Lecture 10

Clustered Longitudinal Studies

Common in Public Health

- 3^+-level: time $\subset$ person $\subset$ family ... $\subset$ city ...

3-level: " $\subset$ tooth $\subset$ mouth/person

Lecture 10 Outline

- Today we will review clustered longitudinal studies

- Basic feature of this data: Three levels
  - Clusters (level 3)
  - Units within clusters (level 2)
  - Time within Units (level 1)

- Apply random effects models
  - May include multiple intercepts
  - May define random slopes (these can quantify differences across level 2 or level 3 units).

Common Design

- Cluster randomized trial
  - Treatment is randomized to the village, medical center, school, etc.
  - Outcome is measured at the individual level over time

- Longitudinal school-based study
  - Randomly selected schools
  - Students nested within schools
  - Students are assessed at multiple time points during the academic year.

Example: Dental Veneer Data

- Chapter 7 from West B, Welch K and Galecki A. Linear Mixed Models: A Practical Guide Using Statistical Software. Chapman Hall / CRC Press, first edition, 2006.

- Ocampo, J. Data from M.S. Thesis, Effect of Porcelain Laminate Contour on Gingival Inflammation, University of Michigan School of Dentistry, 2005.

Example: Dental Veneer Data

- Study aimed to investigate the impact of veneer placement on subsequent gingival (gum) health among adult patients

- Ceramic veneers are commonly applied to hide discoloration

- The treatment process involved removing some of the surface of each treated tooth, and then attaching the veneer to the tooth with an adhesive

ouch!

- Target was to place the veneer at the same contour as the original tooth.

Model 1

- Level 1 (Time) model

time (k) for tooth (ij)

$$Y_{ijk} = \beta_{0ij} + \beta_{1ij} \times I(Time_{ijk} = 6) + \varepsilon_{ijk}$$

To allow 6 mo + 3 mo to have different results

$$\varepsilon_{ijk} \sim N(0, \sigma^2)$$

- Level 2 (Tooth) model

tooth (ij) for mouth (i); no time (k)

$$\beta_{0ij} = \beta_{0i} + \beta_2 \times CDA_{ij} + b_{0ij}$$

$$\beta_{1ij} = \beta_{1i} + \beta_3 \times CDA_{ij}$$

$b_{0ij} \sim N(0, \tau^2)$, independent of ε_{ijk}

- Level 3 (patient) model

mouth (i); no time/tooth (jk)

$$\beta_{0i} = \beta_0 + \beta_4 \times Age_i + b_{0i}$$

$$\beta_{1i} = \beta_1 + \beta_5 \times Age_i + b_{1i}$$

$b_{0i} \sim N(0, {\gamma_0}^2)$, $b_{1i} \sim N(0, {\gamma_1}^2)$, $Corr(b_{0i}, b_{1i}) = \rho$

b_{0i}, b_{1i} independent of $\varepsilon_{ijk}, b_{0ij}$

Alternative expression of Model 1

- You can combine the level-1, level-2 and leve-3 models on the prior slide into a single formula

$$\begin{aligned} \boxed{Y_{ijk}} = (\beta_0 + \boxed{b_{0i} + b_{oij}}) + (\beta_1 + \boxed{b_{1i}}) \times I(Time_{ijk} = 6) + \\ \beta_2 \times CDA_{ij} + \beta_3 \times CDA_{ij} \times I(Time_{ijk} = 6) + \\ \beta_4 \times age_i + \beta_5 \times age_i \times I(Time_{ijk} = 6) + \boxed{\varepsilon_{ijk}} \end{aligned}$$

[ ] - fixed;

O - random effect

- From the above, you can see clearly how the random intercept for patient and tooth are incorporated

- Also, you can see that we will allow for time-specific main effects at the patient level.

Example: Dental Veneer Data

- Goal of the study: determine whether differing amounts of contour difference (CDA) due to placement of the veneer might affect gingival health in the treated teeth over time.

- Gingival health was the amount of gingival crevicular fluid (GCF) in pockets of the gum adjacent to the treated teeth.
  - Production of GCF is associated with inflammatory process
  - Higher values here mean poorer gum health
  - GCF was measured for each tooth at baseline and then 3 months and at 6 months post treatment.

- Available data are a subset of 55 treated teeth located in the maxillary arches of 12 patients.

2 obs on (changes)

