STAT3401: Advanced data analysis
Week 10: Models for Clustered Longitudinal Data

Berwin Turlach
School of Mathematics and Statistics
Berwin.Turlach@gmail.com

The University of Western Australia

Models for Longitudinal Data (WWG Chapter 7)

We illustrate fitting LMMs to *clustered longitudinal data*

- *Unit of analysis* are nested with *clusters*
- *Repeated measures* are collected on the units of analysis over time
- Such data can be considered to have three levels

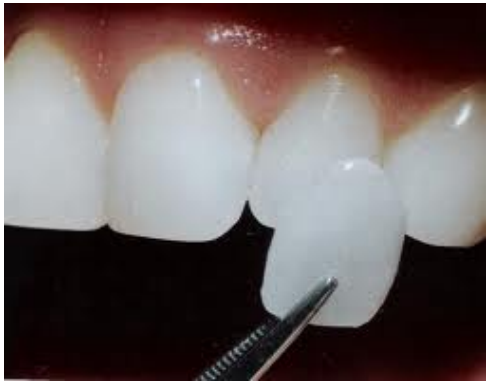
Examples of clustered longitudinal data (WWG 7.1)

TABLE 7.1

Examples of Clustered Longitudinal Data in Different Research Settings

Level of Data		Research Setting		
		Environment	**Education**	**Dentistry**
Cluster of Units (Level 3)	**Cluster ID variable (random factor)**	**Plot**	**Classroom**	**Patient**
	Covariates	Soil minerals, tree crown density in the plot	Teacher years of experience, classroom size	Gender, age
Unit of Analysis (Level 2)	**Unit of Analysis ID variable (random factor)**	**Tree**	**Student**	**Tooth**
	Covariates	Tree size	Gender, age, baseline score	Treatment, tooth type
Time (Level 1)	**Time variable**	**Week**	**Marking period**	**Month**
	Dependent variable	Oxygen yield	Test score	Gingival crevicular fluid (GCF)
	Time-varying covariates	Sunlight exposure, precipitation	Attendance	Frequency of tooth brushing

Clustered Longitudinal Data—The Dental Veneer Study



- Data were collected by researchers at the University of Michigan Dental School
- Study aimed to investigate the impact of veneer placement on subsequent gingival (gum) health among adult patients
- Ceramic veneers were applied to selected teeth 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
- The veneer was placed to match the original contour of the tooth as closely as possible.

Clustered Longitudinal Data—The Dental Veneer Study (ctd)

- The investigators were interested in studying whether differing amounts of contour difference (CDA) due to placement of the veneer might affect gingival health in the treated teeth over time.
- One measure of gingival health was the amount of GCF in pockets of the gum adjacent to the treated teeth. GCF was measured for each tooth at visits 3 months and at 6 months post treatment.
- A total of 88 teeth in 17 patients were prepared for veneer placement, and a baseline measure of GCF was collected for each tooth
- We consider only the 55 treated teeth located in the maxillary arches of 12 patients

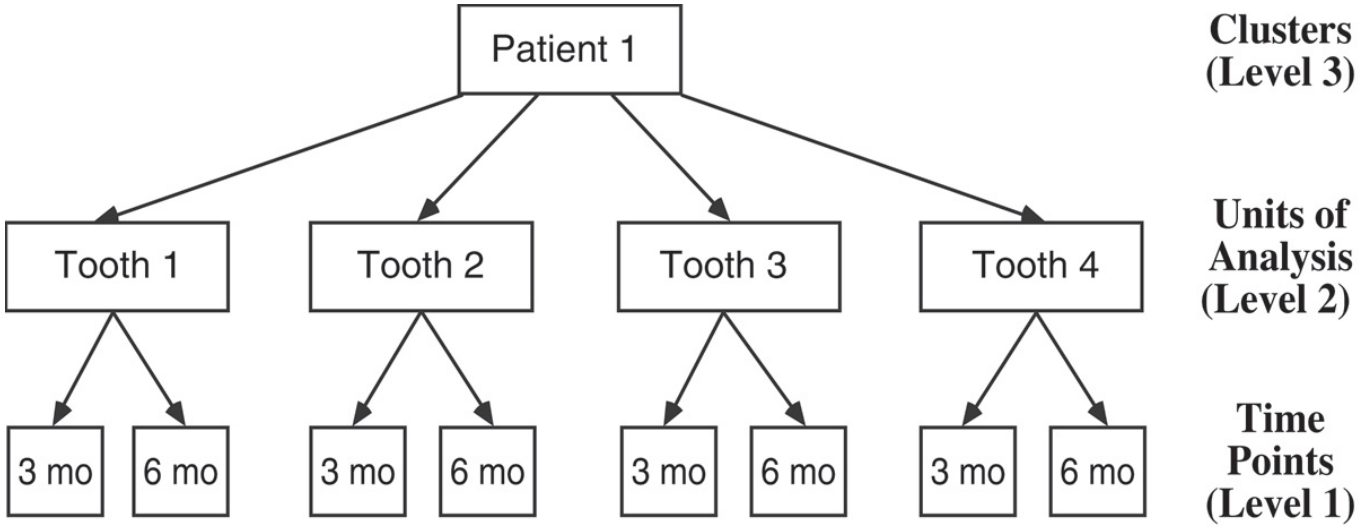


FIGURE 7.1
Structure of the clustered longitudinal data for the first patient in the Dental Veneer data set.

Structure of the data

Patient (Level 3) Variables

PATIENT = Patient ID variable (Level 3 ID)

AGE = Age of patient when veneer was placed, constant for all observations on the same patient

Tooth (Level 2) Variables

TOOTH = Tooth number (Level 2 ID)

BASE_GCF = Baseline measure of GCF 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

Time-Varying (Level 1) Variables

TIME = Time points of longitudinal measures (3 = 3 Months, 6 = 6 Months)

GCF = Gingival crevicular fluid adjacent to the tooth, collected at each time point (dependent variable)

Basic data manipulation

```
library(nlme)
library(lattice)
```

```
veneer <- read.table("PATH_TO_DATA/veneer.dat", header = TRUE)
```

```
sapply(veneer, data.class)
```

```
##   patient     tooth       age  base_gcf       cda      time       gcf
## "numeric" "numeric" "numeric" "numeric" "numeric" "numeric" "numeric"
```

```
head(veneer)
```

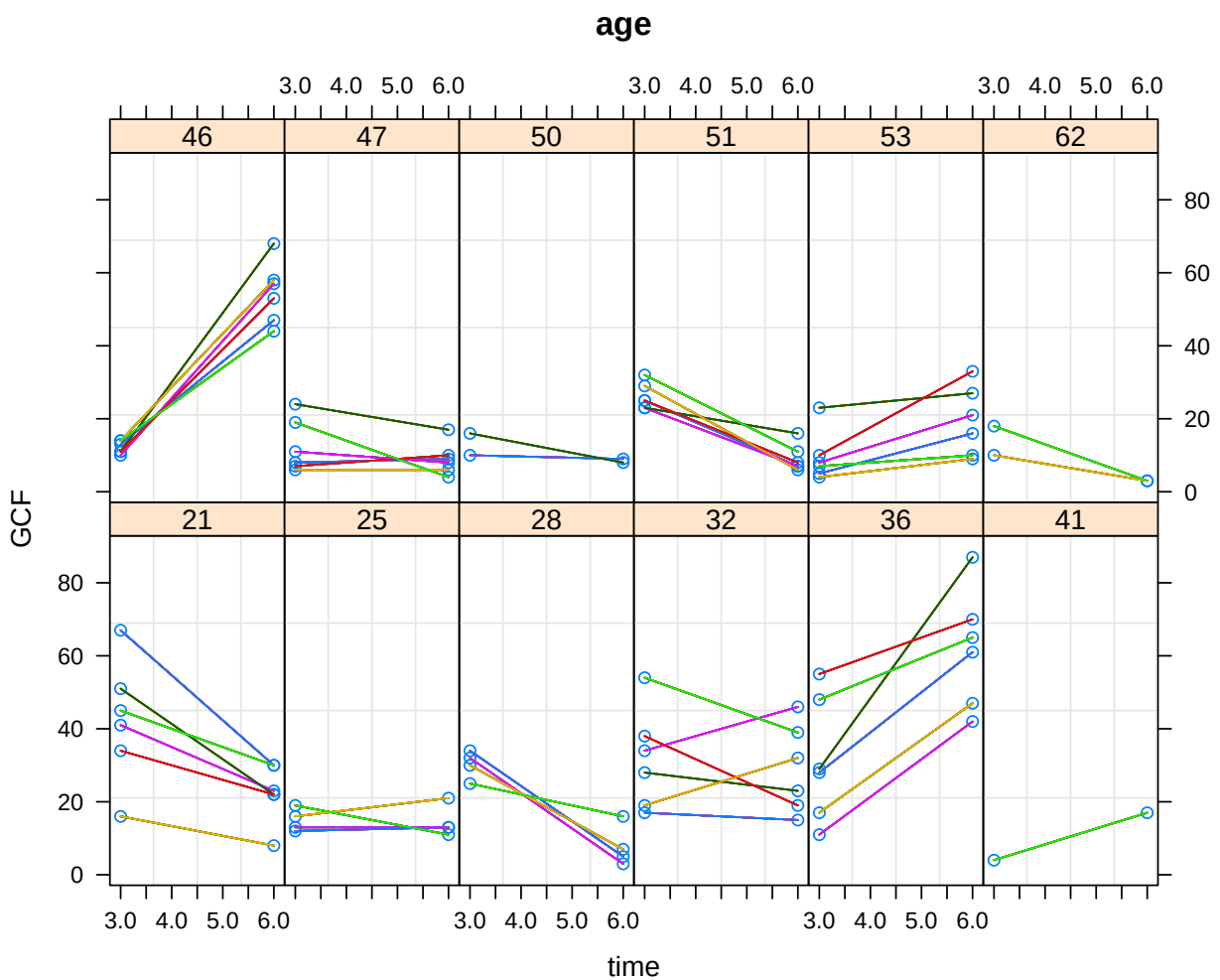
```
##   patient tooth age base_gcf   cda time gcf
## 1       1     6  46       17 4.667    3  11
## 2       1     6  46       17 4.667    6  68
## 3       1     7  46       22 4.667    3  13
## 4       1     7  46       22 4.667    6  47
## 5       1     8  46       18 5.000    3  14
## 6       1     8  46       18 5.000    6  58
```

```
veneer <- within(veneer, {
    age.f <- factor(age)
    time.f <- factor(time)
    tooth.f <- factor(tooth)
})
```

