- two measurement occasions (pre-post)
  - one group: $t_{n-1} = \frac{\bar{d}}{s_d/\sqrt{n}}$, $d_i = \beta_0 + \epsilon_i$, $H_0 : \beta_0 = 0$ (within-variation due to treatment effect)
  - two group:
    - change score analysis: $d_i = \beta_0 + \beta_1 \cdot \text{grp}_i + \epsilon_i$, or $y_{i2} = \beta_0 + \beta_1 \cdot \text{grp}_i + 1 \cdot y_{i1} + \epsilon_i$ (between variation due to treatment effect)
    - ANCOVA: $y_{i2} = \beta_0 + \beta_1 \cdot \text{grp}_i + \beta_2 \cdot y_{i1} + \epsilon_i$ (between ind variation due to treatment effect and between-subject baseline variation, control for baseline variations $\beta_2 \cdot y_{i1}$)
- more than two measures: model mean + covariance

Response profile analysis (saturated model, time as discrete)

$$\boldsymbol{Y_i} = \boldsymbol{X_i}\boldsymbol{\beta} + \boldsymbol{\epsilon_i} \quad \text{or} \quad Y_{ij} = \beta_1 x_{ij1} + \ldots + \beta_p x_{ijp} + e_{ij}$$

where $\boldsymbol{e_i} \sim MVN(\mathbf{0}, \Sigma_i)$, Σ_i(**unknown**, no assump.) is the var-cov matrix of individual i . param of interest: $\boldsymbol{\beta}$ and Σ.

```
glsmodel <− gls(Lead ~ Treatment * Week,corr=corSymm(form = ~ Time | ID),
weights = varIdent(form = ~ 1 | Week), method = "REML", data = tlclong)
```

- Are the mean response profiles similar in the groups
- ignore time ordering (discrete t), for balanced data, for 4-5 occassions

Linear/quadratic trend model

$$\text{Linear: } E(Y_{ij}) = \beta_0 + \beta_1 \cdot t_{ij} + \beta_2 \cdot \text{grp}_{ij} + \beta_3 \cdot t_{ij} \cdot \text{grp}_{ij}$$

```
linmodel <− gls(fev1 ~ smoking*year, corr=corSymm(form= ~ t | id),
weights = varIdent(form = ~ 1 | t), data = smoking)
```

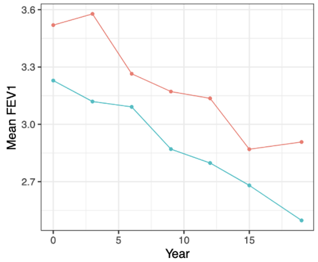
$$\text{Quadratic: } E(Y_{ij}) = \beta_0 + \beta_1 \cdot t_{ij} + \beta_2 \cdot \text{grp}_{ij} + \beta_3 \cdot t_{ij} \cdot \text{grp}_{ij} + \beta_4 \cdot t_{ij}^2 + \cdot t_{ij}^2 \cdot \text{grp}_{ij}$$

- rate of change dep on t_{ij} in quadratic model
- center t to avoid collinearity in quadratic model (yrvec = unique(smoking$year), meanyr = mean(yrvec), yearc = year - meanyr)
- LRT: testing mean use "ML", default is "REML"

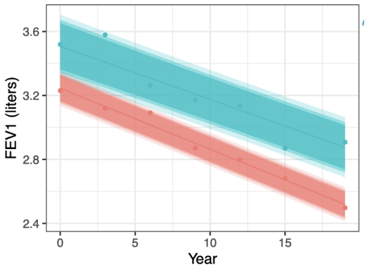
Linear spline

$$E(Y_{ij}) = \beta_0 + \beta_1 \cdot t_{ij} + \beta_2 \cdot \text{grp}_{ij} + \beta_3 \cdot t_{ij} \cdot \text{grp}_{ij} + \beta_4 \cdot (t_{ij} - t^\star)_+ + \beta_5 \cdot (t_{ij} - t^\star) \cdot \text{grp}_{ij}$$

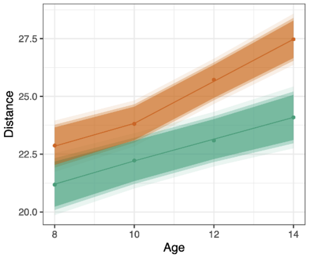
where $(t_{ij} - t^\star)_+ = (t_{ij} - t^\star)$ if $t_{ij} > t^\star$ 0 otherwise.