teeth from 12 patients

Example: Dental Veneer Data

change from base

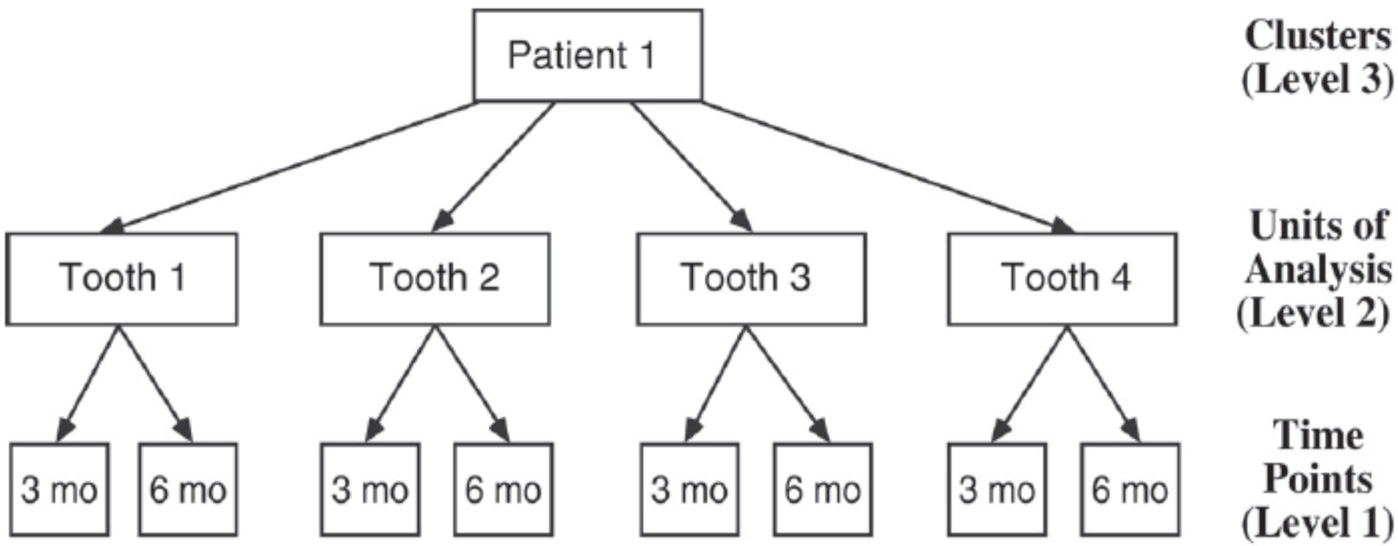


FIGURE 7.1
Structure of the clustered longitudinal data for the first patient in the Dental Veneer data set.

Example: Dental Veneer Data

patient : Patient ID variable (Level 3 ID)

tooth : Tooth number (Level 2 ID)

age : Age of patient when veneer was placed, constant for all observations on the same patient

base.gcf : Baseline measure of Gingival Crevicular Fluid for the tooth, constant for all observations on the same tooth

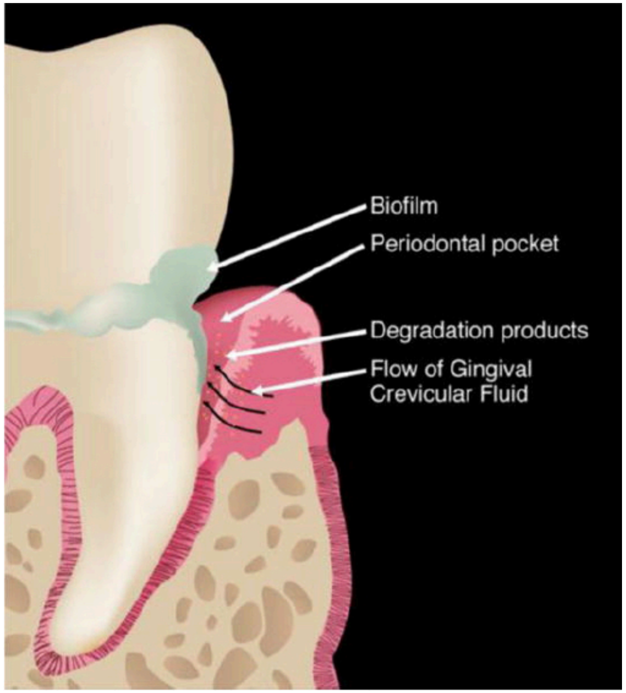
cda : Average contour difference in the tooth after veneer placement, constant for all observations on the same tooth

Smaller is better?

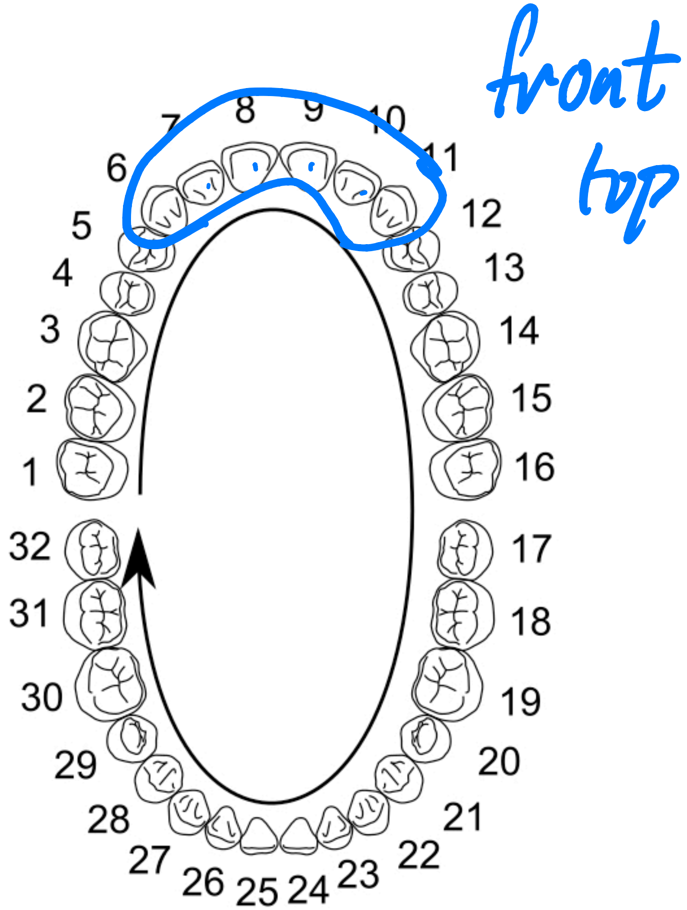
time : Time points of longitudinal measures (3 = Three Months, 6 = Six Months)

