Analysis and Sensitivity Analysis of Incomplete Data

Geert Molenberghs
geert.molenberghs@med.kuleuven.be
geert.molenberghs@uhasselt.be

Interuniversity Institute for Biostatistics and statistical Bioinformatics (I-BioStat)
KU Leuven, Belgium & Universiteit Hasselt
www.ibiostat.be

RBras, Salvador, May 2016
# Contents

1 Related References ........................................................................................................................................... 1

## Longitudinal Data

2 The Rat Data .................................................................................................................................................. 9

3 The Toenail Data ............................................................................................................................................. 13

4 A Model for Longitudinal Data ...................................................................................................................... 17

5 The General Linear Mixed Model .................................................................................................................. 22

6 Estimation and Inference .................................................................................................................................. 25
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>Generalized Estimating Equations</td>
<td>28</td>
</tr>
<tr>
<td>8</td>
<td>Generalized Linear Mixed Models (GLMM)</td>
<td>36</td>
</tr>
<tr>
<td>9</td>
<td>Fitting GLMM’s in SAS</td>
<td>43</td>
</tr>
<tr>
<td>10</td>
<td>Marginal Versus Random-effects Models</td>
<td>46</td>
</tr>
<tr>
<td>II</td>
<td>Incomplete Data</td>
<td>52</td>
</tr>
<tr>
<td>11</td>
<td>A Gentle Tour</td>
<td>53</td>
</tr>
<tr>
<td>12</td>
<td>Direct Likelihood / Ignorable Likelihood</td>
<td>75</td>
</tr>
<tr>
<td>13</td>
<td>Multiple Imputation</td>
<td>82</td>
</tr>
<tr>
<td>14</td>
<td>The Analgesic Trial</td>
<td>86</td>
</tr>
<tr>
<td>15</td>
<td>Creating Monotone Missingness</td>
<td>96</td>
</tr>
<tr>
<td>16</td>
<td>Case Study: The Dataset</td>
<td>109</td>
</tr>
<tr>
<td>17</td>
<td>Case Study: Weighted Generalized Estimating Equations</td>
<td>113</td>
</tr>
</tbody>
</table>
Chapter 1
Related References


Part I

Longitudinal Data
Chapter 2
The Rat Data

- Research question (Dentistry, K.U.Leuven):
  How does craniofacial growth depend on testosterone production?

- Randomized experiment in which 50 male Wistar rats are randomized to:
  - Control (15 rats)
  - Low dose of Decapeptyl (18 rats)
  - High dose of Decapeptyl (17 rats)
• Treatment starts at the age of 45 days; measurements taken every 10 days, from day 50 on.

• The responses are distances (pixels) between well defined points on x-ray pictures of the skull of each rat:
• Measurements with respect to the roof, base and height of the skull. Here, we consider only one response, reflecting the height of the skull.

• Individual profiles:
• Complication: Dropout due to anaesthesia (56%):

<table>
<thead>
<tr>
<th>Age (days)</th>
<th># Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control</td>
</tr>
<tr>
<td>50</td>
<td>15</td>
</tr>
<tr>
<td>60</td>
<td>13</td>
</tr>
<tr>
<td>70</td>
<td>13</td>
</tr>
<tr>
<td>80</td>
<td>10</td>
</tr>
<tr>
<td>90</td>
<td>7</td>
</tr>
<tr>
<td>100</td>
<td>4</td>
</tr>
<tr>
<td>110</td>
<td>4</td>
</tr>
</tbody>
</table>

• Remarks:
  ▶ A lot of variability between rats, much less variability within rats
  ▶ Fixed number of measurements scheduled per subject, but not all measurements available due to dropout, for known reason.
  ▶ Measurements taken at fixed time points
Chapter 3
The Toenail Data

- **Toenail Dermatophyte Onychomycosis**: Common toenail infection, difficult to treat, affecting more than 2% of population.

- Classical treatments with antifungal compounds need to be administered until the whole nail has grown out healthy.

- New compounds have been developed which reduce treatment to 3 months.

- Randomized, double-blind, parallel group, multicenter study for the comparison of two such new compounds (A and B) for oral treatment.
Research question:

Severity relative to treatment of TDO?

- 2 × 189 patients randomized, 36 centers
- 48 weeks of total follow up (12 months)
- 12 weeks of treatment (3 months)
- Measurements at months 0, 1, 2, 3, 6, 9, 12.
- Frequencies at each visit (both treatments):
3.1 Repeated Measures / Longitudinal Data

Revised measures are obtained when a response is measured repeatedly on a set of units

- **Units:**
  - Subjects, patients, participants, …
  - Animals, plants, …
  - Clusters: families, towns, branches of a company,…
  - …

- **Special case:** *Longitudinal data*
Chapter 4
A Model for Longitudinal Data

• In practice: often unbalanced data:
  ▶ unequal number of measurements per subject
  ▶ measurements not taken at fixed time points

• Therefore, multivariate regression techniques are often not applicable

• Often, subject-specific longitudinal profiles can be well approximated by linear regression functions

• This leads to a 2-stage model formulation:
  ▶ **Stage 1**: Linear regression model for each subject separately
  ▶ **Stage 2**: Explain variability in the subject-specific regression coefficients using known covariates
4.1 Example: The Rat Data

- Individual profiles:

- Transformation of the time scale to linearize the profiles:

\[
\text{Age}_{ij} \quad \rightarrow \quad t_{ij} = \ln\left[1 + \left(\text{Age}_{ij} - 45\right)/10\right]
\]
• Note that $t = 0$ corresponds to the start of the treatment (moment of randomization)

• **Stage 1 model:** \[ Y_{ij} = \beta_{1i} + \beta_{2i} t_{ij} + \epsilon_{ij}, \quad j = 1, \ldots, n_i \]

• In the second stage, the subject-specific intercepts and time effects are related to the treatment of the rats

• **Stage 2 model:**

\[
\begin{align*}
\beta_{1i} &= \beta_0 + b_{1i}, \\
\beta_{2i} &= \beta_1 L_i + \beta_2 H_i + \beta_3 C_i + b_{2i},
\end{align*}
\]
• $L_i$, $H_i$, and $C_i$ are indicator variables:

$$L_i = \begin{cases} 1 & \text{if low dose} \\ 0 & \text{otherwise} \end{cases} \quad H_i = \begin{cases} 1 & \text{if high dose} \\ 0 & \text{otherwise} \end{cases} \quad C_i = \begin{cases} 1 & \text{if control} \\ 0 & \text{otherwise} \end{cases}$$

• Parameter interpretation:
  - $\beta_0$: average response at the start of the treatment (independent of treatment)
  - $\beta_1$, $\beta_2$, and $\beta_3$: average time effect for each treatment group
4.2 The Linear Mixed-effects Model

- **Stage 1 model:** 
  \[ Y_{ij} = \beta_1 + \beta_2 t_{ij} + \varepsilon_{ij}, \quad j = 1, \ldots, n_i \]

- **Stage 2 model:**
  \[
  \begin{cases}
  \beta_{1i} = \beta_0 + b_{1i}, \\
  \beta_{2i} = \beta_1 L_i + \beta_2 H_i + \beta_3 C_i + b_{2i},
  \end{cases}
  \]

- **Combined:**
  \[ Y_{ij} = (\beta_0 + b_{1i}) + (\beta_1 L_i + \beta_2 H_i + \beta_3 C_i + b_{2i}) t_{ij} + \varepsilon_{ij} \]
  \[
  = \begin{cases}
  \beta_0 + b_{1i} + (\beta_1 + b_{2i}) t_{ij} + \varepsilon_{ij}, & \text{if low dose} \\
  \beta_0 + b_{1i} + (\beta_2 + b_{2i}) t_{ij} + \varepsilon_{ij}, & \text{if high dose} \\
  \beta_0 + b_{1i} + (\beta_3 + b_{2i}) t_{ij} + \varepsilon_{ij}, & \text{if control.}
  \end{cases}
  \]
Chapter 5
The General Linear Mixed Model

\[ Y_i = X_i \beta + Z_i b_i + \varepsilon_i \]

\[ b_i \sim N(0, D), \]
\[ \varepsilon_i \sim N(0, \Sigma_i), \]

\[ b_1, \ldots, b_N, \varepsilon_1, \ldots, \varepsilon_N \]

Terminology:

- Fixed effects: \( \beta \)
- Random effects: \( b_i \)
- Variance components: elements in \( D \) and \( \Sigma_i \)

\[ \text{independent} \]
5.1 Hierarchical versus Marginal Model

- The general linear mixed model is given by:
  \[ Y_i = X_i\beta + Z_i b_i + \varepsilon_i \]

\[
\begin{align*}
  b_i &\sim N(0, D), \\
  \varepsilon_i &\sim N(0, \Sigma_i), \\
  b_1, \ldots, b_N, \varepsilon_1, \ldots, \varepsilon_N \text{ independent},
\end{align*}
\]

- It can be rewritten as:
  \[ Y_i | b_i \sim N(X_i\beta + Z_i b_i, \Sigma_i) \]
  \[ b_i \sim N(0, D) \]
• It is therefore also called a hierarchical model:
  ▶ A model for $Y_i$ given $b_i$
  ▶ A model for $b_i$

• Marginally, we have that $Y_i$ is distributed as: $Y_i \sim N(X_i\beta, Z_iDZ_i' + \Sigma_i)$

• Hence, very specific assumptions are made about the dependence of mean and covariance on the covariates $X_i$ and $Z_i$:
  ▶ **Implied mean**: $X_i\beta$
  ▶ **Implied covariance**: $V_i = Z_iDZ_i' + \Sigma_i$

• Note that the hierarchical model implies the marginal one, **not** vice versa
Chapter 6
Estimation and Inference

- Notation:
  - $\beta$: vector of fixed effects (as before)
  - $\alpha$: vector of all variance components in $D$ and $\Sigma_i$
  - $\theta = (\beta', \alpha')'$: vector of all parameters in marginal model

- In most cases, $\alpha$ is not known, and needs to be replaced by an estimate $\hat{\alpha}$

- Two frequently used estimation methods for $\alpha$:
  - Maximum likelihood
  - Restricted maximum likelihood
6.1 Inference

- Inference for $\beta$:
  - Wald tests, $t$- and $F$-tests
  - LR tests (not with REML)

- Inference for $\alpha$:
  - Wald tests
  - LR tests (even with REML)
  - Caution: Boundary problems!

