Chapter 14

Generalized Linear Mixed Effects Models

14.1 INTRODUCTION

In the previous chapter we described marginal models for longitudinal data. Marginal models can be considered an extension of generalized linear models that *directly* incorporate the within-subject association among the repeated measurements. To estimate the regression parameters in a marginal model, we made some assumptions about the marginal distribution of the response at each occasion (e.g., assumptions about the mean, and its dependence on the covariates, and the variance of each Y_{ij}). We also made assumptions about the pairwise within-subject associations among the responses, thereby linking repeated observations of the same subject. A notable feature of marginal models is that the mean response and the covariance are modeled separately. This separation ensures that the interpretation of the regression coefficients in a marginal model does not rely on the assumed model for the covariance among the responses. In specifying the marginal means, variances, and pairwise associations, we did not fully specify the joint distribution of the vector of responses. However, these assumptions were sufficient for estimating and constructing confidence intervals for the regression parameters using the GEE approach.

An alternative approach for accounting for the within-subject association is via the introduction of random effects. In Chapter 8 we saw how the incorporation of random effects at the individual level induces correlation among the repeated measures at the population level. In this chapter we describe how generalized linear models can be extended to longitudinal data by allowing a subset of the regression coefficients to vary randomly from one individual to another. These models are known as *generalized linear mixed effects models*, and they extend in a natural way the conceptual approach represented by the linear mixed effects models discussed in Chapter 8. However, we must caution the reader at the outset that the introduction of random effects in generalized linear models produces a greater degree of conceptual and analytic complexity relative to marginal models or to random effects in linear models. Although both classes of models account for the within-subject association among repeated measurements, the manner in which they do so has important implications for the interpretation of the regression parameters. In Chapter 16 we highlight the major distinctions between the regression coefficients in marginal and generalized linear mixed models and consider various aspects of interpretation of the regression effects in these two classes of models for longitudinal data.

14.2 INCORPORATING RANDOM EFFECTS IN GENERALIZED LINEAR MODELS

The basic premise underlying the generalized linear mixed effects model for longitudinal data is the assumption of heterogeneity across individuals in the study population in a subset of the regression coefficients from a generalized linear model. That is, a subset of the regression coefficients (e.g., intercepts in a logistic regression model) are assumed to vary across individuals according to some distribution. The random effects can be thought of as reflecting natural heterogeneity due to many unmeasured factors. For mathematical and computational convenience, we ordinarily assume that the random effects have a multivariate normal distribution. Then, conditional on the random effects, we assume that the responses for any particular individual are independent observations from a distribution belonging to the exponential family (e.g., the Bernoulli distribution if Y_{ij} is binary, the Poisson distribution if Y_{ij} is a count, or the multinomial distribution if Y_{ij} is ordinal). The latter assumption is completely analogous to the "conditional independence" assumption ($R_i = \sigma^2 I_{n_i}$) made in the linear mixed effects model described in Chapter 8. In fact the linear mixed effects model is simply a special case of the generalized linear mixed effects model where the conditional mean, given the random effects, is related to the covariates via an identity link function and the conditional distribution of the responses is assumed to be normal. Because the linear mixed effects model is a special case, it is useful for pedagogical purposes to consider its formulation within the framework and terminology of generalized linear models. By doing so, the extensions to other types of response variables will become more apparent.

Linear Mixed Effects Models

In this section we consider the linear mixed effects model as a generalized linear model, albeit one with both fixed and random effects. Recall that the standard generalized linear model formulation requires a three-part specification: (1) a distributional assumption, (2) a systematic component, and (3) a link function. In the linear mixed effects model it is assumed that the *conditional* distribution of each Y_{ij}, given a vector of random effects b_i, has a normal distribution, with $\text{Var}(Y_{ij}|b_i) = \sigma^2$ (i.e., $\phi = \sigma^2$ and $v(\mu_{ij}) = 1$). In addition, given the random effects b_i, it is assumed that the Y_{ij} are independent of one another (i.e., given b_i, Y_{ij} and Y_{ik} are assumed to be independent of each other). This completes the distributional assumptions on the Y_{ij}. Next the conditional mean of Y_{ij} is assumed to depend on both fixed and random effects via the following extended definition of the linear predictor:

$$\eta_{ij} = X'_{ij}\beta + Z'_{ij}b_i,$$

where the definition of the linear predictor has been extended to incorporate both population (or fixed) and subject-specific (or random) effects. In addition the random effects, b_i, are assumed to have a multivariate normal distribution. This specifies the systematic component. Finally, an identity link function relates the conditional mean of Y_{ij} to the linear predictor,

$$E(Y_{ij}|b_i) = \eta_{ij} = X'_{ij}\beta + Z'_{ij}b_i.$$

That is, for the identity link, $\eta_{ij} - g\{E(Y_{ij}|b_i)\} = E(Y_{ij}|b_i)$ and hence,

$$E(Y_{ij}|b_i) = X'_{ij}\beta + Z'_{ij}b_i.$$

In the linear mixed effects model, the response for the i^{th} subject at the j^{th} occasion is assumed to differ from the population mean, $X'_{ij}\beta$, by a subject-specific effect, $Z'_{ij}b_i$, and a within-subject measurement error, ϵ_{ij}. The within-subject measurement errors are independently normally distributed, with zero mean and variance σ^2.

When collected in a vector, $\epsilon_i \sim N(0, R_i)$, where $R_i = \sigma^2 I_{n_i}$ (the "conditional independence" assumption). Recall that $R_i = \text{Cov}(\epsilon_i)$ describes the covariance among observations when we focus on the mean response profile of any *individual;* that is, it is the covariance of the i^{th} subject's deviations from his/her mean response profile, $X_i\beta + Z_i b_i$. Also the b_i are assumed to vary independently from one individual to another, with $b_i \sim N(0, G)$.

When expressed in vector and matrix notation, the linear mixed effects model is

$$Y_i = X_i\beta + Z_i b_i + \epsilon_i,$$

where the vector of regression parameters β (the fixed effects) is assumed to be the same for all individuals and the vector of subject-specific regression coefficients b_i (the random effects) describes how the i^{th} individual's mean response profile deviates from the overall population trend. A distinctive feature of the linear mixed effects model is that it yields simple expressions for both the conditional mean response (for any individual),

$$E(Y_i|b_i) = X_i\beta + Z_i b_i,$$

and the marginal mean response (for the population), averaged over all individuals,

$$E(Y_i) = X_i\beta.$$

Thus the regression coefficients β have population-averaged interpretations in terms of how the mean response changes over time and how these changes relate to covariates.

Finally, the conditional covariance of the responses, given the random effects b_i, is assumed to be a diagonal matrix with

$$\text{Cov}(Y_i|b_i) = \text{Cov}(\epsilon_i) = R_i = \sigma^2 I_{n_i}.$$

On the other hand, the marginal covariance of the responses (the covariance among deviations of the i^{th} individual's responses from the population mean, $X_i\beta$),

$$\begin{aligned} \text{Cov}(Y_i) &= \text{Cov}(Z_i b_i) + \text{Cov}(\epsilon_i) \\ &= Z_i G Z'_i + R_i \\ &= Z_i G Z'_i + \sigma^2 I_{n_i}, \end{aligned}$$

is certainly not diagonal. Thus the introduction of random effects, b_i, in the linear mixed effects model induces correlation (marginally) among the Y_i. This consequence of introducing random effects extends more generally and, in a very natural way, to any generalized linear model with random effects. That is, the correlations among the repeated observations on an individual can be thought of as arising from sharing a set of underlying random effects.

Generalized Linear Mixed Effects Models

Next we consider how the ideas underlying the linear mixed effects model can be extended to generalized linear models. Once again, we can formulate the generalized linear mixed model using a three-part specification:

1. We assume that the conditional distribution of each Y_{ij}, given a $q \times 1$ vector of random effects b_i, belongs to the exponential family of distributions and that $\text{Var}(Y_{ij}|b_j) = v\{E(Y_{ij}|b_i)\}\,\phi$, where $v(\cdot)$ is a known variance function, a function of the conditional mean, $E(Y_{ij}|b_i)$. In addition, given the random effects b_i, it is assumed that the Y_{ij} are independent of one another; this is the "conditional independence" assumption.

2. The conditional mean of Y_{ij} is assumed to depend on fixed and random effects via the following linear predictor:

$$\eta_{ij} = X'_{ij}\beta + Z'_{ij}b_i,$$

with

$$g\{E(Y_{ij}|b_i)\} = \eta_{ij} = X'_{ij}\beta + Z'_{ij}b_i$$

for some known link function, $g(\cdot)$.

3. The random effects are assumed to have some probability distribution. In principle, any multivariate distribution can be assumed for the b_i, in practice, it is common to assume that the b_i have a multivariate normal distribution, with zero mean and $q \times q$ covariance matrix, G. In addition the random effects, b_i, are assumed to be independent of the covariates, X_i.

These three components completely specify a broad class of generalized linear mixed models. Note that in Chapters 12 and 13 we extended generalized linear models by making assumptions about the mean and covariance of Y_i; in particular, we did not make full distributional assumptions about Y_i. In contrast, the three components of a generalized linear mixed model given above completely specify the joint distribution of Y_i. To fix ideas, consider the following four illustrative examples of generalized linear mixed effects models using this three-component specification.

Example 1: Generalized Linear Mixed Model for a Continuous Response

Suppose that Y_{ij} is a continuous response and that it is of interest to relate changes in the mean

response over time to the covariates. An example of a linear mixed effects model for Y_{ij} is given by the following three-part specification:

1. Conditional on a vector of random effects b_i, the Y_{ij} are independent and assumed to have a normal distribution, with $\text{Var}(Y_{ij}|b_i) = \sigma^2$ (i.e., $\phi = \sigma^2$ and the variance does not depend on the conditional mean).

2. The conditional mean of Y_{ij} depends on fixed and random effects via the following linear predictor:

$$\eta_{ij} = X'_{ij}\beta + Z'_{ij}b_i,$$

where $X'_{ij} = Z'_{ij} - (1\ t_{ij})$, with

$$\begin{aligned} E(Y_{ij}|b_i) = \eta_{ij} &= X'_{ij}\beta + Z'_{ij}b_i \\ &= \beta_1 + \beta_2 t_{ij} + b_{1i} + b_{2i}t_{ij} \\ &= (\beta_1 + b_{1i}) + (\beta_2 + b_{2i})t_{ij}. \end{aligned}$$

That is, the conditional mean of Y_{ij} is related to the linear predictor by an identity link function, $\eta_{ij} = g\{E(Y_{ij}|b_i)\} = E(Y_{ij}|b_i)$.

3. The random effects are assumed to have a bivariate normal distribution, with zero mean and 2×2 covariance matrix G.

This illustration of a generalized linear mixed effects model is simply a random intercepts and slopes model and is a special case of the linear mixed effects models considered in Chapter 8. However, when it is viewed as a generalized linear mixed effects model, a much broader class of models for continuous responses can, in principle, be entertained. For example, the mean can be related to the linear predictor by a link function other than the identity. Thus, if the effects of covariates are thought to act multiplicatively on the mean response, a log link function might be more appropriate. Alternatively, the variance can be allowed to depend on any known function of the mean response.

Example 2: Generalized Linear Mixed Model for Counts

Suppose that Y_{ij} is a count. An example of a generalized linear mixed model for Y_{ij} is given by the following three-part specification:

1. Conditional on a vector of random effects b_i, the Y_{ij} are independent and have a Poisson distribution, with $\text{Var}(Y_{ij}|b_i) = E(Y_{ij}|b_i)$, (i.e., $\phi = 1$).

2. The conditional mean of Y_{ij} depends on fixed and random effects via the following linear predictor:

$$\eta_{ij} = X'_{ij}\beta + Z'_{ij}b_i,$$

where $X'_{ij} = Z'_{ij} = (1, t_{ij})$, with

$$\log\{E(Y_{ij}|b_i)\} = \eta_{ij} = X'_{ij}\beta + Z'_{ij}b_i.$$

That is, the conditional mean of Y_{ij} is related to the linear predictor by a log link function; this is an example of a log-linear mixed effects model.

3. The random effects are assumed to have a bivariate normal distribution, with zero mean and 2×2 covariance matrix G.

In this example the model is a log-linear regression model with randomly varying intercepts and slopes. This model posits that there is natural heterogeneity among individuals in both their baseline level and changes in the expected counts over time. Note that in this example the model assumes Poisson variation for the counts, conditional on the random effects. In Section 14.4 we discuss ways to relax this assumption and allow for overdispersion relative to Poisson variability.

Example 3: Generalized Linear Mixed Model for a Binary Response

Suppose that Y_{ij} is a binary response, taking values of 0 or 1. A logistic mixed effects model for Y_{ij} is given by the following three-part specification:

1. Conditional on a single random effect b_i, the Y_{ij} are independent and have a Bernoulli distribution, with $\text{Var}(Y_{ij}|b_i) = E(Y_{ij}|b_i)\{1 - E(Y_{ij}|b_i)\}$, (i.e., $\phi = 1$).

2. The conditional mean of Y_{ij} depends on fixed and random effects via the following linear predictor:

$$\eta_{ij} = X'_{ij}\beta + Z'_{ij}b_i = X'_{ij}\beta + b_i,$$

where $Z_{ij} = 1$ for all $i = 1 \ldots, N$, and $j = 1, \ldots, n_i$, with

$$\log\left\{\frac{\Pr(Y_{ij} = 1|b_i)}{\Pr(Y_{ij} = 0|b_i)}\right\} = \eta_{ij} = X'_{ij}\beta + b_i.$$

That is, the conditional mean of Y_{ij} is related to the linear predictor by a logit link function.

