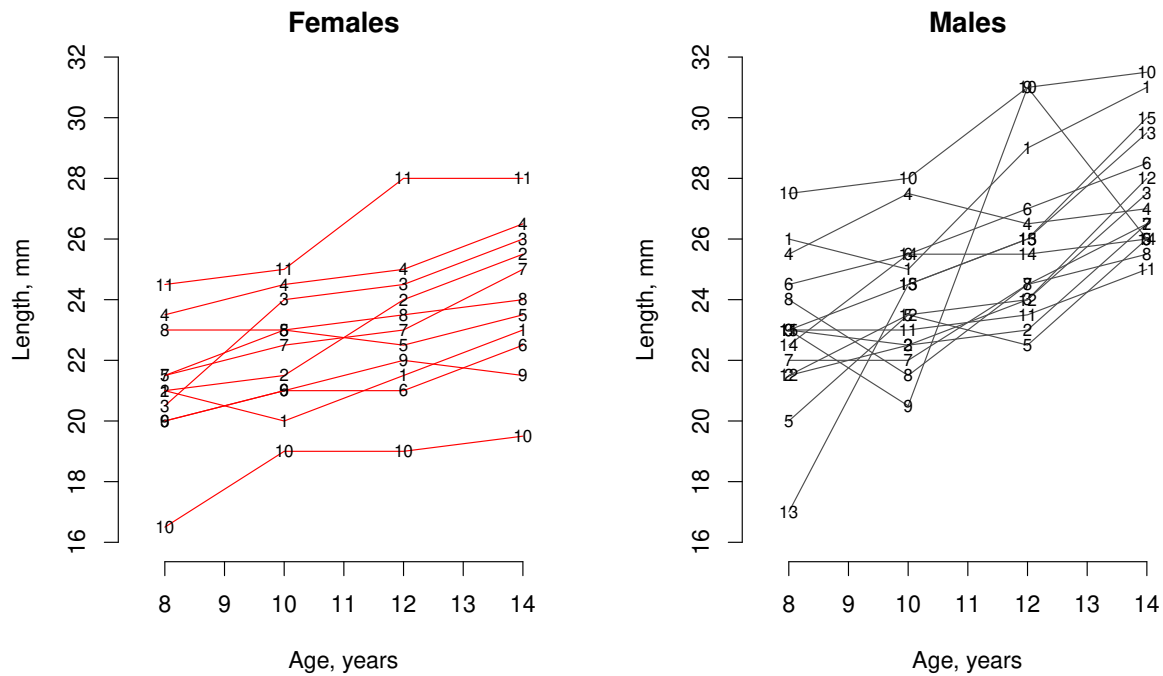
Part V:

Generalized Linear Mixed Models

Review of linear mixed models

Marginal model only cares about population mean
GEE is a discrete version of WLS

- Dental growth data:



- Strategy is to assume that each subject/cluster has its own regression model that is characterized by a combination of
  - ⋆ *fixed effects* that are common to all clusters
  - ⋆ *random effects* that are cluster-specific

- Formally, we can write down the linear mixed model for k^{th} cluster as:

$$\boldsymbol{Y}_k \;=\; \boldsymbol{X}_k\boldsymbol{\beta}^* \;+\; \boldsymbol{Z}_k\boldsymbol{\gamma}_k \;+\; \boldsymbol{\epsilon}_k$$

$$\begin{aligned} \mathrm{E}[\boldsymbol{\gamma}_k] &= \mathbf{0} & \qquad \mathrm{Cov}[\boldsymbol{\gamma}_k] &= \boldsymbol{G}(\boldsymbol{\alpha}) \\ \mathrm{E}[\boldsymbol{\epsilon}_k] &= \mathbf{0} & \qquad \mathrm{Cov}[\boldsymbol{\epsilon}_k] &= \boldsymbol{R}_k(\boldsymbol{\alpha}) \\ \mathrm{Cov}[\boldsymbol{\gamma}_k, \boldsymbol{\epsilon}_k] &= \mathbf{0} \end{aligned}$$

  - ⋆ assume independence across clusters
  - ⋆ if the off-diagonals of $\boldsymbol{R}_k(\boldsymbol{\alpha})$ are set to be zero $\Rightarrow$ assume conditional independence of study units within a cluster, given the random effect(s)

- The introduction of cluster-specific random effects induces correlation structure, marginally, among the study units within a cluster
  - ⋆ alternate structures can be obtained by changing the specification of $\boldsymbol{Z}_k$

- Alternative specifications of $\boldsymbol{R}_k(\boldsymbol{\alpha})$ provide additional flexibility
  - ⋆ e.g. serial dependence in the error terms

- Estimation/inference for $(\boldsymbol{\beta}^*, \boldsymbol{\alpha})$ is typically based on an *integrated* or *marginal* likelihood:

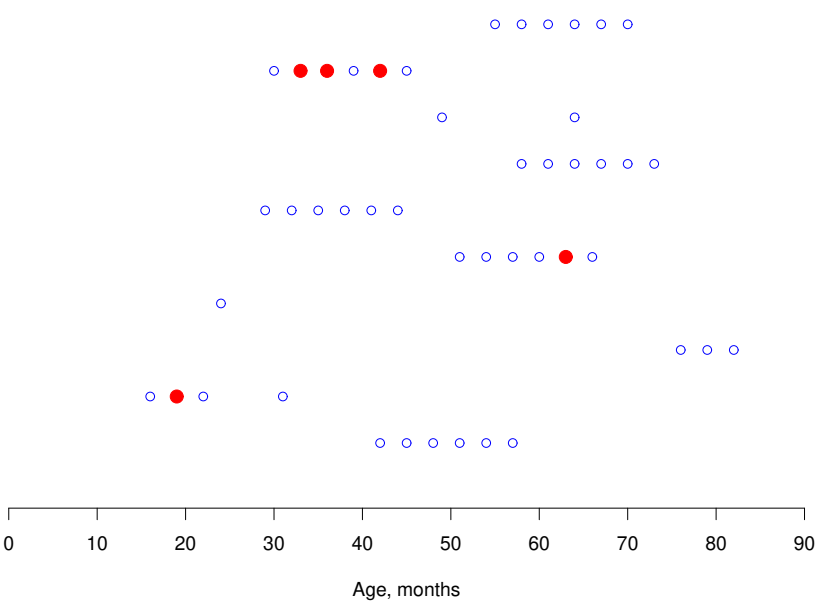
$$\begin{aligned}\mathcal{L}(\boldsymbol{\beta}^*, \boldsymbol{\alpha}) &= \prod_{k=1}^{K} \int f_{\boldsymbol{Y}|\boldsymbol{\gamma}}(\boldsymbol{Y}_k|\ \boldsymbol{\beta}^*, \boldsymbol{\alpha}, \boldsymbol{\gamma}_k)\ f_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}_k|\ \boldsymbol{\alpha})\ \partial\boldsymbol{\gamma}_k \\ &= \prod_{k=1}^{K} \int \left\{ \prod_{i=1}^{n_k} f_{Y|\boldsymbol{\gamma}}(Y_{ki}|\ \boldsymbol{\beta}^*, \boldsymbol{\alpha}, \boldsymbol{\gamma}_k) \right\}\ f_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}_k|\ \boldsymbol{\alpha})\ \partial\boldsymbol{\gamma}_k\end{aligned}$$

  - ⋆ integrate cluster-specific contributions over the distribution of the random effects
  - ⋆ solely a function of the unknown $(\boldsymbol{\beta}^*, \boldsymbol{\alpha})$
  - ⋆ proceed via ML or REML

- Estimates for the cluster-specific random effects, $\boldsymbol{\gamma}_k$, can be obtained via empirical Bayes

Indonesian Child Health Survey data

- Towards extending the mixed effects framework beyond continuous response data, let's consider the ICHS data
  - ⋆ goal is to study the relationship between vitamin A deficiency and risk of respiratory infection
  - ⋆ diagnosis of xerophthalmia serves as a surrogate for vitamin A deficiency
- Longitudinal binary response data on K=275 pre-school children



- In Part IV of the notes we built a series of *marginal models* for the relationship between presence/absence of xerophthalmia and risk of respiratory infection:

$$\mu_{ki} \;=\; \mathrm{E}[Y_{ki}|\; \boldsymbol{X}_{ki}] \;=\; \Pr(Y_{ki}=1|\; \boldsymbol{X}_{ki})$$

  - ⋆ marginal with respect to cluster membership
  - ⋆ separate specification of a working covariance model:

$$\boldsymbol{V}_k(\boldsymbol{\beta},\boldsymbol{\alpha}) \;=\; \boldsymbol{S}_k(\boldsymbol{\beta})^{1/2}\boldsymbol{R}_k(\boldsymbol{\alpha})\boldsymbol{S}_k(\boldsymbol{\beta})^{1/2}$$

  - ⋆ estimation/inference via *generalized estimating equations* (GEE)

Q: Why might we be interested in considering the mixed effects framework?