http://www-personal.umich.edu/~bwest/chapter7.html

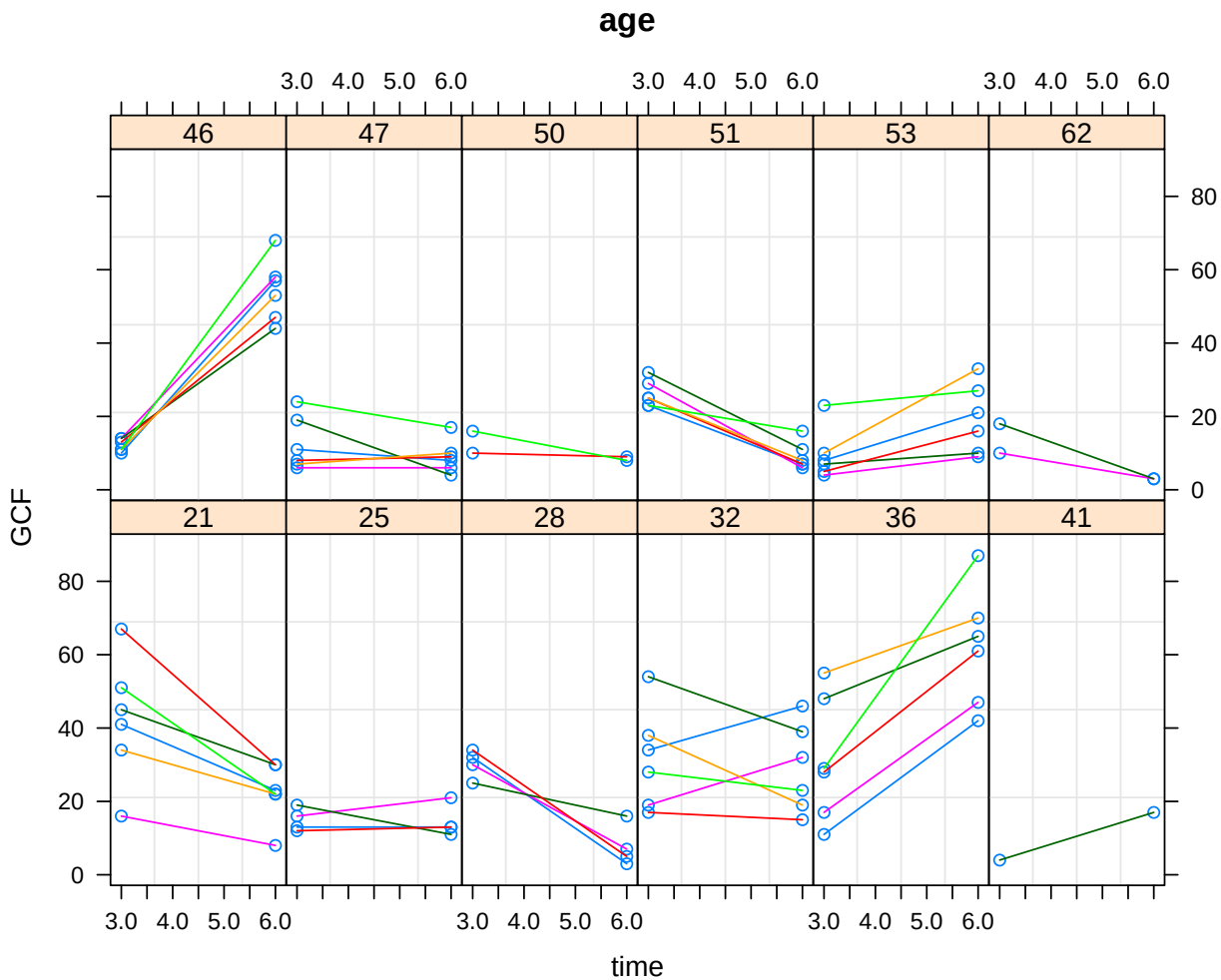
Summarising data graphically

```
veneer.g1 <- groupedData(gcf ~ time | tooth.f, outer = ~age.f, data = veneer)

## Warning:  duplicated levels in factors are deprecated

plot(veneer.g1, outer = TRUE, aspect = 2, key = FALSE, xlab = "time", ylab = "GCF", main = "age",
    layout = c(6, 2))
```



Summarising data graphically (ctd)

```
veneer.g1 <- groupedData(gcf ~ time | tooth.f, data = veneer)
plot(veneer.g1, outer = ~age.f, aspect = 2, key = FALSE, xlab = "time", ylab = "GCF",
    main = "age", layout = c(6, 2))
```



Summarising data—some comments

- Observe that the GCF values for all teeth within a given patient tend to follow the same trend over time (lines are roughly parallel within each patient)
- In some patients, the GCF levels tend to increase, whereas in others the GCF levels tend to decrease or remain relatively constant over time; this pattern suggests that an appropriate model for the data might include random patient-specific time slopes.
- The GCF levels of the teeth also tend to differ by patient, suggesting that a model should also include random patient-specific intercepts
- There is also evidence in most patients that the level of GCF tends to differ by tooth, suggesting that we may want to include random tooth-specific intercepts in the model.

Dental veneer data analysis—overview

We follow a top-down modelling strategy

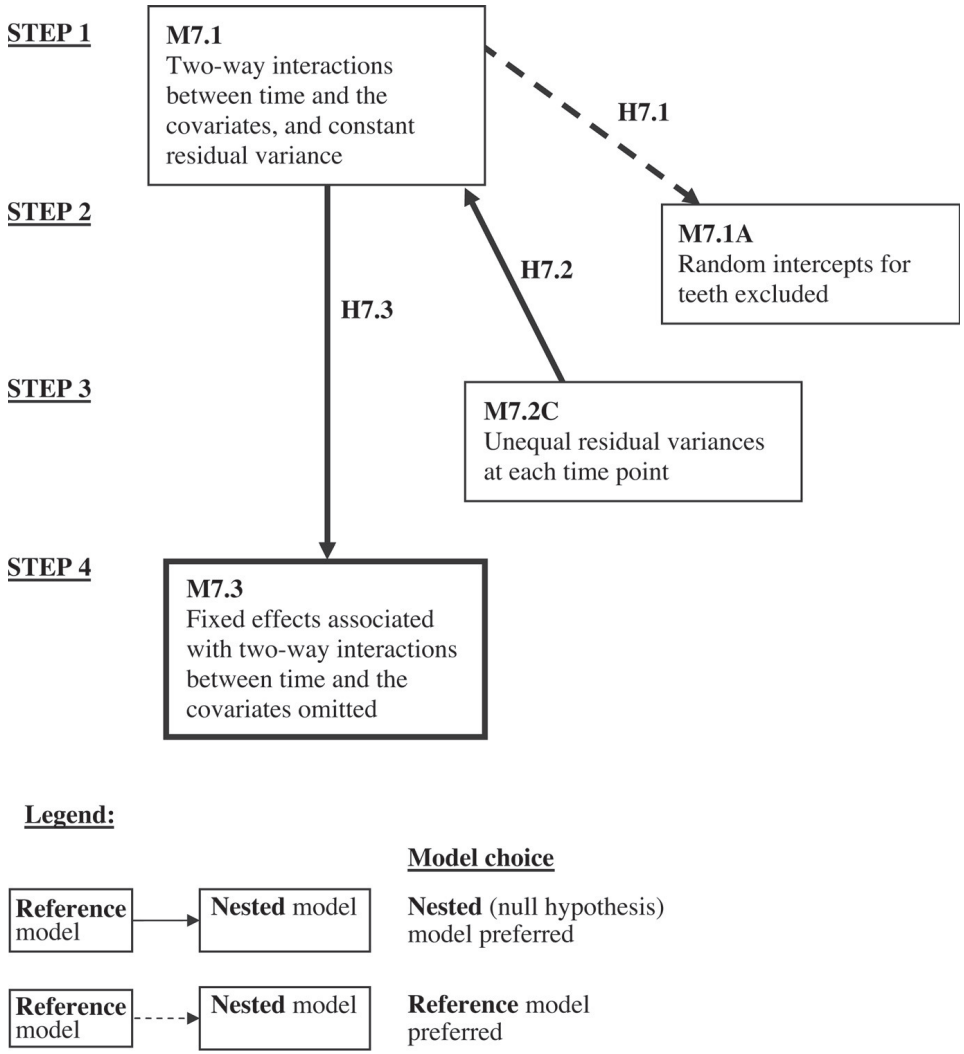


FIGURE 7.3
Guide to model selection and related hypotheses for the analysis of the Dental Veneer data.

