ORGANIZATION OF LECTURES

Introduction and Overview
Longitudinal Data: Basic Concepts
Review of Generalized Linear Models for Longitudinal Data

- Marginal Models and Generalized Estimating Equations (GEE)
- Generalized Linear Mixed Models (GLMMs)
- Contrasting Marginal and Mixed Effects Models
- Examples and Illustrations

Concluding Remarks
Overview

In recent years, there have been remarkable advances in methods for analyzing longitudinal data.

When the response variable is continuous, familiar linear regression models can be extended to handle the correlated outcomes.

For linear models the correlation among repeated measures can be modelled explicitly (e.g., via covariance pattern models) or implicitly (e.g., via introduction of random effects).

A key feature of linear models is that interpretation of regression coefficients remains the same regardless of how the correlation is accounted for.

When the response variable is categorical (e.g., binary and count data), generalized linear models (e.g., logistic regression) can be extended to handle the correlated outcomes.

However, the non-linear transformations of the mean response (e.g., logit) raises additional issues concerning the interpretation of the regression coefficients.

Different approaches for accounting for the correlation lead to models having regression coefficients with distinct interpretations.

Different models for categorical outcomes have somewhat different targets of inference.
LONGITUDINAL DATA: BASIC CONCEPTS

Defining feature of longitudinal studies is that measurements of the same individuals are taken repeatedly through time.

Longitudinal studies allow direct study of change over time.

Objective: primary goal is to characterize the change in response over time and the factors that influence change.

With repeated measures on individuals, we can capture within-individual change.

Complication: repeated measures on individuals are correlated.

Terminology

Individuals/Subjects: Participants in a longitudinal study are referred to as individuals or subjects.

Occasions: In a longitudinal study individuals are measured repeatedly at different occasions or times.

The number of repeated observations, and their timing, can vary widely from one longitudinal study to another.

When number and timing of the repeated measurements are the same for all individuals, study design is said to be “balanced” over time.
MOTIVATING EXAMPLE

Oral Treatment of Toenail Infection

Randomized, double-blind, parallel-group, multicenter study of 294 patients comparing 2 oral treatments (denoted A and B) for toe-nail infection.

Outcome variable: Binary variable indicating presence of onycholysis (separation of the nail plate from the nail bed).

Patients evaluated for degree of onycholysis (separation of the nail plate from the nail-bed) at baseline (week 0) and at weeks 4, 8, 12, 24, 36, and 48.

Interested in the rate of decline of the proportion of patients with onycholysis over time and the effects of treatment on that rate.

Clinical trial of anti-epileptic drug progabide

(Thall and Vail, Biometrics, 1990)

Randomized, placebo-controlled study of treatment of epileptic seizures with progabide.

Patients were randomized to treatment with progabide, or to placebo in addition to standard therapy.

Outcome variable: Count of number of seizures

Measurement schedule: Baseline measurement during 8 weeks prior to randomization. Four measurements during consecutive two-week intervals.

Sample size: 28 epileptics on placebo; 31 epileptics on progabide
REVIEW OF GENERALIZED LINEAR MODELS

Generalized linear models are a class of regression models; they include the standard linear regression model but also many other important models:

- Linear regression/ANOVA for continuous data
- Logistic regression for binary data
- Loglinear models for count data

Generalized linear models extend the methods of regression analysis to settings where the outcome variable can be categorical.

In this workshop, we consider extensions of generalized linear models to longitudinal data.

Notation for Generalized Linear Models

Assume $N$ independent observations of a single response variable, $Y_i$.

Associated with each response, $Y_i$, there are $p$ covariates, $X_{i1}, ..., X_{ip}$.

Goal: Primarily interested in relating the mean of $Y_i$, $\mu_i = E(Y_i|X_{i1}, ..., X_{ip})$, to the covariates.
In generalized linear models:

(i) the distribution of the response is assumed to belong to a family of distributions known as the exponential family, e.g., normal, Bernoulli, binomial, and Poisson distributions.

(ii) A transformation of the mean response, $\mu_i$, is then linearly related to the covariates, via an appropriate link function:

$$g(\mu_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip},$$

where link function $g(\cdot)$ is a known function, e.g., $\log(\mu_i)$.

This implies that it is the transformed mean response that changes linearly with changes in the values of the covariates.

---

Mean and Variance of Exponential Family Distributions

Exponential family distributions share some common statistical properties.

The variance of $Y_i$ can be expressed in terms of

$$\text{Var} (Y_i|X_{i1},...,X_{ip}) = \phi v(\mu_i),$$

where the scale parameter $\phi > 0$.

The variance function, $v(\mu_i)$, describes how the variance of the response is functionally related to $\mu_i$, the mean of $Y_i$. 
Link Function

The link function applies a transformation to the mean and then links the covariates to the transformed mean,

\[ g(\mu_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip}, \]

where link function \( g(\cdot) \) is known function, e.g., \( \log(\mu_i) \).

This implies that it is the transformed mean response that changes linearly with changes in the values of the covariates.