- Inference for the random effects:
  - Empirical Bayes inference based on posterior density $f(b_i|Y_i = y_i)$
  - ‘Empirical Bayes (EB) estimate’: Posterior mean
6.2 Fitting Linear Mixed Models in SAS

- A model for the rat data: \( Y_{ij} = (\beta_0 + b_{1i}) + (\beta_1 L_i + \beta_2 H_i + \beta_3 C_i + b_{2i}) t_{ij} + \varepsilon_{ij} \)

- SAS program:
  ```sas
  proc mixed data=rat method=reml;
  class id group;
  model y = t group*t / solution;
  random intercept t / type=un subject=id ;
  run;
  ```

- Fitted averages:
Chapter 7
Generalized Estimating Equations

- Univariate GLM, score function of the form (scalar $Y_i$):
  \[ S(\beta) = \sum_{i=1}^{N} \frac{\partial \mu_i}{\partial \beta} v_i^{-1} (y_i - \mu_i) = 0, \quad \text{with} \ v_i = \text{Var}(Y_i). \]

- In longitudinal setting: $Y = (Y_1, \ldots, Y_N)$:
  \[ S(\beta) = \sum_{i=1}^{N} D_i' [V_i(\alpha)]^{-1} (y_i - \mu_i) = 0 \]

  where
  - $D_i$ is an $n_i \times p$ matrix with $(i, j)$th elements $\frac{\partial \mu_{ij}}{\partial \beta}$
  - Is $V_i$ an $n_i \times n_i$ diagonal?
  - $y_i$ and $\mu_i$ are $n_i$-vectors with elements $y_{ij}$ and $\mu_{ij}$
7.1 Large Sample Properties: Naive Approach

As \( N \to \infty \)

\[
\sqrt{N}(\hat{\beta} - \beta) \sim N(0, I_0^{-1})
\]

where

\[
I_0 = \sum_{i=1}^{N} D_i' [V_i(\alpha)]^{-1} D_i
\]

- **(Unrealistic) Conditions:**
  - \( \alpha \) is known
  - the parametric form for \( V_i(\alpha) \) is known

- **Solution:** working correlation matrix
7.2 Unknown Covariance Structure

Keep the score equations

\[ S(\beta) = \sum_{i=1}^{N} [D_i]' [V_i(\alpha)]^{-1} (y_i - \mu_i) = 0 \]

**BUT**

- suppose \( V_i(.) \) is not the true variance of \( Y_i \) but only a plausible guess, a so-called working correlation matrix

- specify correlations and not covariances, because the variances follow from the mean structure

- the score equations are solved as before
7.3 Asymptotic Properties: Robust Approach

The asymptotic normality results change to (sandwich estimator)

\[ \sqrt{N}(\hat{\beta} - \beta) \sim N(0, I_0^{-1}I_1I_0^{-1}) \]

\[ I_0 = \sum_{i=1}^{N} D_i' [V_i(\alpha)]^{-1} D_i \]
\[ I_1 = \sum_{i=1}^{N} D_i' [V_i(\alpha)]^{-1} \text{Var}(Y_i)[V_i(\alpha)]^{-1} D_i. \]

- The estimators \( \hat{\beta} \) are consistent even if the working correlation matrix is incorrect.

- An estimate is found by replacing the unknown variance matrix \( \text{Var}(Y_i) \) by

\[ (Y_i - \hat{\mu}_i)(Y_i - \hat{\mu}_i)' \]
7.4 Application to the Toenail Data

7.4.1 The model

• Consider the model:

\[ Y_{ij} \sim \text{Bernoulli}(\mu_{ij}), \quad \log \left( \frac{\mu_{ij}}{1 - \mu_{ij}} \right) = \beta_0 + \beta_1 T_i + \beta_2 t_{ij} + \beta_3 T_i t_{ij} \]

• \( Y_{ij} \): severe infection (yes/no) at occasion \( j \) for patient \( i \)

• \( t_{ij} \): measurement time for occasion \( j \)

• \( T_i \): treatment group
7.4.2 Standard GEE

- **SAS Code:**

  ```sas
  proc genmod data=test descending;
  class idnum timeclss;
  model onyresp = treatn time treatn*time
      / dist=binomial;
  repeated subject=idnum / withinsubject=timeclss
      type=exch covb corrw modelse;
  run;
  ```

- **SAS statements:**

  - The `REPEATED` statements defines the GEE character of the model.
  - `type=`: working correlation specification (UN, AR(1), EXCH, IND,...)
  - `modelse`: model-based s.e.’s on top of default empirically corrected s.e.’s
  - `corrw`: printout of working correlation matrix
  - `withinsubject=`: specification of the ordering within subjects
### Regression parameters:

#### Analysis Of Initial Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Error</th>
<th>Confidence Limits</th>
<th>Chi-Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-0.5571</td>
<td>0.1090</td>
<td>-0.7708 -0.3433</td>
<td>26.10</td>
</tr>
<tr>
<td>treatn</td>
<td>1</td>
<td>0.0240</td>
<td>0.1565</td>
<td>-0.2827 0.3307</td>
<td>0.02</td>
</tr>
<tr>
<td>time</td>
<td>1</td>
<td>-0.1769</td>
<td>0.0246</td>
<td>-0.2251 -0.1288</td>
<td>51.91</td>
</tr>
<tr>
<td>treatn*time</td>
<td>1</td>
<td>-0.0783</td>
<td>0.0394</td>
<td>-0.1556 -0.0010</td>
<td>3.95</td>
</tr>
<tr>
<td>Scale</td>
<td>0</td>
<td>1.0000</td>
<td>0.0000</td>
<td>1.0000 1.0000</td>
<td></td>
</tr>
</tbody>
</table>

#### Analysis Of GEE Parameter Estimates

| Parameter   | Estimate | Error | Confidence Limits | Z Pr > |Z| |
|-------------|----------|-------|-------------------|--------|---|
| Intercept   | -0.5840  | 0.1734| -0.9238 -0.2441   | -3.37  | 0.0008|
| treatn      | 0.0120   | 0.2613| -0.5001 0.5241    | 0.05   | 0.9633|
| time        | -0.1770  | 0.0311| -0.2380 -0.1161   | -5.69  | <.0001|
| treatn*time | -0.0886  | 0.0571| -0.2006 0.0233    | -1.55  | 0.1208|
## Analysis Of GEE Parameter Estimates

Model-Based Standard Error Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Error</th>
<th>95% Confidence Limits</th>
<th>Z</th>
<th>Pr &gt;</th>
<th>Z</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.5840</td>
<td>0.1344</td>
<td>-0.8475 -0.3204</td>
<td>-4.34</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>treatn</td>
<td>0.0120</td>
<td>0.1866</td>
<td>-0.3537 0.3777</td>
<td>0.06</td>
<td>0.9486</td>
<td></td>
<td></td>
</tr>
<tr>
<td>time</td>
<td>-0.1770</td>
<td>0.0209</td>
<td>-0.2180 -0.1361</td>
<td>-8.47</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>treatn*time</td>
<td>-0.0886</td>
<td>0.0362</td>
<td>-0.1596 -0.0177</td>
<td>-2.45</td>
<td>0.0143</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

▷ The working correlation:

- Exchangeable Working Correlation

  Correlation 0.420259237
Chapter 8
Generalized Linear Mixed Models (GLMM)

- Given a vector $b_i$ of random effects for cluster $i$, it is assumed that all responses $Y_{ij}$ are independent, with density

$$f(y_{ij}|\theta_{ij}, \phi) = \exp\{\phi^{-1}[y_{ij}\theta_{ij} - \psi(\theta_{ij})] + c(y_{ij}, \phi)\}$$

- $\theta_{ij}$ is now modelled as $\theta_{ij} = x_{ij}'\beta + z_{ij}'b_i$

- As before, it is assumed that $b_i \sim N(0, D)$
• Let $f_{ij}(y_{ij}|b_i, \beta, \phi)$ denote the conditional density of $Y_{ij}$ given $b_i$, the conditional density of $Y_i$ equals

$$f_i(y_i|b_i, \beta, \phi) = \prod_{j=1}^{n_i} f_{ij}(y_{ij}|b_i, \beta, \phi)$$

• The marginal distribution of $Y_i$ is given by

$$f_i(y_i|\beta, D, \phi) = \int \prod_{j=1}^{n_i} f_{ij}(y_{ij}|b_i, \beta, \phi) \ f(b_i|D) \ db_i$$

where $f(b_i|D)$ is the density of the $N(0, D)$ distribution.

• The likelihood function for $\beta$, $D$, and $\phi$ now equals

$$L(\beta, D, \phi) = \prod_{i=1}^{N} f_i(y_i|\beta, D, \phi)$$

$$= \prod_{i=1}^{N} \int \prod_{j=1}^{n_i} f_{ij}(y_{ij}|b_i, \beta, \phi) \ f(b_i|D) \ db_i$$
• Under the normal linear model, the integral can be worked out analytically.

• In general, approximations are required:
  ▶ Approximation of integrand: Laplace approximation
  ▶ Approximation of data: Taylor series
  ▶ Approximation of integral: (Adaptive) Gaussian quadrature

• Predictions of random effects can be based on the posterior distribution

\[ f(b_i | Y_i = y_i) \]

• ‘Empirical Bayes (EB) estimate’:
  Posterior mode, with unknown parameters replaced by their MLE
8.1 Example: Toenail Data

- $Y_{ij}$ is binary severity indicator for subject $i$ at visit $j$.

- Model:

$$Y_{ij}|b_i \sim \text{Bernoulli}(\pi_{ij}), \quad \log \left( \frac{\pi_{ij}}{1 - \pi_{ij}} \right) = \beta_0 + b_i + \beta_1 T_i + \beta_2 t_{ij} + \beta_3 T_i t_{ij}$$

- Notation:
  - $T_i$: treatment indicator for subject $i$
  - $t_{ij}$: time point at which $j$th measurement is taken for $i$th subject

- Adaptive as well as non-adaptive Gaussian quadrature, for various $Q$. 