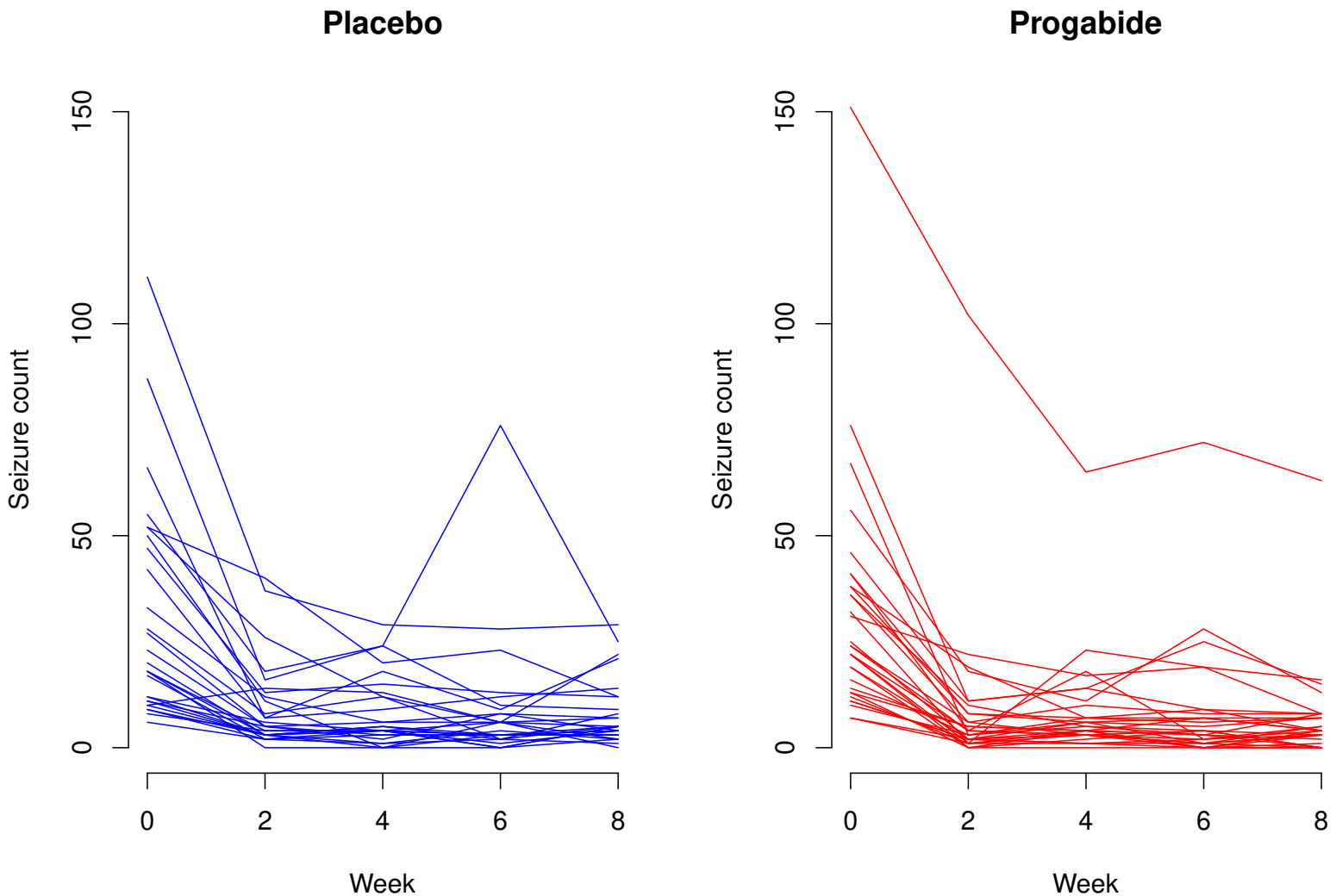
Seizure data

- As a second example, consider the `seizure` data, available on the course website, from a study conducted to evaluate the efficacy of progabide in reducing partial seizures
  - ⋆ Leppik et al (*Neurology*, 1987)
  - ⋆ Thall and Vail (*Biometrics*, 1990)
- Data consists of K=59 individuals
  - ⋆ a total of 75 patients were enrolled but only 59 had no 'protocol violations'
- Each patient had an initial eight-week baseline assessment followed by four consecutive two-week assessments
  - ⋆ actually a crossover trial although we only have data from the first phase

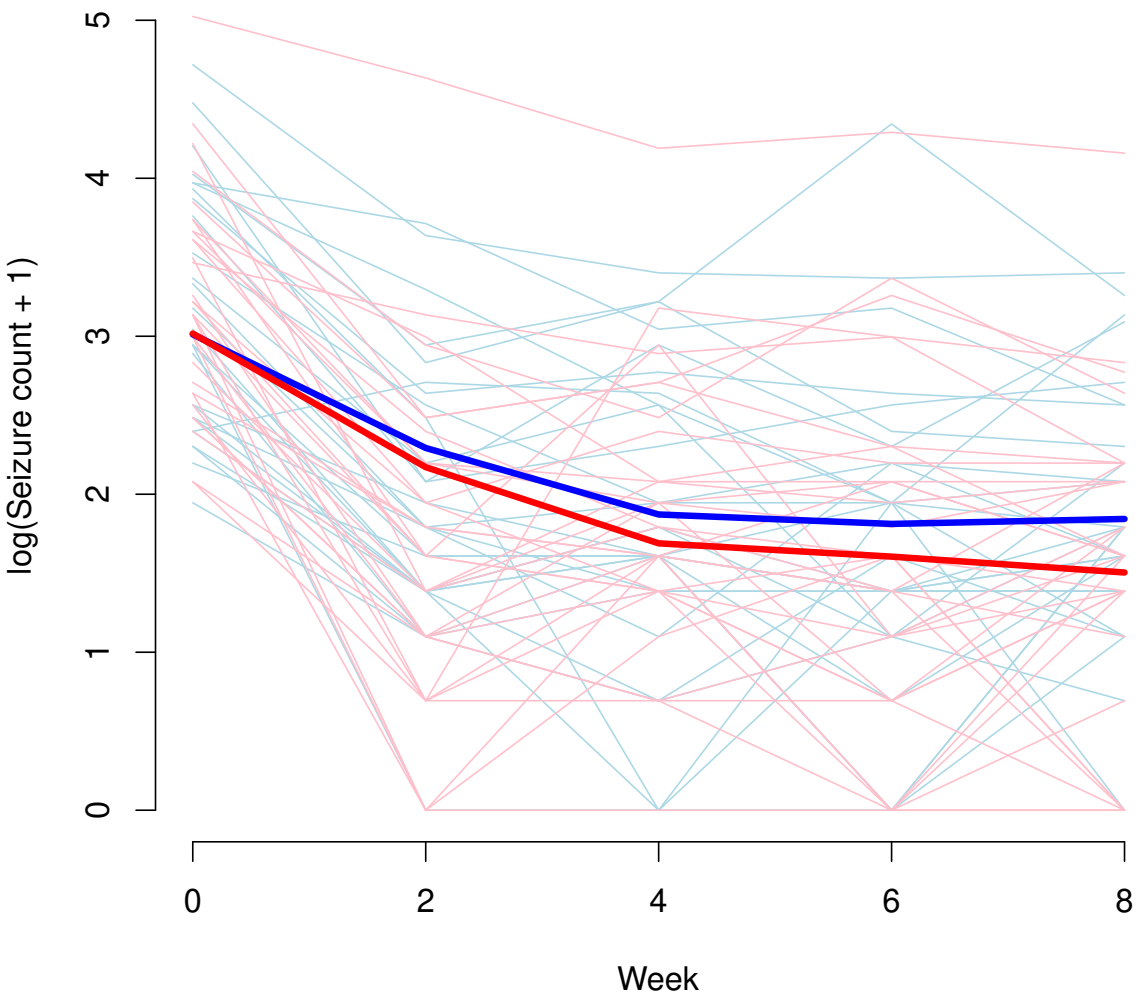
```
> ##
> load("Seizure.RData")
>
> head(seizure)
    id count visit treatment age weeks
1  104    11     0         0  31     8
2  104     5     1         0  31     2
3  104     3     2         0  31     2
4  104     3     3         0  31     2
5  104     3     4         0  31     2
6  106    11     0         0  30     8
>
> table(table(seizure$id))

 5
59
>
> summary(seizure$age[seizure$visit == 0])
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  18.00   23.00   28.00   28.34   32.00   42.00
```