gcf : Gingival Crevicular Fluid adjacent to the tooth, collected at each time point (Dependent Variable)


DECAY AND CROOKED TEETH
OMG!
TEETH
TRANSFORM
Before
After
ley Street Smile Clinic



Gingival crevicular fluid (GCF) is an inflammatory exudate that can be collected at the **gingival** margin or within the **gingival** crevice. The biochemical analysis of the **fluid** offers a non-invasive means of assessing the host response in periodontal disease.



Universal numbering system. This is a dental practitioner view, so tooth number 1, the rear upper tooth on the patient's right, appears on the left of the chart.

Example: Dental Veneer Data

	patient	tooth	age	base_gcf	cda	time	gcf
1.	1	6	46	17	4.666667	3	11
2.	1	6	46	17	4.666667	6	68
3.	1	7	46	22	4.666667	3	13
4.	1	7	46	22	4.666667	6	47
5.	1	8	46	18	5	3	14
6.	1	8	46	18	5	6	58
7.	1	9	46	12	3.333333	3	10
8.	1	9	46	12	3.333333	6	57
9.	1	10	46	10	8.666667	3	14
10.	1	10	46	10	8.666667	6	44
11.	1	11	46	17	5.666667	3	11
12.	1	11	46	17	5.666667	6	53
13.	3	6	32	3	7.666667	3	28
14.	3	6	32	3	7.666667	6	23

Example: Dental Veneer Data

- Outline of exploratory analysis
  - Number of teeth per patient
  - Each patient has GCF measured at 3 and 6 months, so complete data within patient/tooth
  - Summarize patient characteristics
  - Summarize tooth characteristics and how they vary within/between patient
  - Explore GCF over time by patient and tooth (we will also look at change in GCF)
  - Explore the relationship between change in GCF and CDA/age separately for each follow-up time

- After this exploratory analysis, we will specify some models

Dental Veneer Data: Descriptive Stats

1 + 4 + 8 + 42 = 55 teeth

Patients have 1 (n = 1), 2 (n = 2), 4 (n = 2) or 6 (n = 7) teeth. or

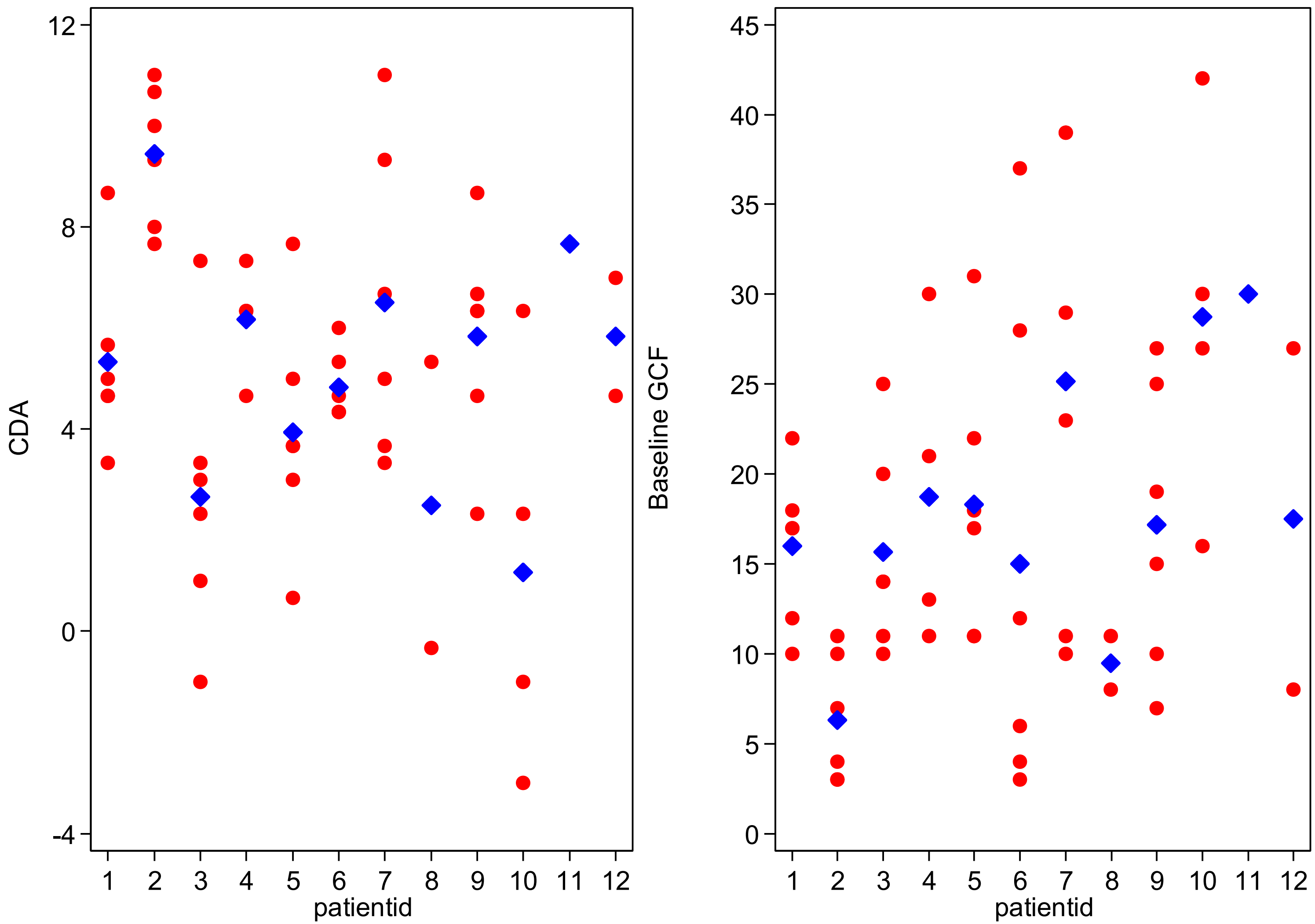
12 people

```
* Patient level characteristics
sum age base_gcf
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         age |         12          41    12.64192         21         62
```