RBras, Salvador, May 2016
• Results:

<table>
<thead>
<tr>
<th></th>
<th>Gaussian quadrature</th>
<th>Adaptive Gaussian quadrature</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$Q = 3$</td>
<td>$Q = 5$</td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>-1.52 (0.31)</td>
<td>-2.49 (0.39)</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-0.39 (0.38)</td>
<td>0.19 (0.36)</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>-0.32 (0.03)</td>
<td>-0.38 (0.04)</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>-0.09 (0.05)</td>
<td>-0.12 (0.07)</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>2.26 (0.12)</td>
<td>3.09 (0.21)</td>
</tr>
<tr>
<td>$-2\ell$</td>
<td>1344.1</td>
<td>1259.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$Q = 3$</td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>-2.05 (0.59)</td>
<td>-1.47 (0.40)</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-0.16 (0.64)</td>
<td>-0.09 (0.54)</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>-0.42 (0.05)</td>
<td>-0.40 (0.04)</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>-0.17 (0.07)</td>
<td>-0.16 (0.07)</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>4.51 (0.62)</td>
<td>3.70 (0.34)</td>
</tr>
<tr>
<td>$-2\ell$</td>
<td>1259.1</td>
<td>1257.1</td>
</tr>
</tbody>
</table>
• Conclusions:
  ▶ (Log-) likelihoods are not comparable
  ▶ Different \( Q \) can lead to considerable differences in estimates and standard errors
  ▶ For example, using non-adaptive quadrature, with \( Q = 3 \), we found no difference in time effect between both treatment groups
    \( (t = -0.09/0.05, p = 0.0833) \).
  ▶ Using adaptive quadrature, with \( Q = 50 \), we find a significant interaction between the time effect and the treatment
    \( (t = -0.16/0.07, p = 0.0255) \).
  ▶ Assuming that \( Q = 50 \) is sufficient, the ‘final’ results are well approximated with smaller \( Q \) under adaptive quadrature, but not under non-adaptive quadrature.
Comparison of fitting algorithms:
  ▶ Adaptive Gaussian Quadrature, \( Q = 50 \)
  ▶ MQL and PQL

Summary of results:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>QUAD</th>
<th>PQL</th>
<th>MQL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept group A</td>
<td>-1.63 (0.44)</td>
<td>-0.72 (0.24)</td>
<td>-0.56 (0.17)</td>
</tr>
<tr>
<td>Intercept group B</td>
<td>-1.75 (0.45)</td>
<td>-0.72 (0.24)</td>
<td>-0.53 (0.17)</td>
</tr>
<tr>
<td>Slope group A</td>
<td>-0.40 (0.05)</td>
<td>-0.29 (0.03)</td>
<td>-0.17 (0.02)</td>
</tr>
<tr>
<td>Slope group B</td>
<td>-0.57 (0.06)</td>
<td>-0.40 (0.04)</td>
<td>-0.26 (0.03)</td>
</tr>
<tr>
<td>Var. random intercepts ( \tau^2 )</td>
<td>15.99 (3.02)</td>
<td>4.71 (0.60)</td>
<td>2.49 (0.29)</td>
</tr>
</tbody>
</table>

Severe differences between QUAD (gold standard ?) and MQL/PQL.

MQL/PQL may yield (very) biased results, especially for binary data.
Chapter 9
Fitting GLMM’s in SAS

- **GLIMMIX**: Laplace, MQL, PQL, adaptive quadrature

- **NLMIXED**: Adaptive and non-adaptive quadrature
  \[\rightarrow\text{not discussed here}\]
9.1 Example: Toenail data

- Re-consider logistic model with random intercepts for toenail data

- SAS code (PQL):
  
  ```sas
  proc glimmix data=test method=RSPL ;
  class idnum;
  model onyresp (event='1') = treatn time treatn*time
         / dist=binary solution;
  random intercept / subject=idnum;
  run;
  ```

- MQL obtained with option ‘method=RMPL’

- Laplace obtained with option ‘method=LAPLACE’
• Adaptive quadrature with option ‘method=QUAD(qpoints=5)’

• Inclusion of random slopes:

  random intercept time / subject=idnum type=un;
We compare our GLMM results for the toenail data with those from fitting GEE’s (unstructured working correlation):

<table>
<thead>
<tr>
<th>Parameter</th>
<th>GLMM Estimate (s.e.)</th>
<th>GEE Estimate (s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept group A</td>
<td>−1.6308 (0.4356)</td>
<td>−0.7219 (0.1656)</td>
</tr>
<tr>
<td>Intercept group B</td>
<td>−1.7454 (0.4478)</td>
<td>−0.6493 (0.1671)</td>
</tr>
<tr>
<td>Slope group A</td>
<td>−0.4043 (0.0460)</td>
<td>−0.1409 (0.0277)</td>
</tr>
<tr>
<td>Slope group B</td>
<td>−0.5657 (0.0601)</td>
<td>−0.2548 (0.0380)</td>
</tr>
</tbody>
</table>
• The strong differences can be explained as follows:
  ▶ Consider the following GLMM:

  \[ Y_{ij} | b_i \sim \text{Bernoulli}(\pi_{ij}), \quad \log \left( \frac{\pi_{ij}}{1 - \pi_{ij}} \right) = \beta_0 + b_i + \beta_1 t_{ij} \]

  ▶ The conditional means \( E(Y_{ij} | b_i) \), as functions of \( t_{ij} \), are given by

  \[
  E(Y_{ij} | b_i) = \frac{\exp(\beta_0 + b_i + \beta_1 t_{ij})}{1 + \exp(\beta_0 + b_i + \beta_1 t_{ij})}
  \]
The marginal average evolution is now obtained from averaging over the random effects:

\[
E(Y_{ij}) = E[E(Y_{ij} | b_i)] = E \left[ \frac{\exp(\beta_0 + b_i + \beta_1 t_{ij})}{1 + \exp(\beta_0 + b_i + \beta_1 t_{ij})} \right] \\
\neq \frac{\exp(\beta_0 + \beta_1 t_{ij})}{1 + \exp(\beta_0 + \beta_1 t_{ij})}
\]
• Hence, the parameter vector $\beta$ in the GEE model needs to be interpreted completely different from the parameter vector $\beta$ in the GLMM:
  
  ⊲ GEE: marginal interpretation
  
  ⊲ GLMM: conditional interpretation, conditionally upon level of random effects

• For logistic mixed models, with normally distributed random random intercepts, it can be shown that the marginal model can be well approximated by again a logistic model, but with parameters approximately satisfying

$$
\frac{\hat{\beta}_{RE}}{\hat{\beta}_M} = \sqrt{c^2 \sigma^2 + 1} > 1,
$$

$$
\sigma^2 = \text{variance random intercepts}
$$

$$
c = 16\sqrt{3}/(15\pi)
$$

• For the toenail application, $\sigma$ was estimated as 4.0164, such that the ratio equals

$$
\sqrt{c^2 \sigma^2 + 1} = 2.5649.
$$
<table>
<thead>
<tr>
<th>Parameter</th>
<th>GLMM</th>
<th>GEE</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept group A</td>
<td>-1.6308 (0.4356)</td>
<td>-0.7219 (0.1656)</td>
<td>2.2590</td>
</tr>
<tr>
<td>Intercept group B</td>
<td>-1.7454 (0.4478)</td>
<td>-0.6493 (0.1671)</td>
<td>2.6881</td>
</tr>
<tr>
<td>Slope group A</td>
<td>-0.4043 (0.0460)</td>
<td>-0.1409 (0.0277)</td>
<td>2.8694</td>
</tr>
<tr>
<td>Slope group B</td>
<td>-0.5657 (0.0601)</td>
<td>-0.2548 (0.0380)</td>
<td>2.2202</td>
</tr>
</tbody>
</table>

• Note that this problem does not occur in linear mixed models:
  
  ▶ Conditional mean:  
  \[ E(Y_i|b_i) = X_i\beta + Z_i b_i \]

  ▶ Specifically:  
  \[ E(Y_i|b_i = 0) = X_i\beta \]

  ▶ Marginal mean:  
  \[ E(Y_i) = X_i\beta \]

• For GLMM: \[ E[g(Y)] \neq g[E(Y)] \]
10.1 Toenail Data: Overview

• Overview of all analyses on toenail data:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>QUAD</th>
<th>PQL</th>
<th>MQL</th>
<th>GEE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept group A</td>
<td>-1.63 (0.44)</td>
<td>-0.72 (0.24)</td>
<td>-0.56 (0.17)</td>
<td>-0.72 (0.17)</td>
</tr>
<tr>
<td>Intercept group B</td>
<td>-1.75 (0.45)</td>
<td>-0.72 (0.24)</td>
<td>-0.53 (0.17)</td>
<td>-0.65 (0.17)</td>
</tr>
<tr>
<td>Slope group A</td>
<td>-0.40 (0.05)</td>
<td>-0.29 (0.03)</td>
<td>-0.17 (0.02)</td>
<td>-0.14 (0.03)</td>
</tr>
<tr>
<td>Slope group B</td>
<td>-0.57 (0.06)</td>
<td>-0.40 (0.04)</td>
<td>-0.26 (0.03)</td>
<td>-0.25 (0.04)</td>
</tr>
<tr>
<td>Var. random intercepts ($\tau^2$)</td>
<td>15.99 (3.02)</td>
<td>4.71 (0.60)</td>
<td>2.49 (0.29)</td>
<td></td>
</tr>
</tbody>
</table>

• Conclusion:

$|\text{GEE}| < |\text{MQL}| < |\text{PQL}| < |\text{QUAD}|$
Part II

Incomplete Data
Chapter 11
A Gentle Tour

- Orthodontic growth data
- Commonly used methods
- Survey of the terrain
11.1 Growth Data: An (Un)balanced Discussion

• Taken from Potthoff and Roy, Biometrika (1964)

• Research question:

Is dental growth related to gender?