---

Canonical link and variance functions for the normal, Bernoulli, and Poisson distributions.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Var. Function, ( v(\mu) )</th>
<th>Canonical Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>( v(\mu) = 1 )</td>
<td>Identity: ( \mu = \eta )</td>
</tr>
<tr>
<td>Bernoulli</td>
<td>( v(\mu) = \mu(1 - \mu) )</td>
<td>Logit: ( \log \left[ \frac{\mu}{1 - \mu} \right] = \eta )</td>
</tr>
<tr>
<td>Poisson</td>
<td>( v(\mu) = \mu )</td>
<td>Log: ( \log(\mu) = \eta )</td>
</tr>
</tbody>
</table>

where \( \eta = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p \).
Common Examples

Normal distribution:
If we assume that \( g(\cdot) \) is the identity function,
\[
g(\mu) = \mu
\]
then
\[
\mu_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip},
\]
gives the standard linear regression model, with \( \text{Var}(Y_i|X_{i1}, ..., X_{ip}) = \phi \).

Note: Variance is unrelated to the mean.

Bernoulli distribution:
For the Bernoulli distribution, \( 0 < \mu_i < 1 \), so we would prefer a link function that transforms the interval \([0, 1]\) on to the entire real line \((-\infty, \infty)\):
\[
\begin{align*}
\text{logit} : \ln \left[ \frac{\mu_i}{1 - \mu_i} \right] \\
\text{probit} : \Phi^{-1}(\mu_i)
\end{align*}
\]
where \( \Phi(\cdot) \) is the standard normal cumulative distribution function.
If we assume a logit link function then
\[
\log \left[ \frac{\mu_i}{1 - \mu_i} \right] = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip},
\]
yields logistic regression model, with \( \text{Var}(Y_i|X_{i1}, ..., X_{ip}) = \mu_i(1 - \mu_i) \) (Bernoulli variance).
**Poisson distribution:**

For the Poisson distribution, $\mu_i > 0$, so we would prefer a link function that transforms the interval $(0, \infty)$ on to the entire real line $(-\infty, \infty)$.

If we assume a log link function then

$$\log(\mu_i) = \beta_0 + \beta_1X_{i1} + \beta_2X_{i2} + \cdots + \beta_pX_{ip},$$

yields Poisson or loglinear regression model, with $\text{Var}(Y_i|X_{i1}, \ldots, X_{ip}) = \mu_i$ (Poisson variance).

---

**GENERALIZED LINEAR MODELS FOR LONGITUDINAL DATA**

Next, we focus on two general approaches for analyzing longitudinal responses.

These approaches can be considered extensions of generalized linear models to correlated data.

The main emphasis will be on discrete response data, e.g., count data or binary responses.

Note: In linear (mixed effects) models for continuous responses, the interpretation of the regression coefficients is independent of the correlation among the responses.
With discrete response data, this is no longer the case.

With non-linear models for discrete data, different approaches for accounting for the correlation leads to models having regression coefficients with distinct interpretations.

We will return to this important issue later.

We will consider two main extensions of generalized linear models:

1. Marginal Models
2. Generalized Linear Mixed Models

MARGINAL MODELS

The basic premise of marginal models is to make inferences about population (or sub-population) averages.

The term ‘marginal’ is used here to emphasize that the mean response modelled is conditional only on covariates and not on other responses or random effects.

A feature of marginal models is that the models for the mean and the ‘within-subject association’ (e.g., covariance) are specified separately.
Notation

Let $Y_{ij}$ denote response variable for $i^{th}$ subject on $j^{th}$ occasion.

$Y_{ij}$ can be continuous, binary, or a count.

We assume there are $n_i$ repeated measurements on the $i^{th}$ subject and each $Y_{ij}$ is observed at time $t_{ij}$.

Associated with each response, $Y_{ij}$, there is a $p \times 1$ vector of covariates, $X_{ij}$.

Covariates can be time-invariant (e.g., gender) or time-varying (e.g., time since baseline).

Can group $Y_{ij}$’s into a $n_i \times 1$ vector $Y_i$, and $X_{ij}$’s into a $n_i \times p$ matrix $X_i$.

Features of Marginal Models:

The focus of marginal models is on inferences about population averages.

The marginal expectation, $\mu_{ij} = E(Y_{ij}|X_{ij})$, of each response is modelled as a function of covariates.

Specifically, marginal models have the following three part specification:
1. The marginal expectation of the response, $\mu_{ij}$, depends on covariates through a known link function

$$g(\mu_{ij}) = \beta_0 + \beta_1 X_{1ij} + \beta_2 X_{2ij} + \cdots + \beta_p X_{pij}.$$ 

2. The marginal variance of $Y_{ij}$ depends on the marginal mean according to

$$\text{Var}(Y_{ij}|X_{ij}) = \phi v(\mu_{ij})$$

where $v(\mu_{ij})$ is a known ‘variance function’ and $\phi$ is a scale parameter that may be fixed and known or may need to be estimated. **Note:** For continuous response, can allow $\text{Var}(Y_{ij}|X_{ij}) = \phi_j v(\mu_{ij})$.

3. The ‘within-subject association’ among the responses is a function of the means and of additional parameters, say $\alpha$, that may also need to be estimated.

For example, when $\alpha$ represents pairwise correlations among responses, the covariances among the responses depend on $\mu_{ij}(\beta)$, $\phi$, and $\alpha$:

$$\text{Cov}(Y_{ij}, Y_{ik}|X_{ij}, X_{ik}) = \text{s.d.}(Y_{ij}) \text{Corr}(Y_{ij}, Y_{ik}|X_{ij}, X_{ik}) \text{s.d.}(Y_{ik})$$

$$= \sqrt{\phi v(\mu_{ij}) \text{Corr}(Y_{ij}, Y_{ik}|X_{ij}, X_{ik})} \sqrt{\phi v(\mu_{ik})}$$

where $\text{s.d.}(Y_{ij})$ is the standard deviation of $Y_{ij}$.