3. The single random effect b_i is assumed to have a univariate normal distribution, with zero mean and variance g_{11}.

In this example the model is a simple logistic regression model with randomly varying intercepts. This model can be considered a discrete data analogue of the "compound symmetry" model discussed in Chapters 7 and 8. The model posits that there is natural heterogeneity in individuals' propensity to respond positively that persists throughout all binary responses obtained on any individual.

Example 4: Generalized Linear Mixed Model for an Ordinal Response

Last, suppose that Y_{ij} is an ordinal response with K categories $(1, \ldots, K)$. A logistic mixed effects model for the *cumulative response probabilities* is given by the following three-part

specification:

1. Conditional on a vector of random effects b_i, the Y_{ij} are independent and have a multinomial distribution (with multinomial covariance determined by the conditional means or conditional response probabilities).

2. The k^{th} cumulative response probability for Y_{ij} depends on fixed and random effects via the following linear predictor:

$$\eta_{ij} = X'_{ij}\beta + Z'_{ij}b_i,$$

where $X'_{ij} = Z'_{ij} = (1, t_{ij})$, with

$$\log\{E(Y_{ij}|b_i)\} = \eta_{ij} = X'_{ij}\beta + Z'_{ij}b_i.$$

That is, the conditional cumulative response probabilities are related to the linear predictor by a logit link function.

3. The random effects are assumed to have a bivariate normal distribution, with zero mean and 2 × 2 covariance matrix G.

In this example the model is a proportional odds regression model with randomly varying intercepts and slopes. This model posits that there is natural heterogeneity among individuals in both their baseline level and changes in the cumulative response probabilities over time.

Although in the first three examples we have chosen canonical link functions to relate the conditional mean of Y_{ij} to η_{ij}, in principle, any suitable link function can be chosen. The four examples of generalized linear mixed effects models considered so far are purely illustrative. They demonstrate how the choices of the three components might differ according to the type of response variable. However, these four examples should not be considered prescriptions for constructing generalized linear mixed effects models.

14.3 INTERPRETATION OF REGRESSION PARAMETERS

Although the introduction of random effects can simply be thought of as a means of accounting for the correlation among longitudinal responses, it has important implications for the interpretation of the regression coefficients in generalized linear mixed models. The regression parameters, β, have somewhat different interpretations than the regression parameters in the marginal models considered in Chapters 12 and 13. In generalized linear mixed models the regression coefficients have subject-specific interpretations. That is, they represent the influence of covariates on a *specific* subject's mean response. In particular, the regression coefficients are interpreted in terms of the effects of *within-subject* changes in covariates on changes in an individual's transformed mean response, while holding the remaining covariates constant. This interpretation for β can be better appreciated by considering the following simple example of a logistic regression model with randomly varying intercepts:

$$\log\left\{\frac{\Pr(Y_{ij}=1|b_i)}{\Pr(Y_{ij}=0|b_i)}\right\}=X'_{ij}\beta+b_i,$$

where b_i is assumed to have a univariate normal distribution, with zero mean and variance g_{11}. The interpretation of a component of β, say β_k, is in terms of changes in any given *individual's* log odds of response for a unit *within-subject* change in the corresponding covariate, say X_{ijk}. That is, when X_{ijk} takes on some value x, the log odds of a positive response is

$$\log\left\{\frac{\Pr(Y_{ij}=1|b_i,X_{ij1},\ldots,X_{ijk}=x,\ldots,X_{ijp})}{\Pr(Y_{ij}=0|b_i,X_{ij1},\ldots,X_{ijk}=x,\ldots,X_{ijp})}\right\}$$

$$=b_i+\beta_1X_{ij1}+\cdots+\beta_kx+\cdots+\beta_pX_{ijp}.$$

Similarly, when X_{ijk} now takes on some value $x + 1$, but all other covariate values are held fixed, the log odds of a positive response is

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

$$=b_i+\beta_1X_{ij1}+\cdots+\beta_k(x+1)+\cdots+\beta_pX_{ijp}.$$

Thus, for any individual, the log odds of a positive response for a unit increase in X_{ijk} is simply β_k. (obtained by subtracting the former log odds from the latter). That is, β_k measures the change in the log odds of response per unit increase in X_{ijk}, for any given individual having some unobservable underlying propensity to respond positively, b_i. Also note that this subject-specific interpretation of β_k is far more natural for a covariate that varies within an individual (i.e., a within-subject or time-varying covariate). In that case, β_k has interpretation as the change in an individual's log odds of response for a unit increase in X_{ijk}, while holding that individual's covariates fixed. Because the components of the fixed effects, β, have interpretations that depend on holding b_i, the i^{th} individual's random effect, fixed, they are often referred to as *subject-specific* regression coefficients. As a result generalized linear mixed models are most useful when the main scientific objective is to make inferences about individuals rather than the population averages; the population averages are the targets of inference in marginal models.

When there are between-subject (or time-invariant) covariates in the model, the interpretation of the corresponding components of β is somewhat less transparent and potentially misleading. If X_{ijk} is a between-subject covariate (e.g., gender, treatment, or exposure group), it is misleading to give it a subject-specific interpretation in terms of the change in the log odds of response for a unit increase in X_{ijk} since there are simply no data that provide any information about such an effect. In a sense, this interpretation of β_k would be a complete extrapolation beyond the observed data. Problems of interpretation with a between-subject covariate arise because a change in the value of the covariate requires also a change in the index i of X_{ijk} to, say, X_{ijk} (for $i \neq i'$). However, β_k then becomes confounded with $b_i - b_{i'}$, the difference between the unobserved random effects for the two individuals indexed by i and

i', respectively. To circumvent this problem, we must assume that $b_i = b_{i'}$. That is, β_k must be given an interpretation in terms of a contrast of the log odds of response for two different individuals who happen to have the same value for the unobserved random effect (i.e., $b_i = b_{i'}$), but who differ by one unit in the covariate X_{ijk} (e.g., one individual is exposed, the other is unexposed, or one individual is randomized to treatment, the other to placebo). Because the random effects are latent, unobserved variables, this effect of the covariate is not directly observable from the data at hand. As a result it is somewhat unclear where the information about β_k is obtained when X_{ijk} is a between-subject covariate. In fact, the estimate of βk is based on comparisons between subjects, not comparisons within subjects, and depends on assumptions about the distribution of the random effects. Because the estimate of β_k is a model-based extrapolation, it may be more sensitive to assumptions concerning the random effects distribution that are difficult to check from the data at hand.

The distinction between the regression coefficients in generalized linear mixed models and marginal models is best understood in terms of the targets of inference. In generalized linear mixed models the target of inference is the individual, since the regression coefficients have interpretation in terms of contrasts of the transformed conditional means,

$$E(Y_{ij}|X_{ij}, b_i).$$

By conditioning on the unobserved random effects, b_i, the target of inference has shifted from the population to the individual. In contrast, in marginal models the target of inference is the population, since the regression coefficients in marginal models have interpretation in terms of contrasts of the transformed population means,

$$E(Y_{ij}|X_{ij}),$$

and describe how the average response varies across different subsets of the study population defined by the covariates (e.g., gender, exposure groups, treatment groups).

Note that the population means,

$$E(Y_{ij}|X_{ij}),$$

in marginal models are averaged over the natural individual-to-individual heterogeneity in the study population (as well as over the measurement or sampling variability in the response).

For the special case where an identity link function has been adopted (i.e., for the special case of linear mixed effects models), the regression coefficients in the model for the conditional means,

$$E(Y_{ij}|X_{ij}, b_i) = X'_{ij}\beta + Z'_{ij}b_i,$$

also happen to have interpretation in terms of the population means, since

$$\begin{aligned} E(Y_{ij}|X_{ij}) &= E\{E(Y_{ij}|X_{ij}, b_i)\} \\ &= E\left(X'_{ij}\beta + Z'_{ij}b_i\right) \\ &= X'_{ij}\beta + Z'_{ij}E(b_i) \\ &= X'_{ij}\beta \end{aligned}$$

when averaged over all individuals in the study population. That is, averaged over the

distribution of the random effects, the population means also follow a linear model with regression coefficients β. However, in general, for the non-linear link functions usually adopted for discrete data, this relationship no longer holds. That is, if

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

where $g(\cdot)$ is a non-linear link function (e.g., logit$(\cdot)$ or log$(\cdot)$), then

$$g\{E(Y_{ij}|X_{ij})\} \neq X'_{ij}\beta$$

for all β, when averaged over the distribution of the random effects. Thus, for non-linear (or non-identity) link functions, the regression coefficients in generalized linear mixed effects and marginal models have quite distinct interpretations, and these two classes of regression models have different targets of inference. In short, these two classes of models address different scientific questions. Marginal models address scientific questions that are concerned with changes in the (transformed) mean response over time in the study population, and the impact of covariates on these changes. In contrast, generalized linear mixed effects models address scientific questions that are concerned with changes in the mean response for any individual, and the impact of covariates on these changes.

The following simple illustration helps highlight the main distinction between the regression coefficients in marginal and generalized linear mixed models. Consider the hypothetical data presented in Table 14.1. It displays the (usually unobserved) true propensity for disease, Pr $(Y_{ij} = 1|b_i)$, for three individuals measured at baseline and following treatment with a new drug intended to reduce the risk of disease. The three individuals are discernibly different in terms of their underlying propensity for disease at baseline. This heterogeneity can be expressed in terms of random effects, b_i. In a sense, individuals A, B, and C have "high," "medium," and "low" underlying risk for disease. Also let us assume that the entire population is composed of an equal number of individuals that fall into these three distinct risk groups. Based on this assumption, the final row of Table 14.1 contains the population averages (obtained as equally weighted means).

Table 14.1 Hypothetical data on the true propensity for disease, at baseline and post-baseline, for three individuals with heterogeneous propensities for disease.

Individual	Baseline	Post-Baseline	Difference	Log(Odds Ratio)
A	0.80	0.67	−0.13	−0.68
B	0.50	0.33	−0.17	−0.71
C	0.20	0.11	−0.09	−0.70
Population Average	0.50	0.37	−0.13	

If we considered a linear model for the probability of disease, the risk difference, or difference between the probabilities of disease at baseline and post-baseline, provides a

measure of the effectiveness of the new drug. These differences (post-baseline – baseline) are displayed in the fourth column of Table 14.1 and vary from −0.09 to −0.17. These can be thought of as subject-specific effects of the drug. We can then consider two possible ways to produce a single-number summary of the effectiveness of the drug. The first summary can be obtained by taking the average of the subject-specific effects (as a single-number summary of the subject-specific effects),

$$\frac{-0.13 - 0.17 - 0.09}{3} = -0.13.$$

Alternatively, the average propensity for disease at baseline (0.5) can be compared to the average propensity for disease post-baseline (0.37). The latter can be thought of as a contrast of population averages and this comparison also yields

$$(0.37 - 0.50) = -0.13.$$

That is, the difference (post-baseline – baseline) between the population averages is identical to the population average of the individual-specific differences. As such, the "difference of the averages" is equal to the "average of the differences." This simple numerical illustration confirms the remark that was made earlier about how the fixed effects regression coefficients in the linear mixed effects model (with identity link function) also happen to have interpretation in terms of population averages.

Let us consider a non-linear function of the propensity for disease (this corresponds to adopting a non-linear link function for the probability of disease). The log odds ratio provides a natural measure of the effectiveness of the drug in reducing the risk of disease from baseline. The log odds ratios (comparing the odds of disease post-baseline to the odds of disease at baseline) for individuals A, B, and C are displayed in the fifth column of Table 14.1. For example, the log odds ratio for individual A is

$$\log\left\{\frac{0.67/(1-0.67)}{0.8/(1-0.8)}\right\} = -0.68.$$

The log odds ratios are all very similar in magnitude, ranging from −0.68 to −0.71. Once again, these can be thought of as subject-specific effects of the drug. We can then consider two possible ways to produce a single-number summary of the effectiveness of the drug. The first can be obtained by taking the average of the subject-specific effects (as a single-number summary of the subject-specific effects),

$$\frac{-0.68 - 0.71 - 0.70}{3} = -0.697.$$

This indicates that the effect of the drug on any individual is to approximately halve the odds of disease (since $e^{-0.697} \approx 0.5$). Alternatively, the effectiveness of the drug can be assessed by comparing the log odds of disease in the population at baseline, log(0.5/0.5) = 0, with the log odds of disease in the population post-baseline, log(0.37/0.63) = −0.532. The latter can be thought of as a contrast of population log odds, and this comparison yields a measure of effect, −0.532, which is approximately 25% smaller than the summary of the subject-specific effect, −0.697. That is, the comparison of population log odds results in a discernibly different measure of the effectiveness of the drug than was found in the comparison of subject-specific

effects. With a non-linear function of the propensity for disease, a "non-linear contrast of the averages" is not equal to the "average of the non-linear contrasts." This highlights the main differences between these two approaches when a non-linear link function is adopted.

For this simple numerical illustration the reader may be curious about which of the two summary statistics, −0.697 or −0.532, provides the more *realistic* estimate of the effectiveness of the drug. The answer is that they both do, although they address somewhat different scientific questions. The estimate of −0.697 provides a measure of the expected change in the odds of disease for any individual treated with the drug. That is, there is an approximately 50% reduction in the odds of disease (since $1 - e^{-0.697} \approx 0.5$) for any individual treated with the drug. This is the estimate that will be of most interest to an individual and his/her physician in the physician-patient context. On the other hand, the estimate of −0.532 provides a measure of the expected change in prevalence of disease in the study population if everyone were to be treated with the drug. That is, there would be an expected reduction in the odds of disease in the population of approximately 40% (since $1 - e^{-0.532} \approx 0.4$). This is the estimate that will be of most interest to public health researchers who are considering the potential benefits of the drug for the study population as a whole.