• The distance from the center of the pituitary to the maxillary fissure was recorded at ages 8, 10, 12, and 14, for 11 girls and 16 boys
- Individual profiles:
  - Unbalanced data
  - Balanced data
# 11.2 LOCF, CC, or Direct Likelihood?

Data:

<table>
<thead>
<tr>
<th></th>
<th>20</th>
<th>30</th>
<th>75</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>40</td>
<td>25</td>
<td></td>
</tr>
</tbody>
</table>

**LOCF:**

<table>
<thead>
<tr>
<th></th>
<th>20</th>
<th>30</th>
<th>75</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>40</td>
<td>0</td>
<td>25</td>
<td></td>
</tr>
</tbody>
</table>

$\Rightarrow \begin{pmatrix} 95 & 30 \\ 10 & 65 \end{pmatrix}$

$\hat{\theta} = \frac{95}{200} = 0.475 [0.406; 0.544]$ (biased & too narrow)

**CC:**

<table>
<thead>
<tr>
<th></th>
<th>20</th>
<th>30</th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>40</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

$\Rightarrow \begin{pmatrix} 20 & 30 \\ 10 & 40 \end{pmatrix}$

$\hat{\theta} = \frac{20}{100} = 0.200 [0.122; 0.278]$ (biased & too wide)

**d.l.(MAR):**

<table>
<thead>
<tr>
<th></th>
<th>20</th>
<th>30</th>
<th>30</th>
<th>45</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>40</td>
<td>5</td>
<td>20</td>
<td></td>
</tr>
</tbody>
</table>

$\Rightarrow \begin{pmatrix} 50 & 75 \\ 15 & 60 \end{pmatrix}$

$\hat{\theta} = \frac{50}{200} = 0.250 [0.163; 0.337] $
### 11.3 Direct Likelihood/Bayesian Inference: Ignorability

**MAR**: \( f(Y_i^0 | X_i, \theta) f(r_i | X_i, Y_i^0, \psi) \)

Mechanism is MAR
\( \theta \) and \( \psi \) distinct
Interest in \( \theta \)
(Use observed information matrix)

\[ \Rightarrow \text{Lik./Bayes inference valid} \]

<table>
<thead>
<tr>
<th>Outcome type</th>
<th>Modeling strategy</th>
<th>Software</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gaussian</strong></td>
<td>Linear mixed model</td>
<td>SAS MIXED</td>
</tr>
<tr>
<td><strong>Non-Gaussian</strong></td>
<td>Gen./Non-linear mixed model</td>
<td>SAS GLIMMIX, NLMIXED</td>
</tr>
</tbody>
</table>
11.4 Rubin, 1976

- Ignorability: Rubin (Biometrika, 1976): 35 years ago!

- Little and Rubin (1976, 2002)

- Why did it take so long?
11.5 A Vicious Triangle

- **Academe:** The $R^2$ principle
- **Regulatory:** Rigid procedures $\leftrightarrow$ scientific developments
- **Industry:** We cannot / do not want to apply new methods
11.6 Terminology & Confusion

- The Ministry of Disinformation:

  ← [All directions]

  [Other directions] →

- MCAR, MAR, MNAR: “What do the terms mean?”

- MAR, random dropout, informative missingness, ignorable, censoring, . . .

- Dropout from the study, dropout from treatment, lost to follow up, . . .

- “Under MAR patients dropping out and patients not dropping out are similar.”
11.7 A Virtuous Triangle

- FDA/Industry Workshops
- DIA/EMA Meetings
- The NAS Experience
11.8 The NAS Experience: A Wholesome Product

- FDA $\rightarrow$ NAS $\rightarrow$ the working group

- Composition

- Encompassing:
  - terminology/taxonomy/concepts
  - prevention
  - treatment
11.9 Taxonomy

- **Missingness pattern:** complete — monotone — non-monotone
- **Dropout pattern:** complete — dropout — intermittent
- **Model framework:** SEM — PMM — SPM
- **Missingness mechanism:** MCAR — MAR — MNAR
- **Ignorability:** ignorable — non-ignorable
- **Inference paradigm:** frequentist — likelihood — Bayes
### 11.10 The NAS Panel

<table>
<thead>
<tr>
<th>Name</th>
<th>Specialty</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rod Little</td>
<td>biostat</td>
<td>U Michigan</td>
</tr>
<tr>
<td>Ralph D’Agostino</td>
<td>biostat</td>
<td>Boston U</td>
</tr>
<tr>
<td>Kay Dickerson</td>
<td>epi</td>
<td>Johns Hopkins</td>
</tr>
<tr>
<td>Scott Emerson</td>
<td>biostat</td>
<td>U Washington</td>
</tr>
<tr>
<td>John Farrar</td>
<td>epi</td>
<td>U Penn</td>
</tr>
<tr>
<td>Constantine Frangakis</td>
<td>biostat</td>
<td>Johns Hopkins</td>
</tr>
<tr>
<td>Joseph Hogan</td>
<td>biostat</td>
<td>Brown U</td>
</tr>
<tr>
<td>Geert Molenberghs</td>
<td>biostat</td>
<td>U Hasselt &amp; K.U.Leuven</td>
</tr>
<tr>
<td>Susan Murphy</td>
<td>stat</td>
<td>U Michigan</td>
</tr>
<tr>
<td>James Neaton</td>
<td>biostat</td>
<td>U Minnesota</td>
</tr>
<tr>
<td>Andrea Rotnitzky</td>
<td>stat</td>
<td>Buenos Aires &amp; Harvard</td>
</tr>
<tr>
<td>Dan Scharfstein</td>
<td>biostat</td>
<td>Johns Hopkins</td>
</tr>
<tr>
<td>Joseph Shih</td>
<td>biostat</td>
<td>New Jersey SPH</td>
</tr>
<tr>
<td>Jay Siegel</td>
<td>biostast</td>
<td>J&amp;J</td>
</tr>
<tr>
<td>Hal Stern</td>
<td>stat</td>
<td>UC at Irvine</td>
</tr>
</tbody>
</table>
11.11 Modeling Frameworks & Missing Data Mechanisms

\[ f(y_i, r_i | X_i, \theta, \psi) \]

**Selection Models:**  
\[ f(y_i | X_i, \theta) \quad f(r_i | X_i, y_i^o, y_i^m, \psi) \]

MCAR \quad \rightarrow \quad MAR \quad \rightarrow \quad MNAR

\[ f(r_i | X_i, \psi) \quad f(r_i | X_i, y_i^o, \psi) \quad f(r_i | X_i, y_i^o, y_i^m, \psi) \]

**Pattern-mixture Models:**  
\[ f(y_i | X_i, r_i, \theta) \quad f(r_i | X_i, \psi) \]

**Shared-parameter Models:**  
\[ f(y_i | X_i, b_i, \theta) \quad f(r_i | X_i, b_i, \psi) \]
11.12 Frameworks and Their Methods

\[ f(y_i, r_i | X_i, \theta, \psi) \]

**Selection Models:**

\[ f(y_i | X_i, \theta) f(r_i | X_i, y_i^o, y_i^m, \psi) \]

- **MCAR/simple** \(\rightarrow\) **MAR** \(\rightarrow\) **MNAR**

- **CC?** \(\rightarrow\) **direct likelihood!** \(\rightarrow\) **joint model!?**
- **LOCF?** \(\rightarrow\) **direct Bayesian!** \(\rightarrow\) **sensitivity analysis?!**
- **single imputation?** \(\rightarrow\) **multiple imputation (MI)!**
- **IPW** \(\supset\) **W-GEE!**
- **d.l. + IPW = double robustness! (consensus)**
11.13 Frameworks and Their Methods: Start

\[ f(y_i, r_i | X_i, \theta, \psi) \]

**Selection Models:**

\[ f(y_i | X_i, \theta) f(r_i | X_i, y_i^o, y_i^m, \psi) \]

\[ \text{MCAR/simple} \rightarrow \text{MAR} \rightarrow \text{MNAR} \]

direct likelihood!
direct Bayesian!
multiple imputation (MI)!
IPW ⊃ W-GEE!
d.l. + IPW = double robustness!
11.14 Frameworks and Their Methods: Next

\[ f(y_i, r_i | X_i, \theta, \psi) \]

**Selection Models:**

\[ f(y_i | X_i, \theta) f(r_i | X_i, y_i^o, y_i^m, \psi) \]

- MCAR/simple → MAR → MNAR

- joint model!?  sensitivity analysis!

- PMM
- MI (MGK, J&J)
- local influence
- interval ignorance
- IPW based
### 11.15 Overview and (Premature) Conclusion

<table>
<thead>
<tr>
<th>Type</th>
<th>Methodology</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCAR/simple</td>
<td>CC</td>
<td>biased</td>
<td>not simpler than MAR methods</td>
</tr>
<tr>
<td></td>
<td>LOCF</td>
<td>inefficient</td>
<td></td>
</tr>
<tr>
<td></td>
<td>single imputation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAR</td>
<td>direct lik./Bayes</td>
<td>easy to conduct</td>
<td>Gaussian &amp; non-Gaussian</td>
</tr>
<tr>
<td></td>
<td>IPW/d.r.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>multiple imputation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MNAR</td>
<td>variety of methods</td>
<td>strong, untestable assumptions</td>
<td>most useful in sensitivity analysis</td>
</tr>
</tbody>
</table>

### Notes
- **MCAR/simple**: Data is missing completely at random (MCAR) and is considered simple.
- **MAR**: Data is missing at random (MAR), usually the most common assumption.
- **MNAR**: Data is missing not at random (MNAR), often requiring strong, untestable assumptions for analysis.

RBras, Salvador, May 2016
11.16 Incomplete Longitudinal Data
Data and Modeling Strategies

[Graph showing various data and modeling strategies including Inc.Obs., Unobserved, LOCF 'data', MAR, LOCF, AC, CC, and Comp.Obs.]
Modeling Strategies

Unobserved

LOCF 'data'

MAR

LOCF

AC

CC

Comp.Obs.

