UMC Utrecht

Epidemiology and Big Data

Mixed Models part 2: Longitudinal Data (Modelling Time)

Rebecca Stellato

Overview Part 2: longitudinal data

- Examples of longitudinal data
- Linear mixed effects (LME) models with linear time effect
- The variance-covariance matrix of repeated measures
- Correlation structures & covariance pattern models
- What to do with baseline measurement?

Longitudinal Data

- Longitudinal data: repeated measures of individuals over time
  - variable measured on individual at several time points
  - observations on one and the same individual will not be independent
  - calls for special analysis techniques

if you take my bloodpressure today, tomorrow and next month we expect measures closer to each other to be more similar than measurements further away from each other.

Examples of Longitudinal Data

- Example (Reisby et al.)
  - 66 patients
  - with or without endogenous depression (more personality (within you))
  - depression scores measured weekly at weeks 0 – 5, using Hamilton Depression Rating Scale (HDRS) (starting before treatment)
  - from week 1 onwards, patients are treated with imipramine (antidepressant)
  - **Research question**: is the pattern of HDRS over time different for patients with endogenous and non-endogenous depression?

more external - external cause

treated as a continuous score of depression rating

Examples of Longitudinal Data

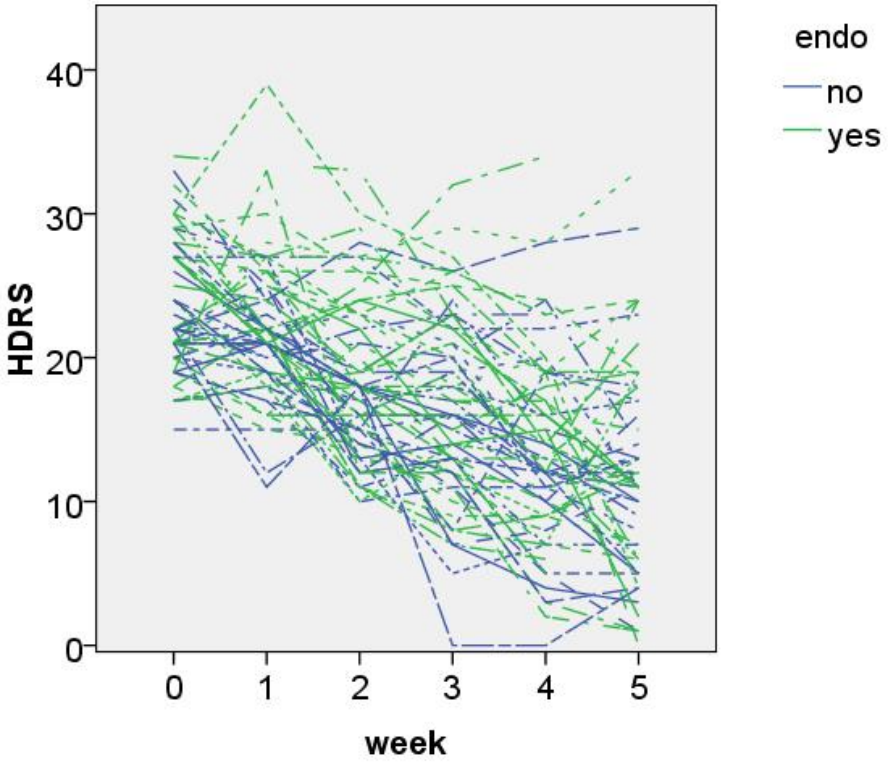
- Example (Stoop et al. 2012)
  - 14 patients with Hurler syndrome
  - after haematopoietic stem cell transplantation
  - various radiologic measurements, including the odontoid/body ratio
  - **Research question**: what is the pattern of orthopedic manifestations after stem cell transplant?

Examples of Longitudinal Data

Reisby et al.

trial type

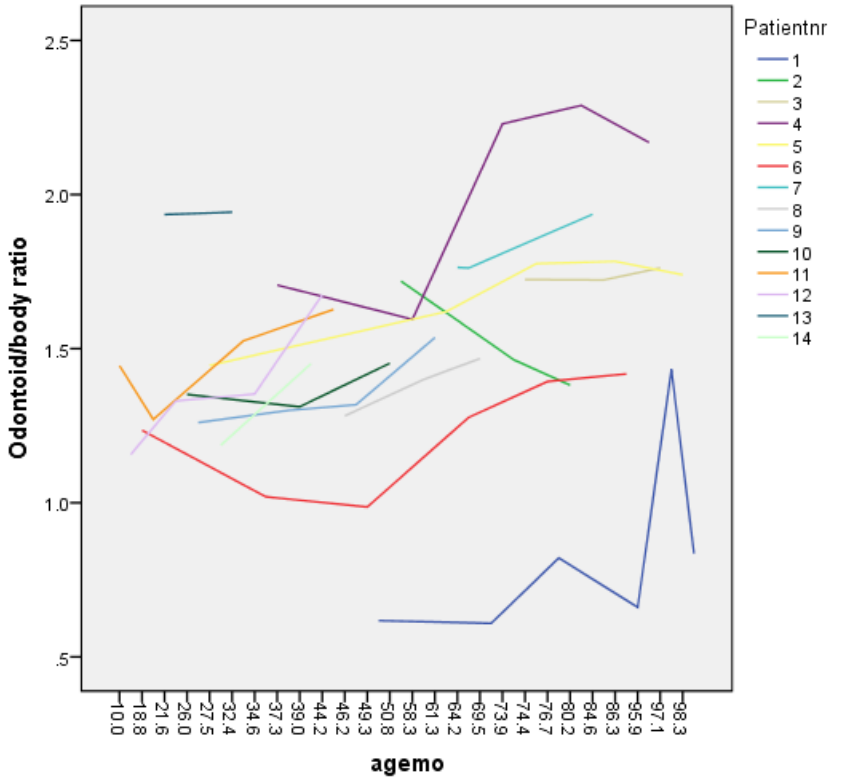
measures at the same points in time



Stoop et al.

observational data

measurements not at the same point in time



Example: Reisby Data

are endo. different from exdo.?