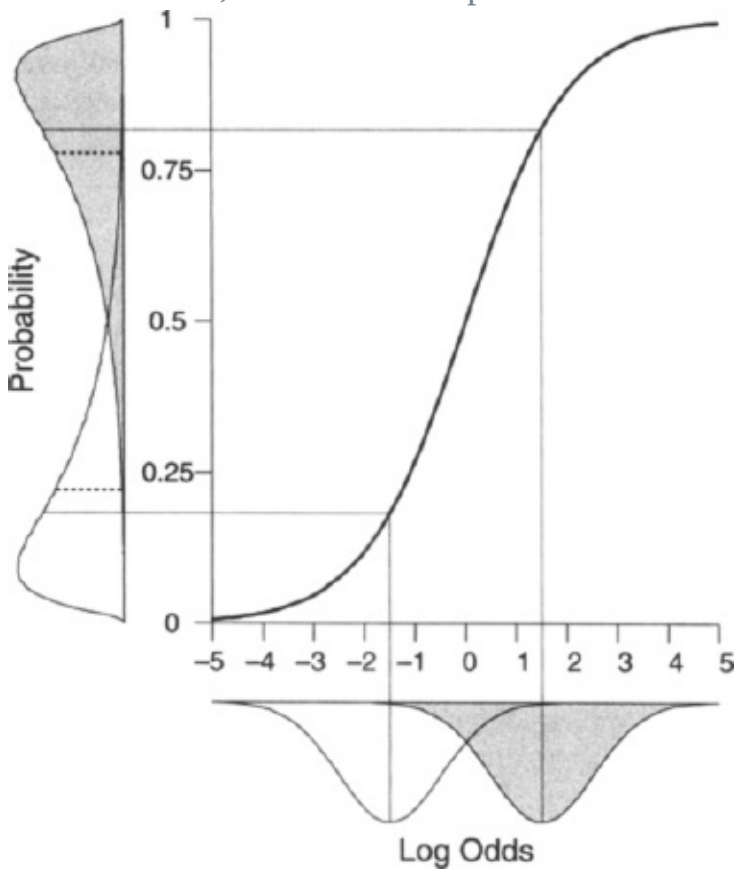
To provide further intuition for why the regression coefficients in generalized linear mixed models and marginal models differ, consider the following example of a logistic regression model, with normally distributed random intercepts:

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

where $t_{ij} = 0$ at baseline and $t_{ij} = 1$ post-baseline. Similar to the illustration in Table 14.1, we assume that individuals are measured at baseline and following treatment with a new drug intended to reduce the risk of disease. Individuals in the population differ in terms of their underlying propensity for disease at baseline; this heterogeneity is expressed in terms of the random effect, b_i. For a "typical" individual from the population (where a "typical" individual is one with unobserved random effect $b_i = 0$, the mean and median of the distribution of b_i), the log odds of disease at baseline is β^*_1; the log odds of disease following treatment with the new drug is $\beta^*_1 + \beta^*_2$.

The log odds of disease at baseline and post-baseline are displayed in Figure 14.1, for the case where $\beta^*_1 = 1.5$, $\beta^*_2 = -3.0$, and $\text{Var}(b_i) = 1.0$. At baseline the log odds has a normal distribution with mean and median of 1.5. (See the shaded density for the log odds in Figure 14.1.) From Figure 14.1 it is clear that there is heterogeneity in risk of disease, with approximately 95% of individuals having a baseline log odds of disease that varies from −0.46 to 3.46 (or $1.5 \pm 1.96\ \sqrt{1.0}$). When the risk of disease is translated from the log odds scale to the probability scale, the baseline probability of disease for a typical individual from the population is approximately 0.82. Furthermore approximately 95% of individuals have a baseline probability of disease that varies from 0.39 to 0.97.

Fig. 14.1 Subject-specific probability of disease as a function of subject-specific log odds of disease at baseline (shaded densities) and post-baseline (unshaded densities). Solid lines represent medians of the distributions; dashed lines represent means of the distributions.



From Figure 14.1 it is transparent that the symmetric, normal distribution for the baseline log odds does not translate into a corresponding symmetric, normal distribution for the probability of disease. Instead, the subject-specific probabilities of disease have a negatively skewed distribution with a median, but not mean, of 0.82. (See solid line in Figure 14.1.) Because of the skewness the mean of the distribution of subject-specific baseline probabilities is pulled toward the tail and is equal to 0.7785. (See dashed line in Figure 14.1.) Thus the 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), due to the non-linearity of the relationship between subject-specific probabilities and log odds.

Similarly the log odds of disease post-baseline has a normal distribution with mean and median of −1.5 (see the unshaded density for the log odds in Figure 14.1); approximately 95% of individuals have a post-baseline log odds of disease that varies from −3.46 to 0.46 (or −1.5 ± 1.96 √1.0.) This shift in the log odds corresponds to a 20-fold decrease (since $1 - e^{-3.0} \approx 0.95$) in the subject-specific odds of disease. When the risk of disease is translated from the log odds scale to the probability scale, the post-baseline probability of disease for a typical individual from the population is approximately 0.18. Furthermore approximately 95% of individuals have a post-baseline probability of disease that varies from 0.03 to 0.61. From Figure 14.1 it is apparent that the subject-specific post-baseline probabilities of disease have a positively skewed distribution with median, but not mean, of 0.18. (See solid line in Figure

14.1.) Because of the skewness, the mean is pulled toward the tail and is equal to 0.2215. (See dashed line in Figure 14.1.)

Figure 14.1 highlights how the effect of treatment on the log odds of disease for a typical individual from the population, $\beta^*_2 = -3.0$, is not the same as the contrast of population log odds. The latter is what is estimated in a marginal model, say

$$\text{logit}\{E(Y_{ij})\} = \beta_1 + \beta_2 t_{ij},$$

and can be obtained by comparing the log odds of disease in the population at baseline, log(0.7785/0.2215) = 1.257, with the log odds of disease in the population post-baseline, log(0.2215/0.7785) = −1.257. This yields a population-averaged measure of effect, $\beta_2 = -2.514$, which is approximately 15% smaller than β^*_2, the subject-specific effect of treatment.

14.4 OVERDISPERSION

In Chapter 11 we mentioned that overdispersion is almost the rule, not the exception, with count data. This can be potentially problematic for a mixed effects model that assumes Poisson variation for counts, conditional on the random effects; similar considerations apply to mixed effects models for binomial counts of the number of successes. Although the inclusion of random coefficients (e.g., random intercepts and slopes) induces overdispersion marginally (when averaged over the distribution of the random effects), the model nonetheless assumes Poisson variation conditional on these subject-specific effects and the covariates. One approach for relaxing the Poisson variability assumption is to extend the model to incorporate an extra source of variability in the subject-specific expected counts (or rates). For example, assuming a log-link function, the model can be extended as follows:

$$\log E(Y_{ij}|b_i, e_{ij}) = \log(T_{ij}) + X'_{ij}\beta + Z'_{ij}b_i + e_{ij},$$

where e_{ij} is an additional random effect that varies over both individuals and measurement occasions. Specifically, if a gamma distribution is assumed for the exponentiated errors, $\exp(e_{ij})$, with mean of 1 and variance θ, then it can be shown that the conditional mean of Y_{ij} (conditional only on b_i and the covariates) is given by

$$\log E(Y_{ij}|b_i) = \log(T_{ij}) + X'_{ij}\beta + Z'_{ij}b_i.$$

That is, the model for the conditional mean is unchanged. However, the inclusion of the random errors implies that the corresponding conditional variance of Y_{ij} is

$$\text{Var}(Y_{ij}|b_i) = E(Y_{ij}|b_i) + \theta\{E(Y_{ij}|b_i)\}^2,$$

which is larger than the conditional mean, $E(Y_{ij}|b_i)$, when $\theta > 0$ thereby allowing for overdispersion. As was mentioned in Chapter 11, one advantage of the assumption of gamma errors is that the distribution of the counts (averaged over the distribution of these errors) has a negative binomial distribution. This makes maximum likelihood (ML) estimation of the model parameters more straightforward. That is, the model with gamma errors can be fit directly to the counts by assuming that they have a conditional negative binomial rather than a Poisson

distribution, given the random effects, b_i, and the covariates.

Finally, we note that a very similar model can be specified by replacing the gamma distribution for the errors with a normal distribution. This leads to an equivalent model for the conditional variance of Y_i (conditional only on b_i and the covariates) that allows for overdispersion; also, with the exception of the intercept, the model for the conditional mean is unchanged. However, relative to a model with gamma errors, the model with normal errors is more computationally challenging to fit because the conditional distribution of Y_i (conditional only on b_i and the covariates, but averaged over the distribution of e_{ij}) does not have a simple closed-form expression. Consequently ML estimation requires numerical integration over the distributions of random effects at two distinct levels, the level of the individual (b_i) and the level of measurements within individuals (e_{ij}). This is an example of a *multilevel* generalized linear mixed model, a class of models that is discussed in Chapter 22.

14.5 ESTIMATION AND INFERENCE

Unlike marginal models, where specification of the marginal means, variances, and pairwise associations does not fully specify the joint distribution of the vector of responses, with generalized linear mixed effects models the joint distributions of both the vector of responses and the vector of random effects are fully specified. As a result we can base estimation and inference on the likelihood function. This has important implications when there are missing data in the response variables. Specifically, conventional likelihood-based analyses of the incomplete data yield valid inferences when data are missing at random (MAR), provided that the likelihood has been correctly specified. Therefore, unlike GEE estimation of marginal models, which requires the stronger assumption that data are missing completely at random (MCAR), likelihood-based estimation of generalized linear models provides valid analyses when data are missing at random (MAR) but not MCAR. Statistical issues concerning the potential impact of missing data on analysis are discussed in detail in Chapters 17 and 18.

In this section we briefly describe maximum likelihood estimation of the fixed effects, β, and the random effects covariance parameters, G. We also discuss prediction of the random effects. Although ML estimation is far less straightforward for generalized linear mixed effects models than it is for the linear mixed effects models considered in Chapter 8, a variety of numerical methods for maximizing the likelihood have recently been implemented in software packages (e.g., PROC GLIMMIX in SAS, the `glmer` function in the `lme4` package in R, and the `xtmelogit` and `xtmepoisson` commands in Stata). In this section we discuss the use of quadrature methods; quadrature methods are simply numerical methods that can be made highly accurate, albeit with substantial computational overhead. In Chapter 15 we discuss two alternative methods of estimation and inference for generalized linear mixed effects models that are far less computationally demanding.

Given the three-part specification of a generalized linear mixed effects model, the joint

probability for Y_i and b_i can be expressed as

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

where

$$f(Y_i|b_i) = f(Y_{i1}|b_i)f(Y_{i2}|b_i)\cdots f(Y_{in_i}|b_i)$$

under the "conditional independence" assumption. Furthermore $f(Y_{ij}|b_i)$ is assumed to have an exponential family distribution, whereas $f(b_i)$ is assumed to have a multivariate normal distribution, with zero mean and covariance matrix G. Since the random effects b_i are unobserved, inference about β and G is based on the so-called marginal or integrated likelihood function,

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

obtained by integrating out or averaging over the distribution of the unobserved random effects, b_i. An integral appears in the marginal likelihood, and this integral denotes the averaging over the distribution of b_i. Since the marginal likelihood has averaged over the b_i, the resulting marginal likelihood function depends only on β, ϕ, and G. That is, the marginal likelihood depends on the covariance of b_i but not on the unobserved b_i.

The ML estimates of β, ϕ, and G are simply those values of β, ϕ, and G that maximize this likelihood function. However, unlike the case of the linear mixed effects model, there are no simple, closed-form solutions. Instead, numerical integration techniques are required for maximizing the likelihood function. Numerical integration techniques, known as Gaussian quadrature, simply approximate the integral appearing in the marginal likelihood function as a weighted sum,

$$L(\beta, \phi, G) \approx \prod_{i=1}^{N} \sum_{k=1}^{K} f(Y_i|b_i = v_k)w_k,$$

where the known quadrature points (the weights, w_k, and the evaluation points, v_k) are chosen to provide an accurate numerical approximation. The number of quadrature points determines the degree of accuracy of the approximation involved in replacing the integral by a weighted sum. The number of quadrature points, K, can be increased or decreased as desired. The more quadrature points used, the more accurate the approximation will be. However, the computational burden also increases with the number of quadrature points, and grows exponentially with the number of random effects. As a result there is a trade-off that must be carefully balanced between the computational burden of quadrature methods and the desired accuracy of the results. In general, computational time is negligible when compared to the time expended in collecting longitudinal data. As a result we recommend increasing the number of quadrature points until there is evidence that all parameter estimates and standard errors are stable.

Given ML estimates of β, ϕ, and G, the random effects b_i for any particular subject can be predicted as follows:

$$\widehat{b}_i = E(b_i|Y_i; \widehat{\beta}, \widehat{\phi}, \widehat{G}).$$

That is, the predicted random effects for the i^{th} subject are simply "estimated" as the conditional mean of b_i given Y_i (and $\widehat{\beta}, \widehat{\phi}, \widehat{G}$); this coincides with the empirical Bayes or BLUP used for prediction of b_i in Chapter 8. Note that $E(b_i|Y_i; \widehat{\beta}, \widehat{\phi}, \widehat{G})$, being a conditional mean, also requires integrating (or averaging) over the distribution of the unobserved random effects, b_i. As a result simple analytic solutions for b_i are rarely available, and numerical integration techniques must also be used here.