Inc.Obs.
11.17 The Depression Trial

- Clinical trial: experimental drug *versus* standard drug

- 170 patients

- Response: change versus baseline in $HAMD_{17}$ score

- 5 post-baseline visits: 4–8
11.18 Analysis of the Depression Trial

- Treatment effect at visit 8 (last follow-up measurement):

<table>
<thead>
<tr>
<th>Method</th>
<th>Estimate</th>
<th>(s.e.)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC</td>
<td>-1.94</td>
<td>(1.17)</td>
<td>0.0995</td>
</tr>
<tr>
<td>LOCF</td>
<td>-1.63</td>
<td>(1.08)</td>
<td>0.1322</td>
</tr>
<tr>
<td>MAR</td>
<td>-2.38</td>
<td>(1.16)</td>
<td>0.0419</td>
</tr>
</tbody>
</table>

Observe the slightly significant $p$-value under the MAR model.
Chapter 12
Direct Likelihood / Ignorable Likelihood

- Simple methods
- Direct likelihood / ignorability
12.1 Simple Methods

Complete case analysis:
⇒ delete incomplete subjects
  • Standard statistical software
  • Loss of information
  • Impact on precision and power
  • Missingness \neq MCAR ⇒ bias

Last observation carried forward:
⇒ impute missing values
  • Standard statistical software
  • Increase of information
  • Constant profile after dropout: unrealistic
  • Usually bias
12.2 Ignorability

Likelihood/Bayesian + MAR

&

Frequentist + MCAR
12.3 Direct Likelihood Maximization

\[
\text{MAR} : f(Y^o_i | \theta) f(D_i | Y^o_i, \psi)
\]

Mechanism is MAR
\[\theta \text{ and } \psi \text{ distinct}\]
Interest in \(\theta\)
(Use observed information matrix)

\[\implies\] Likelihood inference is valid

<table>
<thead>
<tr>
<th>Outcome type</th>
<th>Modeling strategy</th>
<th>Software</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian</td>
<td>Linear mixed model</td>
<td>SAS proc MIXED</td>
</tr>
<tr>
<td>Non-Gaussian</td>
<td>Generalized linear mixed model</td>
<td>SAS proc GLIMMIX, NLMIXED</td>
</tr>
</tbody>
</table>
### 12.4 Original, Complete Orthodontic Growth Data

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Covar</th>
<th># par</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>unstructured</td>
<td>unstructured</td>
<td>18</td>
</tr>
<tr>
<td>2</td>
<td>≠ slopes</td>
<td>unstructured</td>
<td>14</td>
</tr>
<tr>
<td>3</td>
<td>= slopes</td>
<td>unstructured</td>
<td>13</td>
</tr>
<tr>
<td>7</td>
<td>≠ slopes</td>
<td>CS</td>
<td>6</td>
</tr>
</tbody>
</table>

**Growth Data, Model 1**
Unstructured Means, Unstructured Covariance

**Growth Data, Model 2**
Two Lines, Unstructured Covariance

**Growth Data, Model 3**
Parallel Lines, Unstructured Covariance

**Growth Data, Model 7**
Two Lines, Compound Symmetry
## 12.5 Trimmed Growth Data: Simple Methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Model</th>
<th>Mean</th>
<th>Covar</th>
<th># par</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete case</td>
<td>7a</td>
<td>= slopes</td>
<td>CS</td>
<td>5</td>
</tr>
<tr>
<td>LOCF</td>
<td>2a</td>
<td>quadratic</td>
<td>unstructured</td>
<td>16</td>
</tr>
<tr>
<td>Unconditional mean</td>
<td>7a</td>
<td>= slopes</td>
<td>CS</td>
<td>5</td>
</tr>
<tr>
<td>Conditional mean</td>
<td>1</td>
<td>unstructured</td>
<td>unstructured</td>
<td>18</td>
</tr>
</tbody>
</table>
### 12.6 Trimmed Growth Data: Direct Likelihood

<table>
<thead>
<tr>
<th>Mean</th>
<th>Covar</th>
<th># par</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>slopes</td>
<td>CS 6</td>
</tr>
</tbody>
</table>

**Growth Data, Model 1**
- Missing At Random
- Unstructured Means, Unstructured Covariance

**Growth Data, Model 2**
- Missing At Random
- Two Lines, Unstructured Covariance

**Growth Data, Model 3**
- Missing At Random
- Parallel Lines, Unstructured Covariance

**Growth Data, Model 7**
- Missing At Random
- Two Lines, Compound Symmetry
Chapter 13
Multiple Imputation

- Valid under MAR
- An alternative to direct likelihood and WGEE
- Three steps:
  1. The missing values are filled in $M$ times $\Rightarrow M$ complete data sets
  2. The $M$ complete data sets are analyzed by using standard procedures
  3. The results from the $M$ analyses are combined into a single inference
13.1 Use of MI in Practice

- Many analyses of the same incomplete set of data
- A combination of missing outcomes and missing covariates
- As an alternative to WGEE: MI can be combined with classical GEE
- MI in SAS:
  
  **Imputation Task**: PROC MI
  
  **Analysis Task**: PROC "MYFAVORITE"
  
  **Inference Task**: PROC MIANALYZE
13.2 Generalized Estimating Equations

Liang and Zeger (1986)

\[ S(\beta) = \sum_{i=1}^{N} [D_i]^T [V_i(\alpha)]^{-1} (y_i - \mu_i) = 0 \]

- \( V_i(.) \) is not the true variance of \( Y_i \) but only a plausible guess
- the score equations are solved in a standard way
- Asymptotic distribution:
  \[ \sqrt{N}(\hat{\beta} - \beta) \sim N(0, I_0^{-1}I_1I_0^{-1}) \]

\[ I_0 = \sum_{i=1}^{N} D_i^T [V_i(\alpha)]^{-1} D_i \quad I_1 = \sum_{i=1}^{N} D_i^T [V_i(\alpha)]^{-1} \text{Var}(Y_i)[V_i(\alpha)]^{-1} D_i \]
13.3 Weighted GEE

\[
\pi_i = \prod_{\ell=2}^{n_i} (1 - p_{i\ell})
\]

\[
\pi'_i = \left[ d_i - 1 \prod_{\ell=2}^{d_i-1} (1 - p_{i\ell}) \right] \cdot p_{id_i}
\]

\[
p_{i\ell} = P(D_i = \ell | D_i \geq \ell, Y_{i\ell}, X_{i\ell})
\]

\[ R_i = 1 \text{ if subject } i \text{ is complete} \]
\[ R_i = 0 \text{ if subject } i \text{ is incomplete} \]

\[
S(\beta) = \sum_{i=1}^{N} \frac{R_i}{\pi_i} \frac{\partial \mu_i}{\partial \beta} V_i^{-1}(y_i - \mu_i) = 0
\]

\[
S(\beta) = \sum_{i=1}^{N} \frac{1}{\pi'_i} \frac{\partial \mu_i^o}{\partial \beta'} (V_i^o)^{-1}(y_i^o - \mu_i^o) = 0
\]
Chapter 14
The Analgesic Trial

- single-arm trial with 530 patients recruited (491 selected for analysis)
- analgesic treatment for pain caused by chronic nonmalignant disease
- treatment was to be administered for 12 months
- we will focus on Global Satisfaction Assessment (GSA)
- GSA scale goes from 1=very good to 5=very bad
- GSA was rated by each subject 4 times during the trial, at months 3, 6, 9, and 12.
• Research questions:
  ▶ Evolution over time
  ▶ Relation with baseline covariates: age, sex, duration of the pain, type of pain, disease progression, Pain Control Assessment (PCA), . . .
  ▶ Investigation of dropout

• Frequencies:

<table>
<thead>
<tr>
<th>GSA</th>
<th>Month 3</th>
<th>Month 6</th>
<th>Month 9</th>
<th>Month 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>55</td>
<td>38</td>
<td>40</td>
<td>30</td>
</tr>
<tr>
<td>2</td>
<td>112</td>
<td>84</td>
<td>67</td>
<td>66</td>
</tr>
<tr>
<td>3</td>
<td>151</td>
<td>115</td>
<td>76</td>
<td>97</td>
</tr>
<tr>
<td>4</td>
<td>52</td>
<td>51</td>
<td>33</td>
<td>27</td>
</tr>
<tr>
<td>5</td>
<td>15</td>
<td>14</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>Tot</td>
<td>385</td>
<td>302</td>
<td>227</td>
<td>223</td>
</tr>
</tbody>
</table>
### Missingness:

<table>
<thead>
<tr>
<th>Measurement occasion</th>
<th>Month 3</th>
<th>Month 6</th>
<th>Month 9</th>
<th>Month 12</th>
<th>Number</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Completers</strong></td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>163</td>
<td>41.2</td>
</tr>
<tr>
<td><strong>Dropouts</strong></td>
<td>O</td>
<td>O</td>
<td>M</td>
<td>M</td>
<td>51</td>
<td>12.91</td>
</tr>
<tr>
<td></td>
<td>O</td>
<td>O</td>
<td>M</td>
<td>M</td>
<td>51</td>
<td>12.91</td>
</tr>
<tr>
<td></td>
<td>O</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>63</td>
<td>15.95</td>
</tr>
<tr>
<td><strong>Non-monotone missingness</strong></td>
<td>O</td>
<td>O</td>
<td>M</td>
<td>O</td>
<td>30</td>
<td>7.59</td>
</tr>
<tr>
<td></td>
<td>O</td>
<td>M</td>
<td>O</td>
<td>O</td>
<td>7</td>
<td>1.77</td>
</tr>
<tr>
<td></td>
<td>O</td>
<td>M</td>
<td>O</td>
<td>M</td>
<td>2</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>O</td>
<td>M</td>
<td>M</td>
<td>O</td>
<td>18</td>
<td>4.56</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>2</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>O</td>
<td>O</td>
<td>M</td>
<td>1</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>O</td>
<td>M</td>
<td>O</td>
<td>1</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>O</td>
<td>M</td>
<td>M</td>
<td>3</td>
<td>0.76</td>
</tr>
</tbody>
</table>
14.1 Analysis of the Analgesic Trial

