Module 9

Generalized Estimating Equations and Mixed-Effects Models for Longitudinal Data Analysis

Anna Plantinga, PhD
Department of Mathematics and Statistics, Williams College
Katie Wilson, PhD
Department of Biostatistics, University of Washington

SISCER 2022
July 18-19, 2022

Overview

Review of Generalized Linear Models

Generalized linear models (GLMs): class of models for regression analysis of observations that includes **linear regression models** for continuous responses, but also others:

- **Logistic regression** for binary response (e.g. yes/no or 0/1)
- **Log-linear** or **Poisson regression** for counts
- Others. For example ordinal models

Review of Generalized Linear Models

- Assume the outcomes are independent of each other
- Suppose we have N independent observations of a response variable Y
- Let Y_i denote the response variable for the ith subject
- $\mathbf{X}_i = (X_{i1}, \ldots, X_{ip})$ where X_{ik} denotes the kth covariate for the ith subject

Two part specification:

1. Random component (usually a distributional assumption)
2. Systematic component (how the mean relates to covariates)

Review of GLMs: Random Component

Assume we know the distribution of the outcome,

$$Y_i|\mu_i \sim \text{exponential family distribution}$$

- Linear regression: $Y_i \sim N(\mu_i, \sigma^2)$
- Logistic regression: $Y_i \sim \text{Bernoulli}(\mu_i)$
- Poisson regression: $Y_i \sim \text{Poisson}(\mu_i)$

Distribution	Mean	Variance Function
Normal	μ_i	σ^2
Bernoulli	μ_i	$\mu_i(1-\mu_i)$
Poisson	μ_i	μ_i

Review of GLMs: Systematic Component

Assume the transformed mean of Y_i given $\mathbf{X}_i$ ($\mu_i = E[Y_i|\mathbf{X}_i]$) are related to the covariates in the following manner:

$$g(\mu_i) = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip} = \mathbf{X}_i\boldsymbol{\beta}$$

The **link function** $g(\cdot)$ describes the relationship between the linear predictor ($\mathbf{X}_i\boldsymbol{\beta}$) and the expected value of Y_i (i.e., μ_i)

Distribution	Typical link function $g(\cdot)$
Normal	Identity: $g(\mu_i) = \mu_i$
Bernoulli	Logit: $g(\mu_i) = \text{logit}(\mu_i) = \log\left(\frac{\mu_i}{1-\mu_i}\right)$
Poisson	Log: $g(\mu_i) = \log(\mu_i)$

Binary Outcome: Example

Clinical Trial of Contracepting Women:

- 1151 women randomized to either 100mg or 150mg of DMPA
- **Outcome of interest**: Amenorrhea during each 3-month interval after an injection
  - Since participants were measured multiple times we would want to use correlated data techniques to analyze all responses
  - We will focus on the response at the 4th (last scheduled) visit
- Is there a difference in the odds of amenorrhea after 1 year of injections by dose?

Data:

- 100mg: 50.1% (181 out of 361) experienced amenorrhea at 1 year
- 150mg: 53.5% (189 out of 353) experienced amenorrhea at 1 year

Binary Outcome: Example

We will use logistic regression:

- **Random component**: $Y_i \sim \text{Bernoulli}(\mu_i) \rightarrow \text{Var}(Y_i) = \mu_i(1-\mu_i)$
- **Systematic component**:

$$\text{logit}(\mu_i) = \log\left(\frac{\mu_i}{1-\mu_i}\right) = \beta_0 + \beta_1 \text{Dose}_i$$

where $\mu_i = \Pr(Y_i = 1)$ where Y_i is an indicator for whether person i experienced amenorrhea at 1 year ($1 =$ yes; $0 =$ no)

In R:

```
glm(amenorrhea ~ dose, family = binomial(link = "logit"),
data = amenorrhea[amenorrhea$visit==4,])
```

Binary Outcome: Example

```
glm(formula = amenorrhea ~ dose, family = binomial(link = "logit"),
data = amenorrhea[amenorrhea$visit == 4, ])

Deviance Residuals:
Min       1Q  Median      3Q     Max
-1.24   -1.18    1.12    1.18    1.18

Coefficients:
Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.00554    0.10526    0.05     0.96
dose         0.13634    0.14990    0.91     0.36

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 988.87  on 713  degrees of freedom
Residual deviance: 988.04  on 712  degrees of freedom
(437 observations deleted due to missingness)
AIC: 992

Number of Fisher Scoring iterations: 3
```

Binary Outcome: Example

- Estimate for `(Intercept)` = estimate of the log odds of amenorrhea for 100mg dose
- Estimate for `dose` = estimate of the log odds ratio of amenorrhea comparing 150mg dose to 100mg dose

Interpretation: the odds of amenorrhea is estimated to be 15% ($= \exp(0.136) - 1$) higher for those on the 150mg dose as compared to those on the 100mg dose. We do not have evidence that amenorrhea after 1 year differs by dose ($p = 0.36$).

Overview

GEE

- ⋆ Contrast average outcome values across populations of individuals defined by covariate values, while accounting for correlation
  - Focus on a generalized linear model with regression parameters $\boldsymbol{\beta}$, which characterize the systemic variation in Y across covariates X

$$
\begin{aligned}
\mathbf{Y}_i &= \{Y_{i1}, Y_{i2}, \ldots, Y_{in_i}\}^{\mathsf{T}} && \text{Outcomes} \\
\mathbf{X}_{ij} &= \{1, X_{ij1}, X_{ij2}, \ldots, X_{ijp}\} && \text{Covariates} \\
\mathbf{X}_i &= \{\mathbf{X}_{i1}, \mathbf{X}_{i2}, \ldots, \mathbf{X}_{in_i}\}^{\mathsf{T}} && \text{Design matrix} \\
\boldsymbol{\beta} &= \{\beta_0, \beta_1, \beta_2, \ldots, \beta_p\}^{\mathsf{T}} && \text{Regression parameters}
\end{aligned}
$$

  for $i = 1, \ldots, N$ and $j = 1, \ldots, n_i$

  - Longitudinal correlation structure is a nuisance feature of the data

(Liang and Zeger, 1986)

Mean model

Assumptions

- Observations are independent across subjects
- Observations may be correlated within subjects

Mean model: Primary focus of the analysis

$$\begin{aligned} \mathrm{E}[Y_{ij} \mid \mathbf{X}_{ij}] &= \mu_{ij} \\ g(\mu_{ij}) &= \mathbf{X}_{ij}\boldsymbol{\beta} \end{aligned}$$

- May correspond to any generalized linear model with link $g(\cdot)$

Continuous outcome	Count outcome	Binary outcome
$\mathrm{E}[Y_{ij} \mid \mathbf{X}_{ij}] = \mu_{ij}$	$\mathrm{E}[Y_{ij} \mid \mathbf{X}_{ij}] = \mu_{ij}$	$\mathrm{P}[Y_{ij} = 1 \mid \mathbf{X}_{ij}] = \mu_{ij}$
$\mu_{ij} = \mathbf{X}_{ij}\beta$	$\log(\mu_{ij}) = \mathbf{X}_{ij}\beta$	$\mathrm{logit}(\mu_{ij}) = \mathbf{X}_{ij}\beta$

- Characterizes a marginal mean regression model
  - μ_{ij} does not condition on anything other than $\mathbf{X}_{ij}$

Covariance model

Longitudinal correlation is a nuisance; secondary to mean model of interest

1. Assume a form for variance that may depend on μ_{ij}

$$\begin{aligned} \text{Continuous outcome:} \quad & \text{Var}[Y_{ij} \mid \mathbf{X}_{ij}] = \sigma^2 \\ \text{Count outcome:} \quad & \text{Var}[Y_{ij} \mid \mathbf{X}_{ij}] = \mu_{ij} \\ \text{Binary outcome:} \quad & \text{Var}[Y_{ij} \mid \mathbf{X}_{ij}] = \mu_{ij}(1-\mu_{ij}) \end{aligned}$$

   which may also include a scale or dispersion parameter $\phi > 0$

2. Select a model for longitudinal correlation with parameters α

$$\begin{aligned} \text{Independence:} \quad & \text{Corr}[Y_{ij}, Y_{ij'} \mid \mathbf{X}_i] = 0 \\ \text{Exchangeable:} \quad & \text{Corr}[Y_{ij}, Y_{ij'} \mid \mathbf{X}_i] = \alpha \\ \text{Auto-regressive:} \quad & \text{Corr}[Y_{ij}, Y_{ij'} \mid \mathbf{X}_i] = \alpha^{|j-j'|} \\ \text{Unstructured:} \quad & \text{Corr}[Y_{ij}, Y_{ij'} \mid \mathbf{X}_i] = \alpha_{jj'} \end{aligned}$$

Covariance model

Longitudinal correlation is a nuisance; secondary to mean model of interest

- Assume a form for variance that depends on μ
- Select a model for longitudinal correlation with parameters α

$$\text{Var}[Y_{ij} \mid \mathbf{X}_{ij}] = V(\mu_{ij}) \qquad \rightarrow \qquad \mathbf{S}_i(\mu_i) = \text{diag } V(\mu_{ij})$$

$$\text{Corr}[Y_{ij},\ Y_{ij'} \mid \mathbf{X}_i] = \rho_{ijj'}(\boldsymbol{\alpha}) \qquad \rightarrow \qquad \mathbf{R}_i(\boldsymbol{\alpha}) = \text{matrix } \rho(\boldsymbol{\alpha})$$

$$\text{Cov}[\mathbf{Y}_i \mid \mathbf{X}_i] = \mathbf{V}_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \mathbf{S}_i^{1/2}\mathbf{R}_i\mathbf{S}_i^{1/2}$$