- Observed seizure count data by treatment group:



- Observed log-transformed seizure count data by treatment group, together with a lowess smoother:



- Exploratory data analysis is suggestive of:
  - ⋆ **trends**
    - ∗ substantial decrease until week 4
    - ∗ leveling off after week 4
    - ∗ small differences between the two treatment groups, especially after week 6
  - ⋆ **variability**
    - ∗ large variation at baseline
    - ∗ variation persists over time
  - ⋆ **outliers**
    - ∗ subject 227 (placebo) had a large increase between weeks 4 and 6
    - ∗ subject 207 (progabide) had consistently high seizure counts

Q: Moving beyond EDA, why might we be interested in considering the mixed effects framework?

Generalized linear mixed models

Q: Can we extend the linear mixed model framework for continuous responses to:

- ⋆ binary responses, as in the ICHS data?
- ⋆ count responses, as in the seizure data?

- For marginal modeling, we extended weighted least squares for linear models to generalized estimating equations for generalized linear models
  - ⋆ link function
  - ⋆ heteroskedasticity, including mean-variance relationships
  - ⋆ estimation/inference followed on the basis of the same principles

- We can follow the same strategy when extending linear mixed models

Overarching strategy

- Assume each cluster has a regression model characterized by a link function and a set of covariates
  - ⋆ as in GLMs from Methods I
  - ⋆ as in the marginal model specifications from Part IV
- Structure the cluster-specific regressions across the population of clusters via a series of
  - ⋆ *fixed effects* parameters that are common to all clusters in the population
  - ⋆ *random effects* parameters that permit cluster-specific perturbations
- Perform likelihood-based estimation/inference

Notation

$$\boldsymbol{Y}_k = (Y_{k1},\ Y_{k2},\ \ldots,\ Y_{kn_k})^T \qquad \text{Response vector}$$

$$\boldsymbol{\beta}^* = (\beta_1^*,\ \beta_2^*,\ \ldots,\ \beta_p^*)^T \qquad \text{Fixed effects}$$

$$\boldsymbol{X}_{ki} = (X_{ki,1},\ X_{ki,2},\ \ldots,\ X_{ki,p})$$

$$\boldsymbol{X}_k = (\boldsymbol{X}_{k1},\ \boldsymbol{X}_{k2},\ \ldots,\ \boldsymbol{X}_{kn_k})^T \qquad \text{Design matrix for the fixed effects}$$

- $\star$ $n_k \times p$

$$\boldsymbol{\gamma}_k = (\gamma_{k1},\ \gamma_{k2},\ \ldots,\ \gamma_{kq})^T \qquad \text{Random effects}$$

$$\boldsymbol{Z}_{ki} = (Z_{ki,1},\ Z_{ki,2},\ \ldots,\ Z_{ki,q})$$

$$\boldsymbol{Z}_k = (\boldsymbol{Z}_{k1},\ \boldsymbol{Z}_{k2},\ \ldots,\ \boldsymbol{Z}_{kn_k})^T \qquad \text{Design matrix for the random effects}$$

- $\star$ $n_k \times q$

- Notice the use of the '$*$' superscript to distinguish the parameters from those in Part IV of the notes
  - $\star$ distinguish marginal vs. conditional parameters

Definition

- By a *generalized linear mixed effects model*, we mean a statistical model with the following components/assumptions:

(1) a conditional mean model:

$$\mu_{ki} \;=\; g^{-1}(\boldsymbol{X}_{ki}\boldsymbol{\beta}^* \;+\; \boldsymbol{Z}_{ki}\boldsymbol{\gamma}_k)$$

(2) assumptions regarding the random effects, specifically that the $\boldsymbol{\gamma}_k$ are i.i.d from some distribution F_γ with $\mathsf{E}[\boldsymbol{\gamma}_k] = \mathbf{0}$ and $\mathsf{Cov}[\boldsymbol{\gamma}_k] = \boldsymbol{G}(\boldsymbol{\alpha})$

(3) Y_{ki} is distributed according to some member of the *exponential dispersion family*

$$f_Y(y_{ki}; \theta_{ki}, \phi) \;=\; \exp\left\{\frac{y_{ki}\theta_{ki} - b(\theta_{ki})}{\phi} \;+\; c(y_{ki}, \phi)\right\}$$

(4) assumption of conditional independence for the elements of $\boldsymbol{Y}_k$, given $\boldsymbol{\gamma}_k$

Comments

- As in linear mixed models, the components of the linear predictor serve different purposes:
  - ⋆ $\boldsymbol{\beta}^*$: *fixed effects* that are common to all subjects
    - ∗ determine the shape of the underlying 'population' regression
  - ⋆ $\boldsymbol{\gamma}_k$: *random effects* that are cluster-specific
    - ∗ determine perturbations around the 'population' regression

- By restricting attention to the exponential dispersion family, the first two conditional moments can be conveniently written as:

$$\begin{aligned} \mathsf{E}[Y_{ki} | \ \boldsymbol{\gamma}_k] &= \mu_{ki} \\ &= b'(\theta_{ki}) \\ \mathsf{V}[Y_{ki} | \ \boldsymbol{\gamma}_k] &= \phi V(\mu_{ki}) \\ &= \phi b''(\theta_{ki}) \end{aligned}$$

Estimation/inference

- We are going to consider two likelihood-based strategies

  1. **Conditional likelihood:**
     - treat the random effects as if they are fixed (unknown) parameters and *eliminate* them by conditioning on their sufficient statistics
     - does not require specification or even consideration of the distribution F_γ

  2. **Maximum likelihood:**
     - treat the random effects as unobserved latent variables and *integrate* over their assumed distribution
     - requires specification of the distribution F_γ
     - typically assume $F_\gamma \equiv \text{MVN}(\mathbf{0}, \boldsymbol{G}(\boldsymbol{\alpha}))$

Conditional likelihood

- For $\phi = 1$, treating $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_K)$ as fixed (unknown) parameters and given that the distribution of Y_{ki} belongs to the exponential dispersion family, we have that the likelihood is proportional to

$$\mathcal{L}(\boldsymbol{\beta}^*, \boldsymbol{\gamma}) \;=\; \prod_{k=1}^{K} \prod_{i=1}^{n_k} f_Y(Y_{ki}; \theta_{ki}) \;\propto\; \prod_{k=1}^{K} \prod_{i=1}^{n_k} \exp\{Y_{ki}\theta_{ki} - b(\theta_{ki})\}$$

  where $\theta_{ki} \equiv \theta_{ki}(\boldsymbol{\beta}^*, \boldsymbol{\gamma}_k)$