Dental veneer data analysis—top down modelling strategy

- Fit a model with loaded mean structure
- Select a structure for the random effects
- Select a structure for the residual covariance structure
- Reduce the model by removing non-significant fixed effects and check model diagnostics

Model Specification (WWG 7.3.2)

The general model specification for an individual response GCF_{tij}, on tooth i nested within patient j at visit t ($t = 1, 2$), is:

$$\begin{aligned} GCF_{tij} = {} & \beta_0 + \beta_1 \times TIME_t + \beta_2 \times BASE\_GCF_{ij} + \beta_3 \times CDA_{ij} + \beta_4 \times AGE_j \\ & + \beta_5 \times TIME_t \times BASE\_GCF_{ij} + \beta_6 \times TIME_t \times CDA_{ij} \\ & + \beta_7 \times TIME_t \times AGE_j \\ & + u_{0j} + u_{1j} \times TIME_t + u_{0i|j} + \varepsilon_{tij} \end{aligned}$$

Model Specification (ctd)

Using this model specification we note:

- The parameters β_0 through β_7 represent the fixed effects associated with the intercept, TIME, the patient-level and tooth-level covariates, and their two-way interactions
- u_{0j} and u_{1j} are random patient effects associated with the intercept and time slope, respectively
- $u_{0i|j}$ is the random effect associated with a tooth nested within a patient
- ε_{tij} represents a residual
- We assume that the random effects, u_{0j} and u_{1j}, associated with patients are independent of the random effects, $u_{0i|j}$, associated with teeth nested within patients, and that all random effects are independent of the residuals

Random effects model specification

We assume that the distribution of the random effects associated with patient j, u_{0j} and u_{1j}, is multivariate normal:

$$\mathbf{u}_j = \begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{D}^{(2)}\right),$$

$$\mathbf{D}^{(2)} = \begin{pmatrix} \text{var}[u_{0j}] & \text{cov}[u_{0j}, u_{1j}] \\ \text{cov}[u_{0j}, u_{1j}] & \text{var}[u_{1j}] \end{pmatrix} = \begin{pmatrix} \sigma^2_{int:patient} & \sigma_{int,time:patient} \\ \sigma_{int,time:patient} & \sigma^2_{time:patient} \end{pmatrix}$$

The distribution of the random effects associated with tooth i nested within patient j is assumed to be

$$u_{0i|j} \sim N\left(0, D^{(1)}\right), \qquad D^{(1)} = \text{var}[u_{0i|j}] = \sigma^2_{int:tooth(patient)}$$

Random effects model specification (ctd)

The distribution of the residuals, ε_{tij}, associated with observations on the same tooth is assumed to be multivariate normal:

$$\boldsymbol{\varepsilon}_{ij} = \begin{pmatrix} \varepsilon_{1ij} \\ \varepsilon_{2ij} \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{R}_{ij}\right), \quad \mathbf{R}_{ij} = \begin{pmatrix} \text{var}[\varepsilon_{1ij}] & \text{cov}[\varepsilon_{1ij}, \varepsilon_{2ij}] \\ \text{cov}[\varepsilon_{1ij}, \varepsilon_{2ij}] & \text{var}[\varepsilon_{2ij}] \end{pmatrix}$$

Model 7.1: $\mathbf{R}_{ij} = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} = \sigma^2 \mathbf{I}_2$

Model 7.2A: $\mathbf{R}_{ij} = \begin{pmatrix} \sigma^2_{t1} & \sigma_{t1,t2} \\ \sigma_{t1,t2} & \sigma^2_{t2} \end{pmatrix}$

Model 7.2B: $\mathbf{R}_{ij} = \begin{pmatrix} \sigma^2+\sigma_{t1,t2} & \sigma_{t1,t2} \\ \sigma_{t1,t2} & \sigma^2+\sigma_{t1,t2} \end{pmatrix}$

Model 7.2C: $\mathbf{R}_{ij} = \begin{pmatrix} \sigma^2_{t1} & 0 \\ 0 & \sigma^2_{t2} \end{pmatrix}$

Model specification: Multilevel notation (WWG 7.3.2.2 & 7.4.5)

Specification for Model 7.1:

LEVEL 1 MODEL (Time):

$$GCF_{tij} = \pi_{0ij} + \pi_{1ij} \times TIME_{tij} + \varepsilon_{tij}$$

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

LEVEL 2 MODEL (Tooth):

$$\pi_{0ij} = \beta_{00j} + \beta_{01j} \times BASE\_GCF_{ij} + \beta_{02j} \times CDA_{ij} + r_{0ij}$$
$$\pi_{1ij} = \beta_{10j} + \beta_{11j} \times BASE\_GCF_{ij} + \beta_{12j} \times CDA_{ij}$$

where $r_{0ij} \sim N(0, \sigma^2_{int:tooth(patient)})$, independent of the ε_{tij}

Model specification: Multilevel notation (ctd)

LEVEL 3 MODEL (Patient):

$$
\begin{aligned}
\beta_{00j} &= \gamma_{000} + \gamma_{001} \times AGE_j + u_{00j} \\
\beta_{01j} &= \gamma_{010} \\
\beta_{02j} &= \gamma_{020} \\
\beta_{10j} &= \gamma_{100} + \gamma_{101} \times AGE_j + u_{10j} \\
\beta_{11j} &= \gamma_{110} \\
\beta_{12j} &= \gamma_{120}
\end{aligned}
$$

where $\begin{pmatrix} u_{00j} \\ u_{10j} \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{D}^{(2)}\right)$, independent of the r_{0ij} and ε_{tij}

Summary of models considered for the dental veneer data

TABLE 7.2

Summary of Models Considered for the Dental Veneer Data

			Notation		Model				
		Term/Variable	General	HLM[a]	7.1	7.2A[b]	7.2B[b]	7.2C	7.3
Fixed effects		Intercept	β_0	γ_{000}	√	√	√	√	√
		TIME	β_1	γ_{100}	√	√	√	√	√
		BASE_GCF	β_2	γ_{010}	√	√	√	√	√
		CDA	β_3	γ_{020}	√	√	√	√	√
		AGE	β_4	γ_{001}	√	√	√	√	√
		TIME × BASE_GCF	β_5	γ_{110}	√	√	√	√	
		TIME × CDA	β_6	γ_{120}	√	√	√	√	
		TIME × AGE	β_7	γ_{101}	√	√	√	√	
Random effects	Patient (*j*)	Intercept	u_{0j}	u_{00k}	√	√	√	√	√
		TIME	u_{1j}	u_{10k}	√	√	√	√	√
	Tooth (*i*) within Patient (*j*)	Intercept	$u_{0i\,1\,j}$	r_{0jk}	√	√	√	√	√
Residuals	Visit (*t*)		ε_{tij}	ε_{ijk}	√	√	√	√	√

Summary of models considered for the dental veneer data (ctd)

TABLE 7.2

Summary of Models Considered for the Dental Veneer Data

		Term/Variable	Notation		Model				
			General	HLM[a]	7.1	7.2A[b]	7.2B[b]	7.2C	7.3
Covariance Parameters (θ_D) for D Matrix	Patient level	Variance of intercepts	$\sigma^2_{int\,:\,patient}$	$\tau_\beta[1,1]$	√	√	√	√	√
		Variance of slopes	$\sigma^2_{time\,:\,patient}$	$\tau_\beta[2,2]$	√	√	√	√	√
		Covariance of intercepts, slopes	$\sigma_{int,time\,:\,patient}$	$\tau_\beta[2,1]$	√	√	√	√	√
		Structure[c]	$D^{(2)}$	τ_β	UN	UN	UN	UN	UN
	Tooth Level	Variance of intercepts	$\sigma^2_{int\,:\,tooth(patient)}$	τ_π	√	√	√	√	√
Covariance Parameters (θ_R) for R_{ij} Matrix	Time Level	Variances at Time 1, Time 2	σ^2_{t1} σ^2_{t2}	σ^2_1 σ^2_2	Equal	Unequal	Equal	Unequal	Equal
		Covariance of Time 1, Time 2	$\sigma_{t1,t2}$	Varies[d]	0	√	√	0	0
		Structure[c]	R_{ij}	S	$\sigma^2 I_2$	UN	CS	HET	$\sigma^2 I_2$

[a] The notation for the HLM software is described in more detail in Subsection 7.4.5.