Correlation models

Independence: $\text{Corr}[Y_{ij}, Y_{ij'} \mid \mathbf{X}_i] = 0$

$$\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

Exchangeable: $\text{Corr}[Y_{ij}, Y_{ij'} \mid \mathbf{X}_i] = \alpha$

$$\begin{bmatrix} 1 & \alpha & \alpha & \cdots & \alpha \\ \alpha & 1 & \alpha & \cdots & \alpha \\ \alpha & \alpha & 1 & \cdots & \alpha \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha & \alpha & \alpha & \cdots & 1 \end{bmatrix}$$

Correlation models

Auto-regressive: $\text{Corr}[Y_{ij}, Y_{ij'} \mid \mathbf{X}_i] = \alpha^{|j-j'|}$

$$\begin{bmatrix} 1 & \alpha & \alpha^2 & \cdots & \alpha^{m-1} \\ \alpha & 1 & \alpha & \cdots & \alpha^{m-2} \\ \alpha^2 & \alpha & 1 & \cdots & \alpha^{m-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha^{m-1} & \alpha^{m-2} & \alpha^{m-3} & \cdots & 1 \end{bmatrix}$$

Unstructured: $\text{Corr}[Y_{ij}, Y_{ij'} \mid \mathbf{X}_i] = \alpha_{jj'}$

$$\begin{bmatrix} 1 & \alpha_{21} & \alpha_{31} & \cdots & \alpha_{m1} \\ \alpha_{12} & 1 & \alpha_{32} & \cdots & \alpha_{m2} \\ \alpha_{13} & \alpha_{23} & 1 & \cdots & \alpha_{m3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha_{1m} & \alpha_{2m} & \alpha_{3m} & \cdots & 1 \end{bmatrix}$$

Correlation models

Correlation between any two observations on the same subject. . .

- **Independence**: . . . is assumed to be zero
  - Appropriate with use of robust variance estimator (large N)
- **Exchangeable**: . . . is assumed to be constant
  - More appropriate for clustered data
- **Auto-regressive**: . . . is assumed to depend on time or distance
  - More appropriate for equally-spaced longitudinal data
- **Unstructured**: . . . is assumed to be distinct for each pair
  - Only appropriate for short series (small n) on many subjects (large N)

Semi-parametric

- Specification of a mean model and correlation model does not identify a complete probability model for the outcomes
- The [mean, correlation] model is semi-parametric because it only specifies the first two moments of the outcomes

Question: Without a likelihood function, how do we estimate $\boldsymbol{\beta}$ and generate valid statistical inference, while accounting for correlation?

Answer: Construct an unbiased estimating function

Estimating functions

The estimating function for estimation of $\boldsymbol{\beta}$ is given by

$$\mathcal{U}(\boldsymbol{\beta}) = \sum_{i=1}^{N} \mathbf{D}_i^{\top}(\boldsymbol{\beta}) \mathbf{V}_i^{-1}(\boldsymbol{\beta}, \boldsymbol{\alpha})[\mathbf{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})]$$

$$g(\boldsymbol{\mu}_i) = \mathbf{X}_i \boldsymbol{\beta}$$

$$\mathbf{D}_i(\boldsymbol{\beta}) = \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}$$

$$\mathbf{D}_i(j,k) = \frac{\partial \mu_{ij}}{\partial \beta_k}$$

- $\mathbf{V}_i$ is the 'working' variance-covariance matrix: $\text{Cov}[\mathbf{Y}_i \mid \mathbf{X}_i]$
  - Depends on the assumed form for the variance: $\text{Var}[Y_{ij} \mid \mathbf{X}_{ij}]$
  - Depends on the specified correlation model: $\text{Corr}[Y_{ij}, Y_{ij'} \mid \mathbf{X}_i]$
- $\mathcal{U}(\boldsymbol{\beta})$ depends on the model or value for $\boldsymbol{\alpha}$

Generalized estimating equations

Setting an **estimating function** equal to 0 defines an **estimating equation**

$$\begin{aligned} \mathbf{0} &= \mathcal{U}(\hat{\boldsymbol{\beta}}) \\ &= \sum_{i=1}^{N} \mathbf{D}_i^{\mathsf{T}}(\hat{\boldsymbol{\beta}})\mathbf{V}_i^{-1}(\hat{\boldsymbol{\beta}}, \boldsymbol{\alpha})[\mathbf{Y}_i - \boldsymbol{\mu}_i(\hat{\boldsymbol{\beta}})] \end{aligned}$$

- 'Generalized' because it corresponds to a GLM with link function $g(\cdot)$
- Solution to the estimation equation defines an estimator $\hat{\boldsymbol{\beta}}$
- Note $\mathcal{U}(\hat{\boldsymbol{\beta}})$ depends on the model or value for $\boldsymbol{\alpha}$

Generalized estimating equations: Intuition

$$\mathbf{0} = \sum_{i=1}^{N} \underbrace{\mathbf{D}_i^{\top}(\hat{\boldsymbol{\beta}})}_{\boxed{3}} \underbrace{\mathbf{V}_i^{-1}(\hat{\boldsymbol{\beta}}, \boldsymbol{\alpha})}_{\boxed{2}} \underbrace{[\mathbf{Y}_i - \boldsymbol{\mu}_i(\hat{\boldsymbol{\beta}})]}_{\boxed{1}}$$

1. The model for the mean $\boldsymbol{\mu}_i(\boldsymbol{\beta})$, is compared to the observed data $\mathbf{Y}_i$. Setting the functions equal to $\mathbf{0}$ tries to minimize the difference between the **observed** and **expected**
2. Estimation uses the inverse of the variance (covariance) to weight the data from subject i; more weight is given to differences between observed and expected for those subjects who contribute more information
3. This is simply a 'change of scale' from the scale of the mean, $\boldsymbol{\mu}_i(\boldsymbol{\beta})$, to the scale of the regression coefficients (covariates)

Because the GEE depends on both $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, an iterative procedure is used

GEE: Properties of $\hat{\beta}$

Question: What are the properties of $\hat{\beta}$, the regression estimate?

Answer:

- The regression coefficient estimate will be correct (in large samples) even if you choose the wrong dependence model
- However, the variance of the regression estimate must capture the correlation in the data, either through choosing the correct covariance model, or using an **alternative variance estimate**
- Correctly specified (or close) covariance model will yield regression estimate that is most (or more) efficient (smaller variance)

GEE: Sandwich variance estimator

The empirical estimator is:

$$\widehat{\text{Cov}}(\hat{\beta}) = \underbrace{\left(\sum_{i=1}^{N} \hat{\mathbf{D}}_i^\top \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{D}}_i\right)^{-1}}_{\text{bread}} \underbrace{\left(\sum_{i=1}^{N} \hat{\mathbf{D}}_i^\top \hat{\mathbf{V}}_i^{-1} \widetilde{\text{Cov}}(\mathbf{Y}_i) \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{D}}_i\right)}_{\text{cheese}} \underbrace{\left(\sum_{i=1}^{N} \hat{\mathbf{D}}_i^\top \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{D}}\right)^{-1}}_{\text{bread}}$$

where we can use the residuals for $\widetilde{\text{Cov}}(\mathbf{Y_i})$, i.e., has the entries $e_{ij}e_{ik}$ where $e_{ij} = Y_{ij} - \hat{\mu}_{ij}$.

Note that with linear regression (Gaussian family with identity link), $\mathbf{D}_i = \mathbf{X}_i$ and $\mathbf{V}_i = \phi \mathbf{R}_i$ where $\mathbf{R}_i$ is our working correlation matrix and $\phi = \sigma^2$

GEE: Sandwich variance estimator

$$\widehat{\text{Cov}}(\hat{\beta}) = \underbrace{\left(\sum_{i=1}^{N} \hat{\mathbf{D}}_i^\top \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{D}}_i\right)^{-1}}_{\text{bread}} \underbrace{\left(\sum_{i=1}^{N} \hat{\mathbf{D}}_i^\top \hat{\mathbf{V}}_i^{-1} \widetilde{\text{Cov}}(\mathbf{Y}_i) \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{D}}_i\right)}_{\text{cheese}} \underbrace{\left(\sum_{i=1}^{N} \hat{\mathbf{D}}_i^\top \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{D}}\right)^{-1}}_{\text{bread}}$$

- Also known as sandwich, robust, or Huber-White variance estimator
- Requires large enough sample size ($N \geq 40$)
- Requires large enough sample size relative to cluster size ($N \gg n$)
- **Gives valid standard errors for the estimated regression coefficients even if the correlation model is wrong!**

GEE: Inference

Can perform Wald test and construct a Wald confidence interval:

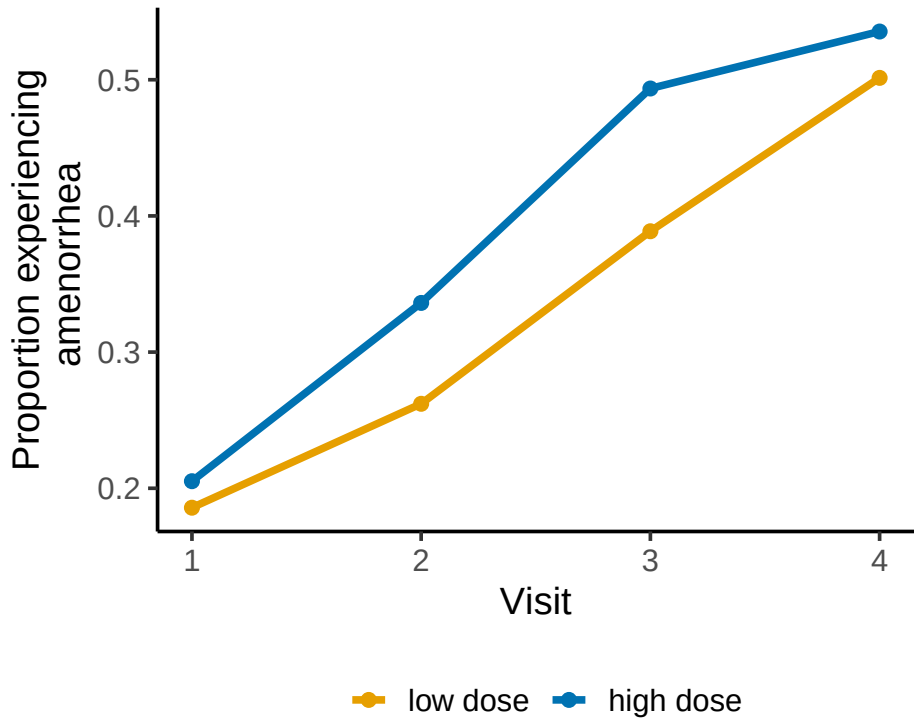
- $\hat{\beta}_k$/s.e. - valid test
- $\hat{\beta}_k \pm 1.96 \times$ s.e. - valid 95% confidence interval

Cannot perform a likelihood ratio test (no fully specified probability model) so no AIC or BIC either.