Finally, in our discussion of generalized linear mixed models we have assumed the distribution of the random effects is multivariate normal. Distributional assumptions about the random effects are difficult to assess from the data at hand. In particular, when the response variable is discrete, the data often contain little information to distinguish between competing distributions for the random effects. As mentioned at the end of Section 10.5, predictions of the random effects (i.e., the empirical BLUPs) are known to be heavily influenced by the normal distribution assumption for the random effects. Because the distribution of the empirical BLUPs inherits much of its shape from the assumed distribution for the random effects, histograms and normal quantile plots of the empirical BLUPs cannot be relied on for assessing the adequacy of the normal distribution assumption for the random effects. However, in general, the estimates of the fixed effects are much less sensitive to misspecification of the random effects distribution. That is, assuming that the random effects have a normal distribution when the true distribution of the random effects is non-normal (e.g., a skewed distribution) does not produce discernibly biased estimates of the fixed effects. The fixed effects estimates are, however, sensitive to a different kind of misspecification of the random effects distribution. When the assumption that the random effects are independent of the covariates, X_i, does not hold, the estimates of the fixed effects can be severely biased. This type of misspecification might arise, for example, in a study where one exposure group is more heterogeneous than another (i.e., the variance of the random effects depends on the exposure group).

14.6 A NOTE ON CONDITIONAL MAXIMUM LIKELIHOOD

In the previous section we described maximum likelihood estimation under the assumption that the distribution of the random effects is normal. There is an alternative approach to estimation of the fixed effects, β, that considers the b_i to be an additional set of fixed parameters. For example, instead of introducing randomly varying intercepts for each individual in the logistic regression model,

$$\log\left\{\frac{\Pr(Y_{ij}=1|b_i)}{\Pr(Y_{ij}=0|b_i)}\right\} = X'_{ij}\beta + b_i,$$

we can incorporate fixed intercepts via the inclusion of indicator variables for each individual (in addition to the usual covariates of interest),

$$\log\left\{\frac{\Pr(Y_{ij}=1|\alpha_i)}{\Pr(Y_{ij}=0|\alpha_i)}\right\} = X'_{ij}\beta + \alpha_i,$$

where the α_i denote *fixed effects* representing stable (i.e., time-invariant) characteristics of individuals that are not otherwise accounted for by the covariates in the model. Although such a model is no longer a generalized linear mixed effects model (GLMM), the regression parameters, β, do have similar interpretations in terms of *subject-specific* effects of the covariates. This model is completely analogous to the linear *fixed effects* model described in Chapter 9. However, unlike in the linear fixed effects model, maximum likelihood estimation of the logistic regression parameters breaks down when the number of repeated measurements on each individual is relatively small. In general, optimal properties of maximum likelihood estimation require that the sample size, N, is large relative to the number of model parameters to be estimated. Herein lies the problem with maximum likelihood estimation in this setting: the number of fixed intercepts to be estimated grows as N does, while the amount of information about each parameter remains fixed. This problem is well-recognized in the statistical literature where it is referred to as the "incidental parameters problem." Fortunately, there is a variant of maximum likelihood, known as *conditional* maximum likelihood, that can be used for inference in this setting.

The main idea behind the conditional likelihood approach is to eliminate the fixed subject-specific intercepts, α_i, by constructing a likelihood that is conditional on the "sufficient statistics" for these parameters. In doing so, the "incidental parameters problem" is circumvented because the conditional likelihood no longer depends on the subject-specific intercepts. In addition the conditional likelihood has a relatively simple closed-form expression. In all other respects, though, the conditional likelihood can be used for estimation of β in the usual way; that is, the conditional ML estimates of β are those values maximizing the conditional likelihood. For example, in the logistic regression model with subject-specific intercepts, the sufficient statistic for α_i is the total number of successes for each subject, say $S_i = \sum_{j=1}^{n_i} Y_{ij}$ (for i = 1, …, N). As an aside, note that conditioning on the total number of successes provides justification for the use of data on discordant pairs only in the familiar McNemar's test for paired (n_i = 2) binary data; in the general matched pair design, the responses for the two members of the pair are similar to repeated measures. Conditional maximum likelihood can be used for estimation of β in any generalized linear model with fixed subject-specific effects provided that a canonical link function is adopted and the model is restricted to having subject-specific intercepts only. We note that the fixed effects estimator of within-subject effects in the linear fixed effects model discussed in Chapter 9 (Section 9.2) can be derived as the conditional ML estimator.

There are two potential advantages of the conditional likelihood approach. First, the conditional likelihood method makes no assumptions about the distribution of the subject-specific effects. In contrast, GLMMs assume that the distribution of b_i is normal, and

inferences may be sensitive to any misspecification of the random effects distribution. Moreover it is difficult to test the validity of the distributional assumption for b_i in GLMMs. Second, similar to the linear fixed effects model, the within-subject estimator of β from the conditional likelihood method is not subject to confounding by omission of between-subject covariates. Therefore the conditional ML estimators are less sensitive to model misspecification than the GLMM estimators of β. Recall that the latter require that the random effects, b_i, be independent of the covariates, X_i. However, these two advantages are offset by the following limitations. First, the conditional likelihood approach cannot be generalized to more complex models with both subject-specific intercepts and subject-specific slopes. Second, the method can only be applied in models that assume the canonical link function (e.g., logistic regression models for Bernoulli or binomial outcomes, loglinear regression models for Poisson count data); for other (non-canonical) link functions, the "sufficient statistics" for the subject-specific effects do not exist. Third, conditional maximum likelihood does not permit estimation of the subject-specific effects or their variability. Fourth, when data are incomplete, the standard conditional ML estimator is biased if data are missing at random (MAR) but not missing completely at random (MCAR); see Section 4.3 for the definitions of, and the distinction between, MCAR and MAR. That is, the conditional ML estimator yields valid inferences when either the data are complete or any missing data are MCAR; when data are MAR, it can produce badly biased estimates of the effects of time-varying covariates. Fifth, the conditional ML estimator is usually less efficient than the GLMM estimator of β; this is the price to be paid for making fewer assumptions. Finally, the conditional likelihood method can only estimate the effects of within-subject or time-varying covariates. The effects of time-invariant covariates cannot be estimated by conditional ML; this is the same limitation that was noted for linear fixed effects models in Chapter 9. This restriction to estimating only the effects of time-varying covariates is unappealing when there is scientific interest in the effects of both time-varying and time-invariant effects. Indeed, in many longitudinal studies the primary covariates of scientific interest are time-invariant, such as fixed treatment or exposure groups, or various background characteristics of individuals (e.g., gender or socioeconomic status).

In summary, conditional inference based on generalized linear models that treat the subject-specific effects as fixed rather than random can be regarded as more robust than maximum likelihood estimation under the assumptions of a corresponding GLMM. The validity of the conditional ML estimators of β requires fewer assumptions. On the other hand, the potential use of conditional inference for longitudinal data analysis is far more limited, restricted to the effects of within-subject covariates. Because the validity of the conditional ML estimators requires fewer assumptions, a diagnostic check on the assumptions for the random effects in GLMMs can be based on a comparison of the estimates of the within-subject effects obtained from these two approaches. Specifically, when the assumptions for the random effects in GLMMs hold, the differences between these estimates should be small; discernible differences, beyond those due to sampling variation, suggest that the assumptions regarding the random effects are not valid. This comparison can be formalized by constructing a statistical test based on the standardized differences between the two sets of estimates for the within-

subject effects; this test, developed for GLMMs by Tchetgen and Coull (2006), is completely analogous to the "Hausman test" discussed in Chapter 9. We illustrate this strategy of using the conditional ML estimates as a diagnostic assessment of the assumptions for the random effects in GLMMs in Section 16.5.

14.7 CASE STUDIES

In this section we illustrate the main ideas presented earlier by considering GLMMs for analyzing longitudinal data from three studies. The first illustration considers a logistic regression model, with random effects, for analyzing data on amenorrhea from a randomized clinical trial of contracepting women. The second illustration considers a Poisson regression model, with random effects, for analyzing count data on epileptic seizures from a clinical trial of the anti-epileptic drug, progabide. The third example considers a proportional odds model, with random effects, for analyzing ordinal data from a clinical trial of patients with rheumatoid arthritis.

Clinical Trial of Contracepting Women

The first example is from a longitudinal clinical trial of contracepting women reported by Machin et al. (1988). In this trial women received an injection of either 100 mg or 150 mg of depot-medroxyprogesterone acetate (DMPA) on the day of randomization and three additional injections at 90-day intervals. There was a final follow-up visit 90 days after the fourth injection, that is, one year after the first injection. Throughout the study each woman completed a menstrual diary that recorded any vaginal bleeding pattern disturbances. The diary data were used to determine whether a woman experienced amenorrhea, the absence of menstrual bleeding for a specified number of days.

A total of 1151 women completed the menstrual diaries, and the diary data were used to generate a binary sequence for each woman, according to whether she had experienced amenorrhea in the four successive three-month intervals. A feature of this clinical trial is that there was substantial dropout. More than one-third of the women dropped out before the completion of the trial; 17% dropped out after receiving only one injection of DMPA, 13% dropped out after receiving only two injections, and 7% dropped out after receiving three injections. For women who dropped out before the end of the 90-day injection interval, a determination of whether they experienced amenorrhea was made, on a proportionate basis, using their existing menstrual diary data for that interval. Statistical issues concerning the potential impact of missing data on the analysis are discussed in Chapters 17 and 18.

In clinical trials of modern hormonal contraceptives, pregnancy is exceedingly rare (and would be regarded as a failure of the contraceptive method), and this is not the main outcome of interest in this study (Machin et al., 1988). The outcome of interest is instead a binary response indicating whether a woman experienced amenorrhea in the four successive three-

month intervals. The goal of the analyses presented here is to determine subject-specific changes in the risk of amenorrhea over the course of the study (12 months), and the influence of dosage of DMPA on changes in a woman's risk of amenorrhea. Of note, the treatment covariate (high versus low dosage of DMPA) is time-invariant.

Let $Y_{ij} = 1$ if the i^{th} woman experienced amenorrhea in the j^{th} injection interval ($j = 1, \ldots, 4$), and $Y_{ij} = 0$ otherwise. The following mixed effects logistic regression model for Y_{ij} was fit to the data:

$$\begin{aligned}\text{logit}\{E(Y_{ij}|b_i)\} &= \beta_1 + \beta_2\,\text{Time}_{ij} + \beta_3\,\text{Time}_{ij}^2 + \beta_4\,\text{Dose}_i \times \text{Time}_{ij} \\ &\quad + \beta_5\,\text{Dose}_i \times \text{Time}_{ij}^2 + b_i,\end{aligned}$$

where Time = 1, 2, 3, 4 for the four consecutive 90-day injection intervals, and Dose = 1 if randomized to 150 mg of DMPA, and Dose = 0 otherwise. Note that there is no baseline measure of amenorrhea prior to receiving the first contraceptive injection. However, due to randomization, we assume that the baseline risk (at Time = 0) is the same in both dosage groups and omit a main effect of dose from the model.

Given b_i, it is assumed that the Y_{ij} are independent and have a Bernoulli distribution, with $\text{Var}(Y_{ij}|b_i) = E(Y_{ij}|b_i)\,\{1 - E(Y_{ij}|b_i)\}$, and $\phi = 1$. Finally, we assume that the single random effect b_i has a univariate normal distribution, with zero mean and variance σ^2_b, $b_i \sim N(0, \sigma^2_b)$. This mixed effects model posits natural heterogeneity in women's propensity or underlying risk of amenorrhea that persists throughout all binary responses obtained over the duration of the study.

The ML estimates of the regression parameters for this model are presented in Table 14.2. These results provide evidence that the subject-specific log odds of amenorrhea increases over the 12 months of the trial, and that subject-specific changes in the risk of amenorrhea depend on the dose of DMPA. For example, for a woman assigned to the low dose of DMPA, the log odds of amenorrhea increases approximately linearly, with an increase in the log odds of 1.09 (or $1.1332 - 0.0419$) at 3 months, 2.10 (or $2 \times 1.1332 - 4 \times 0.0419$) at 6 months, 3.02 (or $3 \times 1.1332 - 9 \times 0.0419$) at 9 months, and 3.86 (or $4 \times 1.1332 - 16 \times 0.0419$) at 12 months. These increases in risk correspond to odds ratios of 3.0 (or $e^{1.09}$), 8.2 (or $e^{2.10}$), 20.5 (or $e^{3.02}$), and 47.5 (or $e^{3.86}$) at 3, 6, 9, and 12 months, respectively. On the other hand, for a woman assigned to the high dose of DMPA, the log odds of amenorrhea increases quadratically, with an increase of 1.55 at 3 months, 2.79 at 6 months, 3.73 at 9 months, and 4.37 at 12 months. That is, the early linear trend shows a decline toward the end. These increases in risk correspond to odds ratios of 4.7 (or $e^{1.55}$), 16.3 (or $e^{2.79}$), 41.7 (or $e^{3.73}$), and 79.0 (or $e^{4.37}$) at 3, 6, 9, and 12 months, respectively. For both groups, all of these increases in the odds ratios are significant at the 0.05 level.

Table 14.2 Parameter estimates and standard errors from a mixed effects logistic regression model, with randomly varying intercepts, for the amenorrhea data.

Variable	Estimate	SE	Z
Intercept	−3.8057	0.3050	−12.48
Time	1.1332	0.2682	4.22
Time^2	−0.0419	0.0548	−0.76
Dose × Time	0.5644	0.1922	2.94
Dose × Time^2	−0.1096	0.0496	−2.21
σ_b^2	5.0646	0.5840	

Note: ML estimation is based on 50-point adaptive Gaussian quadrature.