[b] In Model 7.2A and Model 7.2B, the residual covariance parameters are aliased with the variance of the random tooth-level intercepts.

[c] UN = unstructured, CS = compound symmetry, HET = diagonal with heterogeneous variances.

[d] The notation for this covariance parameter varies in HLM, depending on the structure specified.

Model selection using a top down approach

We follow the top down strategy to select the final model

- Fit a model with a loaded mean structure (Model 7.1)
- Select a structure for the random effects (Model 7.1 vs Model 7.1A)
- Select a structure for the residual covariance structure (Model 7.1 vs Model 7.1C)
- Reduce the model by removing non-significant fixed effects (Model 7.1 vs Model 7.3)

Fit a model with a loaded mean structure (Model 7.1)

```
model7.1.fit <- lme(gcf ~ time + base_gcf + cda + age + time:base_gcf + time:cda + time:age,
    random = list(patient = ~time, tooth = ~1), data = veneer, method = "REML")
summary(model7.1.fit)
```

```
## Linear mixed-effects model fit by REML
##  Data: veneer
##     AIC   BIC logLik
##   873.1 907.2 -423.6
##
## Random effects:
##  Formula: ~time | patient
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev Corr
## (Intercept) 23.567 (Intr)
## time         6.688 -0.95
##
##  Formula: ~1 | tooth %in% patient
##         (Intercept) Residual
## StdDev:       6.853    7.049
##
## Fixed effects: gcf ~ time + base_gcf + cda + age + time:base_gcf + time:cda +      time:age
##               Value Std.Error DF t-value p-value
## (Intercept)   69.92    28.399 51  2.4620  0.0172
## time          -6.02     7.446 51 -0.8078  0.4229
## base_gcf      -0.32     0.292 41 -1.0805  0.2862
## cda           -0.88     1.082 41 -0.8153  0.4196
## age           -0.97     0.608 10 -1.5951  0.1418
## time:base_gcf  0.07     0.058 51  1.1676  0.2484
## time:cda       0.13     0.218 51  0.5763  0.5670
## time:age       0.11     0.166 51  0.6576  0.5138
....
```

Fit a model with a loaded mean structure (Model 7.1, ctd)

```
summary(model7.1.fit)

....
##  Correlation:
##                (Intr) time   bs_gcf cda    age    tm:bs_ tim:cd
## time           -0.923
## base_gcf       -0.349  0.229
## cda            -0.361  0.240  0.326
## age            -0.917  0.880  0.117  0.121
## time:base_gcf   0.304 -0.264 -0.871 -0.283 -0.101
## time:cda        0.311 -0.272 -0.277 -0.872 -0.103  0.317
## time:age        0.846 -0.934 -0.074 -0.076 -0.936  0.084  0.085
##
## Standardized Within-Group Residuals:
##      Min       Q1      Med       Q3      Max
## -1.46355 -0.45282 -0.08649  0.35348  2.51322
##
## Number of Observations: 110
## Number of Groups:
##             patient tooth %in% patient
##                  12                 55

VarCorr(model7.1.fit)

##             Variance        StdDev Corr
## patient =   pdLogChol(time)
## (Intercept) 555.39          23.567 (Intr)
## time         44.72           6.688 -0.95
## tooth =     pdLogChol(1)
## (Intercept)  46.96           6.853
## Residual     49.69           7.049
```

Fit a model with a loaded mean structure (Model 7.1, ctd)

```
intervals(model7.1.fit)
```

```
## Approximate 95% confidence intervals
##
##  Fixed effects:
##                    lower     est.    upper
## (Intercept)    12.90439 69.91682 126.9293
## time          -20.96440 -6.01532   8.9338
## base_gcf       -0.90416 -0.31514   0.2739
## cda            -3.06820 -0.88239   1.3034
## age            -2.32399 -0.96961   0.3848
## time:base_gcf  -0.04846  0.06736   0.1832
## time:cda       -0.31253  0.12583   0.5642
## time:age       -0.22384  0.10903   0.4419
## attr(,"label")
## [1] "Fixed effects:"
##
##  Random Effects:
##   Level: patient
##                           lower    est.   upper
## sd((Intercept))        14.3852 23.5667 38.6081
## sd(time)                4.2072  6.6875 10.6301
## cor((Intercept),time) -0.9891 -0.9503 -0.7886
##   Level: tooth
##                  lower  est. upper
## sd((Intercept)) 4.839 6.853 9.704
##
##  Within-group standard error:
## lower  est. upper
## 5.683 7.049 8.743
```

Fit a model with a loaded mean structure (Model 7.1, ctd)

```
random.effects(model7.1.fit)
```

```
## Level: patient
##    (Intercept)    time
## 1      -43.239 12.8229
## 3        3.965  0.1962
## 4       26.467 -6.1154
## 5      -12.230  0.4196
## 6       -8.186  1.6070
## 7        2.101 -1.9919
## 8       22.854 -4.9477
## 9        3.710 -1.9946
## 10     -25.523 10.3561
## 12      19.401 -6.3370
## 13      -9.579  1.3856
## 14      20.259 -5.4007
##
## Level: tooth %in% patient
##       (Intercept)
## 1/6        3.8595
## 1/7       -2.3142
## 1/8        1.6473
## 1/9       -0.3795
## 1/10      -2.2355
## 1/11      -0.8387
## 3/6       -2.8833
## 3/7       -8.3992
## 3/8       -2.2630
## 3/9        7.1372
....
```

```
ranef(model7.1.fit)
```

```
....
## 3/9        7.1372
## 3/10      11.2582
## 3/11      -0.8210
## 4/6        1.9869
## 4/7        0.8183
## 4/8        0.8168
## 4/9       -0.7807
## 4/10       2.5821
## 4/11      -0.2036
## 5/7       -3.5713
## 5/8        0.2167
## 5/9       -3.8586
## 5/10      -2.2217
## 6/6        6.4461
## 6/7       -2.8084
## 6/8       -4.8179
## 6/9       -0.4681
## 6/10      -4.9259
## 6/11       4.1294
## 7/6        5.7824
## 7/7       -2.0675
## 7/8       -3.6234
## 7/9       -1.6711
## 7/10      -0.1406
## 7/11      -2.2622
## 8/6        2.9589
....
```

Select a random effects structure (Model 7.1 vs Model 7.1A)

To decide whether to keep the random effects associated with tooth within patients we fit

```
model7.1A.fit <- lme(gcf ~ time + base_gcf + cda + age + time:base_gcf + time:cda + time:age,
    random = list(patient = ~time), data = veneer, method = "REML")
anova(model7.1.fit, model7.1A.fit)
```

```
##                Model df   AIC   BIC logLik   Test L.Ratio p-value
## model7.1.fit       1 13 873.1 907.2 -423.6
## model7.1A.fit      2 12 882.3 913.8 -429.1 1 vs 2   11.17   8e-04
```

Then test the hypothesis:

$$H_0: \quad \sigma^2_{int:tooth(patient)} = 0$$
$$H_1: \quad \sigma^2_{int:tooth(patient)} > 0$$

```
h7.1.pvalue <- 0.5 * (1 - pchisq(11.17, 1))
h7.1.pvalue
```

```
## [1] 0.0004157
```

11.17 is the difference in -2REML log-likelihood between the two models which follows the usual mixture of χ^2 distributions.

P-value is ≈ 0.0004 and hence the random effects associated with tooth within patient should be retained in the model

Fit a model with a loaded mean structure (Model 7.1, ctd)

```
getVarCov(model7.1.fit)
```

```
## Error:  not implemented for multiple levels of nesting
```

```
xtabs(~tooth + patient, veneer)
```

```
##       patient
## tooth 1 3 4 5 6 7 8 9 10 12 13 14
##    6  2 2 2 0 2 2 2 2  2  0  0  0
##    7  2 2 2 2 2 2 2 2  2  2  0  0
##    8  2 2 2 2 2 2 2 0  2  2  0  2
##    9  2 2 2 2 2 2 2 0  2  2  0  0
##    10 2 2 2 2 2 2 2 0  2  2  2  2
##    11 2 2 2 0 2 2 2 0  2  0  0  0
```

```
veneer <- within(veneer, toothid <- factor(paste(patient, tooth, sep = ".")))
str(veneer)
```

```
## 'data.frame': 110 obs. of  11 variables:
##  $ patient : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ tooth   : int  6 6 7 7 8 8 9 9 10 10 ...
##  $ age     : int  46 46 46 46 46 46 46 46 46 46 ...
##  $ base_gcf: int  17 17 22 22 18 18 12 12 10 10 ...
##  $ cda     : num  4.67 4.67 4.67 4.67 5 ...
##  $ time    : int  3 6 3 6 3 6 3 6 3 6 ...
##  $ gcf     : int  11 68 13 47 14 58 10 57 14 44 ...
##  $ tooth.f : Factor w/ 6 levels "6","7","8","9",..: 1 1 2 2 3 3 4 4 5 5 ...
##  $ time.f  : Factor w/ 2 levels "3","6": 1 2 1 2 1 2 1 2 1 2 ...
##  $ age.f   : Factor w/ 12 levels "21","25","28",..: 7 7 7 7 7 7 7 7 7 7 ...
##  $ toothid : Factor w/ 55 levels "1.10","1.11",..: 3 3 4 4 5 5 6 6 1 1 ...
```