Case Study: Clinical Trial of Contracepting Women

Longitudinal clinical trial where people who menstruate received an injection of either 100 mg or 150 mg of DMPA at randomization and then every 90 days. Final follow-up visit after the 4th injection, 1 year after the randomization.

- $N = 1151$ people completed menstrual diaries
- Response: whether the person experienced amenorrhea in the previous 3 months
- Substantial dropout: more than 1/3 dropped out before completing the trial



Case Study: Clinical Trial of Contracepting Women

$$\log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) = \beta_0 + \beta_1 \mathrm{t}_{ij} + \beta_2 t_{ij}^2 + \beta_3 \mathrm{Dose}_i \times t_{ij} + \beta_4 \mathrm{Dose}_i \times t_{ij}^2$$

where $\mu_{ij} = \Pr(Y_{ij} = 1)$

- Y_{ij}: indicator for whether person i experienced amenorrhea in the jth injection interval (1 = yes; 0 = no)
- t_{ij}: measurement occasion (corresponds to the four consecutive 90-day injection intervals)
- Dose_i: indicator for whether the person i was randomized to 150 mg of DMPA or 100 mg (1 = 150 mg; 0 = 100 mg)

Case Study: Clinical Trial of Contracepting Women

```
m_ind <- geeglm(amenorrhea ~ visit + visit:dose + I(visit^2) + I(visit^2):dose,
id = id, data = amenorrhea, family = binomial, waves = visit,
corstr = "independence")
m_exc <- geeglm(amenorrhea ~ visit + visit:dose + I(visit^2) + I(visit^2):dose,
id = id, data = amenorrhea, family = binomial, waves = visit,
corstr = "exchangeable")
m_ar1 <- geeglm(amenorrhea ~ visit + visit:dose + visit2 + visit2:dose,
id = id, data = amenorrhea, family = binomial, waves = visit,
corstr = "ar1")
m_uns <- geeglm(amenorrhea ~ visit + visit:dose + I(visit^2) + I(visit^2):dose,
id = id, data = amenorrhea, family = binomial, waves = visit,
corstr = "unstructured")

m_glm <- glm(amenorrhea ~ visit + visit:dose + I(visit^2) + I(visit^2):dose,
data = amenorrhea, family = binomial)
```

Case Study: Clinical Trial of Contracepting Women

```
geeglm(formula = amenorrhea ~ visit + visit:dose + I(visit^2) +
I(visit^2):dose, family = binomial, data = amenorrhea, id = id,
waves = visit, corstr = "independence")

Coefficients:
Estimate Std.err   Wald Pr(>|W|)
(Intercept)      -2.1955  0.1784 151.37  < 2e-16 ***
visit             0.6698  0.1622  17.04  3.7e-05 ***
I(visit^2)       -0.0303  0.0328   0.86   0.3549
visit:dose        0.2973  0.1135   6.87   0.0088 **
dose:I(visit^2)  -0.0624  0.0296   4.44   0.0351 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation structure = independence
Estimated Scale Parameters:

Estimate Std.err
(Intercept)        1  0.0289
Number of clusters:   1151  Maximum cluster size: 4
```

Case Study: Clinical Trial of Contracepting Women

```
geeglm(formula = amenorrhea ~ visit + visit:dose + I(visit^2) +
I(visit^2):dose, family = binomial, data = amenorrhea, id = id,
waves = visit, corstr = "exchangeable")

Coefficients:
Estimate Std.err   Wald Pr(>|W|)
(Intercept)      -2.2370  0.1765 160.64  < 2e-16 ***
visit             0.6967  0.1586  19.31  1.1e-05 ***
I(visit^2)       -0.0328  0.0320   1.05   0.3055
visit:dose        0.3284  0.1100   8.91   0.0028 **
dose:I(visit^2)  -0.0637  0.0286   4.97   0.0259 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation structure = exchangeable
Estimated Scale Parameters:

Estimate Std.err
(Intercept)        1  0.0287
Link = identity

Estimated Correlation Parameters:
Estimate Std.err
alpha    0.363  0.0243
Number of clusters:   1151  Maximum cluster size: 4
```

Case Study: Clinical Trial of Contracepting Women

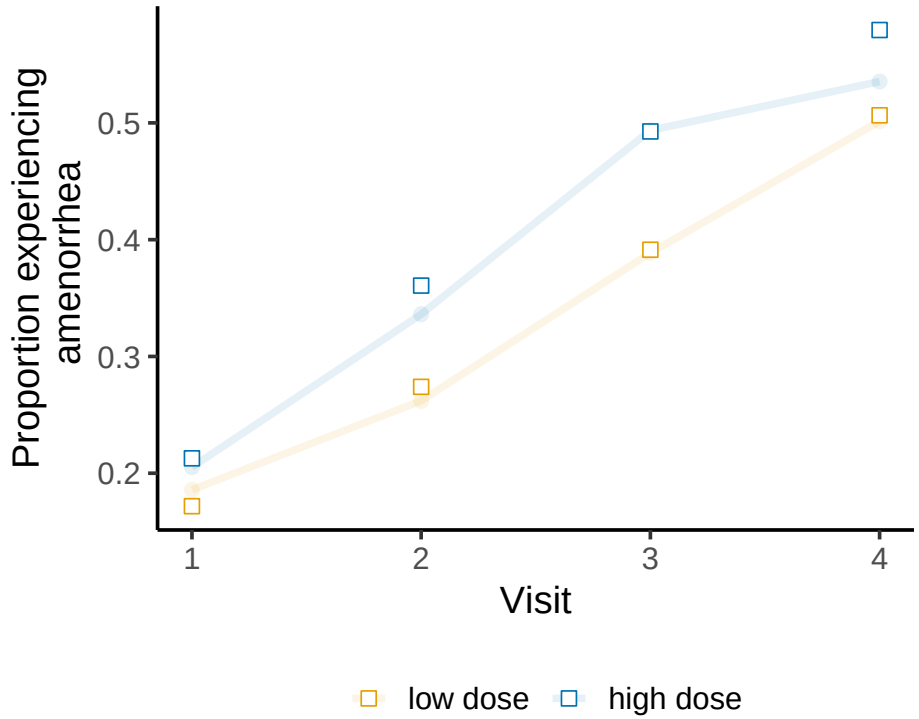
	$\hat{\beta}_0$ (SE)	$\hat{\beta}_1$ (SE)	$\hat{\beta}_2$ (SE)	$\hat{\beta}_3$ (SE)	$\hat{\beta}_4$ (SE)
Ind	-2.2 (0.18)	0.67 (0.16)	-0.03 (0.03)	0.30 (0.11)	-0.06 (0.03)
Exch	-2.2 (0.18)	0.70 (0.16)	-0.03 (0.03)	0.33 (0.11)	-0.06 (0.03)
AR1	-2.2 (0.18)	0.71 (0.16)	-0.03 (0.03)	0.36 (0.11)	-0.08 (0.03)
Unst	-2.2 (0.18)	0.70 (0.16)	-0.03 (0.03)	0.34 (0.11)	-0.07 (0.03)
GLM	-2.2 (0.21)	0.67 (0.19)	-0.03 (0.04)	0.30 (0.11)	-0.06 (0.03)

- Working independence and using `glm()` give exactly the same point estimates
- `glm()` standard errors are not correct (here, too large)
- Working independence, exchangeable, AR1, and unstructured provide similar results

Case Study: Clinical Trial of Contracepting Women

Using working exchangeable:

- The ratio of the population odds of amenorrhea at 12 months comparing high dose to low dose is estimated to be 1.34 (95% CI: 1.01 - 1.78)
- Conducting a multivariate Wald test ($H_0 : \beta_3 = \beta_4 = 0$) we find a statistically significant effect of dose, p-value $= 0.002$



GEE: Which Correlation Model to Choose?

Question: Which correlation model should I choose?

Answer: Ideally, to preserve CI statements, Type I error, the choice of working correlation should be based on external information or substantive grounds rather than exploratory analysis

Question: If the correlation model does not need to be correctly specified to obtain a consistent estimator for β or valid standard errors for $\hat{\beta}$, why not always use an independence working correlation model?