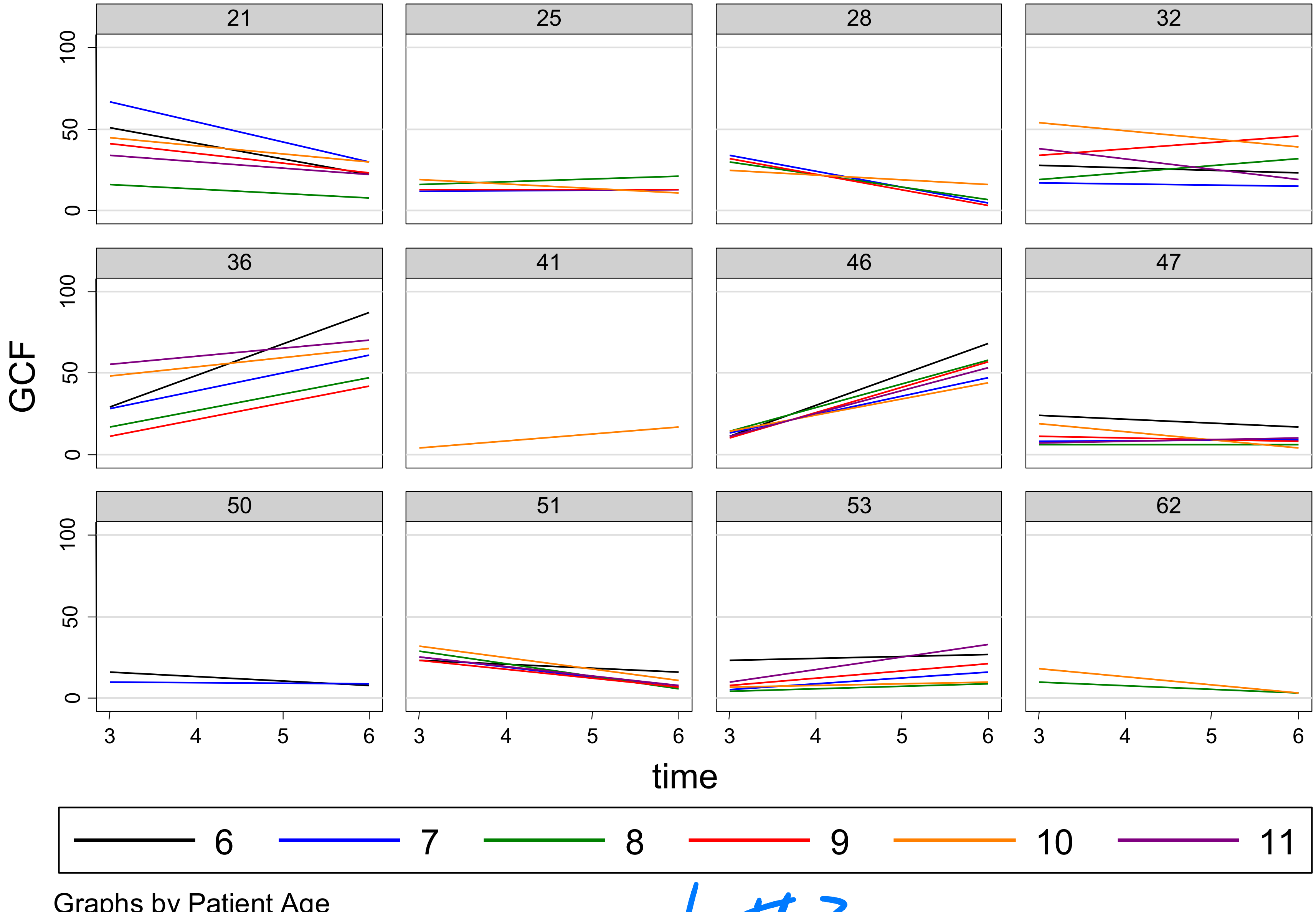
```
 * Tooth level characteristics
 sum cda if tooth_counter==1

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
    base_gcf |         55    17.38182    10.08573          3         42
         cda |        55    5.181818    3.034655         -3         11
```