Fit a model with a loaded mean structure (Model 7.1, ctd)

```
model7.1c.fit <- lme(gcf ~ time + base_gcf + cda + age + time:base_gcf + time:cda + time:age,
    random = list(patient = pdBlocked(list(pdSymm(~time), pdIdent(~toothid - 1)))), data = veneer,
    method = "REML")

## Warning:  fewer observations than random effects in all level 1 groups
```

The estimated $\hat{\mathbf{V}}_j = \mathbf{Z}_j \hat{\mathbf{D}} \mathbf{Z}_j^T + \hat{\mathbf{R}}_j$ for patient j is:

```
getVarCov(model7.1c.fit, individual = 11, type = "marginal")

## patient 13
## Marginal variance covariance matrix
##       1     2
## 1 156.0  59.5
## 2  59.5 464.9
##   Standard Deviations: 12.49 21.56
```

The estimated $\hat{\mathbf{R}}_j$ for $\mathbf{R}_j$ for patient j is:

```
getVarCov(model7.1c.fit, individual = 11, type = "conditional")

## patient 13
## Conditional variance covariance matrix
##       1     2
## 1 49.69  0.00
## 2  0.00 49.69
##   Standard Deviations: 7.049 7.049
```

Fit a model with a loaded mean structure (Model 7.1, ctd)

The estimated $\hat{\mathbf{V}}_j = \mathbf{Z}_j \hat{\mathbf{D}} \mathbf{Z}_j^{\mathsf{T}} + \hat{\mathbf{R}}_j$ for patient j is:

```
getVarCov(model7.1c.fit, individual = 12, type = "marginal")
```

```
## patient 14
## Marginal variance covariance matrix
##        1      2      3      4
## 1 155.97  59.50  59.32  12.54
## 2  59.50 464.92  12.54 368.27
## 3  59.32  12.54 155.97  59.50
## 4  12.54 368.27  59.50 464.92
##   Standard Deviations: 12.49 21.56 12.49 21.56
```

The estimated $\hat{\mathbf{R}}_j$ for $\mathbf{R}_j$ for patient j is:

```
getVarCov(model7.1c.fit, individual = 12, type = "conditional")
```

```
## patient 14
## Conditional variance covariance matrix
##       1     2     3     4
## 1 49.69  0.00  0.00  0.00
## 2  0.00 49.69  0.00  0.00
## 3  0.00  0.00 49.69  0.00
## 4  0.00  0.00  0.00 49.69
##   Standard Deviations: 7.049 7.049 7.049 7.049
```

Select residual covariance structure (Model 7.2A)

```
model7.2A.fit <- lme(gcf ~ time + base_gcf + cda + age + time:base_gcf + time:cda + time:age,
    random = list(patient = ~time, tooth = ~1), corr = corCompSymm(0.5, form = ~1 | patient/tooth),
    weights = varIdent(form = ~1 | time), data = veneer, method = "REML")

summary(model7.2A.fit)
```

```
## Linear mixed-effects model fit by REML
##  Data: veneer
##      AIC   BIC logLik
##   876.2 915.5 -423.1
##
## Random effects:
##  Formula: ~time | patient
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev Corr
## (Intercept) 23.380 (Intr)
## time         6.681 -0.952
##
##  Formula: ~1 | tooth %in% patient
##         (Intercept) Residual
## StdDev:       6.235    8.392
##
## Correlation Structure: Compound symmetry
##  Formula: ~1 | patient/tooth
##  Parameter estimate(s):
##    Rho
## 0.1429
....
```

Select residual covariance structure (Model 7.2A, ctd)

```
summary(model7.2A.fit)

....
## Variance function:
##  Structure: Different standard deviations per stratum
##  Formula: ~1 | time
##  Parameter estimates:
##      3      6
## 1.0000 0.7993
## Fixed effects: gcf ~ time + base_gcf + cda + age + time:base_gcf + time:cda +      time:age
##                Value Std.Error DF t-value p-value
## (Intercept)   70.32    28.696 51  2.4507  0.0177
## time          -6.07     7.434 51 -0.8163  0.4181
## base_gcf      -0.32     0.308 41 -1.0380  0.3053
## cda           -0.87     1.134 41 -0.7651  0.4486
## age           -0.98     0.609 10 -1.6034  0.1399
## time:base_gcf  0.07     0.057 51  1.1921  0.2387
## time:cda       0.12     0.216 51  0.5581  0.5792
## time:age       0.11     0.166 51  0.6633  0.5101
##  Correlation:
##                (Intr) time   bs_gcf cda    age    tm:bs_ tim:cd
## time          -0.922
## base_gcf      -0.365  0.233
## cda           -0.377  0.243  0.331
## age           -0.914  0.882  0.124  0.129
## time:base_gcf  0.324 -0.263 -0.887 -0.292 -0.109
## time:cda       0.331 -0.271 -0.286 -0.887 -0.112  0.320
## time:age       0.841 -0.935 -0.075 -0.078 -0.936  0.084  0.086
##
## Standardized Within-Group Residuals:
##      Min       Q1      Med       Q3      Max
## -1.48928 -0.45592 -0.06955  0.31550  2.50850
```

Select residual covariance structure (Model 7.2A, ctd)

```
intervals(model7.2A.fit)

## Approximate 95% confidence intervals
##
##  Fixed effects:
##                    lower     est.    upper
## (Intercept)    12.71464 70.32426 127.9339
## time          -20.99283 -6.06830   8.8562
## base_gcf       -0.94041 -0.31927   0.3019
## cda            -3.15682 -0.86740   1.4220
## age            -2.33339 -0.97645   0.3805
## time:base_gcf  -0.04676  0.06836   0.1835
## time:cda       -0.31307  0.12054   0.5542
## time:age       -0.22264  0.10986   0.4424
## attr(,"label")
## [1] "Fixed effects:"
##
##  Random Effects:
##   Level: patient
##                           lower    est.   upper
## sd((Intercept))       14.1700 23.3796 38.5750
## sd(time)               4.2049  6.6815 10.6168
## cor((Intercept),time) -0.9897 -0.9516 -0.7865
##   Level: tooth
##                    lower  est. upper
## sd((Intercept)) 0.09582 6.235 405.7
##
....
```

```
....
##  Correlation structure:
##       lower   est.  upper
## Rho -0.9999 0.1429 0.9999
## attr(,"label")
## [1] "Correlation structure:"
##
##  Variance function:
##    lower   est. upper
## 6 0.2033 0.7993 3.143
## attr(,"label")
## [1] "Variance function:"
##
##  Within-group standard error:
##   lower    est.   upper
##  0.8293  8.3916 84.9175
```

Select residual covariance structure (Model 7.2B)

```
model7.2B.fit <- lme(gcf ~ time + base_gcf + cda + age + time:base_gcf + time:cda + time:age,
    random = list(patient = ~time, tooth = ~1), corr = corCompSymm(0.5, form = ~1 | patient/tooth),
    data = veneer, method = "REML")

summary(model7.2B.fit)
```

```
## Linear mixed-effects model fit by REML
##  Data: veneer
##      AIC   BIC logLik
##   875.1 911.9 -423.6
##
## Random effects:
##  Formula: ~time | patient
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev Corr
## (Intercept) 23.567 (Intr)
## time         6.688 -0.95
##
##  Formula: ~1 | tooth %in% patient
##         (Intercept) Residual
## StdDev:       5.563    8.105
##
## Correlation Structure: Compound symmetry
##  Formula: ~1 | patient/tooth
##  Parameter estimate(s):
##    Rho
## 0.2437
....
```

Select residual covariance structure (Model 7.2B, ctd)

```
summary(model7.2B.fit)

....
## Fixed effects: gcf ~ time + base_gcf + cda + age + time:base_gcf + time:cda +      time:age
##                Value Std.Error DF t-value p-value
## (Intercept)   69.92    28.399 51  2.4620  0.0172
## time          -6.02     7.446 51 -0.8078  0.4229
## base_gcf      -0.32     0.292 41 -1.0805  0.2862
## cda           -0.88     1.082 41 -0.8153  0.4196
## age           -0.97     0.608 10 -1.5951  0.1418
## time:base_gcf  0.07     0.058 51  1.1676  0.2484
## time:cda       0.13     0.218 51  0.5763  0.5670
## time:age       0.11     0.166 51  0.6576  0.5138
##  Correlation:
##                (Intr) time   bs_gcf cda    age    tm:bs_ tim:cd
## time          -0.923
## base_gcf      -0.349  0.229
## cda           -0.361  0.240  0.326
## age           -0.917  0.880  0.117  0.121
## time:base_gcf  0.304 -0.264 -0.871 -0.283 -0.101
## time:cda       0.311 -0.272 -0.277 -0.872 -0.103  0.317
## time:age       0.846 -0.934 -0.074 -0.076 -0.936  0.084  0.085
##
## Standardized Within-Group Residuals:
##      Min       Q1      Med       Q3      Max
## -1.78996 -0.48228 -0.07406  0.32602  2.52616
##
## Number of Observations: 110
## Number of Groups:
##            patient tooth %in% patient
##                 12                 55
```