Answer: Selecting a non-independence or weighted correlation model

- Permits use of the model-based variance estimator
- May provide improved efficiency for $\hat{\beta}$

GEE: Summary

- Primary focus of the analysis is a marginal mean regression model that corresponds to any GLM
- Longitudinal correlation is secondary to the mean model of interest and is treated as a nuisance feature of the data
- Requires selection of a 'working' correlation model
- Lack of a likelihood function implies that likelihood ratio test statistics are unavailable; hypothesis testing with GEE uses Wald statistics
- Working correlation model does not need to be correctly specified to obtain a consistent estimator for β or valid standard errors for $\hat{\beta}$, but efficiency gains are possible if the correlation model is correct

Issues

- Accommodates only one source of correlation: Longitudinal **or** cluster
- GEE requires that any missing data are missing completely at random
- Issues arise with time-dependent exposures and covariance weighting

Overview

LMM: What's Beneath the Hood?

Maximum Likelihood:

$$\boldsymbol{Y}_i|\boldsymbol{b}_i \sim MVN(\boldsymbol{X}_i\beta + \boldsymbol{Z}_i\boldsymbol{b}_i, \epsilon_i)$$
$$\boldsymbol{b}_i \sim MVN(0, \boldsymbol{D})$$

So the likelihood function is a product of normal densities,

$$P(\boldsymbol{Y}_i, \boldsymbol{b}_i) = P(\boldsymbol{Y}_i|\boldsymbol{b}_i)P(\boldsymbol{b}_i),$$

which is easy to integrate and maximize (continuous outcomes only!).

Restricted Maximum Likelihood:

- ML estimation of Σ is biased (e.g., n vs. $n-p$ in denominator for σ^2)
- REML (default) provides less-biased estimation of Σ
  - Details are beyond the scope of this module
  - Can't use LRT to compare models with different **fixed effects** if fit with REML – use ML instead
  - Okay to compare **random effects** structures (same fixed effects)

Choosing a Random Effects Structure

- Covariance model choice determines the standard error estimates for $\hat{\beta}$; correct model is required for correct standard error estimates
- Suppose we want to decide between two candidate random effect structures:

$$H_0 : \boldsymbol{D} = \begin{bmatrix} D_{11} & 0 \\ 0 & 0 \end{bmatrix} \quad \text{versus} \quad H_1 : \boldsymbol{D} = \begin{bmatrix} D_{11} & D_{21} \\ D_{12} & D_{22} \end{bmatrix},$$

- Could we formally test whether random intercepts are adequate, e.g., with LRT?

Likelihood-based inference for D

- Consider testing $H_0 : D_{12} = D_{22} = 0$
  - Fit both models with REML, compare with LRT
- This is possible in R

```
> anova(mod.ri, mod.rs)
       Model df AIC BIC logLik   Test L.Ratio p-value
mod.ri     1  6 446 462   -217
mod.rs     2  8 449 470   -216 1 vs 2    1.18   0.556
```

- Seems to suggest random intercept is sufficient (AIC is lower for RI than RS, p-value is big)
  - Consistent with our prior exploratory analysis

Choosing a Random Effects Structure

Problem: testing $D_{22} = 0$ is a nonstandard problem

- P-values tend to be conservative, could lead to over-simplifying correlation structure
- (Ad hoc fix proposed by Fitzmaurice, Laird, Ware: use $\alpha = 0.1$)

Alternatives:

- If only interested in inference on β:
  - Choose based on *a priori* scientific knowledge and exploratory analysis
  - Use robust standard errors (see module code; R package clubSandwich)
- If this is an exploratory analysis and you're interested in correlation structure, ok to test structures
  - Can cause type 1 error problems if your focus is inference on β

Dental Growth: Robust SE, Random Intercepts

Variable	$\hat{\beta}$	Estimation	P-value	CI
Intercept	21.2	Model-based	1.28×10^{-47}	(19.9, 22.5)
		Robust	6.35×10^{-12}	(19.9, 22.5)
(Age-8)	0.48	Model-based	2.02×10^{-6}	(0.29, 0.67)
		Robust	2.78×10^{-5}	(0.33, 0.63)
Male	1.41	Model-based	p=0.108	(-0.33, 3.14)
		Robust	p=0.095	(-0.27, 3.08)
(Age-8)*Male	0.30	Model-based	p=0.014	(0.06, 0.55)
		Robust	p=0.020	(0.05, 0.56)

Choosing a Random Effects Structure

A priori considerations:

- Random intercepts only:
  - Simpler (+/-) and easy to interpret (e.g., ICC)
  - Induces exchangeable correlation (+/-)
- Random intercepts and slopes:
  - More information about individual trajectories and covariance (+)
  - More complex (+/-)
    - ★ Over-parameterization of covariance $\Longrightarrow$ inefficient estimation of fixed-effects parameters β

Non-Continuous Outcomes: What Changes?

- We've seen specification of GLMs earlier today:

$$\boldsymbol{Y}_i|\mu_i \sim \text{exponential family distribution}$$
$$g(\mu_i) = \boldsymbol{X}_i\boldsymbol{\beta}$$

- We'll focus on binary outcomes, but distribution and link function may be swapped out for other outcome types

$$\boldsymbol{Y}_i|\mu_i \sim \text{Bernoulli}(\mu_i)$$
$$\text{logit}(\mu_i) = \boldsymbol{\eta}_i = \boldsymbol{X}_i\boldsymbol{\beta}$$

- Can't we just add $\boldsymbol{Z}_i\boldsymbol{b}_i$ to $\boldsymbol{\eta}_i$, and update our likelihood function?
  - ... yes and no.

Generalized Linear Mixed Effects Models

LMM:

$$
\begin{aligned}
Y_i|\mu_i, \boldsymbol{b}_i &\sim \text{Normal}(\mu_i, \sigma^2 \boldsymbol{I}) \\
\mu_i|\boldsymbol{b}_i &= \boldsymbol{X}_i\beta + \boldsymbol{Z}_i\boldsymbol{b}_i \\
\boldsymbol{b}_i &\sim \text{Normal}(0, \boldsymbol{D})
\end{aligned}
$$

GLMM:

$$
\begin{aligned}
Y_i|\mu_i, \boldsymbol{b}_i &\sim \text{Bernoulli}(\mu_i) \\
\text{logit}(\mu_i|\boldsymbol{b}_i) &= \boldsymbol{X}_i\beta + \boldsymbol{Z}_i\boldsymbol{b}_i \\
\boldsymbol{b}_i &\sim \text{Normal}(0, \boldsymbol{D})
\end{aligned}
$$

Generalized Linear Mixed Effects Models

- Conceptually very similar!
- Computationally, GLMM is much more complicated
  - Likelihood: product of non-Normal exponential density (for $\boldsymbol{Y}_i|\boldsymbol{b}_i$) and Normal density (for $\boldsymbol{b}_i$)
  - Typically fit using approximation or numerical techniques
  - (NB: centering and scaling predictors can help with convergence issues)
- Interpretation is (importantly) different

Conditional and Marginal Effects

- Parameter estimates obtained from a marginal model (as obtained via GEE) estimate population-averaged contrasts
- Parameter estimates obtained from a conditional model (as obtained via GLMM) estimate subject-specific contrasts
- In a linear model for a Gaussian outcome with an identity link, these are equivalent; not the case with non-linear models
  - Depends on the outcome distribution
  - Depends on the specified random effects

Conditional = Marginal for LMM, Not GLMM. Why?

For an LMM, we use the identity link function. Then:

- Taking the expected value over $\boldsymbol{b}_i$:

$$\mathrm{E}_b\left\{\mathrm{E}(Y_{ij}|X_{ij},\boldsymbol{b}_i)\right\} = \mathrm{E}_b\left\{X_{ij}^\top\boldsymbol{\beta} + Z_{ij}^\top\boldsymbol{b}_i\right\} = Z_{ij}^\top\boldsymbol{\beta} = \mathrm{E}(Y_{ij}|X_{ij})$$

- So the **average conditional mean** $\mathrm{E}_b\left\{\mathrm{E}(Y_{ij}|X_{ij},\boldsymbol{b}_i)\right\}$ is the same as the **marginal mean** $\mathrm{E}(Y_{ij}|X_{ij})$