Response profile



Linear trend



Linear spline

Linear mixed effect model

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i \cdot b_i + \boldsymbol{\epsilon}_i$$

- $b_i \sim MVN(\mathbf{0}, \boldsymbol{G})$, $\boldsymbol{\epsilon}_i \sim MVN(\mathbf{0}, \boldsymbol{R}_i)$ where $\boldsymbol{R}_i = \sigma^2 I_{n_i}$
- $\boldsymbol{Z}_i(n_i \times q)$, $\boldsymbol{X}_i(n_i \times p)$, $\boldsymbol{\beta}(p \times 1)$, $b_i(q \times 1)$
- between subj variation captured by random intercept/baseline + random slope/gain; within subj variation captured by error term.
- COV$(Y_i \mid b_i) = R_i$, COV$(Y_i) = Z_i G Z_i^T + \sigma^2 I_{n_i}$ [Z includes time, so cov(Y) is a fun of time]

Random intercept model

$$Y_{ij} = \boldsymbol{X}_{ij}\boldsymbol{\beta} + b_i + \epsilon_{ij} \quad \text{or} \quad E(Y_{ij} \mid b_i, \boldsymbol{X}_{ij}) = \boldsymbol{X}_{ij}\boldsymbol{\beta} + b_i$$

- where $b_i \sim N(0, \sigma_b^2)$, error $\epsilon_{ij} \sim N(0, \sigma^2)$, and $b_i \perp \epsilon_{ij}$, $\epsilon_{ij} \perp \epsilon_{ik}$. $V(Y_{ij}) = \sigma_b^2 + \sigma^2$ and COV$(Y_{ij}, Y_{ik}) = \sigma_b^2$
- between-individual comes from variation in baseline(random intercept), within-individual comes from error term
- intra class/within-subj correlation: $\frac{\sigma_b^2}{\sigma_b^2 + \sigma^2}$ (degree of clustering)
- interpreted as **marginal** and conditional mean response

```
rim <− lme(fev1 ~ year + smoking, random = ~ 1 | id, data = smoking) # nlme
rim2 <− lmer(fev1 ~ year + smoking + (1|id), data = smoking) # lme4
```

Random intercept and slope model

$$Y_{ij} = \beta_1 \cdot t_{ij} + \beta_2 \cdot \text{grp}_i + \beta_3 \cdot t_{ij} \cdot \text{grp}_i + b_{1i} + b_{2i} \cdot t_{ij} + \epsilon_{ij}$$

$$E(Y_{ij}|b_i) = \beta_1 + \beta_2 t_{ij} + \beta_3 t_{ij_+} + b_{1i} + b_{2i} t_{ij} + b_{3i} t_{ij_+} \text{(menopause)}$$

```
    bf <− bf %>% mutate(time0 = time * I(time >= 0)) library(nlme)
    modela <− lme(pbf ~ time + time0, random= ~ time + time0 | id, data = bf)
```

Approximately 95% of subjects have changes in percent body fat between -2.09% and 2.92% ($\beta_2 1.96 \times \sqrt{\sigma_{b_2}^2} = 0.421.961.28$)

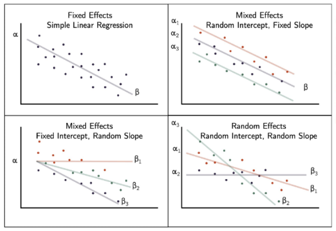
Generalized linear mixed effect model

- $b_i \sim MVN(\mathbf{0}, \boldsymbol{G})$
- conditional mean of Y_{ij}: $g[E(Y_{ij} \mid b_i)] = \eta_{ij} = \boldsymbol{X}_{ij}\boldsymbol{\beta} + \boldsymbol{Z}_{ij} b_i$, $g(\cdot)$ is a function of conditional mean response
- $Var(Y_{ij} \mid b_i) = v[E(Y_{ij} \mid b_i)]\phi$ $v(\cdot)$ is a function of conditional mean response
- subject-specific interpretation: cannot directly get the marginal mean response. β refers to the effect for a unite increase in X_{ijp} for any given individual having some unobservable underlying propensity to repsond positively, b_i
- if X_{ij} is between subject covariate, β change in log odds of response for two different individuals who happen to have the same value for the random effect, but who differ by on unite in the covariate X_{ijp}