Select residual covariance structure (Model 7.2B, ctd)

```
intervals(model7.2B.fit)
```

```
## Error:  cannot get confidence intervals on var-cov components:  Non-positive definite
approximate variance-covariance
```

Select residual covariance structure (Model 7.2C)

```
model7.2C.fit <- lme(gcf ~ time + base_gcf + cda + age + time:base_gcf + time:cda + time:age,
    random = list(patient = ~time, tooth = ~1), weights = varIdent(form = ~1 | time),
    data = veneer, method = "REML")


summary(model7.2C.fit)
```

```
## Linear mixed-effects model fit by REML
##  Data: veneer
##      AIC   BIC logLik
##    874.2 910.9 -423.1
##
## Random effects:
##  Formula: ~time | patient
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev Corr
## (Intercept) 23.380 (Intr)
## time         6.681 -0.952
##
##  Formula: ~1 | tooth %in% patient
##         (Intercept) Residual
## StdDev:       6.849    7.898
##
## Variance function:
##  Structure: Different standard deviations per stratum
##  Formula: ~1 | time
##  Parameter estimates:
##      3      6
## 1.0000 0.7696
....
```

Select residual covariance structure (Model 7.2C, ctd)

```
summary(model7.2C.fit)

....
## Fixed effects: gcf ~ time + base_gcf + cda + age + time:base_gcf + time:cda +      time:age
##                Value Std.Error DF t-value p-value
## (Intercept)   70.32    28.696 51  2.4507  0.0177
## time          -6.07     7.434 51 -0.8163  0.4181
## base_gcf      -0.32     0.308 41 -1.0380  0.3053
## cda           -0.87     1.134 41 -0.7651  0.4486
## age           -0.98     0.609 10 -1.6034  0.1399
## time:base_gcf  0.07     0.057 51  1.1921  0.2387
## time:cda       0.12     0.216 51  0.5581  0.5792
## time:age       0.11     0.166 51  0.6633  0.5101
##  Correlation:
##                (Intr) time   bs_gcf cda    age    tm:bs_ tim:cd
## time          -0.922
## base_gcf      -0.365  0.233
## cda           -0.377  0.243  0.331
## age           -0.914  0.882  0.124  0.129
## time:base_gcf  0.324 -0.263 -0.887 -0.292 -0.109
## time:cda       0.331 -0.271 -0.286 -0.887 -0.112  0.320
## time:age       0.841 -0.935 -0.075 -0.078 -0.936  0.084  0.086
##
## Standardized Within-Group Residuals:
##      Min       Q1      Med       Q3      Max
## -1.44314 -0.49937 -0.05685  0.32971  2.47181
##
## Number of Observations: 110
## Number of Groups:
##             patient tooth %in% patient
##                  12                 55
```

Select residual covariance structure (Model 7.2C, ctd)

```
intervals(model7.2C.fit)

## Approximate 95% confidence intervals
##
##  Fixed effects:
##                    lower     est.    upper
## (Intercept)    12.71466 70.32426 127.9339
## time          -20.99283 -6.06830   8.8562
## base_gcf       -0.94041 -0.31927   0.3019
## cda            -3.15683 -0.86740   1.4220
## age            -2.33339 -0.97645   0.3805
## time:base_gcf  -0.04676  0.06836   0.1835
## time:cda       -0.31307  0.12054   0.5542
## time:age       -0.22264  0.10986   0.4424
## attr(,"label")
## [1] "Fixed effects:"
##
##  Random Effects:
##   Level: patient
##                            lower    est.   upper
## sd((Intercept))        14.1698 23.3796 38.5756
## sd(time)                4.2048  6.6815 10.6170
## cor((Intercept),time) -0.9897 -0.9516 -0.7865
##   Level: tooth
##                 lower  est. upper
## sd((Intercept)) 4.849 6.849 9.675
##
....
```

```
....
##  Variance function:
##   lower   est. upper
## 6 0.444 0.7696 1.334
## attr(,"label")
## [1] "Variance function:"
##
##  Within-group standard error:
##  lower   est.  upper
##  5.872  7.898 10.623
```

Select residual covariance structure (Model 7.1 vs Model 7.2C)

To decide whether the variance of the residuals is constant (homogeneous) across the time points we look at:

```
anova(model7.2C.fit, model7.1.fit)
```

```
##                Model df   AIC   BIC logLik   Test L.Ratio p-value
## model7.2C.fit      1 14 874.2 910.9 -423.1
## model7.1.fit       2 13 873.1 907.2 -423.6 1 vs 2  0.9532  0.3289
```

To test the hypothesis:

$$H_0: \quad \sigma^2_{t1} = \sigma^2_{t2} = \sigma^2$$
$$H_1: \quad \sigma^2_{t1} \neq \sigma^2_{t2}$$

0.9532 is the difference in -2REML log-likelihood between the two models which follows a χ^2_1 distribution.

P-value is 0.3289 and hence there is no evidence in the data that a heterogeneous residual variance structure is necessary.

Removing non-significant fixed effects (Model 7.1 vs Model 7.3)

To test fixed effects, we have to use ML estimates:

```
model7.1.ml.fit <- lme(gcf ~ time + base_gcf + cda + age + time:base_gcf + time:cda +
    time:age, random = list(patient = ~time, tooth = ~1), data = veneer, method = "ML")

model7.3.ml.fit <- lme(gcf ~ time + base_gcf + cda + age, random = list(patient = ~time,
    tooth = ~1), data = veneer, method = "ML")

anova(model7.1.ml.fit, model7.3.ml.fit)
```

```
##                  Model df   AIC   BIC logLik   Test L.Ratio p-value
## model7.1.ml.fit      1 13 869.7 904.8 -421.8
## model7.3.ml.fit      2 10 865.5 892.5 -422.7 1 vs 2   1.841   0.606
```

This tests the hypothesis

$$H_0: \quad \beta_5 = \beta_6 = \beta_7 = 0$$

$$H_1: \quad \beta_5 \neq 0, \text{ or } \beta_6 \neq 0, \text{ or } \beta_7 \neq 0$$

1.841 is the difference in -2ML log-likelihood between the two models which follows a χ^2_3 distribution.
P-value is 0.606 and hence there is no evidence in the data the interaction terms between *TIME* and the other covariates are necessary.

Summary of Hypothesis tests for dental veneer analysis

TABLE 7.4

Summary of Hypotheses Tested for the Dental Veneer Data

Hypothesis Specification			Hypothesis Test				
				Models Compared			
Label	Null (H_0)	Alternative (H_A)	Test	Nested Model (H_0)	Reference Model (H_A)	Estimation Method	Asymptotic/ Approximate Dist. of Test Statistic under H_0
7.1	Drop $u_{0i\|j}$, random tooth-specific intercepts ($\sigma^2_{int:tooth(patient)} = 0$)	Retain $u_{0i\|j}$ ($\sigma^2_{int:tooth(patient)} > 0$)	LRT	Model 7.1A	Model 7.1	REML	$0.5\chi^2_0 + 0.5\chi^2_1$
7.2	Constant residual variance ($\sigma^2_{t1} = \sigma^2_{t2}$)	$\sigma^2_{t1} \neq \sigma^2_{t2}$	LRT	Model 7.1	Model 7.2C	REML	χ^2_1
7.3	Drop fixed effects associated with all two-way interactions ($\beta_5 = \beta_6 = \beta_7 = 0$)	$\beta_5 \neq 0$, or $\beta_6 \neq 0$, or $\beta_7 \neq 0$	LRT	Model 7.3	Model 7.1	ML	χ^2_3

TABLE 7.5

Summary of Hypothesis Test Results for the Dental Veneer Analysis

Hypothesis Label	Test	Estimation Method	Models Compared (Nested vs. Reference)	Test Statistic Value (Calculation)	p-Value
7.1	LRT	REML	7.1A vs. 7.1	$\chi^2(0{:}1) = 11.2$ (858.3 – 847.1)	< .001
7.2	LRT	REML	7.1 vs. 7.2C	$\chi^2(1) = 0.9$ (847.1 – 846.2)	.34
7.3	LRT	ML	7.3 vs. 7.1	$\chi^2(3) = 1.8$ (845.5 – 843.7)	.61

Note: See Table 7.4 for null and alternative hypotheses, and distributions of test statistics under H_0.