Teeth level characteristics

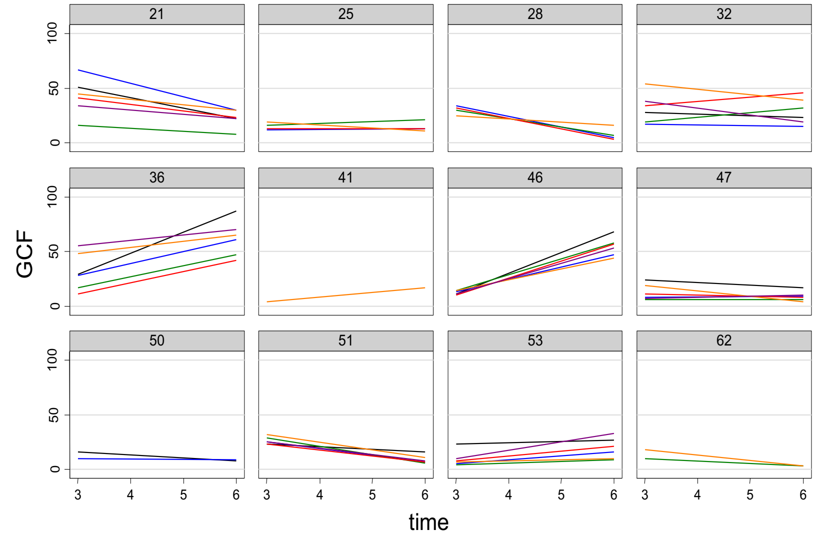


? Sources of variation : time, tooth, mouth (person)



tooth ?

Overall patterns: ignoring covariates

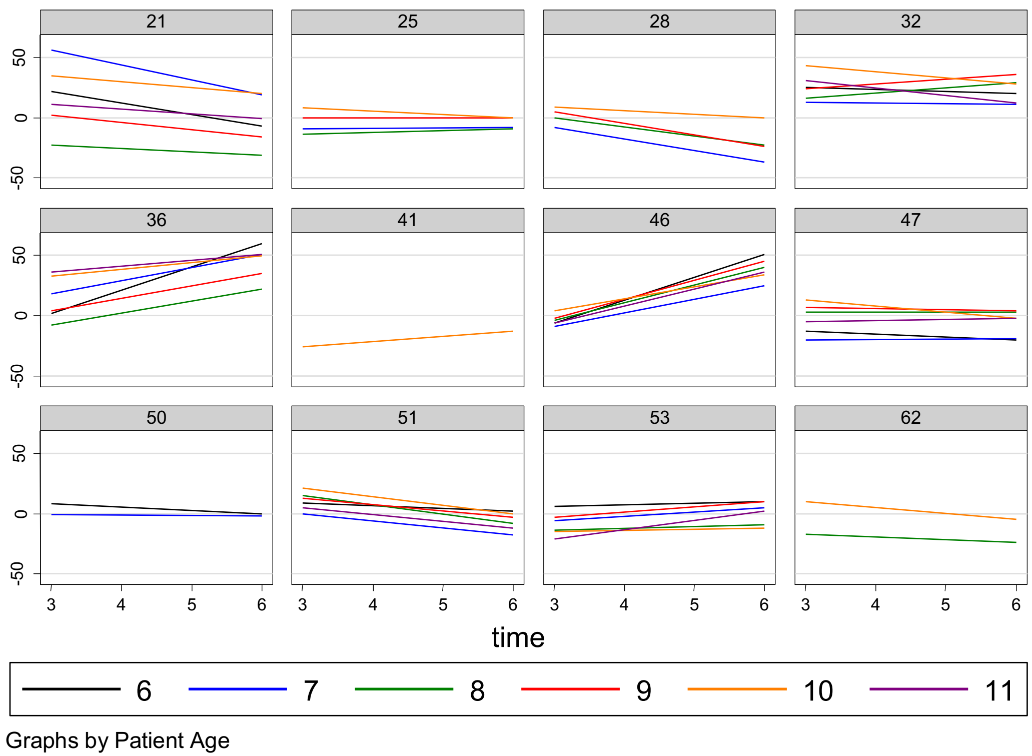


1. Within a patient, the changes over time across teeth are roughly the same (parallel lines within a patient)
2. Patient changes over time are very different; some increase, some decrease, some stay the same.
3. 1 and 2 suggest a random patient intercept and random slope for time at the patient level
4. Within a patient, there are differences across the teeth (roughly parallel lines), suggesting the need for a random intercept for teeth nested within patient.
5. Variation may be different at 3 months and 6 months