Because treatment (low versus high dose of DMPA) is a between-subject variable, the interpretation of the fixed effects for the dose × time interactions is more difficult. The interaction effects must be given an interpretation in terms of a contrast of the increases in log odds of amenorrhea (or the odds ratio) for two different women who happen to have the same underlying risk of experiencing amenorrhea prior to randomization but who differ in terms of dose (i.e., one is assigned to low dose and the other to high dose). From the estimates of the fixed effects in Table 14.2, the ratio of the increased odds of amenorrhea at 12 months for a woman assigned to the high dose, versus another woman with the same risk of amenorrhea prior to randomization who was assigned to the low dose, is 1.66 (or $e^{4.37-3.86}$) with 95% confidence interval: 1.03 to 2.66.

The estimated variance of the random intercepts is relatively large, $\hat{g}_{11} = 5.065$. This implies that there is substantial variability in the propensity to experience amenorrhea, since approximately 95% of the women have a baseline risk of amenorrhea that varies from

$$\frac{\exp(-3.8057 - 1.96\sqrt{5.0646})}{1 + \exp(-3.8057 - 1.96\sqrt{5.0646})}$$

to

$$\frac{\exp(-3.8057 + 1.96\sqrt{5.0646})}{1 + \exp(-3.8057 + 1.96\sqrt{5.0646})},$$

or 0.03% to 64.68%. Alternatively, we can interpret $\hat{g}_{11}$ by appealing to the notion of a latent variable distribution (see Chapter 11, Section 11.3). That is, we can assume a linear mixed effects model for the latent variable L_{ij},

$$\begin{aligned} L_{ij} &= \beta_1 + \beta_2 \,\text{Time}_{ij} + \beta_3 \,\text{Time}_{ij}^2 + \beta_4 \,\text{Dose}_i \times \text{Time}_{ij} \\ &\quad + \beta_5 \,\text{Dose}_i \times \text{Time}_{ij}^2 + b_i + e_{ij}, \end{aligned}$$

where b_i has a normal distribution, with zero mean and variance σ^2_b, and the e_{ij} have a standard logistic distribution, with mean zero and variance $\pi^2/3$. Without loss of generality, we can assume that the threshold for categorizing L_{ij} is zero, with

$$Y_{ij} = 1 \text{ if } L_{ij} > 0,$$

$$Y_{ij} = 0 \text{ if } L_{ij} \leq 0.$$

This model for the latent variable implies the mixed effects logistic regression model for Y_{ij},

$$\begin{aligned}\text{logit}\{E(Y_{ij}|b_i)\} &= \beta_1 + \beta_2 \,\text{Time}_{ij} + \beta_3 \,\text{Time}_{ij}^2 + \beta_4 \,\text{Dose}_i \times \text{Time}_{ij} \\ &\quad + \beta_5 \,\text{Dose}_i \times \text{Time}_{ij}^2 + b_i.\end{aligned}$$

Thus, using the notion of an underlying latent variable distribution, we can compare the magnitudes of the between-subject and within-subject sources of variability of the L_{ij} in terms of the intra-subject correlation (often referred to as the *intra-class correlation*)

$$\rho = \text{Corr}(L_{ij}, L_{ik}) = \frac{\sigma_b^2}{\sigma_b^2 + \pi^2/3}.$$

The estimated intra-subject correlation for the repeated latent responses is

$$\widehat{\rho} = \frac{\widehat{\sigma}_b^2}{\widehat{\sigma}_b^2 + \pi^2/3} = \frac{5.065}{5.065 + 3.290} = 0.61,$$

indicating that there is substantial heterogeneity in the underlying propensity to experience amenorrhea. Note that ρ is the marginal correlation (averaged over the distribution of the random effects) among the unobserved L_{ij}; it is not the marginal correlation among the Y_{ij}.

The mixed effects model considered above includes only a single random effect, b_i. With binary data, and measurements at only four occasions, greater care must be exercised in the specification of the random effects as the limited amount of data may not support estimation of more than a single variance component. Inclusion of both randomly varying intercepts and slopes for the linear time trend in the logistic regression model for the amenorrhea data resulted in convergence problems during estimation. It should not be too surprising that problems arise when fitting this model to the data. Intuitively, attempts to fit a series of logistic regressions to data on at most four observations (the number of repeated measurements on each woman with complete data; recall that due to dropout 37% of the women had fewer than four measurements) are likely to result in numerical problems and/or produce unstable estimates.

This highlights an important feature of longitudinal binary data: there is usually not much information available about random effects, beyond a random subject effect (or random intercept), when the number of repeated measurements is relatively small. Thus convergence problems during estimation are often encountered when random effects beyond a random subject effect are included in logistic regression models for longitudinal data.

The estimates of the fixed effects and variance component reported in Table 14.2 were obtained by maximizing an approximate integrated likelihood, where the integration over the distribution of the random effects was achieved using numerical quadrature. Choice of the number of quadrature points determines the degree of accuracy of the approximation. In Table 14.3 we display the estimate of the variance component, σ^2_b, and the value of the maximized log-likelihood for increasing numbers of quadrature points. The results in Table 14.3 indicate that 5 to 10 quadrature points do not provide sufficient numerical accuracy for the amenorrhea data; this provides an illustration of the dangers of using too few quadrature points. The value

for the maximized log-likelihood and the estimate of σ^2_b become stable once the number of quadrature points exceeds 30. Table 14.3 also provides the CPU time required for fitting the model. As expected, the computational burden increases with the number of quadrature points. In this example there is only a single random effect and the increase in computational burden is relatively minor; however, in general, the computations grow exponentially with the number of random effects. When compared to the time expended in collecting longitudinal data, we regard the time required to accurately fit generalized linear mixed models to be negligible. So, in general, we recommend repeating analyses, with increasing number of quadrature points, until all estimates and standard errors become stable.

Table 14.3 Sensitivity of estimate of variance component to number of quadrature points: mixed effects logistic regression model, with random intercepts, for the amenorrhea data.

Quadrature Points	Log-Likelihood	Estimate of σ^2_b	CPU Time[a]
1	−1957.303	4.3366	2.31
2	−1957.246	3.9811	2.49
3	−1944.100	4.3767	2.79
4	−1933.495	5.1992	3.07
5	−1936.213	4.8369	3.21
10	−1934.514	5.0539	4.45
20	−1934.465	5.0648	6.86
30	−1934.465	5.0646	9.11
40	−1934.465	5.0646	11.40
50	−1934.465	5.0646	13.73
100	−1934.465	5.0646	25.28

[a]CPU time (in seconds) using PROC GLIMMIX in SAS run on a PC with Intel Core 2 Duo 6600 (2.4 GHz) processor. Convergence criterion (GCONV) set equal to 1E-12.

Finally, the estimates of the fixed effects in the mixed effects logistic regression model are larger than those obtained in a similar analysis using a marginal model. For illustrative purposes we fit the following marginal model to the amenorrhea data:

$$\begin{aligned}\text{logit}\{E(Y_{ij})\} &= \beta_1 + \beta_2\,\text{Time}_{ij} + \beta_3\,\text{Time}^2_{ij} + \beta_4\,\text{Dose}_i \times \text{Time}_{ij} \\ &\quad + \beta_5\,\text{Dose}_i \times \text{Time}^2_{ij},\end{aligned}$$

and assume an unstructured log odds ratio pattern for the within-subject association,

$$\log \text{OR}\,(Y_{ij}, Y_{ik}) = \alpha_{jk},$$

where

$$\text{OR}(Y_j, Y_k) = \frac{\Pr(Y_j = 1, Y_k = 1)\Pr(Y_j = 0, Y_k = 0)}{\Pr(Y_j = 1, Y_k = 0)\Pr(Y_j = 0, Y_k = 1)}.$$

The estimated regression coefficients and pairwise log odds ratios for the within-subject

association, obtained using the GEE approach, are presented in Table 14.4. The estimated logistic regression coefficients are smaller (in absolute value) than the estimated fixed effects in Table 14.2. Furthermore, from the estimates of the fixed effects in Table 14.4, the ratio of the population odds of amenorrhea at 12 months for women on the high dose versus low dose is 1.30, with 95% confidence interval: 0.98 to 1.71. These differences in the estimated coefficients and odds ratios are due to the different interpretations of β in the two classes of models; that is, these two classes of models estimate parameters that address substantively different questions.[1] The estimates of the fixed effects of dose in the mixed effects logistic regression model describe the effect of dose on a specific woman's risk of amenorrhea. The corresponding effects in the marginal logistic regression model describe the effects of dose on the prevalence of amenorrhea in the population of women assigned to high versus low doses of DMPA. Although the regression parameters for dose have distinct interpretations, their values coincide when there is no effect of dose. That is, at the null value the same hypothesis concerning the dependence of the risk of amenorrhea on dose is being tested. For example, a multivariate Wald test of $H_0 : \beta_4 = \beta_5 = 0$ based on the marginal model parameter estimates produces $W^2 = 12.3$, with 2 df ($p < 0.005$). The corresponding test from the mixed effects logistic regression parameter estimates produces $W^2 = 12.4$, with 2 df ($p < 0.005$).

Table 14.4 Parameter estimates and standard errors, obtained using GEE approach, from marginal logistic regression model for the amenorrhea data.

Variable	Estimate	SE	Z
Intercept	−2.2461	0.1765	−12.72
Time	0.7030	0.1581	4.45
Time^2	−0.0323	0.0318	−1.02
Dose × Time	0.3380	0.1097	3.08
Dose × Time^2	−0.0683	0.0284	−2.40
α_{12}	1.8475	0.1810	10.21
α_{13}	1.4851	0.1985	7.48
α_{14}	1.7605	0.2482	7.09
α_{23}	2.1610	0.1761	12.27
α_{24}	2.0665	0.2034	10.16
α_{34}	2.2783	0.1827	12.47

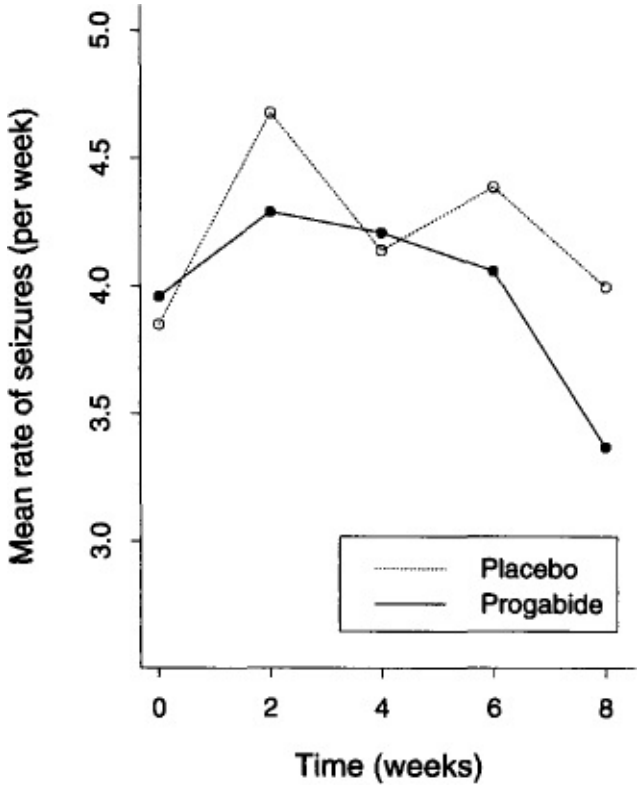
Clinical Trial of an Anti-epileptic Drug

Next we consider data from the placebo-controlled clinical trial of 59 epileptic patients, conducted by Leppik et al. (1987). Patients with partial seizures were enrolled in a

randomized clinical trial of the anti-epileptic drug, progabide. Progabide is an anti-epileptic drug whose primary mechanism of action is to enhance gamma-aminobutyric acid (GABA) content; GABA is the primary inhibitory neurotransmitter in the brain.

Participants in the study were randomized to either progabide or a placebo, as an adjuvant to the standard anti-epileptic chemotherapy. Prior to receiving treatment, baseline data on the number of epileptic seizures during the preceding 8-week interval were recorded. Counts of epileptic seizures during 2-week intervals before each of four successive post-randomization clinic visit were recorded. The average rates of seizures (per week), at baseline and in the four post-randomization visits are displayed in Figure 14.2.

Fig. 14.2 Mean rate of seizures (per week) at baseline, week 2, week 4, week 6, and week 8 in the progabide and placebo groups.



These data contain an extreme observation or outlier: one of the patients (patient 49) reported 151 seizures in the 8-week baseline interval and 302 (102 + 65 + 72 + 63) seizures during the four successive 2-week intervals. This patient was assigned to the progabide group. Since this patient could potentially have an inordinate impact on the analysis, we present results that include and exclude data from this patient.