Final model (Model 7.3, using REML, WWG 7.7)

```
model7.3.fit <- lme(gcf ~ time + base_gcf + cda + age, random = list(patient = ~time,
    tooth = ~1), data = veneer, method = "REML")

summary(model7.3.fit)
```

```
## Linear mixed-effects model fit by REML
##  Data: veneer
##      AIC   BIC logLik
##   861.9 888.4 -420.9
##
## Random effects:
##  Formula: ~time | patient
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev Corr
## (Intercept) 22.913 (Intr)
## time         6.472 -0.947
##
##  Formula: ~1 | tooth %in% patient
##         (Intercept) Residual
## StdDev:       6.889    6.991
##
## Fixed effects: gcf ~ time + base_gcf + cda + age
##             Value Std.Error DF t-value p-value
## (Intercept) 45.74    12.555 54   3.643  0.0006
## time         0.30     1.937 54   0.155  0.8771
## base_gcf    -0.02     0.143 41  -0.128  0.8989
## cda         -0.33     0.529 41  -0.622  0.5373
## age         -0.58     0.214 10  -2.699  0.0224
....
```

Interpreting fixed-effect parameter estimates in final model

We note from the previous slide:

- There appears to be a negative effect of AGE on GCF, after controlling for the effects of TIME, baseline GCF, and CDA; patients who are 1 year older are predicted to have an average value of GCF that is 0.58 units lower than similar patients who are 1 year younger
- There is no significant fixed effect of TIME on GCF overall; this result is not surprising, given that our initial plot showed that the GCF for some patients went up over time, whereas for other patients it decreased over time
- The effect of contour difference (CDA) is not significant, indicating that a greater discrepancy in tooth contour after veneer placement is not associated with a higher mean value of GCF
- The fact that there were no significant interactions between TIME and the other covariates suggests that the effect of TIME on GCF does not tend to differ for different values of AGE, baseline GCF, or contour difference

Interpreting covariance parameter estimates in final model

```
summary(model7.3.fit)

....
## Random effects:
##  Formula: ~time | patient
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev Corr
## (Intercept) 22.913 (Intr)
## time         6.472 -0.947
##
##  Formula: ~1 | tooth %in% patient
##         (Intercept) Residual
## StdDev:       6.889    6.991
....
```

```
VarCorr(model7.3.fit)

##             Variance        StdDev Corr
## patient =   pdLogChol(time)
## (Intercept) 524.99          22.913 (Intr)
## time         41.89           6.472 -0.947
## tooth =     pdLogChol(1)
## (Intercept)  47.46           6.889
## Residual     48.87           6.991
```

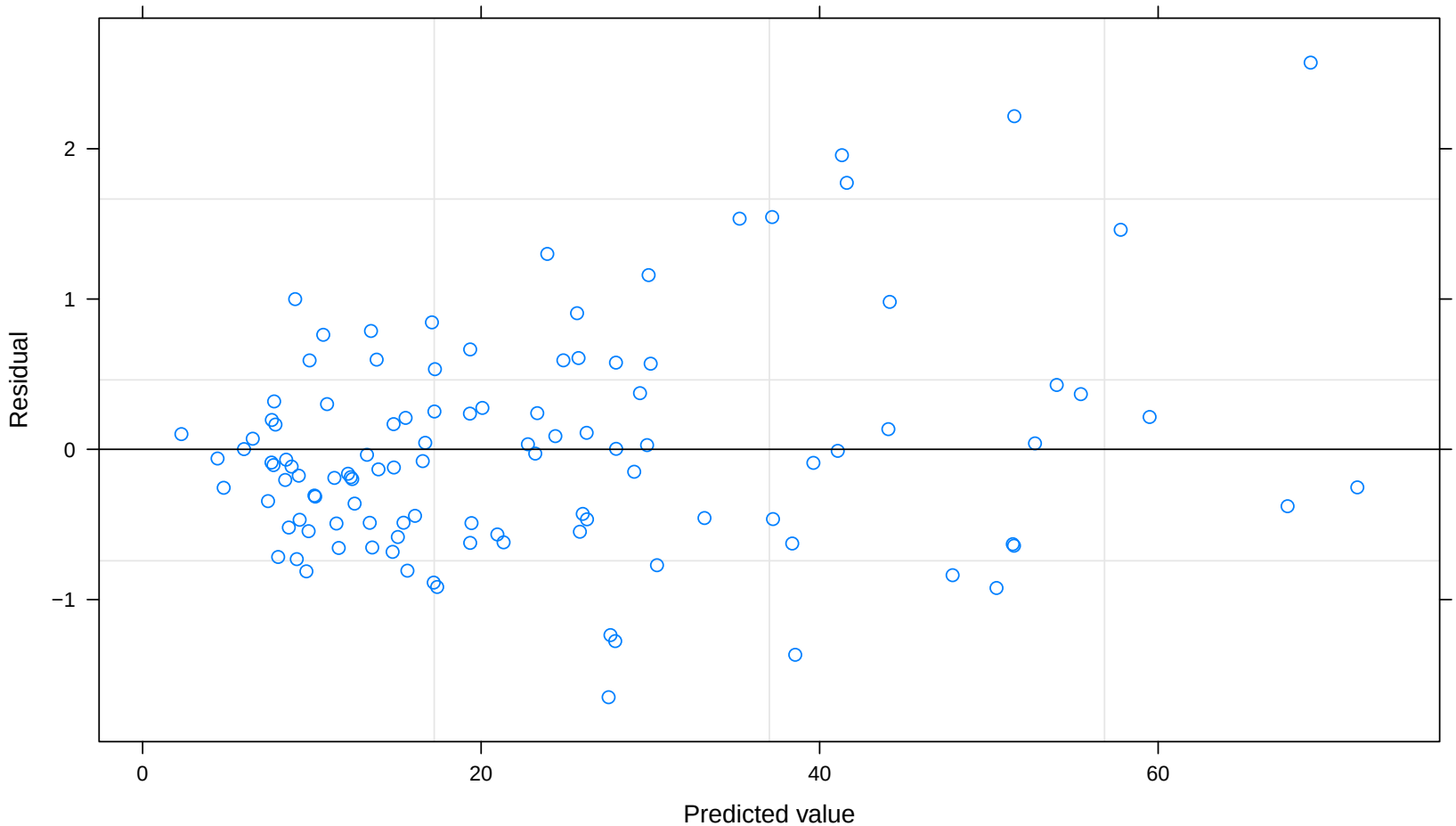
```
intervals(model7.3.fit)

....
##  Random Effects:
##   Level: patient
##                            lower    est.   upper
## sd((Intercept))       14.2877 22.9126 36.7440
## sd(time)               4.1690  6.4721 10.0474
## cor((Intercept),time) -0.9879 -0.9469 -0.7828
##   Level: tooth
##                  lower  est. upper
## sd((Intercept)) 4.887 6.889 9.712
##
##  Within-group standard error:
## lower  est. upper
## 5.663 6.991 8.630
```

Interpreting covariance parameter estimates in final model (ctd)

Based on these results and the formal test of Hypothesis 7.1, we have evidence of between-patient variance and between-tooth variance within the same patient that is not being explained by the fixed effects of the covariates included in Model 7.3.

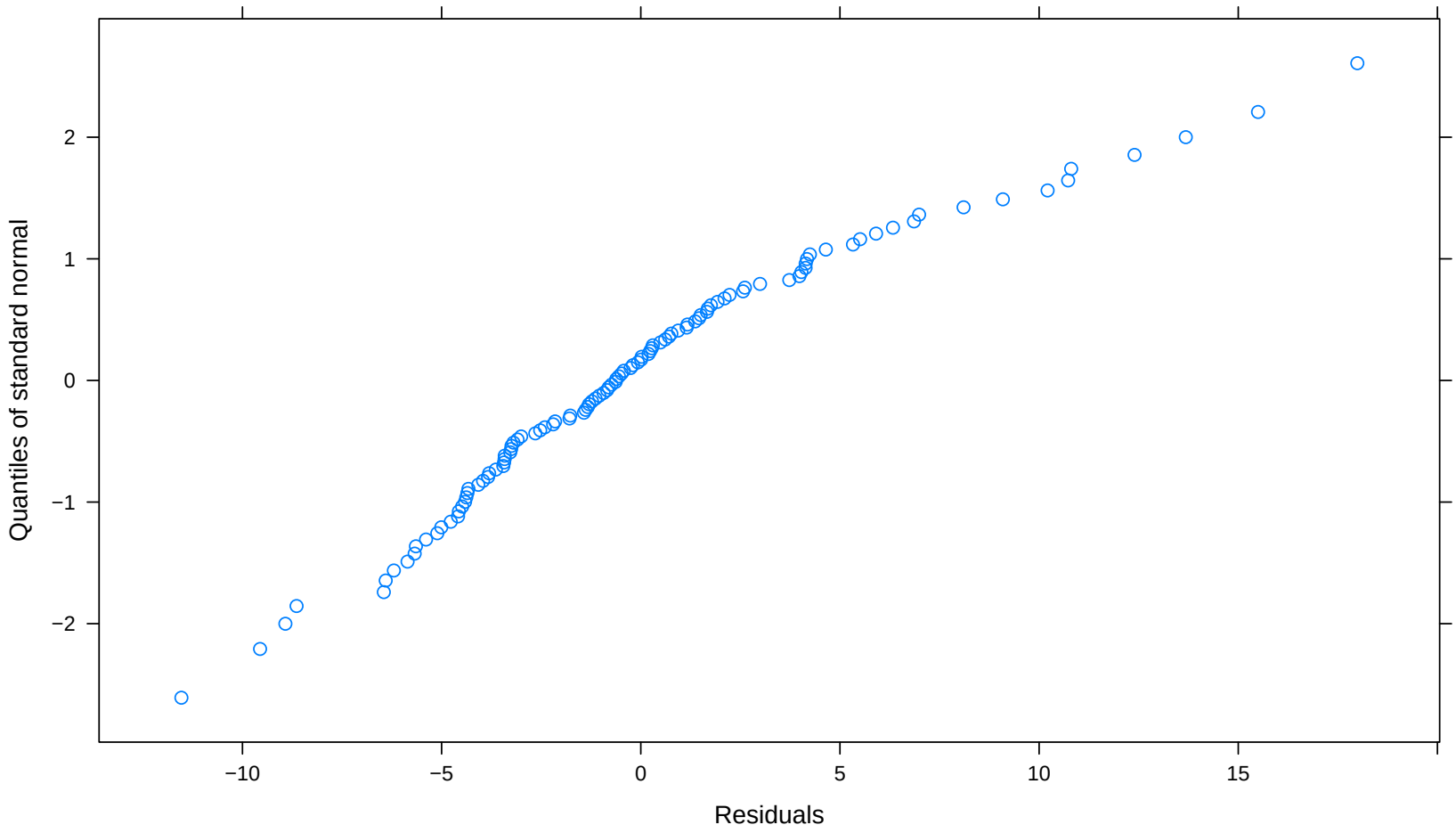
Model diagnostics

```
plot(model7.3.fit, resid(., type = "p") ~ fitted(.), xlab = "Predicted value", ylab = "Residual",
    abline = 0)
```