- Research question: differing patterns over time for the two groups?
  - ➢ fixed effects for intercept, time, group & group*time
  - ○ time continuous or categorical?

need to take time into account (main effect for time), group and group& time

- How to deal with multiple measurements?
  - ➢ Random effects
    - intercept? level 1 measures that are correlated within level 2 units
      - – each patient seems to have a different starting point
    - slope of time? does it look like some people are decreasing more or less quickly than others?
      - – could be patients have differing slopes over time

looks like there might be a random effect/slope for time as there is variation

adding randomness because expecting correlation of repeated measures within the patients

Example: Reisby Data

before I start modeling I look at descriptive statistics showing:

Descriptive Statistics (R)

```
> reisby.wide <- reshape(reisby.long, v.names="hdrs", idvar="id",
timevar="week", direction="wide")
> by(reisby.wide[,3:8], reisby.wide$endo, describe)
endo: 0
        var  n  mean   sd median
hdrs.0    1 28 22.79 4.12   22.0
hdrs.1    2 29 20.48 3.83   21.0
hdrs.2    3 28 17.00 4.35   16.5
hdrs.3    4 29 15.34 6.17   16.0
hdrs.4    5 29 12.62 6.72   12.0
hdrs.5    6 27 11.22 6.34   11.0
------------------------------------
endo: 1
        var  n  mean   sd median
hdrs.0    1 33 24.00 4.85   24.0
hdrs.1    2 34 23.00 5.10   22.0
hdrs.2    3 37 19.30 6.08   18.0
hdrs.3    4 36 17.28 6.56   16.5
hdrs.4    5 34 14.47 7.17   14.0
hdrs.5    6 31 12.58 7.96   11.0
```

less variation in the beginning of the study than in the end of the study (patients tend to be more similar to one another in the beginning of a study)

describe() function in the **'psych'** package (gives more stats than presented here)

Example: Reisby Data

correlations between HDRS measurements (R)

```
> round(cor(reisby.wide[,3:8], use="pairwise.complete.obs"), digits=3)
```

	hdrs.0	hdrs.1	hdrs.2	hdrs.3	hdrs.4	hdrs.5
hdrs.0	1.000	0.493	0.410	0.333	0.227	0.184
hdrs.1	0.493	1.000	0.494	0.412	0.308	0.218
hdrs.2	0.410	0.494	1.000	0.738	0.669	0.461
hdrs.3	0.333	0.412	0.738	1.000	0.817	0.568
hdrs.4	0.227	0.308	0.669	0.817	1.000	0.654
hdrs.5	0.184	0.218	0.461	0.568	0.654	1.000

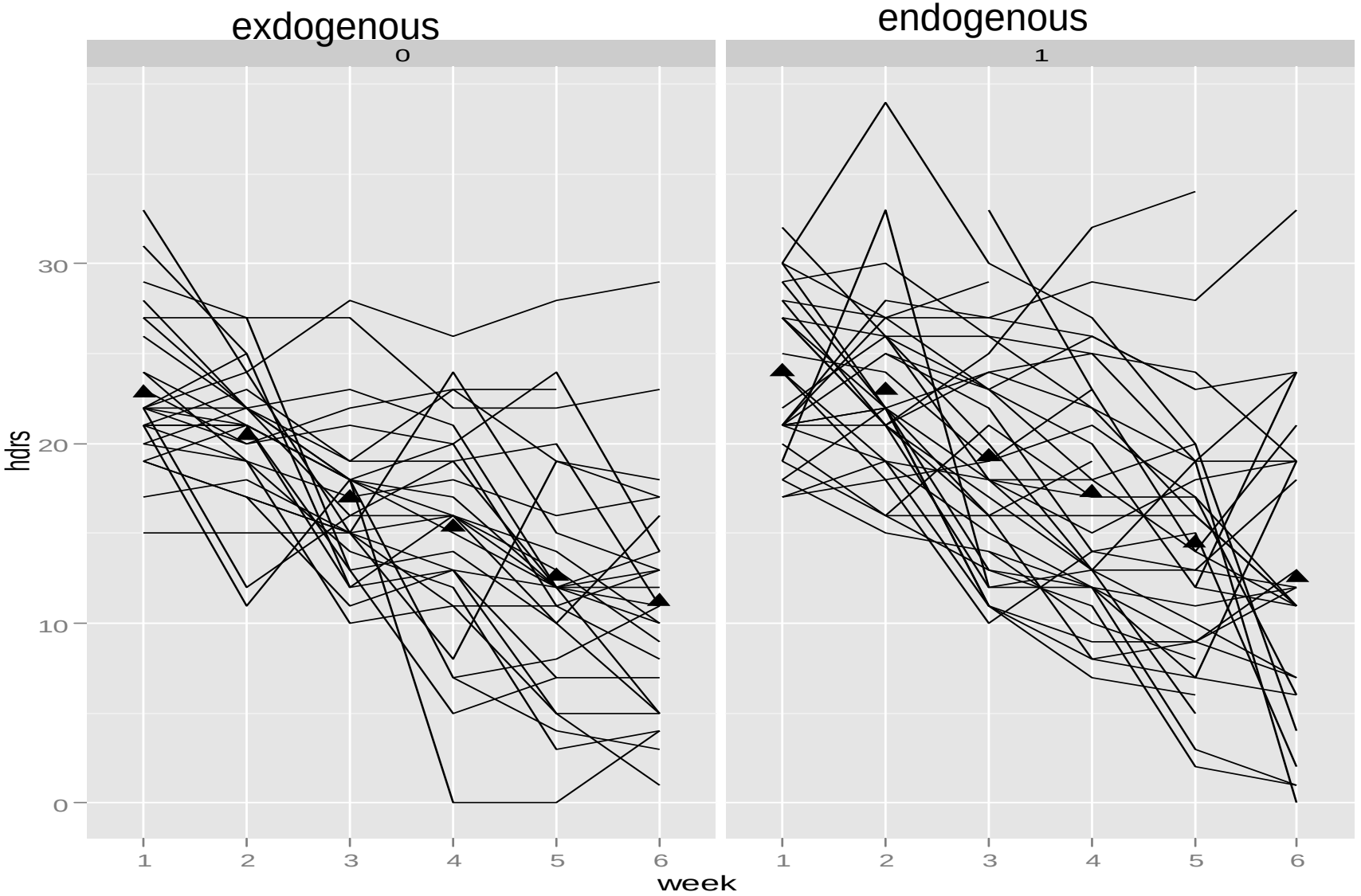
decreasing over time - less correlated the further apart

depression score at time 0 and then for each time

the further the measures are apart (away from diagonal) the less strong is the correlation