- If the canonical link function is adopted, then

$$\theta_{ki} \;=\; \boldsymbol{X}_{ki}\boldsymbol{\beta}^* + \boldsymbol{Z}_{ki}\boldsymbol{\gamma}_k$$

  and the likelihood can be written as

$$\mathcal{L}(\boldsymbol{\beta}^*, \boldsymbol{\gamma}) \;\propto\; \exp\left\{ \boldsymbol{\beta}^* \sum_{k=1}^{K} \sum_{i=1}^{n_k} \boldsymbol{X}_{ki} Y_{ki} \;+\; \sum_{k=1}^{K} \boldsymbol{\gamma}_k \sum_{i=1}^{n_k} \boldsymbol{Z}_{ki} Y_{ki} \;-\; \sum_{k=1}^{K} \sum_{i=1}^{n_k} b(\theta_{ki}) \right\}$$

- We therefore see that the sufficient statistics for $\boldsymbol{\beta}^*$ and $\boldsymbol{\gamma}_k$ are:

$$\begin{aligned} T_{\beta^*} &= \sum_{k=1}^{K} \sum_{i=1}^{n_k} \boldsymbol{X}_{ki} Y_{ki} \\ &= \sum_{k=1}^{K} T_{\beta^*,k} \\ T_{\gamma_k} &= \sum_{i=1}^{n_k} \boldsymbol{Z}_{ki} Y_{ki} \end{aligned}$$

- By conditioning on the observed T_{γ_k}, the resulting density should be independent of $\boldsymbol{\gamma}_k$:

$$\begin{aligned} f(\boldsymbol{Y}_k |\, T_{\gamma_k}, \boldsymbol{\beta}^*) &= \frac{f(\boldsymbol{Y}_k |\, \boldsymbol{\beta}^*, \boldsymbol{\gamma}_k)}{f(T_{\gamma_k} |\, \boldsymbol{\beta}^*, \boldsymbol{\gamma}_k)} \\ &= \frac{f(T_{\beta^*,k}, T_{\gamma_k} |\, \boldsymbol{\beta}^*, \boldsymbol{\gamma}_k)}{f(T_{\gamma_k} |\, \boldsymbol{\beta}^*, \boldsymbol{\gamma}_k)} \end{aligned}$$

- For discrete outcomes, the latter can be written as:

$$\frac{f(T_{\beta^*,k}, T_{\gamma_k} |\, \boldsymbol{\beta}^*, \boldsymbol{\gamma}_k)}{f(T_{\gamma_k} |\, \boldsymbol{\beta}^*, \boldsymbol{\gamma}_k)} = \frac{\sum_{\Omega_{k1}} \exp\{T_{\beta^*,k}\boldsymbol{\beta}^* + T_{\gamma_k}\boldsymbol{\gamma}_k\}}{\sum_{\Omega_{k2}} \exp\{\boldsymbol{\beta}^* \sum_{i=1}^{n_k} \boldsymbol{X}_{ki} Y_{ki} + T_{\gamma_k}\boldsymbol{\gamma}_k\}}$$

where

$$\Omega_{k1} = \left\{ \boldsymbol{Y}_k |\, \sum_{i=1}^{n_k} \boldsymbol{X}_{ki} Y_{ki} = T_{\beta^*,k} \quad \text{and} \quad \sum_{i=1}^{n_k} \boldsymbol{Z}_{ki} Y_{ki} = T_{\gamma_k} \right\}$$

$$\Omega_{k2} = \left\{ \boldsymbol{Y}_k |\, \sum_{i=1}^{n_k} \boldsymbol{Z}_{ki} Y_{ki} = T_{\gamma_k} \right\}$$

- Note, this expression can be reduced to

$$f(\boldsymbol{Y}_k |\, T_{\gamma_k}, \boldsymbol{\beta}^*) = \frac{\sum_{\Omega_{k1}} \exp\{T_{\beta^*,k}\boldsymbol{\beta}^*\}}{\sum_{\Omega_{k1}} \exp\{\boldsymbol{\beta}^* \sum_{i=1}^{n_k} \boldsymbol{X}_{ki} Y_{ki}\}}$$

which is solely a function of $\boldsymbol{\beta}^*$

- Assuming contributions across clusters are independent, estimation/inference can proceed via the *conditional likelihood*:

$$\mathcal{L}^c(\boldsymbol{\beta}^*) \;=\; \prod_{k=1}^{K} \frac{\sum_{\Omega_{k1}} \exp\{T_{\beta^*,k}\boldsymbol{\beta}^*\}}{\sum_{\Omega_{k2}} \exp\{\boldsymbol{\beta}^* \sum_{i=1}^{n_k} \boldsymbol{X}_{ki} Y_{ki}\}}$$

  - ⋆ obtain the log-likelihood and maximize with respect to $\boldsymbol{\beta}^*$
  - ⋆ second partial derivatives to give the information matrix on which inference can be based

Larger information, larger curvature of the likelihood function

Maximum likelihood

- While use of the conditional likelihood is appealing in that one does not have to make any assumptions about the distribution of the γ_k across the population of clusters, it also means that one cannot learn about the distribution of the γ_k across the population of clusters
  - ⋆ cannot quantify variation in the γ_k
  - ⋆ cannot distinguish sources of variation
  - ⋆ cannot estimate cluster-specific profiles

- As an alternative we can follow the strategy used in linear mixed models by treating the γ_k as unobserved latent factors and integrate them out over some adopted distribution:

$$\mathcal{L}(\boldsymbol{\beta}^*, \boldsymbol{\alpha}) \;=\; \prod_{k=1}^{K} \int f_{\boldsymbol{Y}|\boldsymbol{\gamma}}(\boldsymbol{Y}_k|\; \boldsymbol{\beta}^*, \boldsymbol{\alpha}, \boldsymbol{\gamma}_k)\; f_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}_k|\; \boldsymbol{\alpha})\; \partial \boldsymbol{\gamma}_k$$

where $f_{\boldsymbol{\gamma}}(\cdot|\; \boldsymbol{\alpha})$ is the density of the distribution for the $\boldsymbol{\gamma}_k$

- The induced *marginal likelihood* is a function of both $\boldsymbol{\beta}^*$, the regression parameters of interest and $\boldsymbol{\alpha}$, the parameters that index the distribution of the random effects

- For the linear mixed model, if one adopts a multivariate Normal distribution for the $\boldsymbol{\gamma}_k$ then the integral has a closed form expression

- For GLMMs, it will seldom be the case that the integral has a closed form expression
  - ⋆ use approximations such as the Laplace approximation or Gauss-Hermite quadrature

- Estimation/inference then follows as in linear mixed models
  - ⋆ for $\boldsymbol{\beta}^*$, one has all of the usual options (i.e. Wald, score and LRT)
  - ⋆ for variance components, one can use the LRT but with the asymptotic sampling distribution of the test statistic taken as an appropriate mixture of χ^2 distributions

- For estimation of the random effects, recall the empirical Bayes estimates in the linear mixed model was taken to be

$$\widetilde{\boldsymbol{\gamma}}_k \;=\; \mathsf{E}[\boldsymbol{\gamma}_k|\; \boldsymbol{Y}_k, \boldsymbol{\beta}^*, \boldsymbol{\alpha}],$$

  the mean of the posterior distribution of $\boldsymbol{\gamma}_k$ that arises when one views ($\boldsymbol{\beta}^*$, $\boldsymbol{\alpha}$) as known:

$$\pi(\boldsymbol{\gamma}_k|\; \boldsymbol{Y}_k, \boldsymbol{\beta}^*, \boldsymbol{\alpha}) \;\propto\; \mathcal{L}(\boldsymbol{\gamma}_k; \boldsymbol{Y}_k, \boldsymbol{\beta}^*, \boldsymbol{\alpha})\pi(\boldsymbol{\gamma}_k|\; \boldsymbol{\alpha})$$

- While we were able to derive closed-form expression for $\widetilde{\boldsymbol{\gamma}}_k$ in the linear mixed model case, this isn't possible more generally