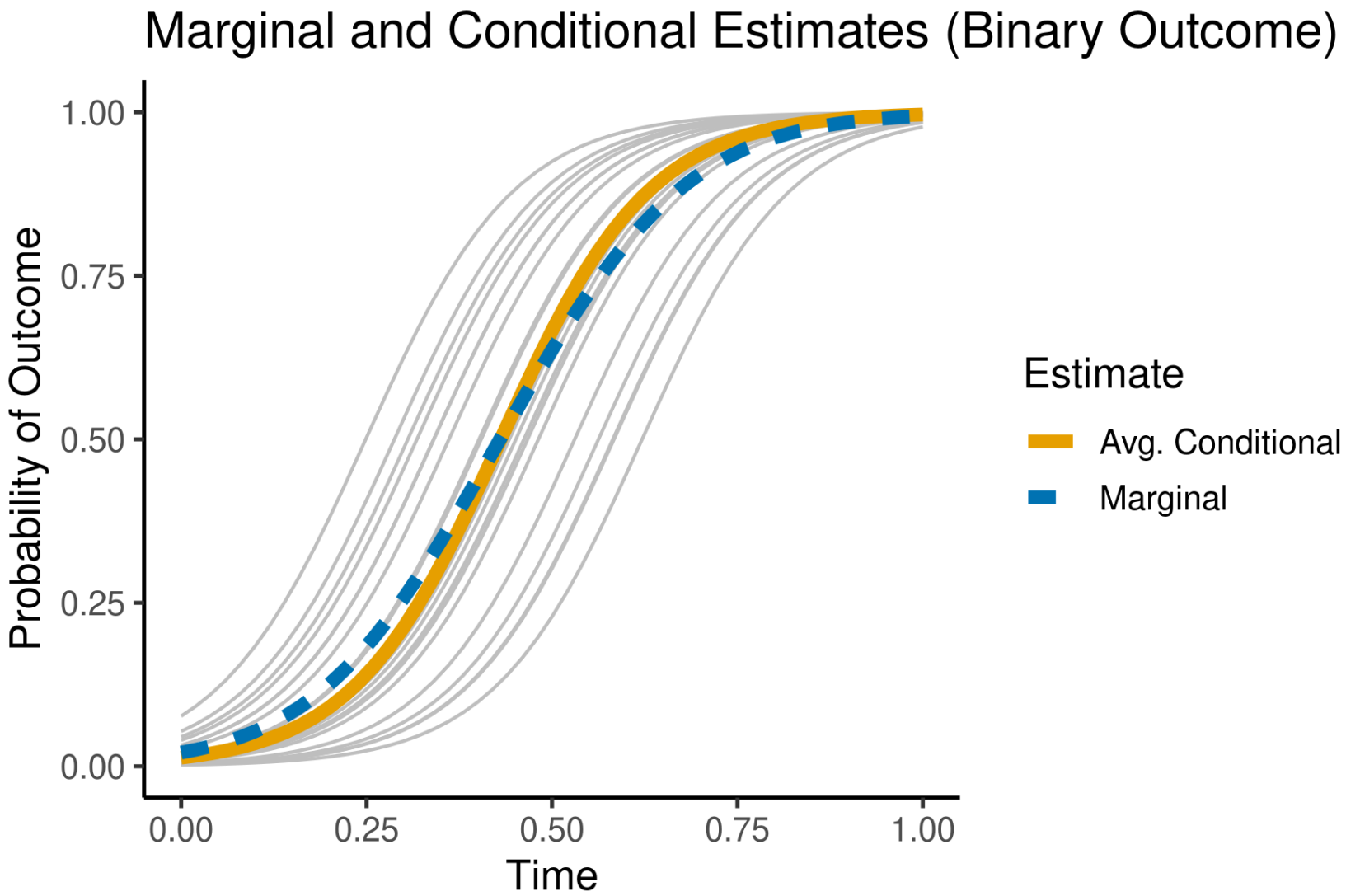
For a GLMM, we use a non-identity link function. Then:

- Taking the expected value over $\boldsymbol{b}_i$:

$$\mathrm{E}_b\left\{g(\mathrm{E}(Y_{ij}|X_{ij},\boldsymbol{b}_i))\right\} = \mathrm{E}_b\left\{g(X_{ij}^\top\boldsymbol{\beta} + Z_{ij}^\top\boldsymbol{b}_i)\right\} \neq g(\mathrm{E}(Y_{ij}|X_{ij}))$$

- So average conditional mean is **not** the same as marginal mean, in general

Conditional $\neq$ Marginal for GLMM



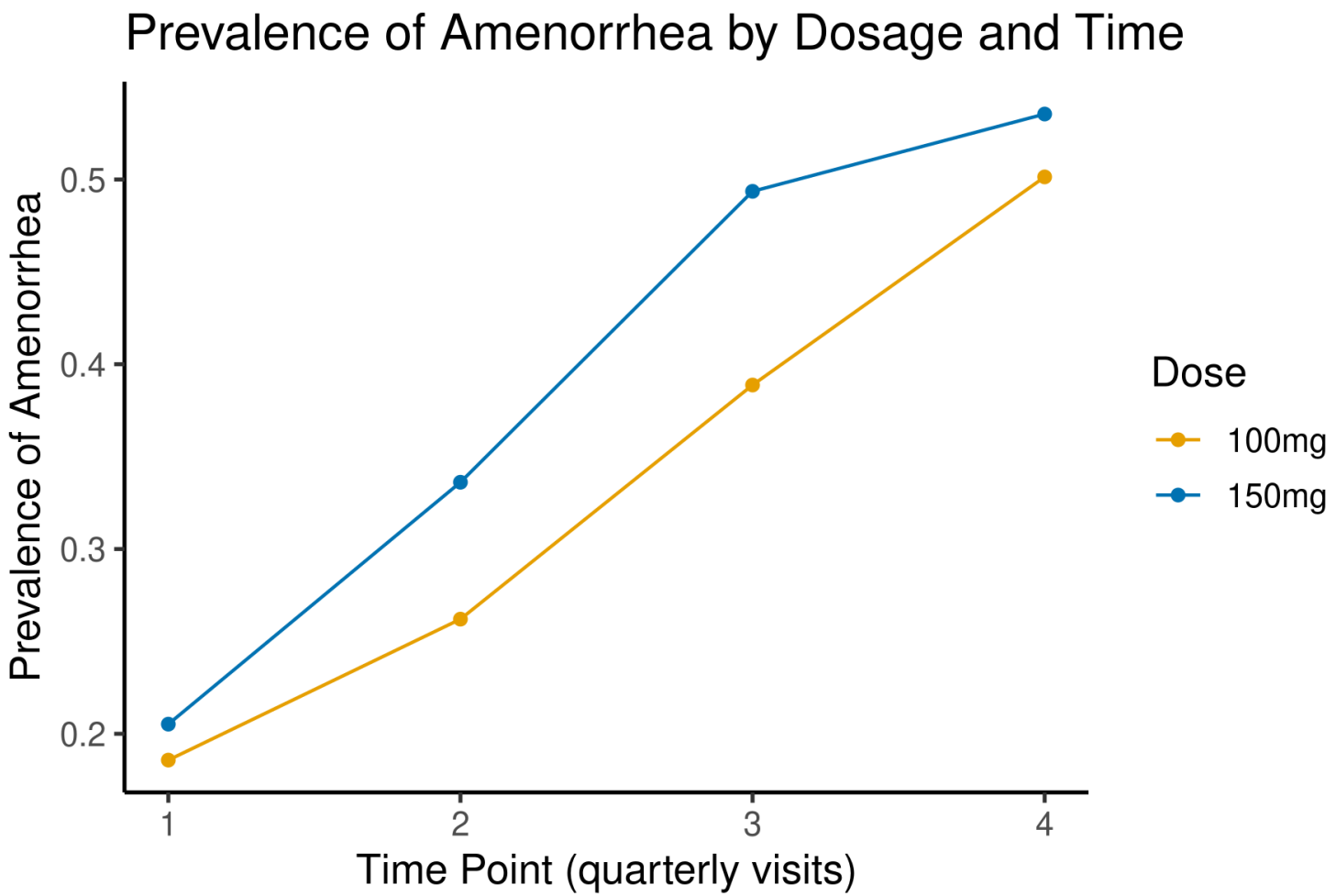
Note: marginal is "attenuated" (smaller estimate) compared to conditional

Interpretation of GLMM

		Fitted model	
Outcome	Coefficient	Random intercept	Random intercept/slope
Continuous	Intercept	Marginal	Marginal
	Slope	Marginal	Marginal
Binary	Intercept	Conditional	Conditional
	Slope	Conditional	Conditional
Count	Intercept	Conditional	Conditional
	Slope	Marginal	Conditional

⋆ Marginal = population-averaged; conditional = subject-specific

Example: Amenorrhea in Contracepting Women



Scientific Questions

- How do subject-specific risks of amenorrhea change over the course of the study?
- What is the influence of dosage on amenorrhea risk?

GLMM for Amenorrhea (random intercepts)

Same fixed effects component of model as GEE, plus random intercepts:

$$\log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) = \beta_0 + \beta_1 \mathrm{t}_{ij} + \beta_2 t_{ij}^2 + \beta_3 \mathrm{Dose}_i \times t_{ij} + \beta_4 \mathrm{Dose}_i \times t_{ij}^2 + b_i$$

where $\mu_{ij} = \Pr(Y_{ij} = 1)$

In R:

```
library(lme4)
glmm.ri <- glmer(amenorrhea ~ visit + visit:dose +
I(visit^2) + I(visit^2):dose + (1 | id),
data = ctcw, nAGQ = 10,
family = "binomial")
```