Example: Reisby Data

"Spaghetti Plot" (R, using ggplot2)



Linear time effect

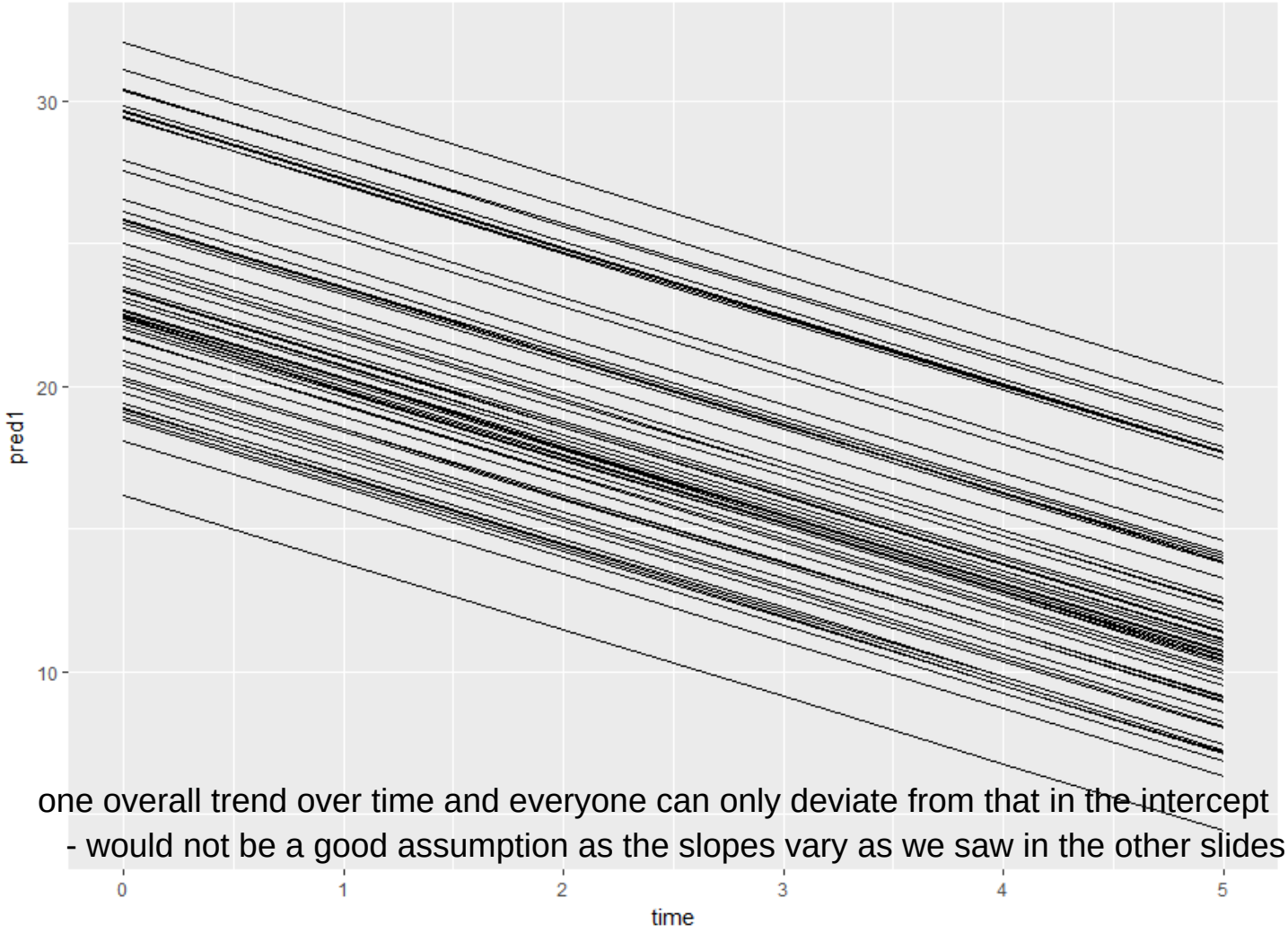
- Time technically measured in categories (weeks 1, 2 …)
- Reasonable to model time as linear? yes
  - 1 parameter for slope of HDRS in time when modeling the trend over time as continuous
  - need to check whether this assumption is reasonable
    - initial data analysis (spaghetti plot, individual plots)
    - model comparison (day 3)

always check if one can use linear time effect (as partly seen in prev. slides)

Random intercept with linear time effect

Predicted values from a LME with random intercept and linear time:

good option



one overall trend over time and everyone can only deviate from that in the intercept - would not be a good assumption as the slopes vary as we saw in the other slides