- However, we can write the empirical Bayes estimate as

$$\widetilde{\boldsymbol{\gamma}}_k \;=\; \frac{\int \boldsymbol{\gamma}_k \mathcal{L}(\boldsymbol{\gamma}_k; \boldsymbol{Y}_k, \boldsymbol{\beta}^*, \boldsymbol{\alpha})\pi(\boldsymbol{\gamma}_k|\; \boldsymbol{\alpha})\; \partial\boldsymbol{\gamma}_k}{\int \mathcal{L}(\boldsymbol{\gamma}_k; \boldsymbol{Y}_k, \boldsymbol{\beta}^*, \boldsymbol{\alpha})\pi(\boldsymbol{\gamma}_k|\; \boldsymbol{\alpha})\; \partial\boldsymbol{\gamma}_k}$$

  which can be computed by approximating each of the integrals using any of the tools we use to evaluate the marginal likelihood

Gauss-Hermite quadrature

- Suppose we wanted to compute an integral of the following form:

$$\int_{-\infty}^{+\infty} h(x) \exp\{-x^2\} \, \partial x$$

Gaussian kernel

where $h(\cdot)$ is an arbitrary function

- A Monte Carlo approximation to this integral would involve:

  (1) selecting a large number of values of $x \in \mathbb{R}$, and

  (2) evaluating '$h(x) \exp\{-x^2\}$' at each selected point

  (3) computing the appropriate sum

- Depending on the form of $h(\cdot)$ this can be highly inefficient
  - ⋆ may need a very large number of x for the approximation to be accurate

- The Gauss-Hermite quadrature approximation to this integral is

$$\int_{-\infty}^{+\infty} h(x) \exp\{-x^2\} \, \partial x \;\approx\; \sum_{j=1}^{m} h(x_j) w_j$$

where

  - ⋆ m is the number of 'nodes'
  - ⋆ $\{x_j; j = 1, \ldots, m\}$ is the set of so-called quadrature points, obtained at the roots of the *physicists' Hermite polynomial*

$$H_m(x) \;=\; (-1)^m \exp\{x^2\} \frac{\partial^m}{\partial x^m} \exp\{-x^2\}$$

    - ∗ it turns out there exists a *probabilists' Hermite polynomial*!
  - ⋆ $\{w_j; j = 1, \ldots, m\}$ is a set of associated weights, each given by

$$w_j \;=\; \frac{2^{m-1} m! \sqrt{(\pi)}}{m^2 \left[H_{m-1}(x_j)\right]^2}$$

- One useful application of Gauss-Hermite quadrature is in approximating integrals that represent expectations with respect to a Normal(μ, σ^2) distribution

$$\mathsf{E}[h(\gamma)] \;=\; \int_{-\infty}^{+\infty} h(\gamma)\; \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(\gamma-\mu)^2}{2\sigma^2}\right\} \partial\gamma$$

- Performing a change of variables

$$x \;=\; \frac{\gamma-\mu}{\sqrt{2}\sigma} \qquad \Longleftrightarrow \qquad \gamma \;=\; \sqrt{2}\sigma x + \mu$$

and using integration by substitution we get

$$\begin{aligned}\mathsf{E}[h(\gamma)] \;&=\; \int_{-\infty}^{+\infty} h(\sqrt{2}\sigma x+\mu)\; \frac{1}{\sqrt{\pi}} \exp\{-x^2\} \partial x \\ &\approx\; \frac{1}{\sqrt{\pi}} \sum_{j=1}^{m} h(\sqrt{2}\sigma x_j + \mu) w_j\end{aligned}$$

- Now consider a random intercepts GLMM:

$$\begin{aligned} Y_{ki}|\ \boldsymbol{X}_{ki}, \gamma_{0k} &\sim \mathsf{F}_{Y|\gamma}(\cdot) \\ g(\mu_{ki}) &= \boldsymbol{X}_{ki}\boldsymbol{\beta}^* + \gamma_{0k} \\ \gamma_{0k} &\sim \mathsf{Normal}(0,\ \sigma^2_\gamma) \end{aligned}$$

- Assuming conditional independence among the components of $\boldsymbol{Y}_k$ given γ_{0k}, the contribution to the marginal likelihood from the k^{th} cluster is:

$$\mathcal{L}_k(\boldsymbol{\beta}^*, \sigma^2_\gamma) = \int \prod_{i=1}^{n_k} f_{\boldsymbol{Y}|\boldsymbol{\gamma}}(Y_{ki}|\ \boldsymbol{\beta}^*, \gamma_{0k})\ f_{\boldsymbol{\gamma}}(\gamma_{0k}|\ \sigma^2_\gamma)\ \partial\gamma_{0k}$$

- We can use Gauss-Hermite quadrature to approximate this integral by noting that

$$\mathcal{L}_k(\boldsymbol{\beta}^*, \sigma^2_\gamma) = \mathsf{E}[h(\gamma_{0k})]$$

where

$$h(\gamma_{0k}) = \prod_{i=1}^{n_k} f_{\boldsymbol{Y}|\boldsymbol{\gamma}}(Y_{ki}|\ \boldsymbol{\beta}^*, \gamma_{0k})$$

- It's worth noting that as one increases the number of nodes there is a trade-off between the accuracy of the approximation and the computational time

- In `R` there are a number of packages that facilitate univariate Gauss-Hermite quadrature
  - ⋆ `fastGHquad`
  - ⋆ `gaussquad`

- The `MultiGHQuad` package has implemented a multivariate version
  - ⋆ grid of quadrature points
  - ⋆ I'd advise some degree of caution, though, as I was unable to download the main article that is cited as a reference

- Finally, although details aren't presented here, it turns out that Gauss-Hermite quadrature with m=1 is equivalent to the Laplace approximation
  - ⋆ Liu and Pierce (Biometrika, 1994)

Binary response data

Q: Returning to the ICHS data, a key question of interest is whether children who are vitamin A deficient have an increased risk of respiratory infection?

- In developing a GLMM for this question, we can follow the same general principles laid out in Parts III and IV of the notes
  - ⋆ choice of the linear predictor
    - ∗ what to include in $\boldsymbol{X}_{ki}$?
    - ∗ what to include in $\boldsymbol{Z}_{ki}$?
  - ⋆ choice of the link function
    - ∗ e.g. logit link or log link
  - ⋆ if necessary, the choice of distributional assumptions regarding the random effects
    - ∗ typically a MVN($\boldsymbol{0}$, $\boldsymbol{G}(\boldsymbol{\alpha})$)

Conditional likelihood

- Consider the random intercepts logistic model:

$$\text{logit}\,\Pr(Y_{ki}=1|\ \boldsymbol{X}_{ki}) \ = \ \beta_0^* \ + \ \beta_1^*\text{Xerophthalmia}_{ki} \ + \ \beta_2^*\text{Age}_{ki} \ + \ \gamma_{0k}$$

- Treating the $\{\gamma_{0k};\ k=1,\ldots,K\}$ as fixed unknown parameters, and given that the logit link is the canonical link function, we can write the likelihood as:

$$\begin{aligned}\mathcal{L}(\boldsymbol{\beta}^*,\boldsymbol{\gamma}) \ &\propto \ \prod_{k=1}^{K}\prod_{i=1}^{n_k}\exp\{Y_{ki}\text{logit}(\mu_{ki}) \ - \ \log(1-\mu_{ki})\} \\ &= \ \prod_{k=1}^{K}\exp\left\{\boldsymbol{\beta}^*\sum_{i=1}^{n_k}\boldsymbol{X}_{ki}Y_{ki} \ + \ \gamma_{0k}\sum_{i=1}^{n_k}Y_{ki}\right. \\ &\qquad\left. - \ \sum_{i=1}^{n_k}\log(1-\exp\{\boldsymbol{X}_{ki}\boldsymbol{\beta}^* \ + \ \gamma_{0k}\})\right\}\end{aligned}$$