- A logistic regression for the dropout indicator:

\[
\text{logit}[P(D_i = j | D_i \geq j, \cdot)] = \psi_0 + \psi_{11} I(GSA_{i,j-1} = 1) + \psi_{12} I(GSA_{i,j-1} = 2) \\
+ \psi_{13} I(GSA_{i,j-1} = 3) + \psi_{14} I(GSA_{i,j-1} = 4) \\
+ \psi_2 \text{PCA0}_i + \psi_3 \text{PF}_i + \psi_4 \text{GD}_i
\]

with

- \( GSA_{i,j-1} \) the 5-point outcome at the previous time
- \( I(\cdot) \) is an indicator function
- \( \text{PCA0}_i \) is pain control assessment at baseline
- \( \text{PF}_i \) is physical functioning at baseline
- \( \text{GD}_i \) is genetic disorder at baseline are used
<table>
<thead>
<tr>
<th>Effect</th>
<th>Par.</th>
<th>Estimate (s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>$\psi_0$</td>
<td>-1.80 (0.49)</td>
</tr>
<tr>
<td>Previous GSA= 1</td>
<td>$\psi_1$</td>
<td>-1.02 (0.41)</td>
</tr>
<tr>
<td>Previous GSA= 2</td>
<td>$\psi_2$</td>
<td>-1.04 (0.38)</td>
</tr>
<tr>
<td>Previous GSA= 3</td>
<td>$\psi_3$</td>
<td>-1.34 (0.37)</td>
</tr>
<tr>
<td>Previous GSA= 4</td>
<td>$\psi_4$</td>
<td>-0.26 (0.38)</td>
</tr>
<tr>
<td>Basel. PCA</td>
<td>$\psi_5$</td>
<td>0.25 (0.10)</td>
</tr>
<tr>
<td>Phys. func.</td>
<td>$\psi_6$</td>
<td>0.009 (0.004)</td>
</tr>
<tr>
<td>Genetic disfunc.</td>
<td>$\psi_7$</td>
<td>0.59 (0.24)</td>
</tr>
</tbody>
</table>

- There is some evidence for MAR: $P(D_i = j|D_i \geq j)$ depends on previous GSA.

- Furthermore: baseline PCA, physical functioning and genetic/congenital disorder.
- GEE and WGEE:

\[
\text{logit}[P(Y_{ij} = 1| t_j, \text{PCA0}_i)] = \beta_1 + \beta_2 t_j + \beta_3 t_j^2 + \beta_4 \text{PCA0}_i
\]

<table>
<thead>
<tr>
<th>Effect</th>
<th>Par.</th>
<th>GEE</th>
<th>WGEE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>$\beta_1$</td>
<td>2.95 (0.47)</td>
<td>2.17 (0.69)</td>
</tr>
<tr>
<td>Time</td>
<td>$\beta_2$</td>
<td>-0.84 (0.33)</td>
<td>-0.44 (0.44)</td>
</tr>
<tr>
<td>$\text{Time}^2$</td>
<td>$\beta_3$</td>
<td>0.18 (0.07)</td>
<td>0.12 (0.09)</td>
</tr>
<tr>
<td>Basel. PCA</td>
<td>$\beta_4$</td>
<td>-0.24 (0.10)</td>
<td>-0.16 (0.13)</td>
</tr>
</tbody>
</table>

- A hint of potentially important differences between both
14.2 Analgesic Trial: Steps for WGEE in SAS

1. Preparatory data manipulation:

   %dropout(…)

2. Logistic regression for weight model:

   proc genmod data=gsac;
   class prevgsa;
   model dropout = prevgsa pca0 physfct gendis / pred dist=b;
   ods output obstats=pred;
   run;

3. Conversion of predicted values to weights:

   …
   %dropwgt(…)
4. Weighted GEE analysis:

```sql
proc genmod data=repbin.gsaw;
   scwgt wi;
   class patid timecls;
   model gsabin = time|time pca0 / dist=b;
   repeated subject=patid / type=un corrw within=timecls;
run;
```
14.3 Analgesic Trial: Steps for WGEE in SAS, Using PROC GEE

- Experimental in SAS 9.4 (SAS/STAT 13.2)

- Preparation:

```sas
data gsaw;
  set gsaw;
  by patid;
  prevgsa = lag(gsa);
  if first.id then prevgsa = 1;
  time = time-1;
  timeclss = time;
run;
```
Weighted GEE analysis:

ods graphics on;
proc gee data=gsaw plots=histogram;
  class patid timecls prevgsa;
  model gsabin = time|time pca0 / dist=bin;
  repeated subject=patid / within=timecls corr=un;
  missmodel prevgsa pca0 physfunt gendist / type=obslevel;
run;
Chapter 15
Creating Monotone Missingness

- When missingness is non-monotone, one might think of several mechanisms operating simultaneously:
  - A simple (MCAR or MAR) mechanism for the intermittent missing values
  - A more complex (MNAR) mechanism for the missing data past the moment of dropout

- Analyzing such data are complicated, especially with methods that apply to dropout only
• Solution:
  ▶ Generate multiple imputations that render the datasets monotone missing, by including into the MI procedure:

    mcmc impute = monotone;

  ▶ Apply method of choice to the so-completed multiple sets of data

• Note: this is different from the monotone method in PROC MI, intended to fully complete already monotone sets of data
15.1 Example: Creating Monotone Missingness to Then Apply Weighted GEE

- Consider again the analgesic trial

- Multiple imputation to create monotone missingness:

  ```
  proc mi data=m.gsa4 seed=459864 simple nimpute=10
  round=0.1 out=m.gsaimput;
  title 'Monotone multiple imputation';
  mcmc impute = monotone;
  var pca0 physfct gsa1 gsa2 gsa3 gsa4;
  run;
  ```
Preparation of the data in vertical format, so that the data can be used in ordinary GEE:

```plaintext
data m.gsa4;
set m.gsa4;
array y (4) gsa1 gsa2 gsa3 gsa4;
do j=1 to 4;
   gsa=y(j);
   time=j;
   timecls=time;
   gsabin=.;
   if gsa=1 then gsabin=1;
   if gsa=2 then gsabin=1;
   if gsa=3 then gsabin=1;
   if gsa=4 then gsabin=0;
   if gsa=5 then gsabin=0;
output;
end;
run;
```
• Standard GEE:

```plaintext
proc gee data=m.gsaw plots=histogram;
  title 'Standard GEE for GSA data';
  class patid timecls;
  model gsabin = time|time pca0 / dist=bin;
  repeated subject=patid / within=timecls corr=un;
run;
```

• Steps to prepare the data for weighted GEE, including definition of the ‘previous’ outcome:

```plaintext
data m.gsaimput02;
set m.gsaimput;
array y (4) gsa1 gsa2 gsa3 gsa4;
do j=1 to 4;
  gsa=y(j);
  time=j;
  timecls=time;
```
\begin{verbatim}
patid2=1000*_imputation_*+patid;
output;
end;
run;

proc sort data=m.gsaimput02;
by _imputation_ patid2;
run;

data m.gsaimput03;
   set m.gsaimput02;
   by patid2;
   prevgsa = lag(gsa);
   if time=1 then prevgsa = 1;
   timeclss = time;
run;
\end{verbatim}
data m.gsaimput03;
    set m.gsaimput03;
    if gsa<=3.5 then gsabin=1;
    if gsa>3.5 then gsabin=0;
    gsabin=gsabin+gsa-gsa;
run;
• Weighted GEE, where weights are created at observation level:

```plaintext
ods graphics on;
proc gee data=m.gsaimput03 plots=histogram;
   title 'Weighted GEE for GSA Data Based on Multiple
         Imputation to Monotonize - OBSLEVEL’;
   by _imputation_;  
   class patid timecls;
   model gsabin = time|time pca0 / dist=bin covb;
   repeated subject=patid / within=timecls corr=un ecovb;
   missmodel prevgsa pca0 physfct / type=obslevel;
   ods output GEEmpPEst=gmparms parminfo=gmpinfo
                    modelinfo=modelinfo GEERCov=gmcovb;
run;

proc mianalyze parms=gmparms parminfo=gmpinfo covb=gmcovb;
   title ’Multiple Imputation Analysis After Weighted GEE for GSA Data’;
   modeleffects intercept time time*time pca0;
run;
```
To use weights at subject rather than observation level:

missmodel prevgsa pca0 physfct / type=sublevel;

Evidently, using these monotonized data, also standard GEE can be used:

ods graphics on;
proc gee data=m.gsaimput03 plots=histogram;
  title 'Standard GEE for GSA Data Based on Multiple Imputation to Monotonize';
  by _imputation_;
  class patid timecls;
  model gsabin = time|time pca0 / dist=bin covb;
  repeated subject=patid / within=timecls corr=un ecovb;
ods output GEEmpPEst=gmparms parminfo=gmpinfo
  modelinfo=modelinfo GEERCov=gmcovb;
run;
• Files:
  ▶ analg11(met-proc-gee).sas
  ▶ analg11(met-proc-gee).lst

• Overview of results:
<table>
<thead>
<tr>
<th>Effect</th>
<th>Par.</th>
<th>Est.(s.e.)</th>
<th>p-value</th>
<th>Est.(s.e.)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Without MI</strong></td>
<td><strong>After MI</strong></td>
<td><strong>Without MI</strong></td>
<td><strong>After MI</strong></td>
</tr>
<tr>
<td>Intercept</td>
<td>$\beta_0$</td>
<td>2.90(0.46)</td>
<td></td>
<td>2.87(0.45)</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>$\beta_1$</td>
<td>-0.81(0.32)</td>
<td>0.0124</td>
<td>-0.83(0.32)</td>
<td>0.0087</td>
</tr>
<tr>
<td>Time$^2$</td>
<td>$\beta_2$</td>
<td>0.17(0.07)</td>
<td>0.0083</td>
<td>0.18(0.06)</td>
<td>0.0058</td>
</tr>
<tr>
<td>PCA$^0$</td>
<td>$\beta_3$</td>
<td>-0.23(0.10)</td>
<td>0.0178</td>
<td>-0.21(0.10)</td>
<td>0.0253</td>
</tr>
</tbody>
</table>

**Weighted GEE (after MI)**

<table>
<thead>
<tr>
<th>Effect</th>
<th>Par.</th>
<th>Est.(s.e.)</th>
<th>p-value</th>
<th>Est.(s.e.)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td><strong>Observation level</strong></td>
<td><strong>Subject level</strong></td>
<td><strong>Observation level</strong></td>
<td><strong>Subject level</strong></td>
</tr>
<tr>
<td>Intercept</td>
<td>$\beta_0$</td>
<td>2.74(0.46)</td>
<td></td>
<td>2.62(0.60)</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>$\beta_1$</td>
<td>-0.76(0.33)</td>
<td>0.0231</td>
<td>-0.71(0.40)</td>
<td>0.0747</td>
</tr>
<tr>
<td>Time$^2$</td>
<td>$\beta_2$</td>
<td>0.17(0.07)</td>
<td>0.0155</td>
<td>0.16(0.08)</td>
<td>0.0444</td>
</tr>
<tr>
<td>PCA$^0$</td>
<td>$\beta_3$</td>
<td>-0.19(0.10)</td>
<td>0.0384</td>
<td>-0.21(0.12)</td>
<td>0.0853</td>
</tr>
</tbody>
</table>
• The dropout model is similar but slightly different than the one used with PROC GENMOD.