Random intercept + random linear time effect

- We can make the assumption that the (linear) time effect is different for each individual by incorporating a random (linear) time effect:
  - $y_{ij} = (\beta_0 + \upsilon_{0i}) + (\beta_1 + \upsilon_{1i}) \cdot time_{ij} + \varepsilon_{ij}$
  - $\varepsilon_{ij} \sim N(0, {\sigma_e}^2)$;
  - $\upsilon_{0i} \sim N(0, {\sigma_{\upsilon 0}}^2)$; $\upsilon_{1i} \sim N(0, {\sigma_{\upsilon 1}}^2)$; $cov(\upsilon_{0i}, \upsilon_{1i}) = \sigma_{\upsilon 01}$

    each with own variance — possibly covariance

    - This last line can also be written: $\begin{bmatrix} \upsilon_{0i} \\ \upsilon_{1i} \end{bmatrix} \sim N(0, \boldsymbol{\Sigma}_\upsilon)$, $\boldsymbol{\Sigma}_\upsilon = \begin{bmatrix} {\sigma_{\upsilon 0}}^2 & \sigma_{\upsilon 01} \\ \sigma_{\upsilon 01} & {\sigma_{\upsilon 1}}^2 \end{bmatrix}$

    - $\boldsymbol{\Sigma}_\upsilon$ is the *variance-covariance matrix* of the random effects

in addition to random intercept we should also add random slope for time assuming residuals to be normally distributed and

R output random intercept + random slope model

Parameter estimates of the fixed part of the previous model:

	Value	Std.Error	DF	t-value	p-value
(Intercept)	22.476263	0.7986132	307	28.144117	0.0000
time	-2.365687	0.3134845	307	-7.546425	0.0000
endo	1.988021	1.0747911	64	1.849681	0.0690
time:endo	-0.027056	0.4217258	307	-0.064155	0.9489

output from model with random intercept and slope for time per person

interaction between endo and exdo is not significant

R output random intercept + random slope model

Parameter estimates of the random part (intercept, slope) of the model:

```
             StdDev   Corr
(Intercept) 3.411893 (Intr)
time        1.441193 -0.285
Residual    3.495500
```

negatively correlated = intercept is higher -> slope is more negative because the higher the participants started (bad depression) the more likely it'll go down whereas participants starting low are more "flexible"

- So:
  - $\hat{\sigma}_e{}^2 = 3.50^2 = 12.22$; $\hat{\sigma}_{\upsilon 0}{}^2 = 3.41^2 = 11.64$; $\hat{\sigma}_{\upsilon 1}{}^2 = 1.44^2 = 2.08$
  - $\widehat{corr}(\upsilon_{0i}, \upsilon_{1i}) = {}^{\hat{\sigma}_{\upsilon 01}}/_{(\hat{\sigma}_{\upsilon 0} \cdot \hat{\sigma}_{\upsilon 1})} = -0.285$
    $\rightarrow \hat{\sigma}_{\upsilon 01} = -0.285 \cdot (3.41 \cdot 1.44) = -1.40$

Note: random intercept and slope are negatively correlated (the higher the intercept the more negative the slope); often true in longitudinal data

Interpretation of model

- Intercept (22.48) is average HDRS score when all variables = 0
  - so for patients with exogenous depression (reference) at time = 0
- Estimate for endo (1.99) is average difference in HDRS between endogenous and exogenous patients at time = 0
  - patients with exogenous depression start with average of 22.48
  - patients with endogenous depression start with average of 22.476 + 1.988 = 24.46
- Estimate of random intercept s.d. 3.41 indicates considerable fluctuation around fixed intercepts:
  - patients can start quite a bit higher/lower than average
- "Average" slope is -2.37 for patients with exogenous depression

3.41 higher or lower start

Average different for groups

considerable fluctuation betweens starting points

Interpretation of model, cont.

slightly steeper (not relevant)

- Interaction endo*time (-0.027) is difference in slope endogenous vs. exogenous
  - per time unit (1 week) the HDRS score decreases on average by 2.37 (exog)
  - per time unit (1 week) the HDRS score decreases on average by 2.39 (endog)

not really relevant and not significant

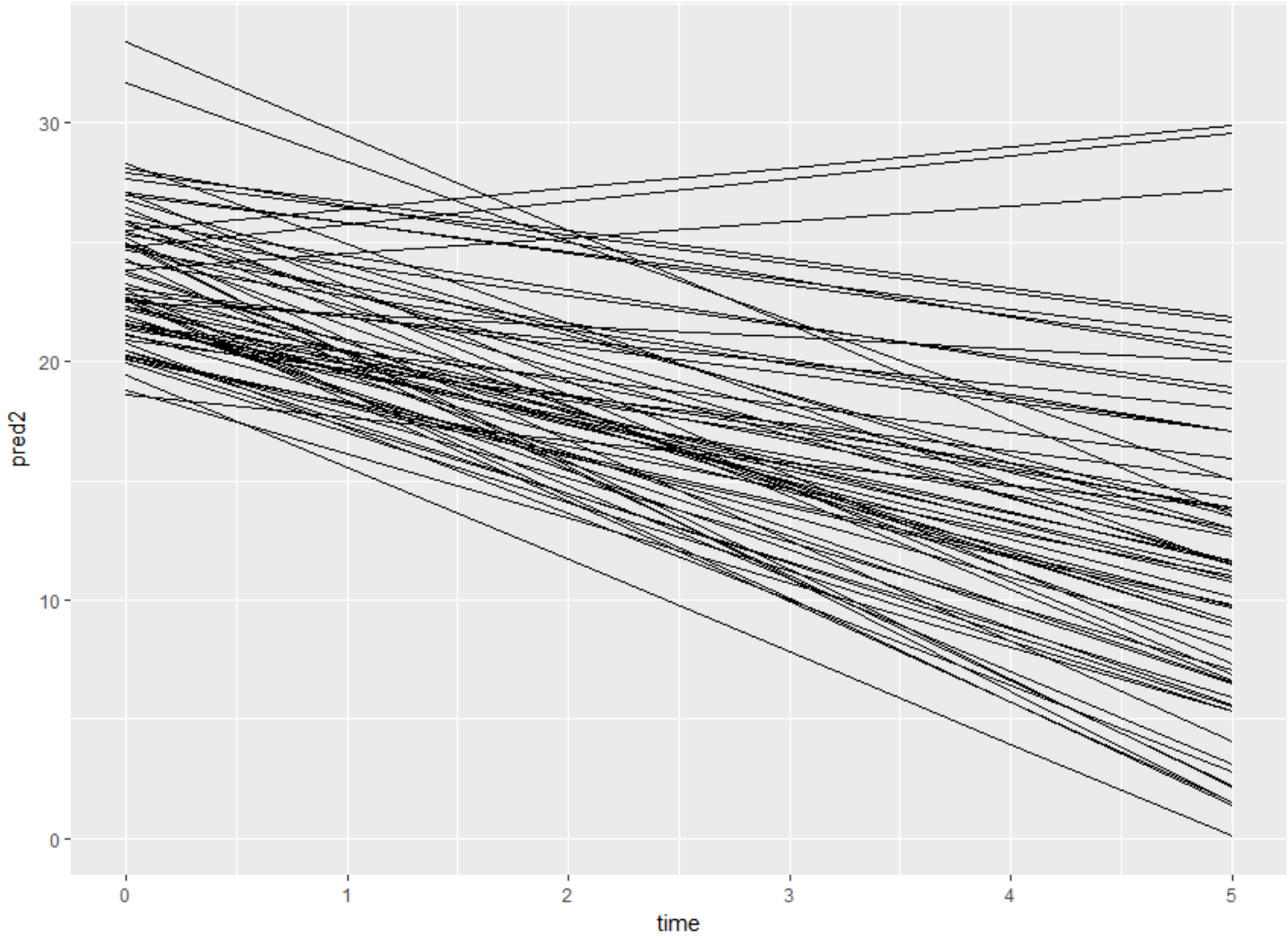
- Estimate of random slope s.d. 1.44, so for individuals the slope can be quite a bit steeper or flatter, may even be positive for some patients (as seen in the plot).