In principle, can also specify higher-order moments.
Examples of Marginal Models

Example 1. Continuous responses:

1. \( \mu_{ij} = \beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \cdots + \beta_p X_{ijp} \).
   (i.e., linear regression)

2. \( \text{Var}(Y_{ij} | X_{ij}) = \phi_j \)
   (i.e., heterogeneous variance, but no dependence of variance on mean)

3. \( \text{Corr}(Y_{ij}, Y_{ik} | X_{ij}, X_{ik}) = \alpha_{|k-j|} \) \((0 \leq \alpha \leq 1)\)
   (i.e., autoregressive correlation)

Example 2. Binary responses:

1. Logit \( (\mu_{ij}) = \beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \cdots + \beta_p X_{ijp} \).
   (i.e., logistic regression)

2. \( \text{Var}(Y_{ij} | X_{ij}) = \mu_{ij} (1 - \mu_{ij}) \)
   (i.e., Bernoulli variance)

3. OR \( (Y_{ij}, Y_{ik} | X_{ij}, X_{ik}) = \alpha_{jk} \)
   (i.e., unstructured odds ratios)

where

\[
\text{OR}(Y_{ij}, Y_{ik} | X_{ij}, X_{ik}) = \\
\frac{\Pr(Y_{ij} = 1, Y_{ik} = 1 | X_{ij}, X_{ik}) \Pr(Y_{ij} = 0, Y_{ik} = 0 | X_{ij}, X_{ik})}{\Pr(Y_{ij} = 1, Y_{ik} = 0 | X_{ij}, X_{ik}) \Pr(Y_{ij} = 0, Y_{ik} = 1 | X_{ij}, X_{ik})},
\]
**Example 3. Count data:**

1. $\log (\mu_{ij}) = \beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \cdots + \beta_p X_{ijp}.$
   (i.e., Poisson regression)

2. $\text{Var} (Y_{ij} | X_{ij}) = \phi \mu_{ij}$
   (i.e., extra-Poisson variance or “overdispersion” when $\phi > 1$)

3. $\text{Corr} (Y_{ij}, Y_{ik} | X_{ij}, X_{ik}) = \alpha$
   (i.e., “compound symmetry” correlation)

---

**Interpretation of Marginal Model Parameters**

The regression parameters, $\beta$, have ‘population-averaged’ interpretations (where ‘averaging’ is over all individuals within subgroups of the population):

- describe effect of covariates on the average responses

- contrast the means in sub-populations that share common covariate values

$\implies$ Marginal models are most useful for population-level inferences.

The regression parameters are directly estimable from the data.

Of note, nature or magnitude of within-subject association (e.g., correlation) does not alter the interpretation of $\beta$. 
For example, consider the following logistic model,

\[
\text{logit}(\mu_{ij}) = \text{logit}\{E(Y_{ij}|X_{ij})\} = \beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \cdots + \beta_p X_{ijp}.
\]

Each element of \(\beta\) measures the change in the log odds of a ‘positive’ response per unit change in the respective covariate, for sub-populations defined by fixed and known covariate values.

The interpretation of any component of \(\beta\), say \(\beta_k\), is in terms of adjusted changes in the transformed mean (or “population-averaged”) response for a unit change in the corresponding covariate, say \(X_{ijk}\).

When \(X_{ijk}\) takes on some value \(x\), the log odds of a positive response is,

\[
\log\left\{ \frac{\Pr(Y_{ij}=1|X_{ij1},...,X_{ijk}=x,...,X_{ijp})}{\Pr(Y_{ij}=0|X_{ij1},...,X_{ijk}=x,...,X_{ijp})} \right\} = \\
\beta_0 + \beta_1 X_{ij1} + \cdots + \beta_k x + \cdots + \beta_p X_{ijp}.
\]

Similarly, when \(X_{ijk}\) now takes on some value \(x + 1\),

\[
\log\left\{ \frac{\Pr(Y_{ij}=1|X_{ij1},...,X_{ijk}=x+1,...,X_{ijp})}{\Pr(Y_{ij}=0|X_{ij1},...,X_{ijk}=x+1,...,X_{ijp})} \right\} = \\
\beta_0 + \beta_1 X_{ij1} + \cdots + \beta_k (x + 1) + \cdots + \beta_p X_{ijp}.
\]

\(\rightarrow \beta_k\) is adjusted change in log odds for subgroups of the study population (defined by any fixed values of \(X_{ij1},...,X_{ij(k-1)},X_{ij(k+1)},...,X_{ijp}\)).
Statistical Inference for Marginal Models

Maximum Likelihood (ML):

Unfortunately, with discrete response data there is no simple analogue of the multivariate normal distribution.

In the absence of a “convenient” likelihood function for discrete data, there is no unified likelihood-based approach for marginal models.


GENERALIZED ESTIMATING EQUATIONS

Avoid making distributional assumptions about \( Y_i \) altogether.

Potential Advantages:

Empirical researcher does not have to be concerned that the distribution of \( Y_i \) closely approximates some multivariate distribution.

It circumvents the need to specify models for the three-way, four-way and higher-way associations (higher-order moments) among the responses.