- Since the model only includes a random intercept, the sufficient statistic for γ_{0k} is the sum of the responses for the k^{th} child:

$$T_{\gamma_k} = \sum_{i=1}^{n_k} Y_{ki}$$

- The conditional likelihood for $\boldsymbol{\beta}^*$ is therefore proportional to the conditional distribution of the response vector $\boldsymbol{Y}_k$ given T_{γ_k}:

$$\mathcal{L}^c(\boldsymbol{\beta}^*) = \prod_{k=1}^{K} \frac{\sum_{\Omega_{k1}} \exp\{T_{\beta^*,k}\boldsymbol{\beta}^*\}}{\sum_{\Omega_{k2}} \exp\{\boldsymbol{\beta}^* \sum_{i=1}^{n_k} \boldsymbol{X}_{ki} Y_{ki}\}}$$

where

$$\Omega_{k1} = \left\{ \boldsymbol{Y}_k | \sum_{i=1}^{n_k} \boldsymbol{X}_{ki} Y_{ki} = T_{\beta^*,k} \quad \text{and} \quad \sum_{i=1}^{n_k} Y_{ki} = T_{\gamma_k} \right\}$$

$$\Omega_{k2} = \left\{ \boldsymbol{Y}_k | \sum_{i=1}^{n_k} Y_{ki} = T_{\gamma_k} \right\}$$

Paired binary data

- To gain some insight into conditional likelihood let's consider the special case of the setting where each of K patients is measured prior to and post the administration of some treatment
  - ⋆ patients, in a sense, serve as their own controls
- Let $\boldsymbol{Y}_k = (Y_{k0}, Y_{k1})$ denote the response vector and $\boldsymbol{X}_k = (X_{k0}, X_{k1}) = (0, 1)$ the corresponding covariate vector
  - ⋆ i.e. $\boldsymbol{X}_k$ is the same $\forall\ k$
- For this setting, one might consider the model:

$$\text{logit}\,\Pr(Y_{ki} = 1|\ \boldsymbol{X}_{ki}) \ = \ \gamma_k + \ \beta^* X_{ki}$$

  for $k = 1, \ldots, K$, and $i = 0, 1$
  - ⋆ between-subject heterogeneity in risk at baseline is represented by the cluster-specific γ_k

- Note, the joint probabilities for any given set of outcomes (i.e. the $\mathsf{P}(Y_{k0},\ Y_{k1})$) given this model are:

	$\boxed{Y_{k1}=0}$	$\boxed{Y_{k1}=1}$
$\boxed{Y_{k0}=0}$	$\frac{1}{1+\exp\{\gamma_k\}}\,\frac{1}{1+\exp\{\gamma_k+\beta^*\}}$	$\frac{1}{1+\exp\{\gamma_k\}}\,\frac{\exp\{\gamma_k+\beta^*\}}{1+\exp\{\gamma_k+\beta^*\}}$
$\boxed{Y_{k0}=1}$	$\frac{\exp\{\gamma_k\}}{1+\exp\{\gamma_k\}}\,\frac{1}{1+\exp\{\gamma_k+\beta^*\}}$	$\frac{\exp\{\gamma_k\}}{1+\exp\{\gamma_k\}}\,\frac{\exp\{\gamma_k+\beta^*\}}{1+\exp\{\gamma_k+\beta^*\}}$

- Treating $\boldsymbol{\gamma}=\{\gamma_1,\ldots,\gamma_K\}$ as fixed parameters, the full likelihood is:

$$\begin{aligned}\mathcal{L}(\beta^*,\boldsymbol{\gamma}) &= \prod_{k=1}^{K}\prod_{i=0}^{1}\frac{\exp\{[\gamma_k+\beta^*X_{ki}]Y_{ki}\}}{1+\exp\{\gamma_k+\beta^*X_{ki}\}}\\ &= \prod_{k=1}^{K}\frac{\exp\{\gamma_kY_{k\cdot}+\beta^*Y_{k1}\}}{[1+\exp\{\gamma_k\}][1+\exp\{\gamma_k+\beta^*\}]}\end{aligned}$$

where $Y_{k\cdot}=Y_{k0}+Y_{k1}$

- As $Y_{k\cdot}$ is the sufficient statistic for γ_k, the conditional likelihood is given by:

$$\mathcal{L}^c(\boldsymbol{\beta}^*) \propto \prod_{k=1}^{K} \mathsf{P}(\boldsymbol{Y}_k | \, Y_{k\cdot}) = \prod_{k=1}^{K} \frac{\mathsf{P}(Y_{k0}, \, Y_{k1})}{\mathsf{P}(Y_{k\cdot})}$$

- Note, since observed values of $Y_{k\cdot}$ can be in $\{0, 1, 2\}$, the contributions to the conditional likelihood simplify considerably for certain observed data patterns for (Y_{k0}, Y_{k1})
  - ⋆ if the observed $Y_{k\cdot}$=0 or 2, there is only one possible combination of the observed (Y_{k0}, Y_{k1}) and

$$\frac{\mathsf{P}(Y_{k0}, \, Y_{k1})}{\mathsf{P}(Y_{k\cdot})} = 1$$

- Hence, only those pairs for which the observed responses are discordant contribute to the observed data conditional likelihood
  - ⋆ similar phenomenon arises in conditional logistic regression for matched case-control studies

- The contribution for a patient with $(Y_{k0}, Y_{k1})=(0,1)$ is:

$$\begin{aligned}
\frac{\mathsf{P}(Y_{k0},\ Y_{k1})}{\mathsf{P}(Y_{k\cdot})} &= \frac{\mathsf{P}(Y_{k0}=0,\ Y_{k1}=1)}{\mathsf{P}(Y_{k0}=0,\ Y_{k1}=1)+\mathsf{P}(Y_{k0}=1,\ Y_{k1}=0)} \\
&= \frac{\frac{1}{1+\exp\{\gamma_k\}}\frac{\exp\{\gamma_k+\beta^*\}}{1+\exp\{\gamma_k+\beta^*\}}}{\frac{1}{1+\exp\{\gamma_k\}}\frac{\exp\{\gamma_k+\beta^*\}}{1+\exp\{\gamma_k+\beta^*\}}+\frac{\exp\{\gamma_k\}}{1+\exp\{\gamma_k\}}\frac{1}{1+\exp\{\gamma_k+\beta^*\}}} \\
&= \frac{\exp\{\beta^*\}}{1+\exp\{\beta^*\}}
\end{aligned}$$

- Similarly, the contribution for a patient with $(Y_{k0}, Y_{k1})=(1,0)$ is:

$$\begin{aligned}
\frac{\mathsf{P}(Y_{k0},\ Y_{k1})}{\mathsf{P}(Y_{k\cdot})} &= \frac{\mathsf{P}(Y_{k0}=1,\ Y_{k1}=0)}{\mathsf{P}(Y_{k0}=0,\ Y_{k1}=1)+\mathsf{P}(Y_{k0}=1,\ Y_{k1}=0)} \\
&= \frac{\frac{\exp\{\gamma_k\}}{1+\exp\{\gamma_k\}}\frac{1}{1+\exp\{\gamma_k+\beta^*\}}}{\frac{1}{1+\exp\{\gamma_k\}}\frac{\exp\{\gamma_k+\beta^*\}}{1+\exp\{\gamma_k+\beta^*\}}+\frac{\exp\{\gamma_k\}}{1+\exp\{\gamma_k\}}\frac{1}{1+\exp\{\gamma_k+\beta^*\}}} \\
&= \frac{1}{1+\exp\{\beta^*\}}
\end{aligned}$$

- Let K_{01} and K_{10} denote the number of patients with $(Y_{k0}, Y_{k1})=(0,1)$ and $(Y_{k0}, Y_{k1})=(1,0)$, respectively

- The conditional likelihood therefore reduces to:

$$\mathcal{L}^c(\beta^*) \;=\; \left(\frac{\exp\{\beta^*\}}{1+\exp\{\beta^*\}}\right)^{K_{01}} \left(\frac{1}{1+\exp\{\beta^*\}}\right)^{K_{10}}$$