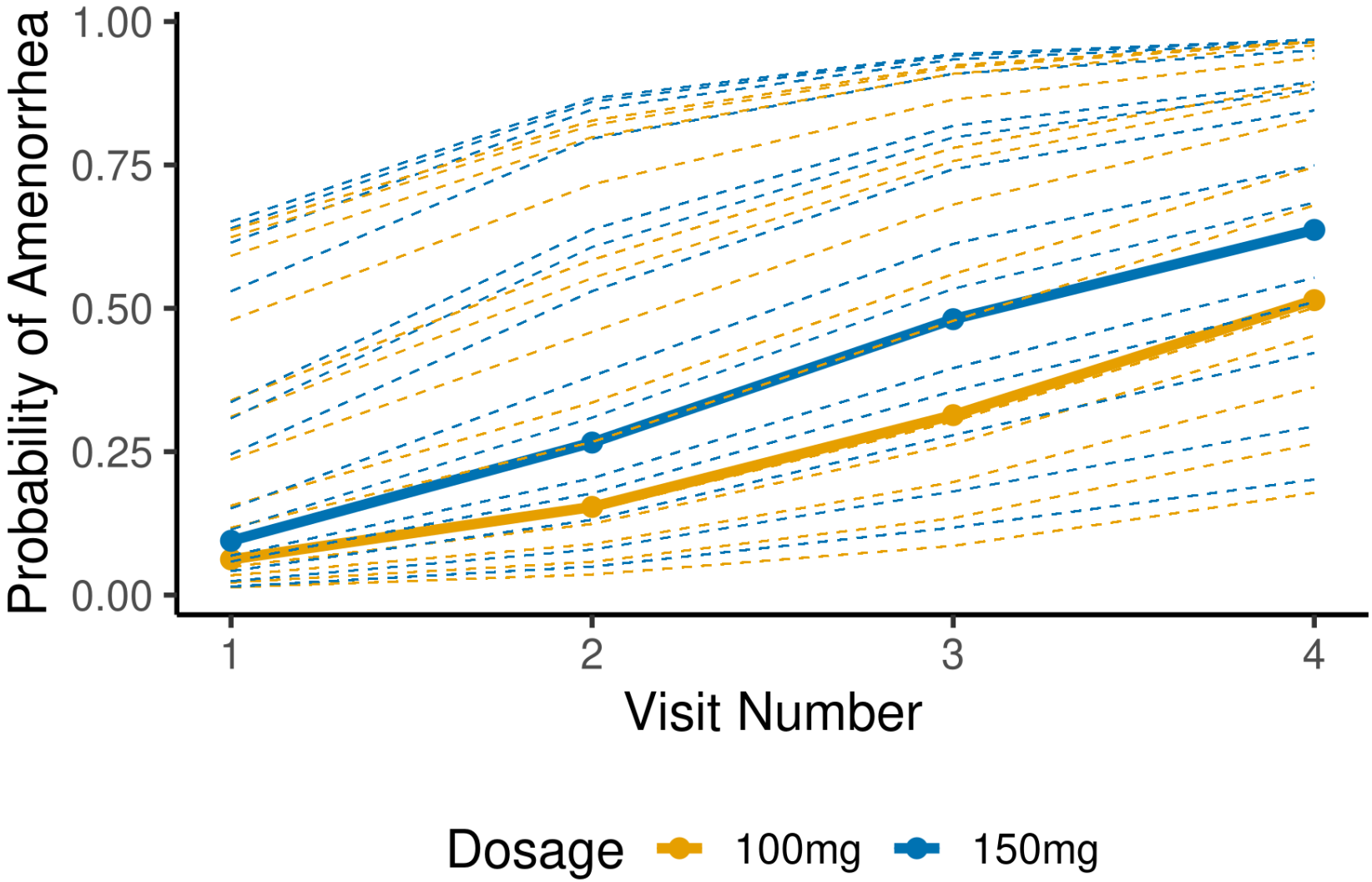
Case Study: Clinical Trial of Contracepting Women

```
Generalized linear mixed model fit by maximum likelihood
(Adaptive Gauss-Hermite Quadrature, nAGQ = 10) ['glmerMod']
Family: binomial  ( logit )
Formula: amenorrhea ~ visit + visit:dose + I(visit^2) + I(visit^2):dose + (1 | id)
Data: ctcw

Random effects:
Groups Name        Variance Std.Dev.
id     (Intercept) 5.054    2.248
Number of obs: 3616, groups:  id, 1151

Fixed effects:
Estimate Std. Error z value Pr(>|z|)
(Intercept)     -3.80348    0.30470 -12.483  < 2e-16 ***
visit            1.13259    0.26811   4.224  2.4e-05 ***
I(visit^2)      -0.04187    0.05479  -0.764  0.44475
visit:dose       0.56417    0.19214   2.936  0.00332 **
dose:I(visit^2) -0.10952    0.04959  -2.209  0.02721 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Case Study: Clinical Trial of Contracepting Women



GLMM for Amenorrhea (random intercepts)

$$\text{logit}(P(\text{Amenorrhea}_i|\text{Dose}_i, \text{Visit}_{ij}, \boldsymbol{b}_i)) = -3.80 + 1.13 \times \text{Visit}_{ij}$$
$$-0.04 \times \text{Visit}_{ij}^2 + 0.56 \times \text{Dose}_i \times \text{Visit}_{ij} - 0.11 \times \text{Dose}_i \times \text{Visit}_{ij}^2 + b_i$$

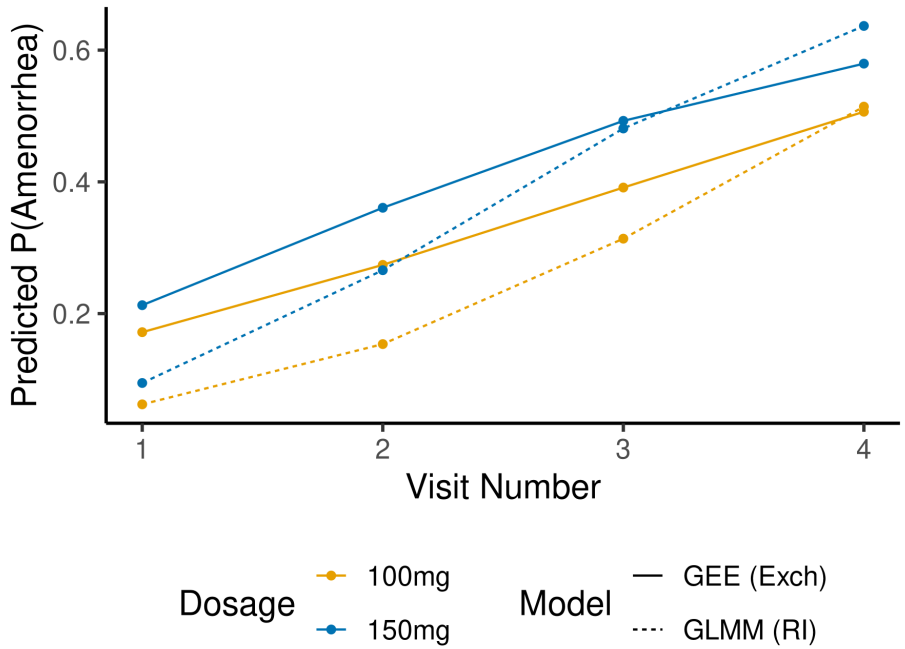
- On average, the odds of amenorrhea at 12 months (Visit 4) for a woman who took the high dose is 1.66 times higher than a woman with the same baseline risk who took the low dose (95% CI: 1.03-2.65).
- Conducting a likelihood ratio test ($H_0 : \beta_3 = \beta_4 = 0$), we find a statistically significant effect of dose (p=0.002)

(NB: conditional interpretations for all parameters in logistic GLMM)

GLMM vs. GEE for Amenorrhea

- GEE: marginal estimates (population-averaged odds ratios)
- GLMM: conditional estimates (within-subject odds ratios)
- Marginal is attenuated relative to conditional
- Note: unlike LMM, random intercept in GLMM does not induce exchangeable correlation
  - So these models do not assume the same correlation structure

Model	GEE (Exch)	GLMM (RI)
$\hat{\beta}_0$ (SE)	-2.2 (0.18)	-3.8 (0.30)
$\hat{\beta}_1$ (SE)	0.70 (0.16)	1.13 (0.27)
$\hat{\beta}_2$ (SE)	-0.03 (0.03)	-0.04 (0.05)
$\hat{\beta}_3$ (SE)	0.33 (0.11)	0.56 (0.19)
$\hat{\beta}_4$ (SE)	-0.06 (0.03)	-0.11 (0.05)



Assumptions and Notes

- Same set of assumptions as LMM:
  - Correct mean model (fixed effects + random effects)
  - Correct covariance specification (random effects)
  - Correct distributional assumptions for $\boldsymbol{Y}_i|\boldsymbol{b}_i$ and $\boldsymbol{b}_i$
- Interpretations:
  - Conditional = marginal for LMM, so can interpret either way
  - Conditional $\neq$ marginal for GLMM; interpret appropriately (usually conditional/subject-specific interpretation)

Overview

Missing data

- Missing values arise in longitudinal studies whenever the intended measurements are not obtained
  - Collect fewer data than planned $\Rightarrow$ decreased efficiency (power)
  - Missingness can depend on outcome values $\Rightarrow$ potential bias
- Important to distinguish between missing data and unbalanced data, although missing data necessarily result in unbalanced data
- Missing data require consideration of the factors that influence the missingness of intended observations
- Also important to distinguish between intermittent missing values (non-monotone) and dropouts in which all observations are missing after subjects are lost to follow-up (monotone)

Pattern	t_1	t_2	t_3	t_4	t_5
Monotone	3.8	3.1	2.0		
Non-monotone	4.1		3.8		

Mechanisms

In order to obtain valid inference from incomplete data, the mechanism producing the missingness must be considered

- **Missing completely at random** (MCAR)
  Missingness does not depend on **either** the observed or missing responses
- **Missing at random** (MAR)
  Missingness depends **only** on the observed responses
- **Missing not at random** (MNAR)
  Missingness depends on the missing responses

MNAR also referred to as informative or non-ignorable missingness; thus MAR and MCAR as non-informative or ignorable missingness (Rubin, 1976)

Examples and implications

- **MCAR**: Administrative censoring at a fixed calendar time
  - Generalized estimating equations are valid
  - Mixed-effects models are valid
  - Note: If missingness depends on covariates, the covariates need to be incorporated into the analysis (missingness would be a problem if you do not condition on them)
- **MAR**: Study protocol that a subject be removed once the value of an outcome variable is below a certain threshold
  - Generalized estimating equations are not valid
  - Mixed-effects models are valid (as long as model is correctly specified)
- **MNAR**: Outcome is a measure of 'quality-of-life' and subjects fail to complete the questionnaire when their quality-of-life is compromises
  - Generalized estimating equations are not valid
  - Mixed-effects models are not valid

$\star$ Without knowing the reasons for missingness or having the missing data cannot know which mechanism for sure.

Missing Data: Analytic Approaches

1. **Complete-case** analysis: use observations from only those that have complete data
   - Valid if data are MCAR
2. **Available data** analysis: use all available observations
   - Valid if data are MCAR
   - Tends to be more efficient than complete case
   - Likelihood-based methods valid if data are MAR and correctly specified
3. **Imputation**: fill in missing values