It leads to a method of estimation, known as generalized estimating equations (GEE), that is straightforward to implement.
The GEE approach has become an extremely popular method for analyzing discrete longitudinal data.

It provides a flexible approach for modelling the mean and the pairwise within-subject association structure.

It can handle inherently unbalanced designs and missing data with ease (albeit making strong assumptions about missingness).

GEE approach is computationally straightforward and has been implemented in existing, widely-available statistical software.

---

**Case Study 1: Oral Treatment of Toenail Infection**

Randomized, double-blind, parallel-group, multicenter study of 294 patients comparing 2 oral treatments (denoted A and B) for toe-nail infection.

Outcome variable: Binary variable indicating presence of onycholysis (separation of the nail plate from the nail bed).

Patients evaluated for degree of onycholysis (separation of the nail plate from the nail-bed) at baseline (week 0) and at weeks 4, 8, 12, 24, 36, and 48.

Interested in the rate of decline of the proportion of patients with onycholysis over time and the effects of treatment on that rate.
Assume that the marginal probability of onycholysis follows a logistic model,

$$\text{logit}\{E(Y_{ij}|X_{ij})\} = \beta_0 + \beta_1 \text{Month}_{ij} + \beta_2 \text{Trt}_i \times \text{Month}_{ij}$$

where $\text{Trt} = 1$ if treatment group B and 0 otherwise.

Here, we assume that $\text{Var}(Y_{ij}|X_{ij}) = \mu_{ij}(1 - \mu_{ij})$.

We also assume an unstructured correlation for the within-subject association (i.e., estimate all possible pairwise correlations).

---

Table 1: GEE estimates and standard errors (empirical) from marginal logistic regression model for onycholysis data.

<table>
<thead>
<tr>
<th>PARAMETER</th>
<th>ESTIMATE</th>
<th>SE</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTERCEPT</td>
<td>-0.698</td>
<td>0.122</td>
<td>-5.74</td>
</tr>
<tr>
<td>Month</td>
<td>-0.140</td>
<td>0.026</td>
<td>-5.36</td>
</tr>
<tr>
<td>Trt × Month</td>
<td>-0.081</td>
<td>0.042</td>
<td>-1.94</td>
</tr>
</tbody>
</table>
Results

From the output above, we would conclude that:

1. There is a suggestion of a difference in the rate of decline in the two treatment groups ($P = 0.052$).

2. Over 12 months, the odds of onycholysis has decreased by a factor of $0.19 \, [\exp(-0.14*12)]$ in treatment group A.

3. Over 12 months, the odds of onycholysis has decreased by a factor of $0.07 \, [\exp(-0.221*12)]$ in treatment group B.

4. Odds ratio comparing 12 month decreases in risk of onycholysis between treatments A and B is approx 2.6 (or $e^{12 \times 0.081}$).

5. Overall, there is a significant decline over time in the prevalence of onycholysis for all randomized patients.

Summary of Key Points

The focus of marginal models is on inferences about population averages. The regression parameters, $\beta$, have ‘population-averaged’ interpretations (where ‘averaging’ is over all individuals within subgroups of the population):

- describe effect of covariates on marginal expectations or average responses

- contrast means in sub-populations that share common covariate values

$\Rightarrow$ Marginal models are most useful for population-level inferences.

Marginal models should not be used to make inferences about individuals (“ecological fallacy”).
GENERALIZED LINEAR MIXED MODELS

So far, we have discussed marginal models for longitudinal data.

Next, we consider a second type of extension, generalized linear mixed models (GLMMs).

We describe how these models extend the conceptual approach represented by the linear mixed effects model.

We also highlight their greater degree of conceptual and analytic complexity relative to marginal models.

Incorporating Random Effects into Generalized Linear Models

Postulate unobserved latent variables (random effects) shared by the repeated measures on the same subject.

The basic premise is that we assume natural heterogeneity across individuals in a subset of the regression coefficients.

That is, a subset of the regression coefficients (e.g., intercepts and slopes) are assumed to vary across individuals according to some distribution.

Then, conditional on the random effects, it is assumed that the responses for a single individual are independent observations from a distribution belonging to the exponential family.
Generalized Linear Mixed Models

The generalized linear mixed model can be considered in two steps:

**First Step**: Assumes that the conditional distribution of each $Y_{ij}$, given individual-specific effects $b_i$, belongs to the exponential family with conditional mean,

$$g\{E(Y_{ij}|X_{ij}, b_i)\} = X_{ij}'\beta + Z_{ij}'b_i$$

where $g(\cdot)$ is a known link function and $Z_{ij}$ is a known design vector, a subset of $X_{ij}$, linking the random effects $b_i$ to $Y_{ij}$.

The particular subset of the regression parameters $\beta$ that vary randomly is determined by components of $X_{ij}$ that comprise $Z_{ij}$.

**Second-Step**: The $b_i$ are assumed to vary independently from one individual to another and $b_i \sim N(0, G)$.

Here, $G$ is the covariance matrix for the random effects.

Note: There is an additional assumption of ‘conditional independence’.