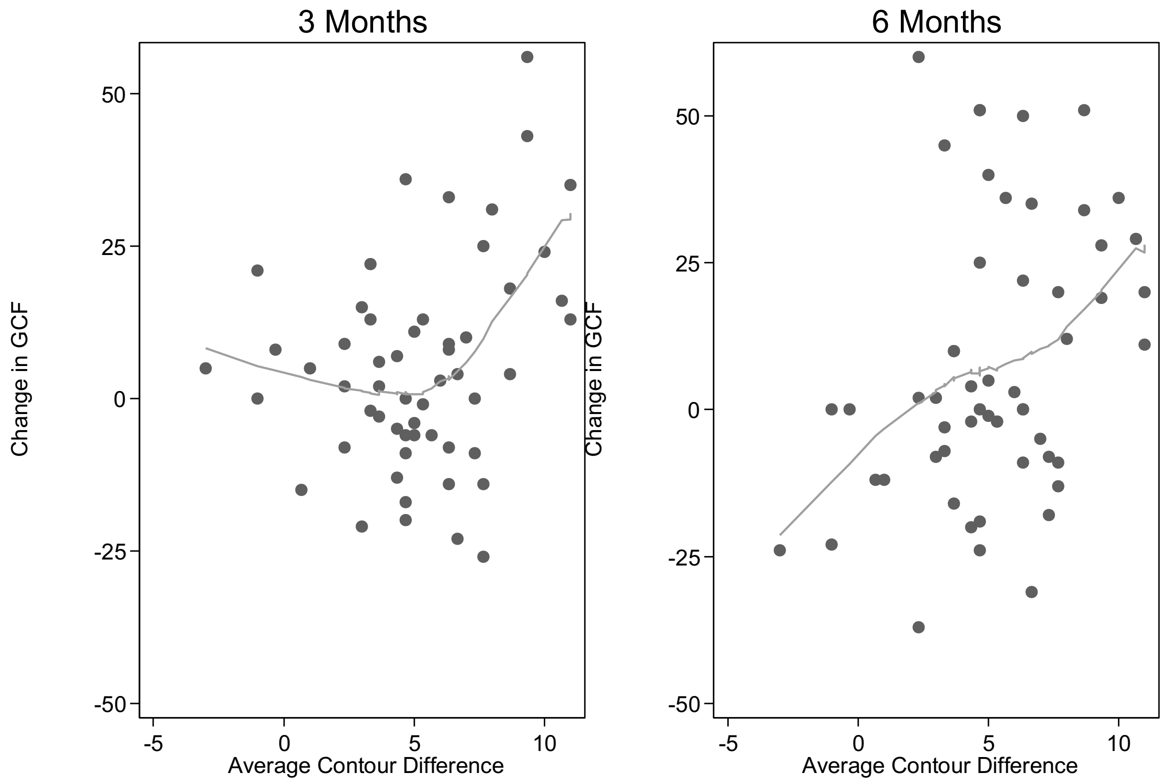
Models to consider

- NOTE: In the text where I extracted this example, the analysis is done on the GCF at 3 and 6 months after adjusting for baseline GCF, CDA and age (separately for each time).

- Let's simplify a bit and look at change in GCF relative to baseline.