We consider an analysis that addresses the question of whether treatment with progabide reduces the rate of epileptic seizures (when compared to placebo). To address this question, we can compare the subject-specific changes, from baseline to follow-up, in the rate of seizures for patients in the two treatment groups. We consider the following mixed effects log-linear regression model for the subject-specific expected counts (or rates) of seizures,

$$\begin{aligned}\log E(Y_{ij}|b_i) &= \log(T_{ij}) + \beta_1 + \beta_2 \,\mathrm{Time}_{ij} + \beta_3 \,\mathrm{Trt}_i + \beta_4 \,\mathrm{Trt}_i \times \mathrm{Time}_{ij} \\ &\quad + b_{1i} + b_{2i} \,\mathrm{Time}_{ij} \\ &= \log(T_{ij}) + (\beta_1 + b_{1i}) + (\beta_2 + b_{2i}) \,\mathrm{Time}_{ij} + \beta_3 \,\mathrm{Trt}_i \\ &\quad + \beta_4 \,\mathrm{Trt}_i \times \mathrm{Time}_{ij},\end{aligned}$$

where Y_{ij} is the number of epileptic seizures for the i^{th} patient in the j^{th} period of observation (j = 0, …, 4), and T_{13} is the length of period j (where T_{ij} = 8 if j = 0 and T_{ij} = 2 if j = 1, 2, 3, 4). The offset, $\log(T_{ij})$, is included because the "time at risk" is not the same in the baseline (8 weeks) and four successive follow-up periods (2-week intervals). The variable Trt is an indicator variable for treatment group, with Trt = 0 if an individual was randomized to the placebo group and Trt = 1 if randomized to the progabide group. The binary variable Time denotes the baseline and follow-up periods, with Time = 0 for the baseline period and Time = 1 for the follow-up periods (periods 1–4). Given b_i, it is assumed that the Y_{ij} are independent and have a Poisson distribution, with $\mathrm{Var}(Y_{ij}|b_i) = E(Y_{ij}|b_i)$, (i.e., $\phi = 1$). Finally, we assume that the random intercepts and slopes, b_i, have a bivariate normal distribution, with zero mean and 2 × 2 covariance matrix G. This mixed effects log-linear regression model posits that there is not only natural heterogeneity among patients in terms of their baseline expected rate of seizures but also heterogeneity in the changes in the expected rates of seizures over time. Unlike the previous example, where there was not much information available in the four repeated binary response about random effects beyond a random subject effect, with these repeated count data there is sufficient information to estimate both random intercepts and slopes.

In Table 14.5 we summarize the interpretation of β in terms of the subject-specific log expected seizure rates (per week) in the two groups at baseline and during post-baseline follow-up. Because all of the covariates in the model are dichotomous, the log-linear fixed effects regression parameters can be given interpretations in terms of subject-specific (log) rate ratios. So, for example, $e^{\beta 2}$ is the rate ratio of seizures, comparing the follow-up periods to baseline, for a "typical" patient in the placebo group (a "typical" patient is one with unobserved random slope b_{2i} = 0, the mean and median of the distribution of b_{2i}). Similarly $e^{\beta 2+\beta 4}$ is the rate ratio of seizures, comparing the follow-up periods to baseline, for a "typical" patient in the progabide group (with unobserved random slope b_{2i} = 0). A direct comparison of the two treatments in terms of changes in the expected rates of seizures is expressible in terms of β_4. That is, β_4 represents the difference between the changes in the log expected rates, comparing a patient from the progabide group to a patient from the placebo group, when the two patients are chosen so that they have the same value for the unobserved slope b_{2i}. That is, $e^{\beta 4}$ is a ratio of rate ratios. If $\beta_4 < 0$, this indicates a greater reduction (or alternatively, a smaller increase) in the seizure rate from baseline for the patient assigned to the progabide group (when compared to the patient assigned to the placebo group).

Table 14.5 Subject-specific log expected seizure rates in the two groups at baseline and during

post-baseline follow-up.

Treatment Group	Period	$\log\left\{\frac{E(Y_{ij}\mid b_i)}{T_{ij}}\right\}$
Placebo	Baseline	$\beta_1 + b_{1i}$
	Follow-up	$(\beta_1 + b_{1i}) + (\beta_2 + b_{2i})$
Progabide	Baseline	$(\beta_1 + b_{1i}) + \beta_3$
	Follow-up	$(\beta_1 + b_{1i}) + (\beta_2 + b_{2i}) + \beta_3 + \beta_4$

For the full sample ($N = 59$) the estimated fixed effects and covariance parameters from the log-linear model are displayed in Table 14.6. A test of the null hypothesis, H_0: $\beta_4 = 0$, indicates that there is a significant time × treatment interaction at the 0.05 level. These results suggest that there are differences between the two treatments in terms of subject-specific changes in the expected rates of seizures. In particular, there is a greater reduction in the expected seizure rate from baseline for patients treated with progabide (when compared to patients treated with the placebo). For a patient receiving the placebo, there is no expected change in the rate of seizures (or $e^{-0.0004} \approx 1.0$), while for a patient treated with progabide the expected decrease in seizures is approximately 26% (or $e^{-0.0004-0.3065} = e^{-0.3069} \approx 0.74$).

Table 14.6 Parameter estimates and standard errors from mixed effects log-linear regression model for the seizure data.

Variable	Estimate	SE	Z
Intercept	1.0707	0.1406	7.62
Time	−0.0004	0.1097	−0.00
Trt	0.0513	0.1931	0.27
Trt × Time	−0.3065	0.1513	−2.03
$g_{11} = \mathrm{Var}(b_{1i})$	0.5010	0.1010	
$g_{22} = \mathrm{Var}(b_{2i})$	0.2334	0.0608	
$g_{12} = \mathrm{Cov}(b_{1i}, b_{2i})$	0.0541	0.0559	

Note: ML estimation is based on 50-point adaptive Gaussian quadrature.

The estimated covariance parameters for the random intercepts and slopes indicate that there is substantial variability in the baseline seizure rate in the study population and also substantial variability in the patient-to-patient changes in the seizure rates in response to treatment. For example, the estimated variance of the random intercepts, $\hat{g}_{11} = 0.501$, implies that there is substantial patient-to-patient variability in terms of their baseline rate of seizures, since approximately 95% of the patients have a baseline seizure rate that varies from

$$\exp(1.071 - 1.96\sqrt{0.501}) \text{ to } \exp(1.071 + 1.96\sqrt{0.501}),$$

or 0.8 to 12.0 seizures per week. Similarly there is discernible heterogeneity in the patient-to-

patient changes in the seizure rates. For example, approximately 95% of patients treated with progabide have changes in the rates of seizures that vary from

$$\exp(1.071 - 1.96\sqrt{0.501}) \text{ to } \exp(1.071 + 1.96\sqrt{0.501}),$$

or changes that vary from a decrease in seizures of 71% to an increase in seizures of 88%. Finally, the correlation among the random intercepts and slopes is weak, indicating that the expected change in the seizure rates in not directly related to the baseline rate of seizures.

As was noted earlier, patient 49 is an outlier with extreme counts at all occasions. While the observations on this patient are likely to inflate the variance of the random effects, especially the variance of b_{1i}, they might also have an inordinate influence on the estimates of the fixed effects parameters. To assess the impact this patient has on the results, we repeated the analysis excluding observations on this patient (N=58). The results of this analysis are displayed in Table 14.7. A test of the null hypothesis, H_0: $\beta_4 = 0$, indicates that there is still a significant time × treatment interaction at the 0.05 level. These results indicate that there is a greater reduction in the expected seizure rate from baseline for patients treated with progabide (when compared to patients treated with the placebo). For a patient receiving the placebo, there is no expected change in the rate of seizures (or $1 - e^{0.0078} \approx 0$), while for a patient treated with progabide the expected decrease in seizures is approximately 30% (or $1 - e^{0.0078-0.3461} = 1 - e^{-0.3383} \approx 0.29$). Qualitatively, the results in Table 14.7 are very similar to those obtained in Table 14.6. As might be expected, the exclusion of patient 49 results in a noticeably smaller estimate of $\text{Var}(b_{1i})$.

Table 14.7 Parameter estimates and standard errors from mixed effects log-linear regression model for the seizure data, excluding patient 49.

Variable	Estimate	SE	Z
Intercept	1.0692	0.1344	7.96
Time	0.0078	0.1070	0.07
Trt	−0.0079	0.1860	−0.04
Trt × Time	−0.3461	0.1489	−2.33
$g_{11} = \text{Var}(b_{1i})$	0.4529	0.0935	
$g_{22} = \text{Var}(b_{2i})$	0.2163	0.0587	
$g_{12} = \text{Cov}(b_{1i}, b_{2i})$	0.0151	0.0529	

Note: ML estimation is based on 50-point adaptive Gaussian quadrature.

Finally, in Section 14.4 we mentioned that overdispersion is almost the rule, not the exception, with count data. To allow for overdispersion, we can extend the log-linear model to incorporate an extra source of variability in the subject-specific expected counts (or rates) of seizures,

$$\log E(Y_{ij}|b_i, e_{ij}) = \log(T_{ij}) + (\beta_1 + b_{1i}) + (\beta_2 + b_{2i})\,\text{Time}_{ij} + \beta_3\,\text{Trt}_i + \beta_4\,\text{Trt}_i \times \text{Time}_{ij} + e_{ij},$$

where e_{ij} is an additional random effect that varies over both individuals and measurement occasions. Assuming that the exponentiated errors, $\exp(e_{ij})$, have a gamma distribution with mean of 1 and variance θ leaves the model for the conditional mean unchanged but implies that the conditional variance of Y_{ij} is

$$\text{Var}(Y_{ij}|b_i) = E(Y_{ij}|b_i) + [E(Y_{ij}|b_i)]^2\theta.$$

This model can be fit directly to the seizure counts data by assuming that they have a negative binomial rather than a Poisson distribution, conditional on the random effects and the covariates.

To avoid any potential problems with model identification, in fitting this extended model, we set the correlation between the random intercepts and slopes to zero. This assumption is strongly supported by the results in Table 14.7; for example, the likelihood ratio test of $H_0 : g_{12} = \text{Cov}(b_{1i}, b_{2i}) = 0$ yields $G^2 = 0.08$, with 1 df($p > 0.95$). The results of fitting the negative binomial mixed effects model to the seizure count data are summarized in Table 14.8. The negative binomial model, through the inclusion of an additional random component to account for overdispersion, has led to a very discernible improvement in fit to these data. For example, the likelihood ratio test comparing the models with and without this additional random component yields $G^2 = 77.9$, with 1 df($p < 0.0001$). A test of the null hypothesis, H_0: $\beta_4 = 0$, indicates that there is a significant time × treatment interaction at the 0.05 level. In these results we see a greater reduction in the expected seizure rate from baseline for patients treated with progabide (when compared to patients treated with the placebo). For a patient receiving the placebo, there is no expected change in the rate of seizures (or $1 - e^{0.0054} \approx 0$), whereas for a patient treated with progabide, the expected decrease in seizures is approximately 30% (or $1 - e^{(0.0054-0.3585)} = 1 - e^{-0.3531} \approx 0.30$). Qualitatively, the results in Table 14.8 for the fixed effects are very similar to those reported in Table 14.7. What differs, though, are the estimates of the variances of the random effects. In the mixed effects Poisson model any excess variability relative to Poisson variation is partially accounted for by the variances of the random intercepts and slopes. Therefore, with the inclusion of an additional random component in the mixed effects negative binomial model to account for this excess variability (with $\widehat{\theta} = 0.1173$), it should not be all that surprising that the estimated variances of the random intercepts and slopes in Table 14.8 are attenuated. For example, when compared to the estimated variance of the slopes in Table 14.7 (where θ is effectively assumed to be zero), the relative magnitude has been reduced by approximately 40%. By allowing for overdispersion ($\theta > 0$), the estimated standard errors in Table 14.8 are somewhat larger than those reported in Table 14.7. The standard errors are approximately 10% larger, which is consistent with an overdispersion factor of approximately 1.2. Thus, while there is evidence of overdispersion, the degree of overdispersion is not substantial.

Table 14.8 Parameter estimates and standard errors from negative binomial mixed effects log-

linear regression model for the seizure data, excluding patient 49.

Variable	Estimate	SE	Z
Intercept	1.0986	0.1481	7.42
Time	0.0054	0.1160	0.05
Trt	0.0033	0.2055	0.02
Trt × Time	−0.3585	0.1629	−2.20
$g_{11} = \text{Var}(b_{1i})$	0.4414	0.1040	
$g_{22} = \text{Var}(b_{2i})$	0.1278	0.0678	
$\theta = \text{Var}\{\exp(e_{ij})\}$	0.1173	0.0254	

Note: ML estimation is based on 50-point adaptive Gaussian quadrature.

In Section 14.4 we remarked that a very similar model can be specified by replacing the gamma distribution for the errors with a normal distribution. This leads to a completely equivalent model for the conditional variance of Y_i (conditional only on b_i and the covariates) that allows for overdispersion. For illustrative purposes we fit the extended model with normal errors, e_{ij}, to the seizure count data. We obtained remarkably similar estimates of the fixed effects and variance components to those reported in Table 14.8. For example, the estimate of β_4, the parameter of main interest, is −0.3525 (SE = 0.1608), which is very similar to the estimate −0.3585 (SE = 0.1629) reported in Table 14.8.

Arthritis Clinical Trial

Finally, we consider longitudinal ordinal data from the longitudinal clinical trial comparing auranofin therapy (3 mg of oral gold, twice daily) and placebo for the treatment of rheumatoid arthritis (Bombardier et al., 1986). Recall, from Section 13.4, that in this six-month, randomized, double-blind trial, 303 patients with classic or definite rheumatoid arthritis were randomized to one of the two treatment groups and followed over time. The outcome variable of interest is a global impression scale (Arthritis Categorical Scale) measured at baseline (month 0), month 2, month 4, and month 6. This is a self-assessment of a patient's current arthritis, measured on a five-level ordinal scale: (1) very good, (2) good, (3) fair, (4) poor, and (5) very poor.

The goal of the analysis is to assess changes in the odds of a more favorable response over the duration of the study, and to determine whether treatment with auranofin has an influence on these changes. Letting Y_{ij} denote the ordinal response for the i^{th} subject at the j^{th} occasion, we assume that the subject-specific log odds of a more favorable response at each occasion follows the proportional odds model

$$\log\left\{\frac{\Pr(Y_{ij} \le k|b_i)}{\Pr(Y_{ij} > k|b_i)}\right\} = \alpha_k + \beta_1 \text{Trt}_i + \beta_2 \sqrt{\text{Month}_{ij}} + \beta_3 \text{Trt}_i \times \sqrt{\text{Month}_{ij}} + b_{1i} + b_{2i}\sqrt{\text{Month}_{ij}},$$

where $\sqrt{\text{Month}_{ij}}$ = the square-root transformation of time, in months, for the i^{th} subject at the j^{th} occasion, $\text{Trt}_i = 1$ if the i^{th} subject is randomized to auranofin, and $\text{Trt}_i = 0$ if randomized to placebo. This mixed effects proportional odds model allows for randomly varying intercepts and slopes for square-root transformed time. It assumes that given b_i, the Y_{ij} are independent and have a multinomial distribution. Last, we assume that the random intercepts and slopes, b_i, have a bivariate normal distribution, with zero mean and 2 × 2 covariance matrix G. This model posits that there is not only heterogeneity among patients in terms of their baseline odds of a more favorable response but also heterogeneity in the changes in the odds of a more favorable response over time.