• Weighted GEE leads to increased standard errors, as observed before.

• This effect is less pronounced when weights are constructed at observation level, rather than at subject level.

• A typical output for one of the imputed datasets takes the form (first imputation out of ten; with weights at observation level):
Parameter Estimates for Response Model
with Empirical Standard Error

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Error</th>
<th>Limits</th>
<th>Z</th>
<th>Pr &gt;</th>
<th>Z</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>2.4299</td>
<td>0.5890</td>
<td>1.2755</td>
<td>3.5843</td>
<td>4.13</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>TIME</td>
<td>-0.5881</td>
<td>0.3912</td>
<td>-1.3548</td>
<td>0.1787</td>
<td>-1.50</td>
<td>0.1328</td>
<td></td>
</tr>
<tr>
<td>TIME*TIME</td>
<td>0.1392</td>
<td>0.0794</td>
<td>-0.0165</td>
<td>0.2949</td>
<td>1.75</td>
<td>0.0796</td>
<td></td>
</tr>
<tr>
<td>PCA0</td>
<td>-0.1797</td>
<td>0.1173</td>
<td>-0.4096</td>
<td>0.0501</td>
<td>-1.53</td>
<td>0.1254</td>
<td></td>
</tr>
</tbody>
</table>

Parameter Estimates for Missingness Model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Error</th>
<th>Limits</th>
<th>Z</th>
<th>Pr &gt;</th>
<th>Z</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>3.1335</td>
<td>0.4060</td>
<td>2.3377</td>
<td>3.9293</td>
<td>7.72</td>
<td>&lt;.0001</td>
<td></td>
</tr>
<tr>
<td>prevgsa</td>
<td>-0.1974</td>
<td>0.0822</td>
<td>-0.3585</td>
<td>-0.0363</td>
<td>-2.40</td>
<td>0.0163</td>
<td></td>
</tr>
<tr>
<td>PCA0</td>
<td>-0.2495</td>
<td>0.0956</td>
<td>-0.4370</td>
<td>-0.0621</td>
<td>-2.61</td>
<td>0.0091</td>
<td></td>
</tr>
<tr>
<td>PHYSFCT</td>
<td>-0.0079</td>
<td>0.0037</td>
<td>-0.0151</td>
<td>-0.0007</td>
<td>-2.16</td>
<td>0.0311</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 16
Case Study: The Dataset

▷ Visual Acuity
▷ Age-related Macular Degeneration
▷ Missingness
16.1 Visual Acuity
16.2 Age-related Macular Degeneration Trial

- Pharmacological Therapy for Macular Degeneration Study Group (1997)
- An ocular pressure disease which makes patients progressively lose vision
- 240 patients enrolled in a multi-center trial (190 completers)
- **Treatment:** Interferon-\(\alpha\) (6 million units) versus placebo
- **Visits:** baseline and follow-up at 4, 12, 24, and 52 weeks
- **Continuous outcome:** visual acuity: \# letters correctly read on a vision chart
- **Binary outcome:** visual acuity versus baseline \(\geq 0\) or \(\leq 0\)
### Missingness:

<table>
<thead>
<tr>
<th>Measurement occasion</th>
<th>4 wks</th>
<th>12 wks</th>
<th>24 wks</th>
<th>52 wks</th>
<th>Number</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O        O        O        O        O</td>
<td>188</td>
<td>78.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dropouts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O        O        O        M        M</td>
<td>24</td>
<td>10.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O        O        M        M        M</td>
<td>8</td>
<td>3.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O        M        M        M        M</td>
<td>6</td>
<td>2.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M        M        M        M        M</td>
<td>6</td>
<td>2.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-monotone missingness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O        O        M        M        O</td>
<td>4</td>
<td>1.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O        M        M        M        O</td>
<td>1</td>
<td>0.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M        O        O        O        O</td>
<td>2</td>
<td>0.83</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M        O        M        M        M</td>
<td>1</td>
<td>0.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Chapter 17
Case Study: Weighted Generalized Estimating Equations

▷ Model for the weights
▷ Incorporating the weights within GEE
17.1 Analysis of the ARMD Trial

• Model for the weights:

\[
\text{logit}[P(D_i = j | D_i \geq j)] = \psi_0 + \psi_1 y_{i,j-1} + \psi_2 T_i + \psi_3 L_{1i} + \psi_4 L_{2i} + \psi_5 L_{3i} \\
+ \psi_6 I(t_j = 2) + \psi_7 I(t_j = 3)
\]

with

▷ \(y_{i,j-1}\) the binary outcome at the previous time \(t_{i,j-1} = t_{j-1}\) (since time is common to all subjects)

▷ \(T_i = 1\) for interferon-\(\alpha\) and \(T_i = 0\) for placebo

▷ \(L_{ki} = 1\) if the patient’s eye lesion is of level \(k = 1, \ldots, 4\) (since one dummy variable is redundant, only three are used)

▷ \(I(\cdot)\) is an indicator function
• Results for the weights model:

<table>
<thead>
<tr>
<th>Effect</th>
<th>Parameter</th>
<th>Estimate (s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>$\psi_0$</td>
<td>0.14 (0.49)</td>
</tr>
<tr>
<td>Previous outcome</td>
<td>$\psi_1$</td>
<td>0.04 (0.38)</td>
</tr>
<tr>
<td>Treatment</td>
<td>$\psi_2$</td>
<td>-0.86 (0.37)</td>
</tr>
<tr>
<td>Lesion level 1</td>
<td>$\psi_{31}$</td>
<td>-1.85 (0.49)</td>
</tr>
<tr>
<td>Lesion level 2</td>
<td>$\psi_{32}$</td>
<td>-1.91 (0.52)</td>
</tr>
<tr>
<td>Lesion level 3</td>
<td>$\psi_{33}$</td>
<td>-2.80 (0.72)</td>
</tr>
<tr>
<td>Time 2</td>
<td>$\psi_{41}$</td>
<td>-1.75 (0.49)</td>
</tr>
<tr>
<td>Time 3</td>
<td>$\psi_{42}$</td>
<td>-1.38 (0.44)</td>
</tr>
</tbody>
</table>
• GEE:

\[
\logit[P(Y_{ij} = 1|T_i, t_j)] = \beta_{j1} + \beta_{j2}T_i
\]

with

- \( T_i = 0 \) for placebo and \( T_i = 1 \) for interferon-\( \alpha \)
- \( t_j (j = 1, \ldots, 4) \) refers to the four follow-up measurements
- Classical GEE and linearization-based GEE
- Comparison between CC, LOCF, and GEE analyses

• SAS code:

```sas
proc genmod data=armdhlp descending;
class trt prev lesion time;
model dropout = prev trt lesion time / pred dist=b;
ods output obstats=pred;
ods listing exclude obstats;
run;
```
... proc genmod data=armdwgee;
title ’data as is - WGEE’;
weight wi;
class time treat subject;
model bindif = time treat*time / noint dist=binomial;
repeated subject=subject / withinsubject=time type=exch modelsel;
run;

proc glimmix data=armdwgee empirical;
title ’data as is - WGEE - linearized version - empirical’;
weight wi;
nloptions maxiter=50 technique=newrap;
class time treat subject;
model bindif = time treat*time / noint solution dist=binary ;
random _residual_ / subject=subject type=cs;
run;
## Results:

<table>
<thead>
<tr>
<th>Effect</th>
<th>Par.</th>
<th>CC</th>
<th>LOCF</th>
<th>Observed data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Unweighted</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Standard GEE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Int.4</td>
<td>$\beta_{11}$</td>
<td>$-1.01(0.24;0.24)$</td>
<td>$-0.87(0.20;0.21)$</td>
<td>$-0.87(0.21;0.21)$</td>
</tr>
<tr>
<td>Int.12</td>
<td>$\beta_{21}$</td>
<td>$-0.89(0.24;0.24)$</td>
<td>$-0.97(0.21;0.21)$</td>
<td>$-1.01(0.21;0.21)$</td>
</tr>
<tr>
<td>Int.24</td>
<td>$\beta_{31}$</td>
<td>$-1.13(0.25;0.25)$</td>
<td>$-1.05(0.21;0.21)$</td>
<td>$-1.07(0.22;0.22)$</td>
</tr>
<tr>
<td>Int.52</td>
<td>$\beta_{41}$</td>
<td>$-1.64(0.29;0.29)$</td>
<td>$-1.51(0.24;0.24)$</td>
<td>$-1.71(0.29;0.29)$</td>
</tr>
<tr>
<td>Tr.4</td>
<td>$\beta_{12}$</td>
<td>$0.40(0.32;0.32)$</td>
<td>$0.22(0.28;0.28)$</td>
<td>$0.22(0.28;0.28)$</td>
</tr>
<tr>
<td>Tr.12</td>
<td>$\beta_{22}$</td>
<td>$0.49(0.31;0.31)$</td>
<td>$0.55(0.28;0.28)$</td>
<td>$0.61(0.29;0.29)$</td>
</tr>
<tr>
<td>Tr.24</td>
<td>$\beta_{32}$</td>
<td>$0.48(0.33;0.33)$</td>
<td>$0.42(0.29;0.29)$</td>
<td>$0.44(0.30;0.30)$</td>
</tr>
<tr>
<td>Tr.52</td>
<td>$\beta_{42}$</td>
<td>$0.40(0.38;0.38)$</td>
<td>$0.34(0.32;0.32)$</td>
<td>$0.44(0.37;0.37)$</td>
</tr>
<tr>
<td>Corr.</td>
<td>$\rho$</td>
<td>$0.39$</td>
<td>$0.44$</td>
<td>$0.39$</td>
</tr>
</tbody>
</table>
Chapter 18
Case Study: Multiple Imputation