a lot of variation around the values

decreases slower/faster quite a bit - variation between change in how they're feeling

Random intercept + random linear time effect

Predicted values from a LME with random intercept + random linear time effect:



looks more like the patterns on p.11 = good why it's more reasonable to add random intercept and random slope for time per person

Random intercept + random linear time effect

- Given what we learned yesterday, the previous analysis (with linear effects for time) is "all" we could do
- What if we think it is unreasonable to use time as continuous?
- Add time as categorical to the fixed effects – and then?
  - some add time as linear to the random effects for time as linear
  - others choose to tackle time as a categorical variable in the random part of the model as well

time categorical as a fixed effect... shown now
(random part of the model when time as categorical in the fixed part of the model)

LMM matrix formulation & var-covar matrix

- Recall model linear mixed model:
  - $y_{ij} = \beta_0 + \beta_1 X_{1ij} + \upsilon_{0i} + \upsilon_{1i} X_{1ij} + \varepsilon_{ij}$
- In matrix formulation:
  - $\boldsymbol{Y}_i = \boldsymbol{X}_i \cdot \boldsymbol{\beta} + \boldsymbol{Z}_i \cdot \upsilon_i + \boldsymbol{\varepsilon}_i$
  - where $\boldsymbol{X}_i$ is the covariate matrix of the fixed effect(s) and $\boldsymbol{Z}_i$ the design matrix of the random effect(s)
- Variance-covariance matrix of repeated measures y:
  - $Var(\boldsymbol{Y}_i) = \boldsymbol{Z}_i \cdot \boldsymbol{\Sigma}_{\upsilon} \cdot \boldsymbol{Z}_i{}' + \sigma_e^2 \cdot \boldsymbol{I}_{n_i}$
  - variance-covariance matrix of outcome for a patient's measurements depends on covariance matrix of random effects $\boldsymbol{\Sigma}_{\upsilon}$ and residual variance σ_e^2
  - the number and variances of the random effects and their correlation(s) determine part of the variance-covariance matrix

LMM matrix formulation & var-covar matrix

$$Var(\boldsymbol{Y}_i) = \boldsymbol{Z}_i \cdot \boldsymbol{\Sigma}_{\cup} \cdot \boldsymbol{Z}_i{}' + {\sigma_e}^2 \cdot \boldsymbol{I}_{n_i}$$

residual part is simple
1 on diagonal 0 on off diagonal

*

$$= \begin{pmatrix} \boldsymbol{s}' & \boldsymbol{t} & \boldsymbol{h} & \boldsymbol{i} & \boldsymbol{n} & \boldsymbol{g} \\ \boldsymbol{m} & \boldsymbol{e} & \boldsymbol{s} & \boldsymbol{s} & \boldsymbol{y} & - \\ (\boldsymbol{d} & \boldsymbol{e} & \boldsymbol{p} & \boldsymbol{e} & \boldsymbol{n} & \boldsymbol{ds} \\ & & \boldsymbol{o} & \boldsymbol{n} & & \\ \boldsymbol{r} & \boldsymbol{a} & \boldsymbol{n} & \boldsymbol{d} & \boldsymbol{o} & \boldsymbol{m} \\ \boldsymbol{e} & \boldsymbol{ff} & \boldsymbol{e} & \boldsymbol{c} & \boldsymbol{t} & \boldsymbol{s}) \end{pmatrix} + \boldsymbol{\sigma_e^2} \cdot \begin{pmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} \end{pmatrix}$$

depending on random effects and their variances and covariances

depending on amount of random effects and covariances

CPMs & var-covar matrix

Another possibility for modelling longitudinal data:

different form lecture 1 because multiple measures on individuals and we are still assuming closer measures to be more similar than the ones further apart

- $\boldsymbol{Y}_i = \boldsymbol{X}_i \cdot \boldsymbol{\beta} + \boldsymbol{\varepsilon}_i$ fixed effects and just some residual variation (matrix) - but this residual matrix has to reflect all of the "messiness" from the original data
  - No random effects!
  - How do we take into account the correlation between measurements on same person?
- Variance-covariance matrix of repeated measures y is now:
  - $Var(\boldsymbol{Y}_i) = Var(\boldsymbol{\varepsilon}_i) = \boldsymbol{\Sigma}$