The ML estimates of the fixed effects and covariance parameters are displayed in Table 14.9. A test of the null hypothesis, H_0: $\beta_3 = 0$, indicates that there is a significant treatment × (square-root transformed) time interaction ($Z = 2.80$, $p < 0.01$). These results indicate that the pattern of subject-specific changes over time in the odds of a more favorable response differs between treatment groups. In particular, relative to baseline, the odds of a more favorable response at week 6 is increased by a factor of 2.66 (or $e^{0.4000 \times \sqrt{6}}$) for a patient receiving placebo but by a factor greater than 7 (or $e^{(0.4000+0.4005) \times \sqrt{6}} = 7.11$) for a similar patient receiving auranofin. At the completion of the study, a patient treated with auranofin is approximately 2 $\frac{1}{2}$ times (or $e^{0.4005 \times \sqrt{6}} = 2.67$) more likely to have a more favorable response when compared to a similar patient treated with placebo. As expected, due to the randomization, $\widehat{\beta}_1 \approx 0$, indicating that patients in the two treatment groups have similar subject-specific log odds of a favorable response at baseline (or month 0).

Table 14.9 Parameter estimates and standard errors from mixed effects proportional odds model for the auranofin clinical trial data.

Variable	Estimate	SE	Z
α_1	−5.2945	0.3289	−16.10
α_2	−2.0085	0.2483	−8.09
α_3	1.0753	0.2382	4.51
α_4	3.8683	0.3006	12.87
Trt	0.1140	0.3184	0.36
$\sqrt{\text{Month}}$	0.4000	0.1018	3.93
Trt × $\sqrt{\text{Month}}$	0.4005	0.1432	2.80
$g_{11} = \text{Var}(b_{1i})$	4.1424	0.8549	
$g_{22} = \text{Var}(b_{2i})$	0.3243	0.1617	
$g_{12} = \text{Cov}(b_{1i}, b_{2i})$	−0.0516	0.2626	

Note: ML estimation is based on 30 point adaptive Gaussian quadrature.

The estimated covariance parameters for the random intercepts and slopes indicate substantial variability in the baseline odds of a more favorable response and also discernible variability in the patient-to-patient changes in the odds in response to treatment. For example, the estimated variance of the random slopes, $\hat{g}_{22} = 0.324$, implies patient-to-patient variability in change in the log odds over time. Approximately 95% of patients treated with placebo have slopes for time that vary over the interval

$$0.400 \pm 1.96\sqrt{0.324} = (-0.72, 1.52),$$

whereas approximately 95% of patients treated with auranofin have slopes for time that vary over the interval

$$(0.400 + 0.401) \pm 1.96\sqrt{0.324} = (-0.32, 1.92).$$

Thus, while 92% of patients treated with auranofin are expected to have positive slopes that correspond to a more favorable response, only 76% of patients treated with placebo are expected to have a more favorable response. The remaining 24% of of patients treated with placebo are expected to have a less favorable response.

14.8 COMPUTING: FITTING GENERALIZED LINEAR MIXED MODELS USING PROC GLIMMIX IN SAS

Until recently a potential limitation of generalized linear mixed models was their computational burden. Because there is no simple closed-form solution for the marginal likelihood, numerical integration techniques are required. Maximum (marginal) likelihood estimation has only recently been implemented in standard statistical software, for example, PROC NLMIXED and PROC GLIMMIX in SAS.

To fit generalized linear mixed models, we use the GLIMMIX procedure in SAS. PROC GLIMMIX can be used to directly maximizes an approximate integrated likelihood, where the integration over the random effects is achieved using numerical quadrature. For example, to fit a logistic regression model with randomly varying intercepts to longitudinal data from two groups (coded 0 and 1), we can use the illustrative SAS commands given in Table 14.10. Similarly, to fit a mixed effects log-linear regression, with randomly varying intercepts and slopes, we can use the illustrative SAS commands given in Table 14.11. To fit a mixed effects log-linear regression that assumes negative binomial variability (or overdispersion relative to Poisson variability), with randomly varying intercepts and slopes, we can use the illustrative SAS commands given in Table 14.12. Finally, to fit a mixed effects proportional odds regression for ordinal responses, with randomly varying intercepts and slopes, we can use the illustrative SAS commands given in Table 14.13.

Table 14.10 Illustrative commands for a mixed effects logistic regression, with randomly varying intercepts, fitted using adaptive quadrature in PROC GLIMMIX in SAS.

```
PROC GLIMMIX METHOD=QUAD(QPOINTS=50); CLASS id group;
  MODEL y=group time group*time / DIST=BINOMIAL LINK=LOGIT S;
  RANDOM INTERCEPT / SUBJECT=id TYPE=UN;
```

Table 14.11 Illustrative commands for a mixed effects log-linear regression, with randomly varying intercepts and slopes, fitted using adaptive quadrature in PROC GLIMMIX in SAS.

```
PROC GLIMMIX METHOD=QUAD(QPOINTS=50);
  CLASS id group;
  MODEL y=group time group*time / DIST=POISSON LINK=LOG S;
  RANDOM INTERCEPT time / SUBJECT=id TYPE=UN;
```

Table 14.12 Illustrative commands for a mixed effects log-linear regression assuming negative binomial variance (overdispersion relative to Poisson variance), with randomly varying intercepts and slopes, fitted using adaptive quadrature in PROC GLIMMIX in SAS.

```
PROC GLIMMIX METHOD=QUAD(QPOINTS=50);
  CLASS id group;
  MODEL y=group time group*time / DIST=NEGBIN LINK=LOG S;
  RANDOM INTERCEPT time / SUBJECT=id TYPE=UN;
```

Table 14.13 Illustrative commands for a mixed effects proportional odds regression, with randomly varying intercepts and slopes, fitted using adaptive quadrature in PROC GLIMMIX in SAS.

```
PROC GLIMMIX METHOD=QUAD(QPOINTS=50);
```

```
CLASS id group;
MODEL y=group time group*time / DIST=MULT LINK=CUMLOGIT S;
RANDOM INTERCEPT time / SUBJECT=id TYPE=UN;
```

PROC GLIMMIX in SAS is a procedure for fitting generalized linear mixed models (among other models) using maximum likelihood (via adaptive quadrature) or using alternative methods of estimation that are less computationally demanding. In particular, PROC GLIMMIX provides two approximate methods of estimation of the model parameters known as *penalized quasi-likelihood* (PQL) and *marginal quasi-likelihood* (MQL). In this section we focus only on maximum likelihood (ML) estimation via adaptive quadrature; at the end of Chapter 15 we discuss the PQL and MQL options in PROC GLIMMIX.

The GLIMMIX procedure is very versatile and the syntax is remarkably similar to that used in PROC MIXED. For example, PROC GLIMMIX has a RANDOM statement that is used in a similar way as in PROC MIXED for introducing random effects. No attempt is made here to give a comprehensive review of the main features of PROC GLIMMIX. Instead, we present illustrative commands for fitting generalized linear mixed effects models in general terms (see Tables 14.10 through 14.13) and then describe the most salient parts of the command syntax for these illustrations. Next we present a brief description of each of the command statements used in Tables 14.10 through 14.13.

PROC GLIMMIX <options>;

This statement calls the procedure GLIMMIX in SAS. It includes an option for specifying the method of estimation, using METHOD=<options>.

The default is METHOD=RSPL, where RSPL denotes a version of *penalized quasi-likelihood* (PQL) estimation. Penalized quasi-likelihood estimation is discussed in Chapter 15. PROC GLIMMIX also includes an implementation of adaptive Gaussian quadrature using the METHOD=QUAD option. When the METHOD=QUAD option is used, the accuracy of the numerical approximation can be increased by specifying the number of quadrature points used during evaluation of integrals for the marginal likelihood. For example, METHOD=QUAD(QPOINTS=50) specifies that 50 quadrature points be used for each random effect, resulting in a total of 50^q quadrature points (where q is the number of random effects, the dimension of bi). A note of caution, the likelihood approximation may not be accurate if too few quadrature points are used.

CLASS variables;

The CLASS statement is used to defien all variables that are to be regarded as categorical or factors. By default, this statement will create indicator variables for each factor using a reference group coding, with the last level (where “last” here refers to the level with the largest alphanumeric value) regarded as the reference group. Different sort orders for the CLASS variables can be requested by the ORDER=<option> on the PROC GLIMMIX statement.

MODEL response = <effects> / <options>;

MODEL events/trials = <effects> / <options>;

The MODEL statement specifies the response variable and the covariate effects. The second form of the MODEL statement, with the `events/trials` syntax, allows the response to be in the form of a ratio of two variables (e.g., counts of the number of successes and the number of trials) and is used for binomial response data. The linear predictor can include both discrete (defined in the CLASS statement) and quantitative (excluded from the CLASS statement) covariates. By default, PROC GLIMMIX includes a column of 1's for the intercept in the model. The SOLUTION or S option is used to produce estimates of the fixed (or covariate) effects.

The option DIST=*keyword* specifies the conditional distribution of the response given the random effects in a GLMM.

The LINK=*keyword* specifies the choice of built-in link function relating the mean response to the linear predictor. If the LINK=*keyword* is omitted, the default link function is the canonical link function associated with the particular exponential family distribution specified on DIST=*keyword*.

A final option often required when modeling count data is an offset. The OFFSET=*variable* specifies a variable to be used as an offset. For example, in modeling count data, the rate is often of more direct interest and the denominator for the counts or "population at risk" (or more specifically, the log of the denominator) can be included as an offset. Note that this variable cannot be a CLASS variable and that it should not be included as one of the covariates listed on the MODEL statement.

RANDOM <random-effects> / SUBJECT=subject-effect <options>;

In a generalized linear mixed model, the RANDOM statement is used to define the covariates in the design matrix, Z_i, for the random effects, b_i. Ordinarily these will be a subset of the covariates included on the MODEL statement. While the MODEL statement is used to define the design matrix for the fixed effects and the RANDOM statement is used to define the design matrix for the random effects, note that an intercept is included by default in the former but not the latter. That is, unlike the MODEL statement, PROC GLIMMIX does not include an intercept in the RANDOM statement by default. However, you can specify INTERCEPT (or INT) as a random effect on the RANDOM statement. The SUBJECT option on the RANDOM statement is used to denote a variable that distinguishes clusters of correlated responses. By including a variable in the SUBJECT option (e.g., a subject identifier), pairs of observations with distinct values of that variable are regarded as independent. Pairs of observations with the same values of that variable share common values of the random effects. The RANDOM statement is also used to specify the structure of the covariance matrix for the random effects, G. The structure of G is specified using the TYPE=option. The random effects can be assumed to be correlated (TYPE=UN) or uncorrelated (TYPE=VC); ordinarily, covariance pattern models are not used to account for the covariance among the random effects. Finally, the SOLUTION option on the RANDOM statement produces the empirical Bayes predictions of the random effects.

We conclude by noting that PROC GLIMMIX can fit a broader class of generalized linear mixed models than have been considered in this chapter. To do so, PROC GLIMMIX makes a distinction between two sources of variation and covariation in the model for the data: (1) variation due to random effects, b_i, and (2) "residual" (co)variation. To distinguish these two sources of (co)variation, PROC GLIMMIX refers to (1) as "G-side" effects and (2) as "R-side" effects. These two non-standard terms are derived from syntax for the covariance matrices for the random effects (denoted G) and the "residual errors" (denoted R) in PROC MIXED for linear mixed effects models. PROC GLIMMIX is quite versatile in allowing models to be fit with various combinations of "G-side" and "R-side" effects. For example, in a standard generalized linear mixed model, the introduction of random effects is handled by including "G-side" effects and the "conditional independence" assumption is handled by implicitly assuming a simple structure for the "R-side" effects (i.e., uncorrelated residual errors, the default option for the "R-side" effects). In principle, it is possible to fit models that relax the "conditional independence" assumption by allowing for a more general structure for the "R-side" effects (e.g., autoregressive residual errors); however, we caution that as with linear mixed effects models, there can be subtle issues of model identification when a more general structure for the "R-side" effects is assumed because it may not be possible to estimate both the "G-side" and the more general "R-side" effects from the data at hand. A more detailed description of "G-side" and "R-side" effects can be found at the end of Chapter 15.

Finally, a word of caution concerning the use of PROC GLIMMIX. Our limited experience with this procedure indicates that it can be sensitive to poor choices of starting values for the covariance parameters for the random effects and/or the numerical accuracy of the quadrature used. Convergence of the algorithm implemented in PROC GLIMMIX should never be taken for granted; neither should convergence to a global maximum be assumed. Instead, we recommend that users of this procedure provide different initial values for the covariance parameters and/or consider a grid search of values to ensure that a global maximum has been obtained. Accurate initial values for the covariance parameters can be obtained by specifying a grid of feasible values using the PARMS statement in PROC GLIMMIX. The PARMS statement specifies initial values for the covariance or scale parameters. If you specify more than one set of initial values on the PARMS statement, PROC GLIMMIX will first evaluate the marginal likelihood at each grid value and select the grid point that produces the largest value of the marginal likelihood as the initial values for the covariance parameters for the subsequent maximization of the marginal likelihood. Problems with convergence and computation are likely to arise when the model parameters are on scales that vary by more than a few orders of magnitude. The latter problem can be circumvented by appropriately rescaling each parameter. For example, when the variances of random intercepts and slopes differ by more than a few orders of magnitude, the variance of the slopes can be rescaled by multiplying the variable for time by an appropriate constant. The numerical accuracy of the quadrature used can always be enhanced by increasing the number of quadrature points; of course, increasing the number of quadrature points does raise the computational burden.