That is, given $b_i$, the responses $Y_{i1}, Y_{i2}, ..., Y_{in_i}$ are assumed to be mutually independent.
Example 1:

Binary logistic model with random intercepts:

\[
\text{logit}\{E(Y_{ij}|X_{ij}, b_i)\} = \beta_0 + \beta_1 X_{ij1} + \cdots + \beta_p X_{ijp} + b_i
\]

\[
\text{Var}(Y_{ij}|X_{ij}, b_i) = E(Y_{ij}|X_{ij}, b_i)\{1 - E(Y_{ij}|X_{ij}, b_i)\} \text{ (Bernoulli variance),}
\]

and \(b_i \sim N(0, \sigma_b^2)\).

Example 2:

Random coefficients (random intercepts and slopes) Poisson regression model:

\[
\log\{E(Y_{ij}|X_{ij}, b_i)\} = \beta_0 + \beta_1 t_{ij} + b_{0i} + b_{1i} t_{ij}
\]

\[
\text{Var}(Y_{ij}|X_{ij}, b_i) = E(Y_{ij}|X_{ij}, b_i) \text{ (Poisson variance),}
\]

and \(b_i \sim N(0, G)\).

Note: \(G\) is the covariance matrix for \(b_{0i}\) and \(b_{1i}\).
Recall that marginal models consider the consequences of dependence among the repeated measures on the same subject, via a “working” covariance.

In contrast, GLMMs provide a potential explanation for the sources of dependence among the repeated measures on the same subject, via the introduction of random effects.

However, the introduction of random effects also has important implications for the interpretation of the regression parameters in GLMMs.

**Interpretation of Fixed Effects**

GLMMs are most useful when the scientific objective is to make inferences about individuals rather than population averages.

Main focus is on the individual and the influence of covariates on a typical \((b_i = 0)\) individual’s responses.

Regression parameters, \(\beta\), measure the change in expected value of response while holding constant other covariates and the random effects.
For example, consider the following logistic model,

\[ \text{logit}\{E(Y_{ij}|X_{ij}, b_i)\} = \beta_0 + \beta_1 X_{ij1} + \cdots + \beta_p X_{ijp} + b_i \]

with \( b_i \sim N(0, \sigma^2) \).

Each element of \( \beta \) measures the adjusted change in the log odds of a ‘positive’ response per unit change in the respective covariate, for an individual with propensity to respond positively, \( b_i \).

The interpretation of any component of \( \beta \), say \( \beta_k \), is in terms of adjusted changes in a specific individual’s log odds of response for a unit change in the corresponding covariate, say \( X_{ijk} \).

Note: This is not always directly observable from the data.

When \( X_{ijk} \) takes on some value \( x \), the log odds of a positive response is,

\[
\log \left\{ \frac{\Pr(Y_{ij}=1|b_i, X_{ij1} = x, \ldots, X_{ijp})}{\Pr(Y_{ij}=0|b_i, X_{ij1} = x, \ldots, X_{ijp})} \right\} = \beta_0 + b_i + \beta_1 X_{ij1} + \cdots + \beta_k x + \cdots + \beta_p X_{ijp}.
\]

Similarly, when \( X_{ijk} \) now takes on some value \( x + 1 \),

\[
\log \left\{ \frac{\Pr(Y_{ij}=1|b_i, X_{ij1} = x+1, \ldots, X_{ijp})}{\Pr(Y_{ij}=0|b_i, X_{ij1} = x+1, \ldots, X_{ijp})} \right\} = \beta_0 + b_i + \beta_1 X_{ij1} + \cdots + \beta_k (x+1) + \cdots + \beta_p X_{ijp}.
\]

\( \rightarrow \beta_k \) is adjusted change in log odds for individual with propensity to respond, \( b_i \).
This *subject-specific* interpretation of $\beta_k$ is more appealing when $X_{ijk}$ is a *time-varying* covariate.

That is, when it is possible to hold $b_i$ (and remaining covariates) fixed and also change the value of the covariate, $X_{ijk}$.

Recall: Time-varying covariate is one whose value can change over time, e.g., time since baseline, smoking status, and environmental exposures.

When $X_{ijk}$ is *time-invariant* the interpretation of $\beta_k$ is less transparent.

With a time-invariant covariate (e.g., gender), changing the value of the covariate requires also a change in the index $i$ of $X_{ijk}$, say $X_{ij'}$.

When $X_{ijk}$ takes on some value $x$, the log odds of a positive response is,

$$
\log \left\{ \frac{Pr(Y_{ijk} = 1|b_i, X_{ij1}, \ldots, X_{ijk} = x, \ldots, X_{ijp})}{Pr(Y_{ij} = 0|b_i, X_{ij1}, \ldots, X_{ijk} = x, \ldots, X_{ijp})} \right\} = 
\beta_0 + b_i + \beta_1 X_{ij1} + \cdots + \beta_k x + \cdots + \beta_p X_{ijp}.
$$

Similarly, when $X_{ij'}$ now takes on some value $x + 1$,

$$
\log \left\{ \frac{Pr(Y_{ij'} = 1|b_{ij'}, X_{ij'1}, \ldots, X_{ij'k} = x + 1, \ldots, X_{ij'p})}{Pr(Y_{ij'} = 0|b_{ij'}, X_{ij'1}, \ldots, X_{ij'k} = x + 1, \ldots, X_{ij'p})} \right\} = 
\beta_0 + b_{ij'} + \beta_1 X_{ij'1} + \cdots + \beta_k (x + 1) + \cdots + \beta_p X_{ij'p}.
$$
Even when we consider two subjects with identical covariates except for the $k^{th}$, the difference in log odds is

$$\beta_k + (b_i - b_{i'}).$$

That is, $\beta_k$ has become confounded with $b_i - b_{i'}$.