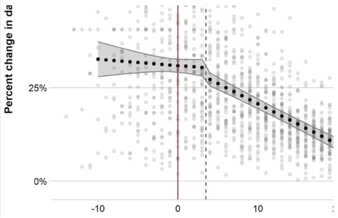
```
    mod1 <− glmer(Count ~ offset(logT) + Post + TrtPost + (1 + Post | PtID),
    data = szrl, family = poisson(link = "log"))
```

Random intercept + covariance pattern model

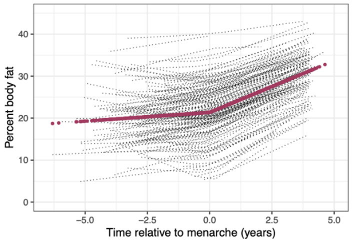
$$Y_{ij} = \boldsymbol{X}_{ij}\boldsymbol{\beta} + b_{0i} + U_i(t_{ij}) + \epsilon_{ij}^\star$$



Random effect models
with fixed covariate



Random intercept with
Time-varying covariate



Random intercept + random slope
with Time-varying covariate

Covariance pattern models

- unstructure: corSymm(form = $\sim t^\star$ | id)
- compound symmetry/exchangeable corCompSymm(form =$\sim t^\star$ | id)
- Toeplitz corARMA(form =$\sim t^\star$ | id,p=1,q=1)
- AR1 corAR1(form = $\sim t^\star$ | id)
- exponential corExp(form =$\sim t$ | id)

where $t^\star$ is the measurement occasion, t is the actual time. LRT and "REML" for comparison of covariance pattern models.

- start with all covariates and decide wich covariance pattern to use (using REML), then ML to select mean resposne model

Compound Symmetry (Exchangeable):ssumes that the variance is constant across occasions, and the correlation is the same between any two measurement occasions.

$$\mathrm{Cov}(\mathrm{Y}_i) = \sigma^2 \begin{pmatrix} 1 & \rho & \rho & \dots & \rho \\ \rho & 1 & \rho & \dots & \rho \\ \rho & \rho & 1 & \dots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \dots & 1 \end{pmatrix}$$

Toeplitz structure assumes that the variance is constant across occasions, and the correlation depends on the distance between the two time points.

$$\mathrm{Cov}(\mathrm{Y}_i) = \sigma^2 \begin{pmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{n-2} \\ \rho_2 & \rho_1 & 1 & \dots & \rho_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \dots & 1 \end{pmatrix}$$

AR1 structure assumes that the correlation decreases as the time difference between measurements increases. The AR1 structure is very parsimonious, only requiring two parameters (variance and correlation).

$$\mathrm{Cov}(\mathrm{Y}_i) = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{pmatrix}$$

Exponential: For measurement occasions that are not equally spaced, an exponential covariance structure is used. This model assumes that the correlation between two time points decays exponentially with the absolute difference in time.

$$\mathrm{Cov}(\mathrm{Y}_i) = \sigma^2 \begin{pmatrix} 1 & \rho^{|t_{i1}-t_{i2}|} & \rho^{|t_{i1}-t_{i3}|} & \dots & \rho^{|t_{i1}-t_{in}|} \\ \rho^{|t_{i2}-t_{i1}|} & 1 & \rho^{|t_{i2}-t_{i3}|} & \dots & \rho^{|t_{i2}-t_{in}|} \\ \rho^{|t_{i3}-t_{i1}|} & \rho^{|t_{i3}-t_{i2}|} & 1 & \dots & \rho^{|t_{i3}-t_{in}|} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{|t_{in}-t_{i1}|} & \rho^{|t_{in}-t_{i2}|} & \rho^{|t_{in}-t_{i3}|} & \dots & 1 \end{pmatrix}$$

Heterogeneous variances: $V(Y_{ij} = \sigma_i^2)$

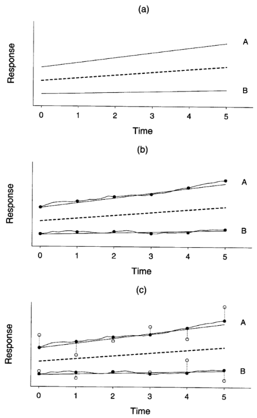


Fig. 2.4 Graphical representation of the cumulative impact of three sources of variability in longitudinal data: (a) between-individual heterogeneity, (b) within-individual biological variation (where ● denotes repeated measure free of measurement error), and (c) measurement error (where ○ denotes observed repeated measure with measurement error).