▷ Settings and Models
▷ Results for GEE
▷ Results for GLMM
18.1 MI Analysis of the ARMD Trial

- $M = 10$ imputations

- **GEE:**
  \[
  \text{logit}[P(Y_{ij} = 1|T_i, t_j)] = \beta_{j1} + \beta_{j2}T_i
  \]

- **GLMM:**
  \[
  \text{logit}[P(Y_{ij} = 1|T_i, t_j, b_i)] = \beta_{j1} + b_i + \beta_{j2}T_i, \quad b_i \sim N(0, \tau^2)
  \]

- $T_i = 0$ for placebo and $T_i = 1$ for interferon-$\alpha$

- $t_j$ ($j = 1, \ldots, 4$) refers to the four follow-up measurements

- Imputation based on the continuous outcome
### Results:

<table>
<thead>
<tr>
<th>Effect</th>
<th>Par.</th>
<th>GEE</th>
<th>GLMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Int.4</td>
<td>$\beta_{11}$</td>
<td>-0.84(0.20)</td>
<td>-1.46(0.36)</td>
</tr>
<tr>
<td>Int.12</td>
<td>$\beta_{21}$</td>
<td>-1.02(0.22)</td>
<td>-1.75(0.38)</td>
</tr>
<tr>
<td>Int.24</td>
<td>$\beta_{31}$</td>
<td>-1.07(0.23)</td>
<td>-1.83(0.38)</td>
</tr>
<tr>
<td>Int.52</td>
<td>$\beta_{41}$</td>
<td>-1.61(0.27)</td>
<td>-2.69(0.45)</td>
</tr>
<tr>
<td>Trt.4</td>
<td>$\beta_{12}$</td>
<td>0.21(0.28)</td>
<td>0.32(0.48)</td>
</tr>
<tr>
<td>Trt.12</td>
<td>$\beta_{22}$</td>
<td>0.60(0.29)</td>
<td>0.99(0.49)</td>
</tr>
<tr>
<td>Trt.24</td>
<td>$\beta_{32}$</td>
<td>0.43(0.30)</td>
<td>0.67(0.51)</td>
</tr>
<tr>
<td>Trt.52</td>
<td>$\beta_{42}$</td>
<td>0.37(0.35)</td>
<td>0.52(0.56)</td>
</tr>
<tr>
<td>R.I. s.d.</td>
<td>$\tau$</td>
<td>2.20(0.26)</td>
<td></td>
</tr>
<tr>
<td>R.I. var.</td>
<td>$\tau^2$</td>
<td>4.85(1.13)</td>
<td></td>
</tr>
</tbody>
</table>
18.2 SAS Code for MI

1. Preparatory data analysis so that there is one line per subject

2. The imputation task:

```sas
proc mi data=armd13 seed=486048 out=armd13a simple nimpute=10 round=0.1;
  var lesion diff4 diff12 diff24 diff52;
  by treat;
run;
```

Note that the imputation task is conducted on the continuous outcome ‘diff’, indicating the difference in number of letters versus baseline

3. Then, data manipulation takes place to define the binary indicators and to create a longitudinal version of the dataset
4. The analysis task (GEE):

```
proc genmod data=armd13c;
  class time subject;
  by _imputation_; 
  model bindif = time1 time2 time3 time4 
       trttime1 trttime2 trttime3 trttime4 
     / noint dist=binomial covb;
  repeated subject=subject / withinsubject=time type=exch modelse;
  ods output ParameterEstimates=gmparms parminfo=gmpinfo CovB=gmcovb;
run;
```
5. The analysis task (GLMM):

```sas
proc nlmixed data=armd13c qpoints=20 maxiter=100 technique=newrap cov ecov;
   by _imputation_;
   eta = beta11*time1+beta12*time2+beta13*time3+beta14*time4+b
       +beta21*trttime1+beta22*trttime2+beta23*trttime3+beta24*trttime4;
   p = exp(eta)/(1+exp(eta));
model bindif ~ binary(p);
random b ~ normal(0,tau*tau) subject=subject;
estimate 'tau2' tau*tau;
ods output ParameterEstimates=nlparms CovMatParmEst=nlcovb
   AdditionalEstimates=nlparmsa CovMatAddEst=nlcovba;
run;
```
6. The inference task (GEE):

```plaintext
proc mianalyze parms=gmparms covb=gmcovb parminfo=gmpinfo wcov bcov tcov;
modeleffects time1 time2 time3 time4 trttime1 trttime2 trttime3 trttime4;
run;
```

7. The inference task (GLMM):

```plaintext
proc mianalyze parms=nlparsms covb=nlcovb wcov bcov tcov;
modeleffects beta11 beta12 beta13 beta14 beta21 beta22 beta23 beta24;
run;
```
18.3 Example of Sensitivity Analysis

- We apply a shift to the treatment group:

```sas
proc mi data=m.armd13 seed=486048 simple out=m.armd13as1
   nimpute=10 round=0.1;
   title 'Shift multiple imputation';
   class treat;
   var lesion diff4 diff12 diff24 diff52;
   fcs reg;
   mnar adjust (diff12 / shift=10 adjustobs=(treat='2'));
   mnar adjust (diff24 / shift=15 adjustobs=(treat='2'));
   mnar adjust (diff52 / shift=20 adjustobs=(treat='2'));
   by treat;
run;
```
- Expanded results (for GLMM only):

<table>
<thead>
<tr>
<th>Effect</th>
<th>Par.</th>
<th>GEE</th>
<th>MAR</th>
<th>shift</th>
</tr>
</thead>
<tbody>
<tr>
<td>Int.4</td>
<td>$\beta_{11}$</td>
<td>$-0.84(0.20)$</td>
<td>$-1.46(0.36)$</td>
<td>$-1.42(0.36)$</td>
</tr>
<tr>
<td>Int.12</td>
<td>$\beta_{21}$</td>
<td>$-1.02(0.22)$</td>
<td>$-1.75(0.38)$</td>
<td>$-1.67(0.38)$</td>
</tr>
<tr>
<td>Int.24</td>
<td>$\beta_{31}$</td>
<td>$-1.07(0.23)$</td>
<td>$-1.83(0.38)$</td>
<td>$-1.83(0.39)$</td>
</tr>
<tr>
<td>Int.52</td>
<td>$\beta_{41}$</td>
<td>$-1.61(0.27)$</td>
<td>$-2.69(0.45)$</td>
<td>$-2.77(0.44)$</td>
</tr>
<tr>
<td>Trt.4</td>
<td>$\beta_{12}$</td>
<td>$0.21(0.28)$</td>
<td>$0.32(0.48)$</td>
<td>$0.24(0.48)$</td>
</tr>
<tr>
<td>Trt.12</td>
<td>$\beta_{22}$</td>
<td>$0.60(0.29)$</td>
<td>$0.99(0.49)$</td>
<td>$0.91(0.50)$</td>
</tr>
<tr>
<td>Trt.24</td>
<td>$\beta_{32}$</td>
<td>$0.43(0.30)$</td>
<td>$0.67(0.51)$</td>
<td>$0.65(0.51)$</td>
</tr>
<tr>
<td>Trt.52</td>
<td>$\beta_{42}$</td>
<td>$0.37(0.35)$</td>
<td>$0.52(0.56)$</td>
<td>$0.60(0.55)$</td>
</tr>
<tr>
<td>R.I. s.d.</td>
<td>$\tau$</td>
<td>$2.20(0.26)$</td>
<td>$2.20(0.25)$</td>
<td></td>
</tr>
<tr>
<td>R.I. var.</td>
<td>$\tau^2$</td>
<td>$4.85(1.13)$</td>
<td>$4.83(1.08)$</td>
<td></td>
</tr>
</tbody>
</table>
A shift applied to one sex group in the orthodontic growth data:

```
proc mi data=m.growthmi seed=459864 simple n impute=10
    round=0.1 out=outmishift;
title 'Shift multiple imputation';
class sex;
by sex;
monotone method=reg;
mnar adjust (meas10 / shift=10 adjustobs=(sex='2'));
var meas8 meas12 meas14 meas10;
run;
```
• A sensitivity analysis following PMM ideas: NCMV-based imputation:

    proc mi data=m.growthmi seed=459864 simple nimpute=10
        round=0.1 out=outmincmv;
    title 'NCMV multiple imputation';
    by sex;
    monotone method=reg;
    mnar model (meas10 / modelobs=ncmv);
    var meas8 meas12 meas14 meas10;
    run;

• The latter is available only for monotone missingness.
### 18.4 Overview and (No Longer Premature) Conclusion

<table>
<thead>
<tr>
<th>MCAR/simple</th>
<th>CC</th>
<th>biased</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LOCF</td>
<td>inefficient</td>
</tr>
<tr>
<td></td>
<td>single imputation</td>
<td>not simpler than MAR methods</td>
</tr>
<tr>
<td>MAR</td>
<td>direct lik./Bayes</td>
<td>easy to conduct</td>
</tr>
<tr>
<td></td>
<td>IPW/d.r.</td>
<td>Gaussian &amp; non-Gaussian</td>
</tr>
<tr>
<td></td>
<td>multiple imputation</td>
<td></td>
</tr>
<tr>
<td>MNAR</td>
<td>variety of methods</td>
<td>strong, untestable assumptions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>most useful in sensitivity analysis</td>
</tr>
</tbody>
</table>