This dilemma can only be resolved by assuming same value for the unobserved random effects, $b_i = b_{i'}$; however, this contrast is not directly observable.

51

**Estimation**

The joint probability density function is given by:

$$f(Y_i|X_i, b_i) f(b_i)$$

Inferences are based on the marginal or integrated likelihood function.

$$L(\beta, \phi, G) = \prod_{i=1}^{N} \int f(Y_i|X_i, b_i) f(b_i) db_i$$

obtained by averaging over the distribution of the unobserved random effects, $b_i$.

However, simple analytic solutions are rarely available and numerical or Monte Carlo integration techniques are required.

52
Case Study 2

Clinical trial of anti-epileptic drug progabide

Randomized, placebo-controlled study of treatment of epileptic seizures with progabide.

Patients were randomized to treatment with progabide, or to placebo in addition to standard therapy.

Response variable: Count of number of seizures

Measurement schedule: Baseline measurement during 8 weeks prior to randomization. Four measurements during consecutive two-week intervals.

Interested in the effect of treatment with progabide on changes in an individual’s rate of seizures?
Assume conditional rate of seizures follows the mixed effects loglinear model,

$$\log \{ E(Y_{ij} | X_{ij}, b_i) \} = \log(t_{ij}) + \beta_0 + b_0i + \beta_1 \text{time}_{ij} + b_1 \text{time}_{ij} + \beta_2 \text{trt}_i + \beta_3 \text{trt}_i \ast \text{time}_{ij}$$

where $t_{ij} =$ length of period; $\text{time}_{ij} =$ 1 if periods 1-4, 0 if baseline; $\text{trt}_i =$ 1 if progabide, 0 if placebo.

$(b_{0i}, b_{1i})$ are assumed to have a bivariate normal distribution with zero mean and covariance $G$.

Also, we assume that

$$\text{Var}(Y_{ij} | X_{ij}, b_i) = E(Y_{ij} | X_{ij}, b_i).$$

### Table 2: Subject-specific log expected seizure rates in the two groups at baseline and during post-baseline follow-up.

| Treatment Group | Period     | log $\left( \frac{E(Y_{ij}|X_{ij},b_i)}{t_{ij}} \right)$ |
|-----------------|------------|----------------------------------------------------------|
| Placebo         | Baseline   | $\beta_0 + b_{0i}$                                       |
|                 | Follow-up  | $(\beta_0 + b_{0i}) + (\beta_1 + b_{1i})$               |
| Progabide       | Baseline   | $(\beta_0 + b_{0i}) + \beta_2$                          |
|                 | Follow-up  | $(\beta_0 + b_{0i}) + (\beta_1 + b_{1i}) + \beta_2 + \beta_3$ |
Parameter estimates and standard errors from mixed effects log-linear regression model for the seizure data.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>SE</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.0707</td>
<td>0.1406</td>
<td>7.62</td>
</tr>
<tr>
<td>$\text{time}_{ij}$</td>
<td>-0.0004</td>
<td>0.1097</td>
<td>-0.00</td>
</tr>
<tr>
<td>$\text{trt}_i$</td>
<td>0.0513</td>
<td>0.1931</td>
<td>0.27</td>
</tr>
<tr>
<td>$\text{trt}<em>i \times \text{time}</em>{ij}$</td>
<td>-0.3065</td>
<td>0.1513</td>
<td>-2.03</td>
</tr>
<tr>
<td>$\text{Var}(b_{0i})$</td>
<td>0.5010</td>
<td>0.1010</td>
<td>4.96</td>
</tr>
<tr>
<td>$\text{Var}(b_{1i})$</td>
<td>0.2334</td>
<td>0.0608</td>
<td>3.84</td>
</tr>
<tr>
<td>$\text{Cov}(b_{0i}, b_{1i})$</td>
<td>0.0541</td>
<td>0.0559</td>
<td>0.97</td>
</tr>
</tbody>
</table>

ML based on 50-point adaptive Gaussian quadrature.

Results of the analysis suggest:

1. A patient treated with placebo has the same expected seizure rate before and after randomization [$\exp(-0.0004) \approx 1$].

2. A patient treated with progabide has expected seizure rate reduced after treatment by approximately 26% [$1 - \exp(-0.0004 - 0.3065) \approx 0.26$].

3. Estimated variance of the random intercepts and slopes is relatively large

4. Heterogeneity should not be ignored
Summary of Key Points

GLMMs extend the conceptual approach represented by the linear mixed effects model.

GLMMs assume natural heterogeneity across individuals in a subset of the regression coefficients.

The focus of GLMMs is on inferences about individuals.

The regression parameters, $\beta$, have ‘subject-specific’ interpretations in terms of changes in the transformed mean response for a specific individual.

Contrasting Marginal and Mixed Effects Models for Longitudinal Data

So far, we have discussed two main extensions of generalized linear models:

1. Marginal Models
2. Generalized Linear Mixed Models

There are important distinctions between these two broad classes of models that go beyond simple differences in approaches for accounting for the within-subject association.

These two classes of models have somewhat different targets of inference and address subtly different questions regarding longitudinal change in the response.
A **marginal model** for the mean response is given by

\[ g(\mu_{ij}) = g\{E(Y_{ij}|X_{ij})\} = X'_{ij}\beta = \beta_0 + \beta_1 X_{ij1} + \cdots + \beta_p X_{ijp}, \]

where \( g(\cdot) \) is an appropriate non-linear link function (e.g., logit or log).