14.9 FURTHER READING

A relatively non-technical discussion of random effects models for binary data can be found in Section 6.7 of Collett (1991). Chapter 12 of Agresti (2002) provides a detailed, although more mathematically challenging, description of generalized linear mixed effects models for categorical data.

Bibliographic Notes

The theoretical foundation for generalized linear mixed effects models can be found in Skellam's (1948) introduction of the beta-binomial distribution. Since then, the statistical literature on generalized linear mixed effects models has grown rapidly. Some key references in the literature include Cox (1970), Pierce and Sands (1975), Williams (1982), Stiratelli et al. (1984), Anderson and Aitkin (1985), Gilmour et al. (1985), Wong and Mason (1985), Schall (1991), Zeger and Karim (1991), Breslow and Clayton (1993), and Hedeker and Gibbons (1994). Molenberghs et al. (2010) present a broad class of generalized linear models that accommodate both overdispersion and correlation via the introduction of two separate sets of random effects.

The marginal maximum likelihood method described previously is based on a numerically integrated likelihood function and requires the computation of the integral over the random effects. A method known as adaptive Gaussian quadrature is commonly used for computing this integral and is described in detail in Pinheiro and Bates (1995); also see Anderson and Aitkin (1985) and Hedeker and Gibbons (1994, 1996). Lesaffre and Spiessens (2001) provide a striking example of the dangers of using too few quadrature points when fitting generalized linear mixed effects models; also see Problem 14.1.10. An alternative approximation, leading to an approach known as penalized quasi-likelihood (PQL), was proposed by Stiratelli et al. (1984). A more accurate approximation, based on higher-order Laplace approximations, is described in Breslow and Lin (1995) and Lin and Breslow (1996).

Finally, Neyman and Scott (1948) defined the so-called incidental parameters problem and showed that maximum likelihood estimation (MLE) may be problematic when the number of model parameters grows with the number of observations. The use of conditional likelihoods, and properties of conditional ML estimators, is discussed in Anderson (1970). In the longitudinal setting, Conaway (1992) and Rathouz (2004) discuss conditional ML estimation with incomplete response data.

Problems

14.1 In a randomized, double-blind, parallel-group, multicenter study comparing two oral anti-fungal treatments (200 mg/day Itraconazole and 250 mg/day Terbinafine) for toenail infection (De Backer et al., 1998; also see Lesaffre and Spiessens, 2001), patients were evaluated for the degree of onycholysis (the degree of separation of the nail plate from the nail-bed) at baseline (week 0) and at weeks 4, 8, 12, 24, 36, and 48 thereafter. The onycholysis outcome

variable is binary (none or mild versus moderate or severe). The binary outcome was evaluated on 294 patients comprising a total of 1908 measurements. The main objective of the analyses is to compare the effects of the two oral anti-fungal treatments (Itraconazole and Terbinafine) on changes in the probability of the binary onycholysis outcome over the duration of the study.

The raw data are stored in an external file: `toenail.dat`

Each row of the data set contains the following five variables:

ID Y Treatment Month Visit

Note: The binary onycholysis outcome variable Y is coded 0 = none or mild, 1 = moderate or severe. The categorical variable Treatment is coded 1 = Terbinafine, 0 = Itraconazole. The variable Month denotes the exact timing of measurements in months. The variable Visit denotes the visit number (visit numbers 1–7 correspond to scheduled visits at 0,4, 8, 12, 24, 36, and 48 weeks).

14.1.1 First, consider a *marginal* model for the log odds of moderate or severe onycholysis. Using GEE, fit a model that assumes linear trends for the log odds over time, with common intercept for the two treatment groups, but different slopes:

$$\text{logit}\{E(Y_{ij})\} = \beta_1 + \beta_2\,\text{Month}_{ij} + \beta_3\,\text{Treatment}_i \times \text{Month}_{ij}.$$

Assume "exchangeable" log odds ratios (or "exchangeable" correlations, if available software does not permit the within-subject association to be parameterized in terms of log odds ratios) for the association among the repeated binary responses.

14.1.2 What is the interpretation of β_2 in this model?

14.1.3 What is the interpretation of β_3 in this model?

14.1.4 From the results of the analysis for Problem 14.1.1, what conclusions do you draw about the effect of treatment on changes in the log odds of moderate or severe onycholysis over time? Provide results that support your conclusions.

14.1.5 Next consider a generalized linear mixed model, with randomly varying intercepts, for the patient-specific log odds of moderate or severe onycholysis. Using maximum likelihood (ML), fit a model with linear trends for the log odds over time and allow the slopes to depend on treatment group:

$$\text{logit}\{E(Y_{ij}|b_i)\} = (\beta_1 + b_i) + \beta_2\,\text{Month}_{ij} + \beta_3\,\text{Treatment}_i \times \text{Month}_{ij},$$

where, given b_i, Y_{ij} is assumed to have a Bernoulli distribution. Assume that $b_i \sim N(0, \sigma^2_b)$.

14.1.6 What is the estimate of σ^2_b? Give an interpretation to the magnitude of the estimated variance?

14.1.7 What is the interpretation of the estimate of β_2?

14.1.8 What is the interpretation of the estimate of β_3?

14.1.9 Compare and contrast the estimates of β_3 from the marginal and mixed effects models. Why might they differ?

14.1.10 Repeat the analysis from Problem 14.1.5 sequentially increasing the number of

quadrature points used. Compare the estimates and standard errors of the model parameters when the number of quadrature points is 2, 5, 10, 20, 30, and 50. Do the results depend on the number of quadrature points?

14.2 The Skin Cancer Prevention Study was a randomized, double-blind, placebo-controlled clinical trial of beta carotene to prevent non-melanoma skin cancer in high-risk subjects (Greenberg et al., 1989, 1990; also see Stukel, 1993). A total of 1805 subjects were randomized to either placebo or 50 mg of beta carotene per day for 5 years. Subjects were examined once a year and biopsied if a cancer was suspected to determine the number of new skin cancers occurring since the last exam. The outcome variable is a count of the number of new skin cancers per year. The outcome was evaluated on 1683 subjects comprising a total of 7081 measurements. The main objective of the analyses is to compare the effects of beta carotene on skin cancer rates.

The raw data are stored in an external file: `skin.dat`

Each row of the data set contains the following 9 variables:

ID Center Age Skin Gender Exposure Y Treatment Year

Note: The outcome variable Y is a count of the of the number of new skin cancers per year. The categorical variable Treatment is coded 1 = beta carotene, 0 = placebo. The variable Year denotes the year of follow-up. The categorical variable Gender is coded 1 = male, 0 = female. The categorical variable Skin denotes skin type and is coded 1 = burns, 0 = otherwise. The variable Exposure is a count of the number of previous skin cancers. The variable Age is the age (in years) of each subject at randomization.

14.2.1 Consider a generalized linear mixed model, with randomly varying intercepts, for the subject-specific log rate of skin cancers.

Using maximum likelihood (ML), fit a model with linear trends for the log rate over time and allow the slopes to depend on treatment group:

$$\log\{E(Y_{ij}|b_i)\} = (\beta_1 + b_i) + \beta_2\,\mathrm{Year}_{ij} + \beta_3\,\mathrm{Treatment}_i \times \mathrm{Year}_{ij},$$

where, given b_i, Y_{ij} is assumed to have a Poisson distribution. Assume that b_i, $\sim N(0, \sigma^2_b)$.

14.2.2 What is the estimate of σ^2_b? Give an interpretation to the magnitude of the estimated variance?

14.2.3 What is the interpretation of the estimate of β_2?

14.2.4 What is the interpretation of the estimate of β_3?

14.2.5 From the results of the analysis for Problem 14.2.1, what conclusions do you draw about the effect of beta carotene on the log rate of skin cancers? Provide results that support your conclusions.

14.2.6 Obtain the predicted (empirical BLUP) random effect for each subject.

(a) Calculate the sample variance of the predictions. How does it compare to the estimate of σ^2_b obtained in Problem 14.2.2? Why might they differ?

(b) Plot the predictions against age and the count of the number of previous skin

cancers. What do you conclude?

14.2.7 Repeat the analysis from Problem 14.2.1 adjusting for skin type, age, and the count of the number of previous skin cancers. What conclusions do you draw about the effect of beta carotene on the adjusted log rate of skin cancers?

14.3 In the U.S. National Institute of Mental Health (NIMH) Schizophrenia Collaborative Study, 437 patients were randomly assigned to receive one of four medications: placebo, chlorpromazine, fluphenazine, or thioridazine; the latter three medications are anti-psychotic drugs. The study protocol called for longitudinal measurements to be made at weeks 0, 1, 3, and 6. A very small subset of patients were additionally measured at weeks 2, 4, and 5. The outcome variable of interest is a 4-level ordinal scale measuring "severity of illness," derived from item 79 of the Inpatient Multidimensional Psychiatric Scale (Lorr and Klett, 1966). The four categories of the ordinal scale correspond to: 1 = "normal or borderline mentally ill," 2 = "mildly or moderately ill," 3 = "markedly ill," and 4= "severely or among the most extremely ill" (Gibbons and Hedeker, 1994; also see Hedeker and Gibbons, 2006). In this study there was substantial attrition, especially in the placebo group; however, for the purpose of this exercise, you can ignore the potential impact of missing data on the analysis. The main objective of the analyses is to assess changes in the odds of a more favorable response over the duration of the study, and to determine whether treatment with anti-psychotic drugs has an influence of these changes.

The raw data are stored in an external file: `schizophrenia.dat`

Each row of the data set contains the following four variables:

ID Y Trt Week

Note: The ordinal outcome variable Y is coded 1 = "normal or borderline mentally ill," 2 = "mildly or moderately ill," 3 = "markedly ill," and 4 = "severely or among the most extremely ill," The categorical treatment variable Trt is coded 1 = anti-psychotic drug, 0 = placebo; that is, the three anti-psychotic medications (chlorpromazine, fluphenazine, or thioridazine) are combined into a single group to be compared to placebo. The variable Week denotes the timing of measurements in weeks.

14.3.1 Consider a proportional odds model, with randomly varying intercepts and slopes, for the patient-specific cumulative log odds of response.

Using maximum likelihood (ML), fit a model with linear trends for the cumulative log odds over time and allow the slopes to depend on treatment group:

$$\log\left\{\frac{\Pr(Y_{ij} \le k|b_i)}{\Pr(Y_{ij} > k|b_i)}\right\} = \alpha_k + \beta_1 \,\mathrm{Trt}_i + \beta_2 \,\mathrm{Week}_{ij} + \beta_3 \,\mathrm{Trt}_i \times \mathrm{Week}_{ij} + b_{1i} + b_{2i} \,\mathrm{Week}_{ij},$$

where, given b_{1i} and b_{2i}, Y_{ij} is assumed to have a multinomial distribution. Assume that b_{1i} and b_{2i} have a bivariate normal distribution, with zero mean and 2 × 2 covariance matrix G.

14.3.2 What is the interpretation of the estimate of β_2?

14.3.3 What is the interpretation of the estimate of β_3?

14.3.4 From the results of the analysis from Problem 14.3.1, what conclusions do you draw about the effect of treatment with anti-psychotic medications on changes in the log odds of a more favorable response over time? Provide results that support your conclusions.

14.3.5 Repeat the analysis from Problem 14.3.1 replacing time (in weeks) with square-root transformed time:

$$\log\left\{\frac{\Pr(Y_{ij} \leq k|b_i)}{\Pr(Y_{ij} > k|b_i)}\right\} = \alpha_k + \beta_1 \,\mathrm{Trt}_i + \beta_2 \sqrt{\mathrm{Week}_{ij}} + \beta_3 \,\mathrm{Trt}_i \times \sqrt{\mathrm{Week}_{ij}} + b_{1i} + b_{2i} \sqrt{\mathrm{Week}_{ij}}.$$

14.3.6 Based on the results from Problem 14.3.5, estimate the change in the odds of a more favorable response at week 6, relative to baseline, for a patient receiving placebo.

14.3.7 Based on the results from Problem 14.3.5, estimate the change in the odds of a more favorable response at week 6, relative to baseline, for a patient receiving an anti-psychotic medication.

14.3.8 Based on the results from Problem 14.3.5, approximately what percentage of patients receiving placebo are expected to show improvement over time?

14.3.9 Based on the results from Problem 14.3.5, approximately what percentage of patients receiving anti-psychotic medications are expected to show improvement over time?

14.3.10 From the results of the analysis for Problem 14.3.5, what conclusions do you draw about the effect of treatment with anti-psychotic medications on changes in the log odds of a more favorable response over time? Provide results that support your conclusions.

14.3.11 In Problem 14.3.1 it is assumed that changes in the cumulative log odds are linear in time (in weeks). In Problem 14.3.5 it is assumed that changes in the cumulative log odds are linear in square-root transformed time ($\sqrt{\mathrm{weeks}}$). Which of the two models appears to fit the data better? Provide results that support your conclusions.

14.3.12 Repeat the analysis from Problem 14.3.5 sequentially increasing the number of quadrature points used. Compare the estimates and standard errors of the model parameters when the number of quadrature points is 2, 5, 10, 20, 30, and 50. Do the results depend on the number of quadrature points?