   - Multiple imputation helps to account for uncertainty
4. **Weighting** methods: accounts for under-representation of certain responses in the observed data by weighting the observed data according to probability of remaining in the study
5. Others...

WGEE

- Extend marginal GEE approach to longitudinal studies with MAR dropout
- Define $R_{ij} = 1$ if jth observation from subject i is observed.
- Observations in the estimating equation are weighted inversely proportional to the probability of being observed

$$\sum_{i=1}^{N} \mathbf{D}_i(\hat{\boldsymbol{\beta}})^\top \mathbf{V}_i^{-1}(\hat{\boldsymbol{\beta}}, \boldsymbol{\alpha}) \boldsymbol{\Delta}_i(\boldsymbol{\theta})[\mathbf{Y}_i - \boldsymbol{\mu}_i(\hat{\boldsymbol{\beta}})] = \mathbf{0}$$

where

$$\boldsymbol{\Delta}_i(\boldsymbol{\theta}) = \begin{pmatrix} R_{i1}w_{i1} & 0 & \cdots & 0 \\ 0 & R_{i2}w_{i2} & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & R_{in}w_{in} \end{pmatrix}$$

(Robins et al., 1995)

WGEE

- Valid under MAR provided the model for probability of dropout is correct
  - Requires either *a priori* knowledge of the weights w_{ij} or
  - Correctly specified dropout model (the probability of remaining in the study at the current time point, given dropout not occurring at previous time points):

$$\pi_{ij} = \Pr(R_{ij} = 1|, R_{i,j-1} = 1, \mathbf{X}_i, Y_{i1}, \ldots, Y_{i,j-1}) \quad \rightarrow \quad w_{ij} = \left[\prod_{k=1}^{j} \pi_{ik}\right]^{-1}$$

- Usual comments regarding GEE apply:
  - Correct specification for the mean μ and sufficiently large N
  - Use of robust variance estimator provides robustness to misspecification of the correlation structure
  - Choice of working correlation matrix affects efficiency

Example

$$\log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) = \beta_0 + \beta_1 \mathrm{t}_{ij} + \beta_2 t_{ij}^2 + \beta_3 \mathrm{Dose}_i \times t_{ij} + \beta_4 \mathrm{Dose}_i \times t_{ij}^2$$

	$\hat{\beta}_0$ (SE)	$\hat{\beta}_1$ (SE)	$\hat{\beta}_2$ (SE)	$\hat{\beta}_3$ (SE)	$\hat{\beta}_4$ (SE)
CC	-2.4 (0.22)	0.77 (0.19)	-0.04 (0.04)	0.25 (0.13)	-0.04 (0.03)
AD	-2.2 (0.18)	0.70 (0.16)	-0.03 (0.03)	0.33 (0.11)	-0.06 (0.03)
WGEE	-2.3 (0.18)	0.71 (0.16)	-0.03 (0.03)	0.34 (0.11)	-0.07 (0.03)

- Complete case (CC) and available data (AD) analyses valid if data are MCAR
- For WGEE, assumed the probability of remaining in the study depends on measurement occasion, dose group, and previous response

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \theta_0 + \theta_1 \mathrm{Dose}_i + \theta_2 t_{ij} + \theta_3 Y_{i,j-1} + \theta_4 \mathrm{Dose}_i \times Y_{i,j-1}$$

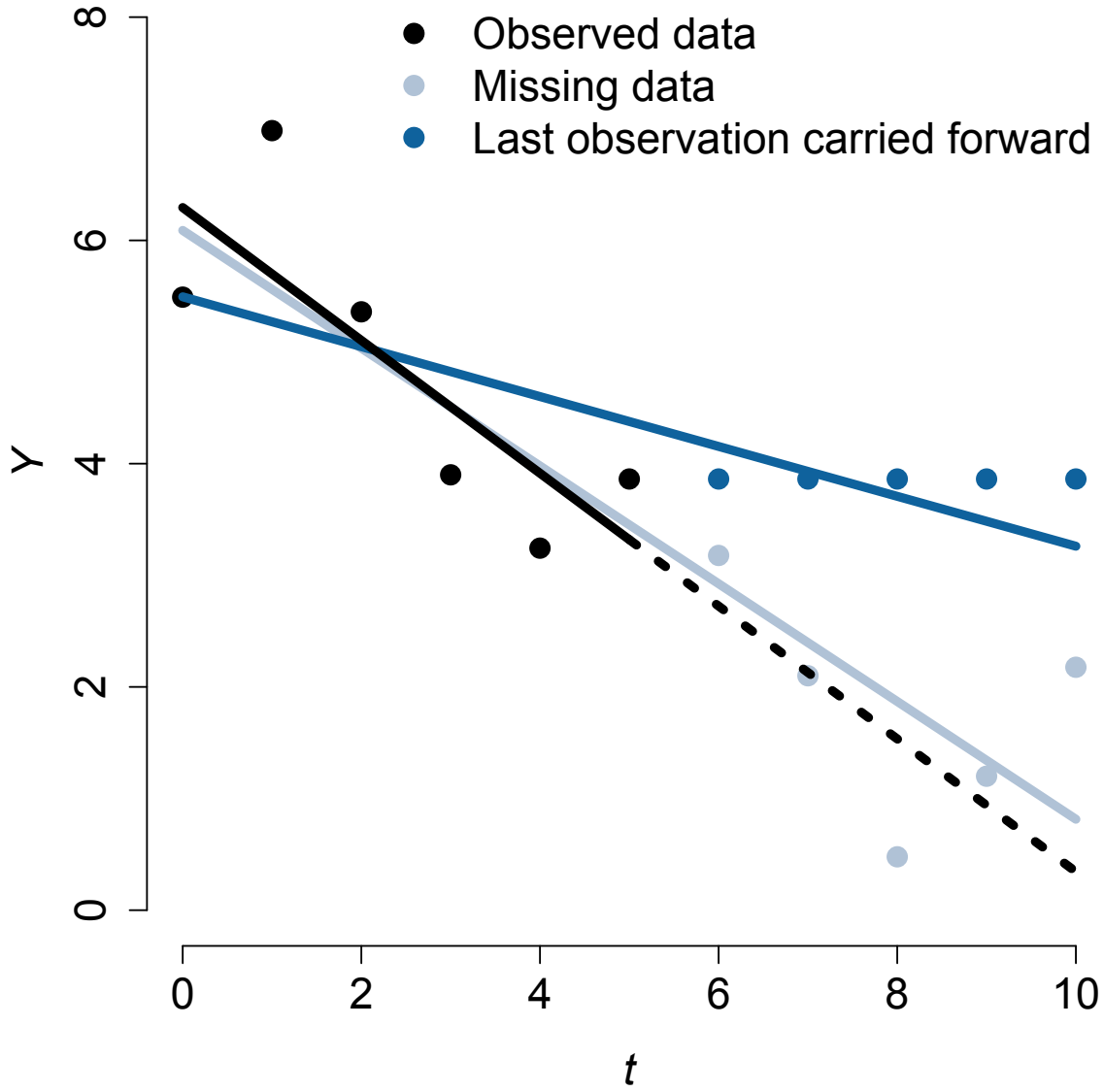
Last observation carried forward

- Extrapolate the last observed measurement to the remainder of the intended serial observations for subjects with any missing data

ID	t_1	t_2	t_3	t_4	t_5
1	3.8	3.1	2.0	2.0	2.0
2	4.1	3.5	3.8	2.4	2.8
3	2.7	2.4	2.9	3.5	3.5

- May result in serious bias in either direction (even when missingness is MCAR)
- May result in anti-conservative p-values; variance is understated
- Has been thoroughly repudiated, but still a standard method used by the pharmaceutical industry and appears in published articles
- A refinement would extrapolate based on a regression model for the average trend, which may reduce bias, but still understates variance

Last observation carried forward



Time-dependent exposures

Important analytical issues arise with time-dependent exposures

1. May be necessary to correctly specify the lag relationship over time between outcome $Y_i(t)$ and exposure $X_i(t)$, $X_i(t-1)$, $X_i(t-2), \ldots$ to characterize the underlying biological latency in the relationship
   - **Example**: Air pollution studies may examine the association between mortality on day t and pollutant levels on days t, $t-1$, $t-2, \ldots$
2. May exist exposure endogeneity in which the outcome at time t predicts the exposure at times $t' > t$; motivates consideration of alternative targets of inference and corresponding estimation methods
   - **Example**: If $Y_i(t)$ is a symptom measure and $X_i(t)$ is an indicator of drug treatment, then past symptoms may influence current treatment

Definitions

Factors that influence $X_i(t)$ require consideration when selecting analysis methods to relate a time-dependent exposure to longitudinal outcomes

- **Exogenous**: An exposure is exogenous w.r.t. the outcome process if the exposure at time t is conditionally independent of the history of the outcome process $\mathcal{Y}_i(t) = \{Y_i(s) \mid s \leq t\}$ given the history of the exposure process $\mathcal{X}_i(t) = \{X_i(s) \mid s \leq t\}$

$$[X_i(t) \mid \mathcal{Y}_i(t), \mathcal{X}_i(t)] = [X_i(t) \mid \mathcal{X}_i(t)]$$

- **Endogenous**: Not exogenous

$$[X_i(t) \mid \mathcal{Y}_i(t), \mathcal{X}_i(t)] \neq [X_i(t) \mid \mathcal{X}_i(t)]$$

Examples

Exogeneity may be assumed based on the design or evaluated empirically

- **Observation time**: Any analysis that uses scheduled observation time as a time-dependent exposure can safely assume exogeneity because time is 'external' to the system under study and thus not stochastic
- **Cross-over trials**: Although treatment assignment over time is random, in a randomized study treatment assignment and treatment order are independent of outcomes by design and therefore exogenous
