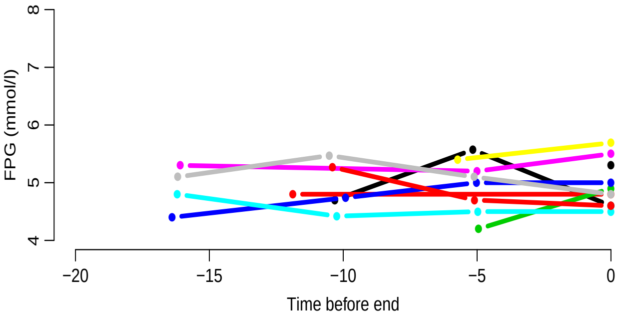

FPG (mmol/l)
4
5
6
7
8
1990
1995
2000
2005
2010
Calendar time
FPG (mmol/l)
4
5
6
7
8
−20
−15
−10
−5
0
Time before end

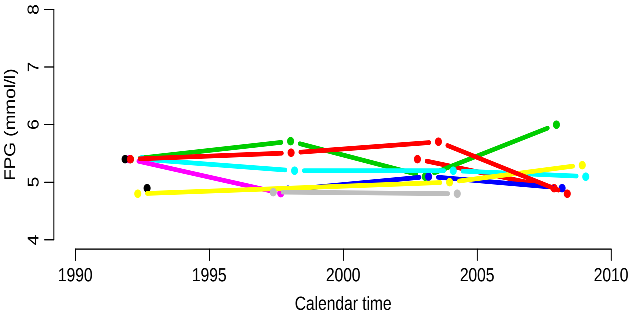

FPG (mmol/l)
4
5
6
7
8
1990
1995
2000
2005
2010
Calendar time
FPG (mmol/l)
4
5
6
7
8
−20
−15
−10
−5
0
Time before end

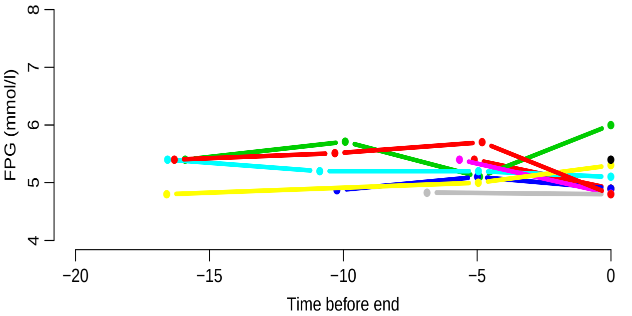

FPG (mmol/l)
4
5
6
7
8
1990
1995
2000
2005
2010
Calendar time
FPG (mmol/l)
4
5
6
7
8
−20
−15
−10
−5
0
Time before end

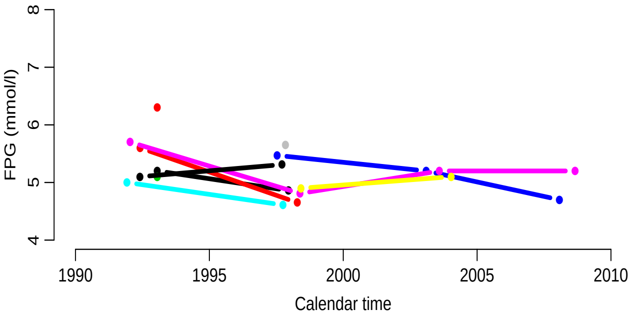

FPG (mmol/l)
4
5
6
7
8
1990
1995
2000
2005
2010
Calendar time
FPG (mmol/l)
4
5
6
7
8
−20
−15
−10
−5
0
Time before end

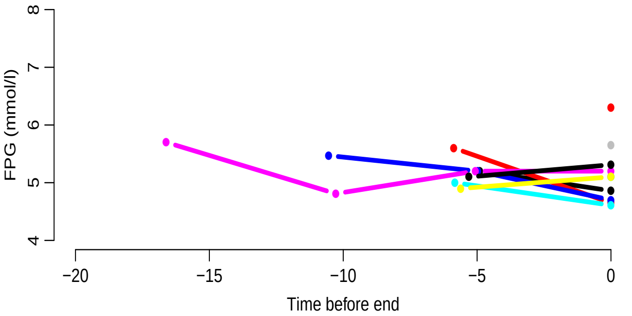

FPG (mmol/l)
4
5
6
7
8
1990
1995
2000
2005
2010
Calendar time


FPG (mmol/l)
4
5
6
7
8
−20
−15
−10
−5
0
Time before end


FPG (mmol/l)
4
5
6
7
8
1990
1995
2000
2005
2010
Calendar time


FPG (mmol/l)
4
5
6
7
8
−20
−15
−10
−5
0
Time before end


FPG (mmol/l)
4
5
6
7
8
1990
1995
2000
2005
2010
Calendar time


FPG (mmol/l)
4
5
6
7
8
−20
−15
−10
−5
0
Time before end

Meaningful timescales

- Time since:

Meaningful timescales

- Time since:
  - Randomization

Meaningful timescales

- Time since:
  - Randomization
  - 1st measurement

Meaningful timescales

- Time since:
  - Randomization
  - 1st measurement
  - Birth

Meaningful timescales

- Time since:
  - Randomization
  - 1st measurement
  - Birth
  - 1 jan. 1900 (calendar time)

Meaningful timescales

- Time since:
  - Randomization
  - 1st measurement
  - Birth
  - 1 jan. 1900 (calendar time)
- Time before:

Meaningful timescales

- Time since:
  - Randomization
  - 1st measurement
  - Birth
  - 1 jan. 1900 (calendar time)
- Time before:
  - DM diagnosis

Meaningful timescales

- Time since:
  - Randomization
  - 1st measurement
  - Birth
  - 1 jan. 1900 (calendar time)
- Time before:
  - DM diagnosis
  - Death

Meaningful timescales

- Time since:
  - Randomization
  - 1st measurement
  - Birth
  - 1 jan. 1900 (calendar time)
- Time before:
  - DM diagnosis
  - Death
  - Last measurement

Meaningful timescales

- Time since:
  - Randomization
  - 1st measurement
  - Birth
  - 1 jan. 1900 (calendar time)
- Time before:
  - DM diagnosis
  - Death
  - Last measurement
  - A random point in time — what is this?

Meaningful timescales

- Time since:
  - Randomization
  - 1st measurement
  - Birth
  - 1 jan. 1900 (calendar time)
- Time before:
  - DM diagnosis
  - Death
  - Last measurement
  - A random point in time — what is this?
- Meaningful to condition on the future?

Conditioning on future — validity

(Tentative arguments)

Conditioning on future — validity

(Tentative arguments)

Meaningful for outcomes:

- we are just making inference in a different (conditional) distribution.

Conditioning on future — validity

(Tentative arguments)

Meaningful for outcomes:

- we are just making inference in a different (conditional) distribution.
- the conditional distribution must not be singular.

Conditioning on future — validity

(Tentative arguments)

Meaningful for outcomes:

- we are just making inference in a different (conditional) distribution.
- the conditional distribution must not be singular.
- generalizable to the unconditional distribution?

Conditioning on future — validity

(Tentative arguments)

Meaningful for outcomes:

- we are just making inference in a different (conditional) distribution.
- the conditional distribution must not be singular.
- generalizable to the unconditional distribution?
- comparable to the unconditional dist'n?

Conditioning on future — validity

(Tentative arguments, cont'd)

Conditioning on future — validity

(Tentative arguments, cont'd)

Not meaningful for covariates:

- **Immortal time bias**:
  Conditioning on future change of exposure, and **hence also** on future survival. So the outcome (death) is deterministic — it will not occur till exposure change.

Conditioning on future — validity

(Tentative arguments, cont'd)

Not meaningful for covariates:

- **Immortal time bias**:
  Conditioning on future change of exposure, and **hence also** on future survival. So the outcome (death) is deterministic — it will not occur till exposure change.
- The joint distribution of (response, predictors) **conditional** on a future value of a covariate may not be what we want.

Conditioning on future — validity

(Tentative arguments, cont'd)

Not meaningful for covariates:

- **Immortal time bias**:
  Conditioning on future change of exposure, and **hence also** on future survival. So the outcome (death) is deterministic — it will not occur till exposure change.
- The joint distribution of (response, predictors) **conditional** on a future value of a covariate may not be what we want.
- ... some may even think it is the unconditional.

Conditioning on future — validity

- Meaningful comparisons conditioning on a future event:

Conditioning on future — validity

- Meaningful comparisons conditioning on a future event:
- the comparison should be conditional on:

Conditioning on future — validity

- Meaningful comparisons conditioning on a future event:
- the comparison should be conditional on:
  - **not** seeing a future event (impossible)

Conditioning on future — validity

- Meaningful comparisons conditioning on a future event:
- the comparison should be conditional on:
  - **not** seeing a future event (impossible)
  - **not having seen** an event . . .