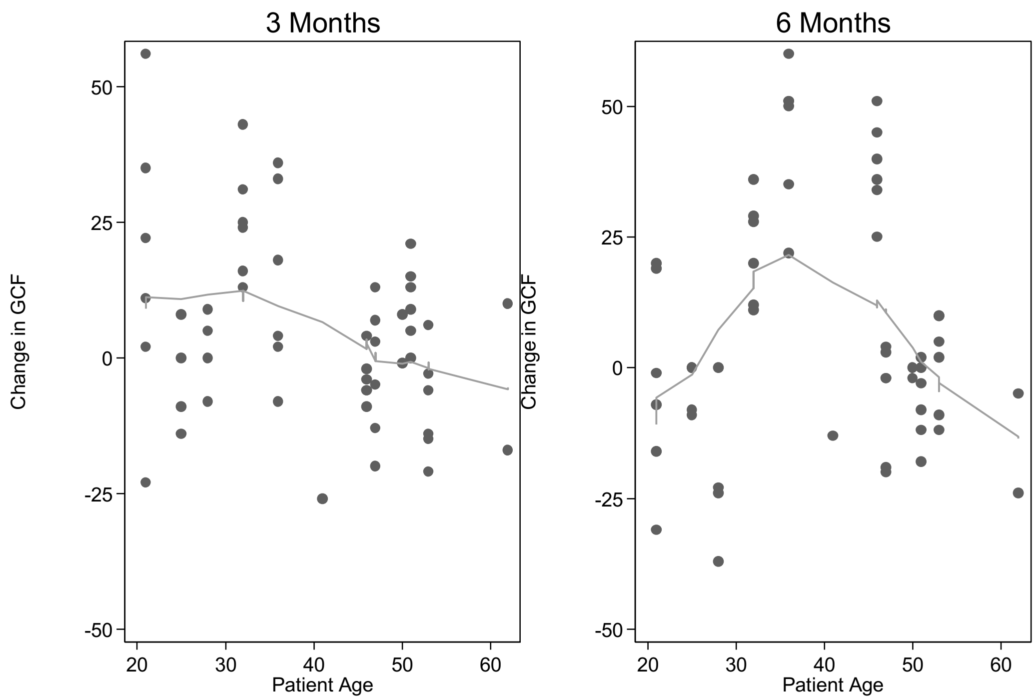


Change in GCF vs. CDA



The relationship between change in GCF and CDA appears to be different at 3 months and 6 months, demonstrating the need for an interaction between CDA and time in our model.

Change in GCF vs. Age



The relationship between change in GCF and age appears to be different at 3 months and 6 months, demonstrating the need for an interaction between age and time in our model.

Summary of exploratory analysis

- Fixed effects:
  - Time
  - CDA
  - Age
  - Interactions: Time x CDA and Time x Age

- Random effects:
  - Random intercept for Patient
  - Random slope for time at the Patient level
  - Random intercept for Tooth

Alternative expression of Model 1

- You can also think of Model 1 as specifying two models; one for the 3-month and 6-month measure of change in GCF

- At 3-months:

$$Y_{ijk} = (\beta_0 + b_{0i} + b_{oij}) + \beta_2 \times CDA_{ij} + \beta_4 \times age_i + \varepsilon_{ijk}$$

- At 6-months:

$$Y_{ijk} = (\beta_0 + b_{0i} + b_{oij}) + (\beta_1 + b_{1i}) + (\beta_2 + \beta_3) \times CDA_{ij} + (\beta_4 + \beta_5) \times age_i + \varepsilon_{ijk}$$

```
. mixed change sixM cda age time_cda time_age || patientid: sixM, //
cov(unstruct) || tooth: , ml stddev

Mixed-effects ML regression                     Number of obs      =       110

-----------------------------------------------------------
                |   No. of       Observations per Group
 Group Variable |   Groups    Minimum    Average    Maximum
----------------+------------------------------------------
      patientid |       12          2        9.2         12
          tooth |       55          2        2.0          2
-----------------------------------------------------------

                                                Wald chi2(5)       =      5.52
Log likelihood =  -441.0811                     Prob > chi2        =    0.3555

------------------------------------------------------------------------------
      change |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        sixM |  -11.35129   19.66105    -0.58   0.564    -49.88625    27.18366
         cda |   .9343684   .7182499     1.30   0.193    -.4733755    2.342112
         age |  -.3726317   .2414621    -1.54   0.123    -.8458887    .1006252
    time_cda |   .1555184   .6162395     0.25   0.801    -1.052289    1.363326
    time_age |   .2787436   .4508155     0.62   0.536    -.6048386    1.162326
       _cons |   14.19252   11.32596     1.25   0.210    -8.005961      36.391
------------------------------------------------------------------------------
```

fixed effects results

change	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
sixM	-11.35129	19.66105	-0.58	0.564	-49.88625	27.18366
cda	.9343684	.7182499	1.30	0.193	-.4733755	2.342112
age	-.3726317	.2414621	-1.54	0.123	-.8458887	.1006252
time_cda	.1555184	.6162395	0.25	0.801	-1.052289	1.363326
time_age	.2787436	.4508155	0.62	0.536	-.6048386	1.162326
_cons	14.19252	11.32596	1.25	0.210	-8.005961	36.391

Main effect of time:

- The change in GCF is expected to decrease by -11.35 comparing 6 months to 3 months among teeth with no average contour difference and patients with age 0.

Interpret the main effect of CDA and Age:

- At 3-months, the change in GCF increases by 0.93 (95% CI: -0.47, 2.34) per unit increase in the average contour difference, among patients with the same age.
- At 3-months, the change in GCF decreases by -0.37 (95% CI: -0.85, 0.10) per year increase in patient age, among teeth with similar average contour difference.

change	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
sixM	-11.35129	19.66105	-0.58	0.564	-49.88625	27.18366
cda	.9343684	.7182499	1.30	0.193	-.4733755	2.342112
age	-.3726317	.2414621	-1.54	0.123	-.8458887	.1006252
time_cda	.1555184	.6162395	0.25	0.801	-1.052289	1.363326
time_age	.2787436	.4508155	0.62	0.536	-.6048386	1.162326
_cons	14.19252	11.32596	1.25	0.210	-8.005961	36.391

Interpret the main effect of CDA and Age:

- At 3-months, the change in GCF increases by 0.93 (95% CI: -0.47, 2.34) per unit increase in the average contour difference, among patients with the same age.
- At 3-months, the change in GCF decreases by -0.37 (95% CI: -0.85, 0.10) per year increase in patient age, among teeth with similar average contour difference.