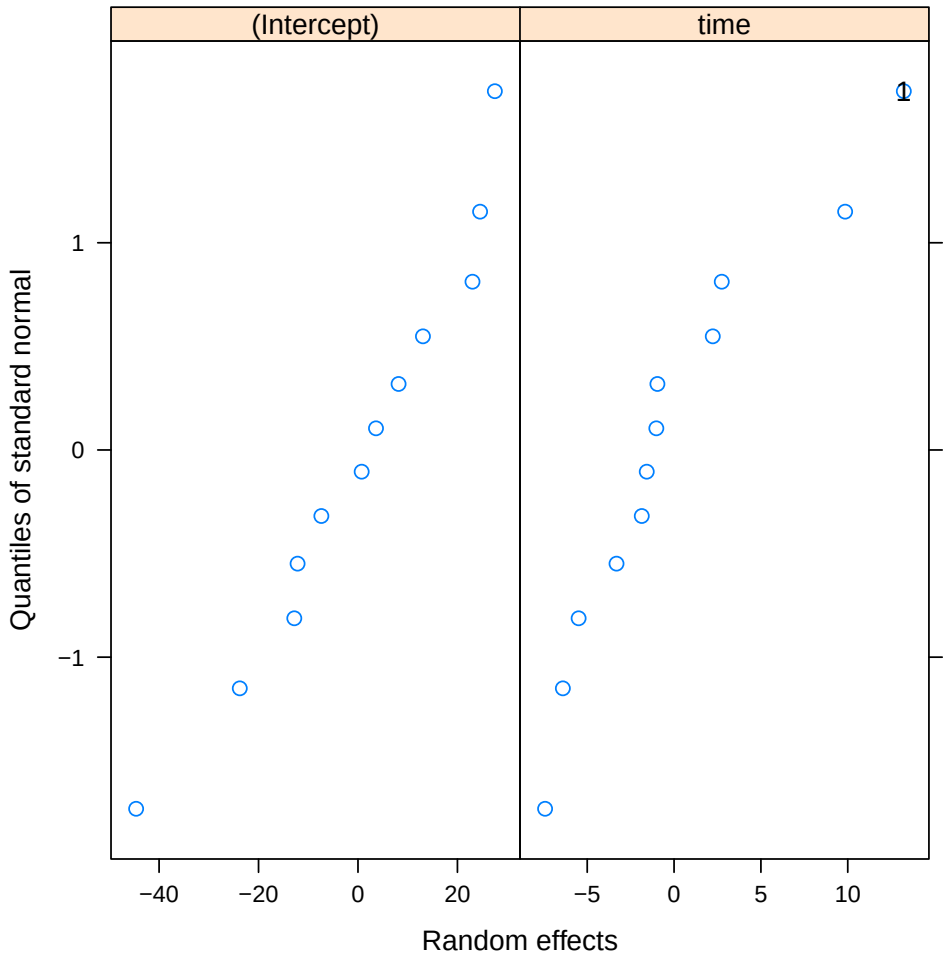
Model diagnostics (ctd)

```
qqnorm(model7.3.fit, ~resid(.))
```



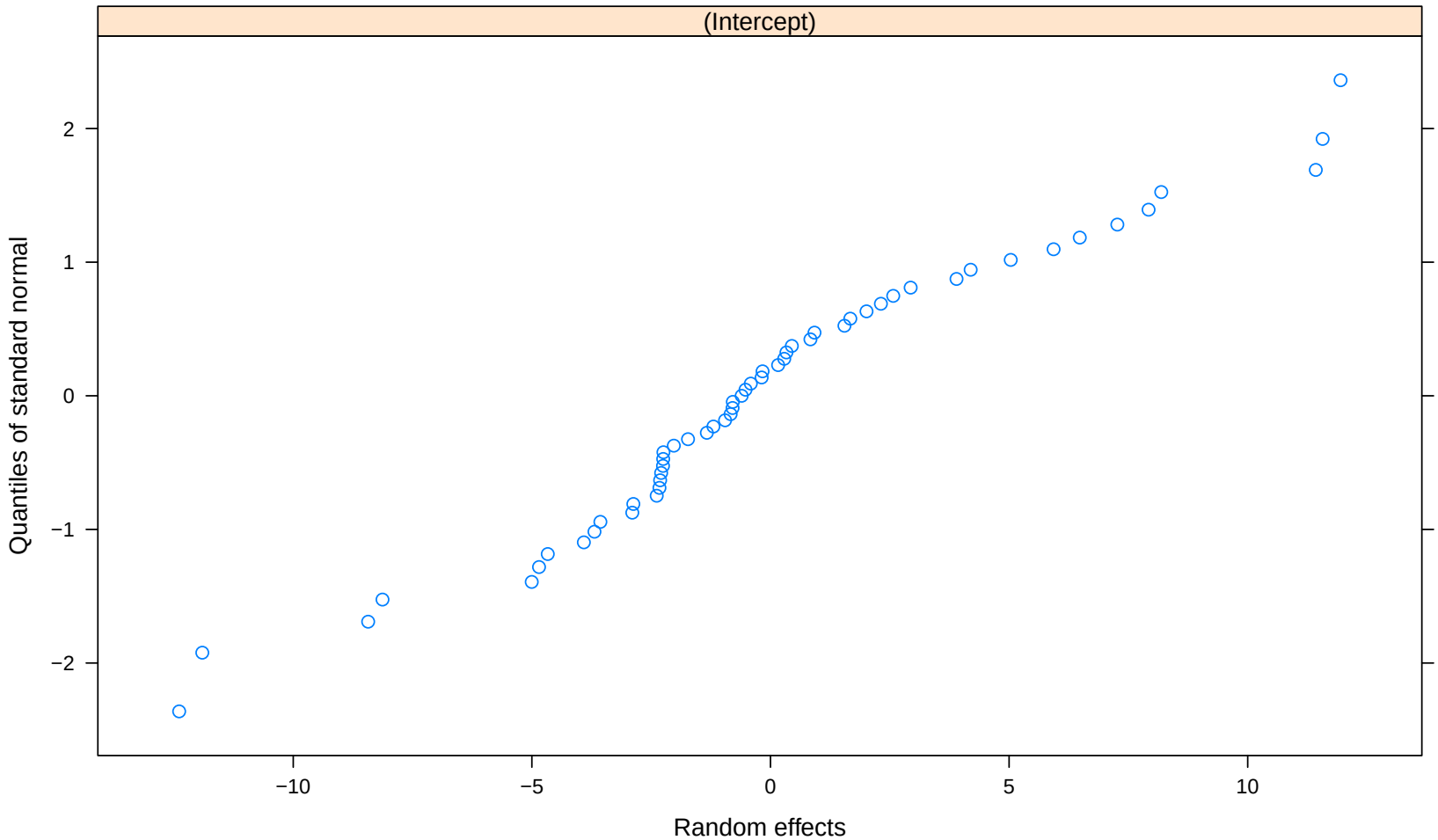
Model diagnostics (ctd)

```
qqnorm(model7.3.fit, ~ranef(., level = 1), layout = c(2, 1), aspect = 2, id = 0.05)
```



Model diagnostics (ctd)

```
qqnorm(model7.3.fit, ~ranef(., level = 2), id = 0.05)
```



Model diagnostics (ctd)

```
plot(model7.3.fit, gcf ~ fitted(.), id = 0.05, abline = c(0, 1), aspect = 1)
```