CPMs & var-covar matrix

- We know the residuals are not independent, so we need to assume correlation for $\boldsymbol{\Sigma}$:

$$\boldsymbol{\Sigma} = \begin{pmatrix} & t & h & i & s & \\ s & h & o & u & l & d \\ b & e & & c & o & r- \\ r & e & c & t & l & y \\ c & o & m & p & l & - \\ i & c & a & t & e & d! \end{pmatrix}$$

- We can use different correlation structures to model $\boldsymbol{\Sigma}$ directly
- We call this type of model a *covariance pattern model* (CPM)
- Some also call them "GEE-type covariance structures"

not using random effects in the model which is why the residual covariance - variance model has to be somewhat complicated

Example: Reisby Data

Possibilities for modelling correlated measures

- Model correlation of measurements *implicitly* (linear mixed effects model):
  - repeated observations are level-1 variables nested within patient (=level 2)
    - random intercept per patient or random intercept per patient + random slope for week per patient or...
    - random effects and their covariance determine structure of var-covar matrix
- Model correlation of measurements *explicitly* (CPM):
  - incorporate a covariance structure of the residuals into the model
  - (usually) assumes equally spaced time intervals
- Combination of the two (mixed regression models with autocorrelated errors)

Observed var-cov matrix Reisby dataset

```
> round(var(reisby.wide[,3:8], use="pairwise.complete.obs"), digits = 2)

       hdrs.0 hdrs.1 hdrs.2 hdrs.3 hdrs.4 hdrs.5
hdrs.0  20.55  10.11  10.14  10.09   7.19   6.28
hdrs.1  10.11  22.07  12.28  12.55  10.26   7.72
hdrs.2  10.14  12.28  30.09  25.13  24.63  18.38
hdrs.3  10.09  12.55  25.13  41.15  37.34  23.99
hdrs.4   7.19  10.26  24.63  37.34  48.59  30.51
hdrs.5   6.28   7.72  18.38  23.99  30.51  52.12
```

variance

increasing variances over time

decreasing as measurements are further and further apart

Various correlation structures

not mixed anymore if no random effect

- nlme has numerous correlation structures for linear ~~mixed~~ models
- Most common/realistic for longitudinal data
  - unstructured
  - autoregressive of order 1: AR(1)
- Bad ideas:
  - uncorrelated (independent)
  - compound symmetry
- Correlations pertain to the residuals within each of the subjects after correcting for fixed (and perhaps random) effects

Independent correlation structure

- The independent (scaled identity) correlation structure assumes residuals to be independent, as if they came from different subjects
- All variances are assumed equal, all correlations are assumed 0
- This is the assumption in ordinary linear regression/ANOVA

$$\Sigma = \begin{pmatrix} \sigma^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2 \end{pmatrix} = \sigma^2 \cdot \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

like two way anova with time as categorical
- constant variance for different times and zero correlation within person over time (too simple)

Independent correlation structure

- Analyzing the data from the Reisby example using time (categorical), tx and time*tx and an independent correlation structure amounts to doing a two-way ANOVA (all observations are assumed to be independent)
- Even when observations are in fact (nearly) independent, the design of the study was to take random patients, and measure these multiple times, not to take random samples at each time point

➢ Preferable to analyze data as being repeated (and thus correlated)!

Compound symmetry correlation structure

- The compound symmetry (exchangeable) correlation structure assumes correlations between all time points to be equal, irrespective of the length of the time intervals.
- All variances are assumed equal, all correlations too:

so no matter how far the measures are apart they would have the same correlation

$$\begin{pmatrix} \sigma^2+{\sigma_0}^2 & {\sigma_0}^2 & {\sigma_0}^2 & {\sigma_0}^2 & {\sigma_0}^2 & {\sigma_0}^2 \\ {\sigma_0}^2 & \sigma^2+{\sigma_0}^2 & {\sigma_0}^2 & {\sigma_0}^2 & {\sigma_0}^2 & {\sigma_0}^2 \\ {\sigma_0}^2 & {\sigma_0}^2 & \sigma^2+{\sigma_0}^2 & {\sigma_0}^2 & {\sigma_0}^2 & {\sigma_0}^2 \\ {\sigma_0}^2 & {\sigma_0}^2 & {\sigma_0}^2 & \sigma^2+{\sigma_0}^2 & {\sigma_0}^2 & {\sigma_0}^2 \\ {\sigma_0}^2 & {\sigma_0}^2 & {\sigma_0}^2 & {\sigma_0}^2 & \sigma^2+{\sigma_0}^2 & {\sigma_0}^2 \\ {\sigma_0}^2 & {\sigma_0}^2 & {\sigma_0}^2 & {\sigma_0}^2 & {\sigma_0}^2 & \sigma^2+{\sigma_0}^2 \end{pmatrix} = ({\sigma_0}^2+\sigma^2) \begin{pmatrix} 1 & \rho & \rho & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho & \rho & \rho \\ \rho & \rho & 1 & \rho & \rho & \rho \\ \rho & \rho & \rho & 1 & \rho & \rho \\ \rho & \rho & \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & \rho & \rho & 1 \end{pmatrix}$$

same variance along diagonal and correlation off diagonal (covariance)
but assumed to be just one number

- Note: $\rho = {\sigma_0}^2 / ({\sigma_0}^2+\sigma^2)$, with ${\sigma_0}^2$ the variance within patients
- ρ is then known as the intraclass correlation coefficient, a ratio of individual variance to total variance

Compound symmetry correlation structure

assumption is again that time has constant variance and constant correlation over time

- A covariance pattern model with a compound symmetry pattern for the residuals is equivalent to a linear mixed model with a random intercept per patient p.31= assumption we are making
  - when we treat time as categorical in the fixed parts of both models
- For data without missing values, these two models are also equivalent to a repeated measures ("split-plot") ANOVA

Intercept only structure probably not a good idea especially with longitudinal data