Interpret the interaction of Time and CDA and Age:

- At 6-months, the change in GCF increases by 0.93 + 0.16 per unit increase in the average contour difference, among patients with the same age.
- At 6-months, the change in GCF decreases by -0.10 (-0.37 + 0.27) per year increase in patient age, among teeth with similar average contour difference.

```
------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
patientid: Unstructured      |
                    sd(sixM) |   18.11874   3.954094      11.81322    27.78995
                   sd(_cons) |   7.449773   2.868992      3.502158    15.84712
            corr(sixM,_cons) |  -.2376063   .3683718     -.7646976    .4799816
-----------------------------+------------------------------------------------
tooth: Identity              |
                   sd(_cons) |   10.98997   1.467394      8.459465    14.27742
-----------------------------+------------------------------------------------
                sd(Residual) |   7.003357   .7549124      5.669604     8.65087
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(4) =    77.92   Prob > chi2 = 0.0000
```

What about different residual variance by time?

```
mixed change sixM cda age time_cda time_age || patientid: sixM, cov(unstruct) || tooth:
, residuals(ind, by(time)) ml stddev

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
patientid: Unstructured      |
                    sd(sixM) |   18.09322   3.941528      11.80553    27.72978
                   sd(_cons) |   6.906742   3.033584      2.920169    16.33573
            corr(sixM,_cons) |  -.1999545   .3946216     -.7650749    .5391599
-----------------------------+------------------------------------------------
tooth: Identity              |
                   sd(_cons) |   10.98127   1.449983       8.47732    14.22481
-----------------------------+------------------------------------------------
Residual: Independent,       |
    by time                  |
                    3: sd(e) |   8.823118   1.394463      6.472808    12.02684
                    6: sd(e) |   4.551194   2.008761      1.916154    10.80987
------------------------------------------------------------------------------
LR test vs. linear regression:      chi2(5) =    80.41   Prob > chi2 = 0.0000
```

LRT comparing this model to Model1, p-value = 0.1142

~~No~~ Insufficient evidence that we need time-specific residual variance to conclude.

What about simplifying the fixed effects?

```
quietly mixed change sixM cda age time_cda time_age || patientid: sixM, \\\
cov(unstruct)  || tooth: , ml
est store model1ML

quietly mixed change sixM cda age || patientid: sixM, cov(unstruct) || tooth: , ml
est store model3ML

lrtest model3ML model1ML

Likelihood-ratio test                                 LR chi2(2)  =      0.42
(Assumption: model3ML nested in model1ML)             Prob > chi2 =    0.8126
```

Model without interactions

```
. mixed change sixM cda age || patientid: sixM, cov(unstruct) || tooth: , ml stddev

                                                Wald chi2(3)      =      5.11
Log likelihood = -441.28862                     Prob > chi2       =    0.1638

------------------------------------------------------------------------------
      change |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        sixM |   .8282559   5.567838     0.15   0.882    -10.08451    11.74102
         cda |   1.000847   .6734638     1.49   0.137    -.3191174    2.320812
         age |  -.3341484   .2337244    -1.43   0.153    -.7922398     .123943
       _cons |   12.30153   10.91347     1.13   0.260    -9.088483    33.69154
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
patientid: Unstructured      |
                    sd(sixM) |   18.54124   4.001472       12.1461    28.30353
                   sd(_cons) |   7.444146   2.872382      3.494396    15.85834
             corr(sixM,_cons) |  -.2493583   .3710917     -.7740198    .4783336
-----------------------------+------------------------------------------------
tooth: Identity              |
                   sd(_cons) |   10.99522   1.467373      8.464602    14.28241
-----------------------------+------------------------------------------------
                sd(Residual) |   6.998187   .7534467       5.66686    8.642284
------------------------------------------------------------------------------
LR test vs. linear regression:       chi2(4) =    81.11   Prob > chi2 = 0.0000
```

Example: Dental Veneer Data

- Analysis was guided by:
  - Patterns within the level 3 unit (patients); i.e. similar changes over time within a patient, very different changes over time across patients. Led us to include random intercept and random slope for time
  - Patterns within a level 3 unit (patients) between level 2 units (teeth): lines were parallel within a patient, led us to include a random intercept for tooth random slope?
  - Appeared to be greater variation in change in GCF at 3-months compared to 6-months, led us to consider time specific residual variance (not necessary)

- Fixed effects were included that:
  - Used exploratory analysis of the data; appeared there may be interactions: time x CDA and time x Age
  - The goal of the study!

General comments on analysis strategy

- Similar EDA as performed in this example.
- May start by considering random intercept for level 3 and level 2 units.
- Explore whether additional correlation structure exists within level 2 units over time
  - Use ACF/variogram/lorelogram as we did in LDA course
- Random slopes for time can be added at either level 3 or level 2 depending on what you see in the data.

- Fixed effects:
  - Think carefully here
  - Depends on the level that the covariate is measured.