In marginal models, \( \beta \)'s have interpretation in terms of changes in the transformed mean response in the study population, and their relation to covariates.

The population means can be expressed in terms of the inverse link function, say \( h(\cdot) = g^{-1}(\cdot) \),

\[ h\{g(\mu_{ij})\} = \mu_{ij} = E(Y_{ij}|X_{ij}) = h(\beta_0 + \beta_1 X_{ij1} + \cdots + \beta_p X_{ijp}). \]

Next, consider the **generalized linear mixed model**

\[ g\{E(Y_{ij}|X_{ij}, b_i)\} = X'_{ij}\beta^* + Z'_{ij}b_i, \]

where the random effects \( b_i \) have a distribution with mean zero and covariance matrix \( G \).

The regression coefficients \( \beta^* \) have subject-specific interpretations in terms of changes in the transformed mean response for a specific individual.

\( \beta^* \) do not describe changes in the transformed mean response in the study population.
In GLMMs there is an implied model for the marginal means.

This can be obtained by averaging over distribution of the random effects,

\[
\mu_{ij} = E(Y_{ij} | X_{ij}) \\
= E\{E(Y_{ij} | X_{ij}, b_i)\} \\
= E\{h(X'_{ij}\beta^* + Z'_{ij}b_i)\} \\
= \int_{-\infty}^{\infty} h(X'_{ij}\beta^* + Z'_{ij}b_i)f(b_i)db_i.
\]

However, this expression for \(E(Y_{ij} | X_{ij})\) does not, in general, have a closed-form expression and, moreover,

\[
E(Y_{ij} | X_{ij}) \neq h(X'_{ij}\beta)
\]

for any \(\beta\), e.g., logistic mixed effects model \(\neq\) marginal logistic model.

That is, marginalized model doesn’t satisfy generalized linear model.

Graphical Illustration

Suppose \(Y_i\) is a vector of binary responses and it is of interest to describe changes in the log odds of success over time.

A logistic regression model, with randomly varying intercepts, is given by

\[
\text{logit}\{E(Y_{ij} | t_{ij}, b_i)\} = \beta^*_0 + \beta^*_1 t_{ij} + b_i
\]

where \(t_{ij} = 0\) at baseline and \(t_{ij} = 1\) post-baseline.

The \(b_i\) are assumed to have a normal distribution with zero mean and variance \(\sigma^2_b = \text{Var}(b_i)\).

Let \(\beta^*_0 = 1.5, \beta^*_1 = -3.0\), and \(\text{Var}(b_i) = 1.0\).
At baseline, log odds has a normal distribution with mean = median = 1.5 (see shaded densities).

Note, however, that subject-specific probabilities of disease have a negatively skewed distribution with median, but not mean, of 0.82. The mean of the subject-specific probabilities is 0.78.

Thus, probability of disease for a “typical” individual from the population (0.82) is not the same as the prevalence of disease in the same population (0.78).
Similarly, the log odds of disease post-baseline has a normal distribution with mean = median = \(-1.5\) (see unshaded densities).

However, subject-specific post-baseline probabilities of disease have a positively skewed distribution with median, but not mean, of 0.18. The mean of the subject-specific probabilities is 0.22.

Thus, probability of disease post-baseline for a “typical” individual from the population (0.18) is not the same as the prevalence of disease in the same population (0.22).

The effect of treatment on the log odds of disease for a typical individual from the population, \(\beta_1^* = -3.0\), is not the same as the contrast of population log odds. The latter is what is estimated in a marginal model, say

$$\logit\{E(Y_{ij}|t_{ij})\} = \beta_0 + \beta_1 t_{ij},$$

and can be obtained by comparing the log odds of disease in the population at baseline, \(\log(0.78/0.22) = 1.255\), with the log odds of disease in the population post-baseline, \(\log(0.22/0.78) = -1.255\).

This yields a population-averaged measure of effect, \(\beta_1 = -2.51\), which is approximately 15% smaller than \(\beta_1^*\), the subject-specific effect of treatment.
Case Study 3

Cross-Over Trial of Cerebrovascular Deficiency

- Two-period cross-over trial comparing effects of active drug to placebo on cerebrovascular deficiency
- 67 patients randomly allocated to two treatment sequences
- 34 patients receiving Placebo → Active
- 33 patients receiving Active → Placebo
- Each patient has a bivariate binary response vector, \( Y_i = (Y_{i1}, Y_{i2}) \) denoting whether an electrocardiogram was normal (0) or abnormal (1).

Data from a two-period cross-over trial comparing the effects of active drug to placebo on cerebrovascular deficiency. The response indicates whether an electrocardiogram was normal (0) or abnormal (1).

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Response (Period 1, Period 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1,1)</td>
</tr>
<tr>
<td>Sequence 1 (P → A)</td>
<td>6</td>
</tr>
<tr>
<td>Sequence 2 (A → P)</td>
<td>9</td>
</tr>
</tbody>
</table>

P: Placebo; A: Active drug.
First, consider marginal logistic model

\[ \text{logit}(\mu_{ij}) = \logit\{\Pr(Y_{ij} = 1) \mid X_{ij}\} = \beta_0 + \beta_1 \text{Treatment} + \beta_2 \text{Period} \]

where Treatment (0 = Placebo, 1 = Active drug) and Period (0 = Period 1, 1 = Period 2).