Unstructured correlation

most complicated matrix
sep. variance of the residuals at each time point and different covariance between the time points

$$\begin{pmatrix} \theta_{11} & \theta_{12} & \theta_{13} & \theta_{14} & \theta_{15} & \theta_{16} \\ \theta_{21} & \theta_{22} & \theta_{23} & \theta_{24} & \theta_{25} & \theta_{26} \\ \theta_{31} & \theta_{32} & \theta_{33} & \theta_{34} & \theta_{35} & \theta_{36} \\ \theta_{41} & \theta_{42} & \theta_{43} & \theta_{44} & \theta_{45} & \theta_{46} \\ \theta_{51} & \theta_{52} & \theta_{53} & \theta_{54} & \theta_{55} & \theta_{56} \\ \theta_{61} & \theta_{62} & \theta_{63} & \theta_{64} & \theta_{65} & \theta_{66} \end{pmatrix}$$

16 possible covariances and 6 variances

- Variances at each time point different
- All covariances among time points different (note that $\theta_{12} = \theta_{21}$)
- Costly structure: 21 df needed for 6 time points!
- Flexible structure

parameters

most flexible for covariance and variance

reasonable to do when only 2-3 measurements per person otherwise to many extra parameters

Unstructured correlation for Reisby data

Estimated from model with time (cat), endo & time*endo no random effects

```
      [,1]  [,2]  [,3]  [,4]  [,5]  [,6]
[1,] 19.63 10.40  6.34  8.09  6.73  4.50
[2,] 10.40 20.87 10.26 10.69  8.33  5.02
[3,]  6.34 10.26 25.90 22.36 23.99 20.88
[4,]  8.09 10.69 22.36 38.44 32.03 29.90
[5,]  6.73  8.33 23.99 32.03 46.89 38.77
[6,]  4.50  5.02 20.88 29.90 38.77 60.19
```

looks like a good fit as it seems similar to the initial table - however very costly and very many df

(Homogeneous) autoregressive of order 1 (AR1) correlation

$$\sigma^2 \cdot \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \rho^4 & \rho^5 \\ \rho & 1 & \rho & \rho^2 & \rho^3 & \rho^4 \\ \rho^2 & \rho & 1 & \rho & \rho^2 & \rho^3 \\ \rho^3 & \rho^2 & \rho & 1 & \rho & \rho^2 \\ \rho^4 & \rho^3 & \rho^2 & \rho & 1 & \rho \\ \rho^5 & \rho^4 & \rho^3 & \rho^2 & \rho & 1 \end{pmatrix}$$

- Note: observations per subject assumed to be taken at equally-spaced intervals
- AR(1) assumes all observations 1 time unit apart have same correlations (ρ)
- Observations 2 units apart have corr ρ^2, obs 3 units apart ρ^3, etc.
- Outcome has same variance (σ^2) across all time points

AR1 with homogeneous variances for Reisby data

Estimated from model with time (cat), tx & time*tx, no random effects

```
        [,1]    [,2]   [,3]   [,4]    [,5]    [,6]
[1,] 35.0510 23.1050 15.231 10.040  6.6183  4.3627
[2,] 23.1050 35.0510 23.105 15.231 10.0400  6.6183
[3,] 15.2310 23.1050 35.051 23.105 15.2310 10.0400
[4,] 10.0400 15.2310 23.105 35.051 23.1050 15.2310
[5,]  6.6183 10.0400 15.231 23.105 35.0510 23.1050
[6,]  4.3627  6.6183 10.040 15.231 23.1050 35.0510
  Standard Deviations: 5.9204 5.9204 5.9204 5.9204 5.9204 5.9204
```

decreasing over time

same fixed effects no random effects - resulting pattern for residuals shows that standard deviation and variances are assumed constant over time but decreasing variance and covariance over time makes snse

AR1 with *heterogeneous* variances for Reisby data

Estimated from model with time (cat), tx & time*tx, no random effects

```
      [,1]  [,2]  [,3]  [,4]  [,5]  [,6]
[1,] 22.98 16.29 11.18  7.76  5.42  4.09
[2,] 16.29 27.12 18.62 12.92  9.02  6.81
[3,] 11.18 18.62 30.05 20.85 14.56 10.99
[4,]  7.76 12.92 20.85 33.98 23.73 17.91
[5,]  5.42  9.02 14.56 23.73 38.92 29.38
[6,]  4.09  6.81 10.99 17.91 29.38 52.11
```

allowing variances to differ over time and correlations to be rau squared to the 2, 3, etc.

nice comprimise with 6 variance parameters 1 correlation parameter = 7 which explains a nice amount of variance

- Heterogeneous variances probably fit the data better

"Covariance patterns" of linear mixed models

- A random intercept model implies a compound symmetry structure for all data combined of variances and covariances (assuming constant variance and constant correlation) - assumption when only using random slope?
- A linear mixed model with random intercept and random slope also implies a certain correlation structure for the data, but this is by no means a simple structure
  - recall: $Var(\boldsymbol{Y}_i) = \boldsymbol{Z}_i \cdot \boldsymbol{\Sigma}_{\upsilon} \cdot \boldsymbol{Z}_i{}' + {\sigma_e}^2 \cdot \boldsymbol{I}_{n_i}$
  - structure depends on the estimates for ${\sigma_{\upsilon 0}}^2$, ${\sigma_{\upsilon 1}}^2$, and $\sigma_{\upsilon 01}$, but *usually* the variances increase for later time points and correlations decrease when time points are further apart
  - this is exactly what we observed for our data set, so this model might fit the data quite well