Conditioning on future — validity

- Meaningful comparisons conditioning on a future event:
- the comparison should be conditional on:
  - **not** seeing a future event (impossible)
  - **not having seen** an event ...

Conditioning on future — validity

- Meaningful comparisons conditioning on a future event:
- the comparison should be conditional on:
  - **not** seeing a future event (impossible)
  - **not having seen** an event . . . yet
- Imposes constraints on possible shapes of trajectories for those without event:

Conditioning on future — validity

- Meaningful comparisons conditioning on a future event:
- the comparison should be conditional on:
  - **not** seeing a future event (impossible)
  - **not having seen** an event ... yet
- Imposes constraints on possible shapes of trajectories for those without event:
- Must be invariant under individual translation of time

Conditioning on future — validity

- Meaningful comparisons conditioning on a future event:
- the comparison should be conditional on:
  - **not** seeing a future event (impossible)
  - **not having seen** an event ... yet
- Imposes constraints on possible shapes of trajectories for those without event:
- Must be invariant under individual translation of time
- Only linear (mean) effects meaningful

Conditioning on future — validity

- Meaningful comparisons conditioning on a future event:
- the comparison should be conditional on:
  - **not** seeing a future event (impossible)
  - **not having seen** an event ... yet
- Imposes constraints on possible shapes of trajectories for those without event:
- Must be invariant under individual translation of time
- Only linear (mean) effects meaningful
- Must include random intercept and slope

Conditioning on future — validity

- Meaningful comparisons conditioning on a future event:
- the comparison should be conditional on:
  - **not** seeing a future event (impossible)
  - **not having seen** an event . . . yet
- Imposes constraints on possible shapes of trajectories for those without event:
- Must be invariant under individual translation of time
- Only linear (mean) effects meaningful
- Must include random intercept and slope
- Is time just a surrogate for age???

Conclusions

Bendix Carstensen

LEAD
31 March 2014
LEAD symposium, EDEG 2014
`http://BendixCarstensen.com/SDC/LEAD`

(`concl`)

Conclusions

- Always look at your data:

Conclusions

- Always look at your data:
  - FU vs. Baseline

Conclusions

- Always look at your data:
  - FU vs. Baseline
  - Spaghetti-plots

Conclusions

- Always look at your data:
  - FU vs. Baseline
  - Spaghetti-plots
- Be explicit about the model used.

Conclusions

- Always look at your data:
  - FU vs. Baseline
  - Spaghetti-plots
- Be explicit about the model used.
- Show all estimates, not only the means,

Conclusions

- Always look at your data:
  - FU vs. Baseline
  - Spaghetti-plots
- Be explicit about the model used.
- Show all estimates, not only the means,
- — the variation between and within persons are also important

Reporting models

- There is no such thing as a "mixed model" or a "random effects model"

Reporting models

- There is no such thing as a “mixed model” or a “random effects model”
- Specify the fixed and random effects.

Reporting models

- There is no such thing as a "mixed model" or a "random effects model"
- Specify the fixed and random effects.
- Report them.

Reporting models

- There is no such thing as a "mixed model" or a "random effects model"
- Specify the fixed and random effects.
- Report them.
- All of them — this is scary; you have to get you head around all of them.

Reporting models

- There is no such thing as a “mixed model” or a “random effects model”
- Specify the fixed and random effects.
- Report them.
- All of them — this is scary; you have to get you head around all of them.
- Fit only one or two models

Reporting models

- There is no such thing as a "mixed model" or a "random effects model"
- Specify the fixed and random effects.
- Report them.
- All of them — this is scary; you have to get you head around all of them.
- Fit only one or two models
- — that captures what you want to know about.

What to report

- Mean trajectories — the mean shape of the measurements.

What to report

- Mean trajectories — the mean shape of the measurements.
- — usually by group

What to report

- Mean trajectories — the mean shape of the measurements.
- — usually by group
- Estimated random effect variations

What to report

- Mean trajectories — the mean shape of the measurements.
- — usually by group
- Estimated random effect variations
    - median difference between persons

What to report

- Mean trajectories — the mean shape of the measurements.
- — usually by group
- Estimated random effect variations
  - median difference between persons
  - — possibly varying along the time scale,