The within subject association between the two responses was modelled in terms of a common log odds ratio, \( \alpha \),

\[ \log \frac{\Pr(Y_{i1} = 1, Y_{i2} = 1 \mid X_{i1}, X_{i2}) \Pr(Y_{i1} = 0, Y_{i2} = 0 \mid X_{i1}, X_{i2})}{\Pr(Y_{i1} = 1, Y_{i2} = 0 \mid X_{i1}, X_{i2}) \Pr(Y_{i1} = 0, Y_{i2} = 1 \mid X_{i1}, X_{i2})} = \alpha. \]

Parameter estimates and standard errors from marginal logistic regression model for the cerebrovascular deficiency data.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>SE</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-1.2433</td>
<td>0.2999</td>
<td>-4.15</td>
</tr>
<tr>
<td>Treatment</td>
<td>0.5689</td>
<td>0.2335</td>
<td>2.44</td>
</tr>
<tr>
<td>Period</td>
<td>0.2951</td>
<td>0.2319</td>
<td>1.27</td>
</tr>
<tr>
<td>log OR (( \alpha ))</td>
<td>3.5617</td>
<td>0.8148</td>
<td>4.37</td>
</tr>
</tbody>
</table>
The results indicate that treatment with the active drug is harmful, increasing the rates of abnormal electrocardiograms.

The odds of an abnormal electrocardiogram is 1.77 (or $e^{0.57}$) times higher when treated with active drug versus placebo.

The estimate of the within-subject association is $\hat{\alpha} = 3.56$, indicating that there is very strong positive association (OR= 35.2).

Next, consider logistic regression model with a random patient effect,

$$\logit\{E(Y_{ij}|X_{ij}, b_i)\} = \beta_0^* + \beta_1^*\text{Treatment} + \beta_2^*\text{Period} + b_i$$

where the random effect $b_i$ is assumed to have a normal distribution with zero mean and variance, $\sigma_b^2 = \text{Var}(b_i)$. 
Parameter estimates and standard errors from mixed effects logistic regression model for the cerebrovascular deficiency data.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>SE</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-4.0817</td>
<td>1.6711</td>
<td>-2.44</td>
</tr>
<tr>
<td>Treatment</td>
<td>1.8631</td>
<td>0.9269</td>
<td>2.01</td>
</tr>
<tr>
<td>Period</td>
<td>1.0376</td>
<td>0.8189</td>
<td>1.27</td>
</tr>
</tbody>
</table>

\[
\sigma^2_b = \text{Var}(b_i) = 24.4365 \quad \text{with} \quad \text{SE} = 18.8500 \quad \text{and} \quad Z = 1.30
\]

ML based on 100-point adaptive Gaussian quadrature.

The results also indicate that treatment with the active drug is harmful, increasing the patient-specific risk of an abnormal electrocardiagram.

In particular, a patient’s odds of an abnormal electrocardiogram is 6.4 (or \(e^{1.86}\)) times higher when treated with active drug than when treated with the placebo.

The estimate of the variance of \(b_i\), \(\hat{\sigma}_b^2 = 24.4\), indicates that there is very substantial between-patient variability in their propensity for abnormal electrocardiograms.
Comparison of the two estimated effects of treatment, $e^{\hat{\beta}_1} = 1.8$ and $e^{\hat{\beta}_1^*} = 6.4$, from the marginal and mixed effects logistic regression models highlights the distinction between these two analytic approaches.

$\hat{\beta}_1$ from marginal model describes how the average rates (expressed in terms of odds) of abnormal ECGs could be increased in the study population if patients are treated with the active drug.

$\hat{\beta}_1^*$ from the mixed effects model describes how the odds of an abnormal ECG increases for any patient treated with the active drug.

Thus, a population-level analysis understates the individual risk, and vice versa.

In summary, the answer to the question “what are the side effects of the active drug” will depend on whether scientific interest is in its impact on the study population or on an individual drawn at random from that population.

With marginal models the main focus is on inferences about the study population.

With generalized linear mixed models the main focus is on inferences about individuals.
Aside

Does the very large estimate of variance, $\hat{\sigma}_b^2 = 24.4$, accurately reflect between-patient variability in the risk of abnormal electrocardiogram?

In this example, a large proportion of subjects (82%) had same response, (0,0) or (1,1), at both occasions.

This feature can only be captured by a normal distribution for the log odds with large variance.

When number of repeated binary responses is small, and there is a large proportion of subjects with positive (negative) responses at all occasions, the normal assumption for $b_i$ is questionable.

CONCLUDING REMARKS

Unlike linear models, where the concepts of regression analysis can be applied quite robustly, longitudinal analysis of categorical data raises many subtle issues.

Different models for categorical outcomes can give discernibly different results.

The choice and meaning of longitudinal models for categorical outcomes require somewhat greater care.

With different targets of inference, different models for categorical outcomes address subtly different questions regarding longitudinal change.
Choice among models?

- should be guided by specific scientific question of interest
- answers to different questions will usually demand that different models have to be applied
- different questions will often produce different, albeit compatible, answers
- “one size does not fit all”

FURTHER READING


Also, see web site: www.biostat.harvard.edu/~fitzmaur/ala
FURTHER READING