- Taking the log, differentiating, setting to zero and solving yields:

$$\widehat{\beta^*} \;=\; \log K_{01} \;- \log K_{10}$$

so that $\exp\{\widehat{\beta^*}\} = K_{01}/K_{10}$ is the maximum conditional likelihood estimate for the conditional odds ratio

Comment

- While conditional likelihood is appealing in the sense that one does not have to specify a distribution for the random effects, it does suffer from a number of drawbacks
  - ⋆ clusters with n_k=1 cannot contribute
  - ⋆ clusters with $Y_{k\cdot}$=0 or n_k cannot contribute
  - ⋆ one cannot estimate effects for cluster-specific or time-invariant covariates
- Motivates maximum likelihood for the binomial GLMM

Maximum likelihood

- Consider the (more general) logisitic-Normal GLMM:

$$\text{logit } \mu_{ki} \;=\; \boldsymbol{X}_{ki}\boldsymbol{\beta}^* \;+\; \boldsymbol{Z}_{ki}\boldsymbol{\gamma}_k$$

  where the $\boldsymbol{\gamma}_k$ are i.i.d MVN($\mathbf{0}$, $\boldsymbol{G}(\boldsymbol{\alpha})$)

- Perform estimation/inference for ($\boldsymbol{\beta}^*$, $\boldsymbol{\alpha}$) via

$$\mathcal{L}(\boldsymbol{\beta}^*, \boldsymbol{\alpha}) \;=\; \prod_{k=1}^{K} \int f_{\boldsymbol{Y}|\boldsymbol{\gamma}}(\boldsymbol{Y}_k|\ \boldsymbol{\gamma}_k, \boldsymbol{\beta}^*, \boldsymbol{\alpha})\ f_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}_k|\ \boldsymbol{\alpha})\ \partial\boldsymbol{\gamma}_k$$

$$\propto \prod_{k=1}^{K} \int \left\{ \prod_{i=1}^{n_k} \mu_{ki}^{Y_{ki}} (1-\mu_{ki})^{1-Y_{ki}} \right\} 1/|\boldsymbol{G}| \exp\left\{ -\frac{1}{2}\boldsymbol{\gamma}_k \boldsymbol{G}^{-1} \boldsymbol{\gamma}_k^T \right\}\ \partial\boldsymbol{\gamma}_k$$

  - ⋆ use a Laplace approximation
  - ⋆ use Gauss-Hermite quadrature

Fitting GLMMs in R

- In R one can fit a GLMM using the `glmer()` function in the `lme4` library
- The basic call to `glmer()` is has the following elements:

`formula`	model specification (see below)
`data`	dataframe
`family`	a GLM family object (as in `glm()`)
`nAGQ`	integer scalar indicating how the integration is to be approximated for random intercept models

- The `formula` argument provides the means to specify the design matrix for both the fixed effects (i.e. the columns of $\boldsymbol{X}_{ki}$) and the random effects (i.e. the columns of $\boldsymbol{Z}_{ki}$) :

$$\texttt{Y} \sim \texttt{X1 + ... + Xp + (Z1 + ... + Zq | id)}$$

- The `nAGQ` argument for random intercept models:
  - ⋆ default of '1' which corresponds to a Laplace approximation to the integral
  - ⋆ values greater than '1' correspond to adaptive Gauss-Hermite quadrature with values corresponding to the number of nodes in the quadrature formula
  - ⋆ larger values correspond to:
    - ∗ greater accuracy in the approximation
    - ∗ greater computational time

- Note, when there is more than one random effect `glmer()` ignores what is put into the `nAGQ` argument and only uses a Laplace approximation
  - ⋆ approximating a multidimensional integration is tricky!

ICHS data analysis

- At the outset, consider fitting the random intercepts logistic model:

$$\text{logit}\,\Pr(Y_{ki}=1|\,\boldsymbol{X}_{ki}) \;=\; \beta_0^* \;+\; \beta_1^*\text{Xerophthalmia}_{ki} \;+\; \beta_2^*\text{Age}_{ki} \;+\; \gamma_{0k}$$

with $\gamma_{0k} \;\sim\; \text{Normal}(0,\;\sigma_\gamma^2)$

```
> ##
> library(lme4)
> load("ICHS.RData")
> dim(ichs)
[1] 1200    9
> length(unique(ichs$id))
[1] 275
>
> ## Random intercepts with default nAGQ=1
> ##
> fit.RI.01 <- glmer(infection ~ xerop + age + (1 | id),
+                    data=ichs, family=binomial)
```

```
> summary(fit.RI.01)
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation)
...
     AIC      BIC   logLik deviance df.resid
   697.6    718.0   -344.8    689.6     1196
...
Random effects:
 Groups Name        Variance Std.Dev.
 id     (Intercept) 0.8262   0.9089
Number of obs: 1200, groups:  id, 275

Fixed effects:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.682352   0.182644 -14.686  < 2e-16 ***
xerop        0.510920   0.486274   1.051    0.293
age         -0.027285   0.006713  -4.064 4.82e-05 ***
...
>
> exp(0.510920)
[1] 1.666824
```

- Preliminary conclusions:
  - ⋆ estimated (conditional) odds ratio is fairly large
  - ⋆ insufficient evidence of a statistically significant association
    - ∗ likely due to the small number of cases with xerophthalmia (see below)
  - ⋆ fairly substantial between-child heterogeneity (see below)

```
> ##
> table(ichs$infection, ichs$xerop)

     0    1
  0 1045   48
  1  100    7
>
> ## Consider variation in the 'baseline' probabilities
> ## - 36 month old children w/o Xerophthalmia
> ##
> expit <- function(x) exp(x) / (1 + exp(x))
> expit(-2.68235 + c(-2, 0, 2)*0.9089)
[1] 0.01099 0.06402 0.29639
```

Q: Interpretation of $\exp\{\widehat{\beta}_1^*\} = 1.66$?

- Before answering this question, its instructive to consider the marginal logistic regression model:

$$\text{logit}\,\Pr(Y_{ki} = 1 |\ \boldsymbol{X}_{ki}) \ = \ \beta_0 \ + \ \beta_1 \text{Xerophthalmia}_{ki} \ + \ \beta_2 \text{Age}_{ki}$$

- To interpret $\exp\{\beta_1\}$, one could say:

  the odds of a respiratory infection among children of a given age with xerophthalmia are $\exp\{\beta_1\}$ *times the odds of a respiratory infection among children of the same age but without a diagnosis of xerophthalmia*

- Contrast between two populations of children
  - ⋆ in each, all members have the same age and xerophthalmia status

- Of course other factors are free to vary within each of the populations
  - ⋆ e.g. gender

- The odds in each of the populations therefore represent an 'average' over the distribution of these other factors
  - ⋆ hence, the interpretation of $\exp\{\beta_1\}$ as a *marginal* or *population-level* contrast

- Now consider the random intercepts logistic regression model:

$$\text{logit}\,\Pr(Y_{ki}=1|\,\boldsymbol{X}_{ki}) \;=\; \beta_0^* \;+\; \beta_1^*\text{Xerophthalmia}_{ki} \;+\; \beta_2^*\text{Age}_{ki} \;+\; \gamma_{0k}$$

- As we interpret $\exp\{\widehat{\beta}_1^*\}$, we need to incorporate the fact that γ_{0k} is in the model

- One option is to refine the interpretation given for the marginal model:

  the odds of a respiratory infection among children of a given age and value of γ_{0k} with xerophthalmia are estimated to be 1.66 times higher than the odds of a respiratory infection among children of the same age and the same value of γ_{0k} but without a diagnosis of xerophthalmia

- Since γ_{0k} is continuous, however, it is, in principle, unique to the cluster
  - ⋆ i.e. each child has their own γ_{0k}

- A consequence of this is that by holding γ_{0k} 'fixed', the two populations seemingly pertain to the same child
  - ⋆ $\exp\{\beta_1^*\}$ is therefore a contrast between a child and himself/herself
  - ⋆ a *within-subject* or *conditional* contrast

- Some folks find this unappealing
  - ⋆ too close to 'causal' interpretation
  - ⋆ i.e. difference one would see if xerophthalmia was 'switched' from 0 to 1

Q: What do you think?