Covariance patterns linear mixed models

Random intercept + random slope model (fixed time categorical):

continuous time in random part and categorical time in the fixed part

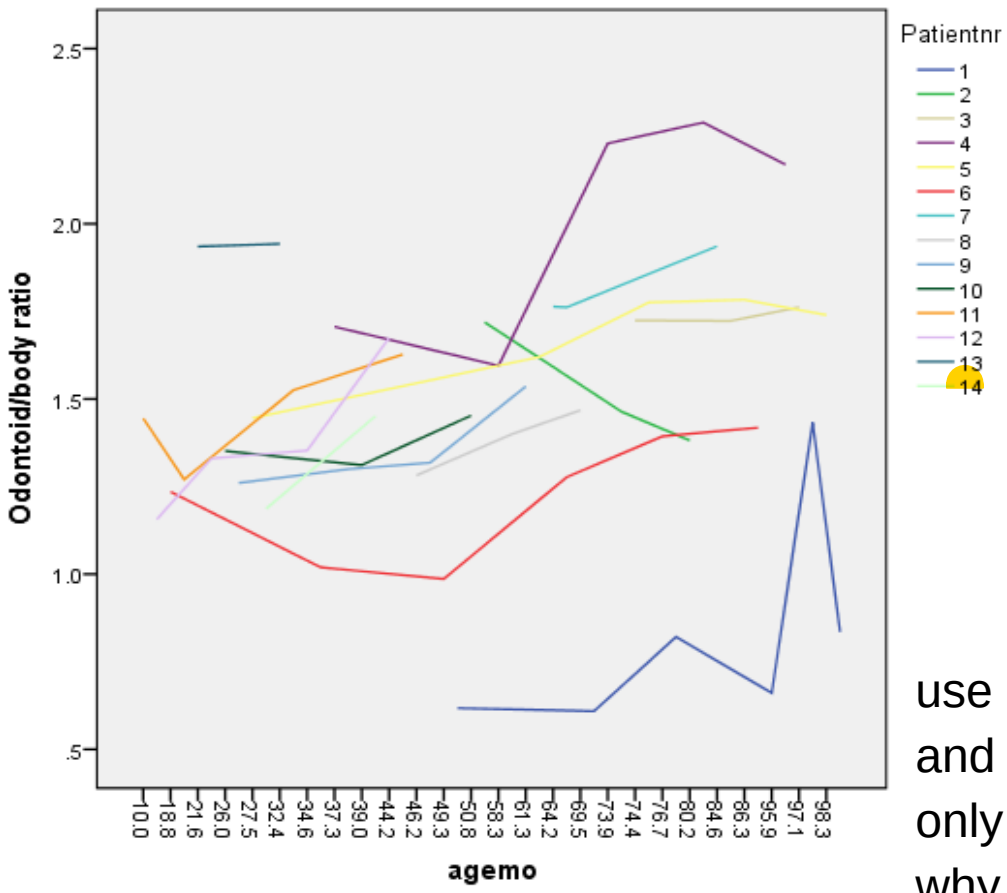
Marginal variance covariance matrix

	1	2	3	4	5	6
1	23.8600	10.239	8.838	7.4364	6.0349	4.6334
2	10.2390	23.134	11.590	12.2660	12.9420	13.6170
3	8.8380	11.590	26.562	17.0960	19.8480	22.6010
4	7.4364	12.266	17.096	34.1440	26.7550	31.5840
5	6.0349	12.942	19.848	26.7550	45.8800	40.5680
6	4.6334	13.617	22.601	31.5840	40.5680	61.7700

does not make a lot of sense to do

but one would get a reasonable fit of the model

Back to Stoop, et al.



- How would we model data from Stoop, et al.?
  - time: discreet or continuous?
  - LME or CPM?
  - time: linear? quadratic?
  - Theory vs. practice....

use a random effect for time and time squared and use parabola for time - problem because most only have two or three measurements which is why it is hard to use par.

What to do with baseline measurement?

usually baseline measure before randomizing the patients

- In clinical trials, a baseline measurement of outcome often taken before randomization
  - Is baseline an "outcome"? because outcome of trial is everything AFTER randomization
    - yes: use as first outcome measurement in mixed model?
    - no: ignore?
    - no: use as covariate in model? and not as an outcome
- In an observational study, there is no experimental intervention not a randomization issue
  - Usually then "baseline" is the first of the measured outcomes
- In Reisby example, baseline HDRS is before patients are treated, but there is no randomization
  - Is baseline HDRS an outcome? or should be use it as a covariate?

Reisby Example, use baseline HDRS as covariate

what if we take it out of the outcome and use it as a covariate

- Select only time >0, use hdrs.base as covariate in mixed model with random intercept, random slope, fixed time & endo
- What do you expect will happen to:
  - ...the estimate of the fixed intercept?
  - ...the estimate of the fixed effect of endo?
  - ...the estimate of the fixed effect of time?

fixed intercept is now average depression score at time zero with baseline zero

average depression score would be lower

difference between endo and exdo

would expect it to disappear but really it only reduces very little

stays pretty much the same

Reisby Example, use baseline HDRS as covariate

controlling for where the treatment started (baseline depression score)

- Select only time >0, use hdrs.base as covariate in mixed model with random intercept, random slope, fixed time & endo
- What do you expect will happen to:
  - ...the variance of the random intercept?

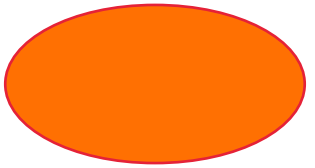
explains variation between individuals and reduces variance between intercepts

  - ...the variance of the random slope?

fixed slope remains the same and variation around slope remains the same but intercept (variation form person to person) is reduced

when I know where they started I explain variation so between ind. variation is contributed to the baseline score - so the differences among people can be explained better

Summary longitudinal data



- Longitudinal data is a specific form of multilevel data
  - measurements within patients, challenge is in modelling time properly
- Time can be continuous or discrete
  - discrete: everyone measured at a few specific time points
    - but, with 3+ measurements per person and approximately linear time trends, you could still consider modelling data as continuous
  - continuous: measurements at different times for different individuals
- We can account for correlation of measurements over time and the covariance
  - explicitly: variance-covariance matrix of residuals (CPMs)
    - primarily when everyone (theoretically) measured at same time points
  - implicitly: random intercept, random slope for time (LMEs) using random effects
  - (both explicitly & implicitly: LMEs with autocorrelated errors)
- "Baseline" measurement of outcome has different meaning depending on study design