- **Empirical evaluation**: Endogeneity may be empirically evaluated using the observed data by regressing current exposure $X_i(t)$ on previous outcomes $Y_i(t-1)$, adjusting for previous exposure $X_i(t-1)$

$$g(\mathrm{E}[X_i(t)]) = \theta_0 + \theta_1 Y_i(t-1) + \theta_2 X_i(t-1)$$

and using a model-based test to evaluate the null hypothesis: $\theta_1 = 0$

Implications

The presence of endogeneity determines specific analysis strategies

- If exposure is exogenous, then the analysis can focus on specifying the lag dependence of $Y_i(t)$ on $X_i(t)$, $X_i(t-1)$, $X_i(t-2), \ldots$
- If exposure is endogenous, then analysts must focus on selecting a meaningful target of inference and valid estimation methods

Targets of inference

With longitudinal outcomes and a time-dependent exposure there are several possible conditional expectations that may be of scientific interest

- **Fully conditional model**: Include the entire exposure process

$$\mathrm{E}[Y_i(t) \mid X_i(1), X_i(2), \ldots, X_i(T_i)]$$

- **Partly conditional models**: Include a subset of exposure process

$$\mathrm{E}[Y_i(t) \mid X_i(t)]$$
$$\mathrm{E}[Y_i(t) \mid X_i(t-k)] \text{ for } k \leq t$$
$$\mathrm{E}[Y_i(t) \mid \mathcal{X}_i(t) = \{X_i(1), X_i(2), \ldots, X_i(t)\}]$$

⋆ An appropriate target of inference that reflects the scientific question of interest must be identified prior to selection of an estimation method

Key assumption

Suppose that primary scientific interest lies in a cross-sectional mean model

$$\mathrm{E}[Y_i(t) \mid X_i(t)] = \beta_0 + \beta_1 X_i(t)$$

To ensure consistency of a generalized estimating equation or likelihood-based mixed-model estimator for β, it is sufficient to assume that

$$\mathrm{E}[Y_i(t) \mid X_i(t)] = \mathrm{E}[Y_i(t) \mid X_i(1), X_i(2), \ldots, X_i(T_i)]$$

Otherwise an **independence** estimating equation should be used

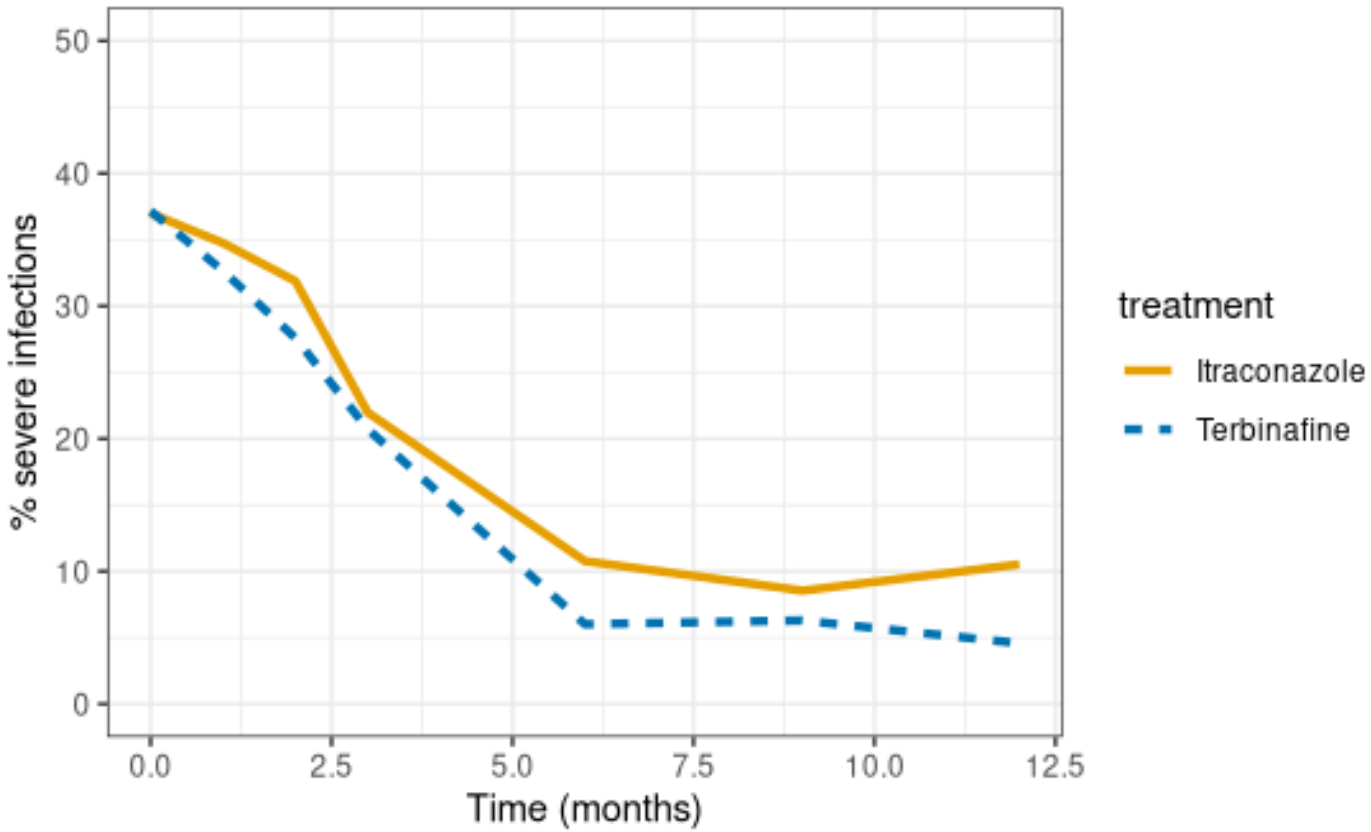
- Known as the full covariate conditional mean assumption
- Implies that with time-dependent exposures must assume exogeneity when using a covariance-weighting estimation method
- The full covariate conditional mean assumption is often overlooked and should be verified as a crucial element of model verification

(Pepe and Anderson, 1994)

Overview

Toenail Infection Analysis

- RCT comparing two oral treatments for a toenail infection
- n=294, most observed at 7 visits (0, 4, 8, 12, 24, 36, 48 weeks)
- Goal: Compare percentage of severe infections over time and between treatment groups



Overview

Summary and Comparison of Methods I

	GEE	**GLMM**
Mean model	Marginal	Conditional
Correlation	Model for correlation	Model for population heterogeneity; correlation induced by random effects
Corr. sources	One source (+ or -)	Multiple sources (+ only)
Model type	Semi-parametric (mean & corr.)	Parametric (exponential family)
Target quantities	Mean model	Mean model and within vs. between-subject heterogeneity

Summary and Comparison of Methods II

	GEE	**GLMM**
Estimation	Unbiased estimating equations	Maximum likelihood
Testing	Wald tests	Likelihood ratio or Wald tests
Inference	Marginal (population-averaged)	Conditional (subject-specific)
Missing data assumptions	Missing completely at random (MCAR)	Missing at random (MAR)
Robustness	Robust to correlation model mis-specification	Requires correct parametric model specification
Other notes	Large sample ($N \geq 40$) Model-based or sandwich variance estimator Efficiency of non-independence correlation models	Induced marginal mean structure and 'attenuation' Typically model-based variance, but can use sandwich (robust) estimator

Summary and Comparison of Methods III: Big Picture

	GEE	**GLMM**
Robustness	Provide valid estimates and standard errors for regression parameters of interest even if the correlation model is incorrectly specified $(+)$ Empirical variance estimator requires large sample size $(-)$	Provide valid estimates and standard errors for regression parameters only under stringent model assumptions $(-)$
Inference	Always provide population-averaged inference regardless of the outcome distribution; ignores subject-level heterogeneity $(+/-)$	Provide population-averaged or subject-specific inference depending on the outcome distribution and specified random effects $(+/-)$
Correlation	Accommodate only one source of correlation $(-/+)$	Accommodate multiple sources of correlation $(+/-)$
Missing Data	Require that any missing data are missing completely at random $(-)$	Require that any missing data are missing at random $(-/+)$

Advice

- Analysis of longitudinal data is often complex and difficult
- You now have versatile methods of analysis at your disposal
- Each of the methods you have learned has strengths and weaknesses
- Do not be afraid to apply different methods as appropriate
- Always be mindful of the scientific question(s) of interest
- Sensitivity analyses are helpful

Resources

Introductory

- Fitzmaurice GM, Laird NM, Ware JH. *Applied Longitudinal Analysis*. Wiley, 2011.
- Gelman A, Hill J. *Data Analysis Using Regression and Multilevel/ Hierarchical Models*. Cambridge University Press, 2007.
- Hedeker D, Gibbons RD. *Longitudinal Data Analysis*. Wiley, 2006.

Advanced

- Diggle PJ, Heagerty P, Liang K-Y, Zeger SL. *Analysis of Longitudinal Data*, 2nd Edition. Oxford University Press, 2002.
- Molenbergs G, Verbeke G. *Models for Discrete Longitudinal Data*. Springer Series in Statistics, 2006.
- Verbeke G, Molenbergs G. *Linear Mixed Models for Longitudinal Data*. Springer Series in Statistics, 2000.