- Returning to the data analysis, let's consider the impact of increasing the accuracy of the approximation to the integrated likelihood

```
> ## Sample code
> ##
> fit.RI.10 <- glmer(infection ~ xerop + age + (1 | id),
+                    data=ichs, family=binomial,
+                    nAGQ=10)
```

- Results suggest that using GH quadrature is wise although the returns quickly diminish as the number of nodes increases:

	nAGQ=1	nAGQ=5	nAGQ=10	nAGQ=25
β_0^*	-2.6824	-2.6217	-2.6213	-2.6213
β_1^*	0.5109	0.5486	0.5488	0.5488
β_2^*	-0.0273	-0.0274	-0.0274	-0.0274
σ_γ	0.9089	0.8126	0.8120	0.8120

- Finally, lets consider a random intercepts/slopes model:

$$\text{logit}\,\Pr(Y_{ki}=1|\ \boldsymbol{X}_{ki}) \ = \ \beta_0^* \ + \ \beta_1^*\text{Xerophthalmia}_{ki} \ + \ \beta_2^*\text{Age}_{ki} \ + \ \gamma_{0k} \ + \ \text{Age}_{ki}\gamma_{1k}$$

with $\boldsymbol{\gamma}_k \ \sim \ \text{MVN}(\mathbf{0},\ \boldsymbol{G}(\boldsymbol{\alpha}))$

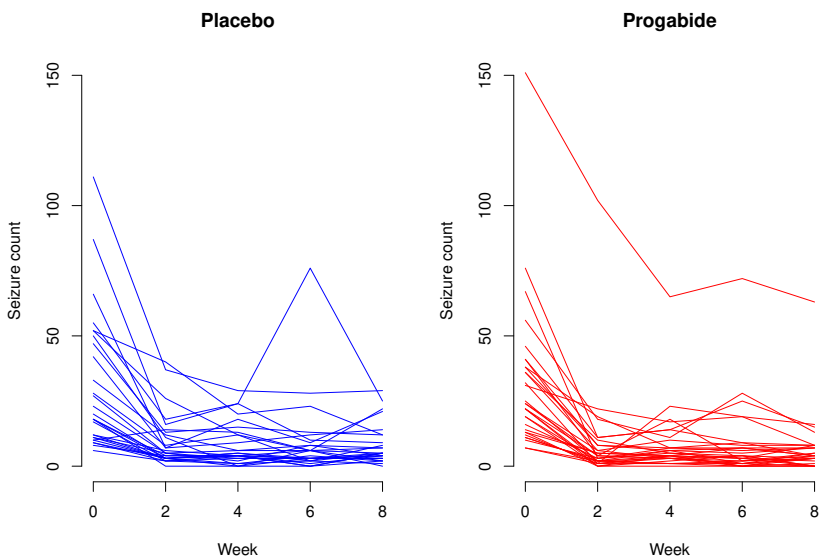
```
> ##
> fit.RIS.01 <- glmer(infection ~ xerop + age + (age | id),
+                     data=ichs, family=binomial)
> summary(fit.RIS.01)
...
    AIC      BIC   logLik deviance df.resid
   700.9    731.5   -344.5    688.9     1194
...
Random effects:
 Groups Name        Variance Std.Dev. Corr
 id     (Intercept) 0.66550  0.8158
        age         0.00013  0.0114   -1.00
...
```

```
Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.62181    0.17881  -14.66   <2e-16 ***
xerop        0.51460    0.47681    1.08   0.2805
age         -0.02247    0.00798   -2.81   0.0049 **
...
>
> ## Compare the LRT statistic to a 1:1 mixture of chi_1^2 and chi_2^2
> ## distributions
> ##
> lrt <- abs(as.numeric(2 * (logLik(fit.RIS.01) - logLik(fit.RI.01))))
> mix.chiSq <- c(rchisq(5e5, df=1), rchisq(5e5, df=2))
> round(c(lrt, mean(mix.chiSq > lrt)), 3)
[1] 0.709 0.552
```

- No evidence to suggest that the introduction random slopes for the age variable improves the fit of the model

Count response data

- Recall the randomized trial of progabide vs placebo:



- From the initial EDA we found indications of
  - ⋆ trends over time
  - ⋆ a moderate treatment effect during follow-up
  - ⋆ large variation at baseline that persisted over time

- Recall, however, that the counts at baseline correspond to an 8 week period, whereas those during follow-up each correspond to 2 week periods

- One might refer to the observed responses as 'Poisson-type' count data
  - ⋆ counts over (possibly varying) time frames
  - ⋆ as opposed to 'Binomial-type' count data, based on a fixed number of trials

When np ~= lambda, a Binomial will converge to a Poisson as n goes to infinity

- When modeling Poisson-type count data, it is common to focus on the rate rather than the mean:

Incidence rate, number of events in one unit of time

1. postulate that:

$$\mathsf{E}[Y_{ki}] \;=\; \mu_{ki} \;=\; \lambda_{ki} t_{ki}$$

   where λ_{ki} is an incidence rate and t_{ki} is the timeframe over which we observe the counts

2. build a regression structure for the λ_{ki} as a function of covariates

- Towards this, let $i \in \{0, 1, 2, 3, 4\}$ indicate time point and consider the following notation:

  Y_{ki} : Number of seizures for patient k during period prior to time point i

  t_{ki} : Number of weeks of observation for patient k during period prior to time point i

  $X_{ki,1}$: Indicator of progabide for patient k

  - $\star$ same $\forall\ i$ within a patient

  $X_{ki,2}$: Binary indicator of pre- vs post-randomization

  $\star$ note that:

$$t_{ki} = \begin{cases} 8 & \text{for } i = 0 \\ 2 & \text{for } i = 1, \ldots, 4 \end{cases}$$

- Using this notation, one GLMM for the seizure count data is:

$$\begin{aligned} Y_{ki} |\ \boldsymbol{X}_{ki}, t_{ki}, \gamma_{0k} &\sim \text{Poisson}(\mu_{ki}) \\ \log \mu_{ki} &= \beta_0^* + \beta_1^* X_{ki,1} + \beta_2^* X_{ki,2} + \beta_3^* X_{ki,1} X_{ki,2} \\ &\quad + \log t_{ki} + \gamma_{0k} \\ \gamma_{0k} &\sim \text{Normal}(0,\ \sigma_\gamma^2) \end{aligned}$$

  - ⋆ a log link function
  - ⋆ interaction term serves to distinguish treatment effects pre- and post-randomization
    - ∗ recall the example used to consider efficiency of repeated measures studies in Part I of the notes
    - ∗ could appeal to randomization and force β_1^* to be zero
  - ⋆ random intercepts to accommodate within-patient correlation

- Fit this model using `glmer()` from the `lme4` package
  - ⋆ use the '`offset`' argument to specify the '$\log t_{ki}$' component

```
> ##
> library(lme4)
> load("Seizure.RData")
> dim(seizure)
[1] 295   6
> head(seizure)
   id count visit treatment age weeks
1 104    11     0         0  31     8
2 104     5     1         0  31     2
3 104     3     2         0  31     2
4 104     3     3         0  31     2
5 104     3     4         0  31     2
6 106    11     0         0  30     8
> length(unique(seizure$id))
[1] 59
>
> ## Data manipulations (to be consistent with the notation in the notes)
> ##
> seizure$X1     <- seizure$treatment
> seizure$X2     <- as.numeric(seizure$visit > 0)
> seizure$X1X2   <- seizure$X1 * seizure$X2
```

- Model that includes a main effect for treatment (i.e. at baseline):

```
>
> fit.RI.0 <- glmer(count ~ X1 * X2 + (1 | id), offset=log(weeks),
+                   family=poisson, data=seizure, nAGQ=25)
>
> summary(fit.RI.0)
...
     AIC      BIC   logLik deviance df.resid
   970.1    988.5   -480.0    960.1      290
...
Random effects:
 Groups Name        Variance Std.Dev.
 id     (Intercept) 0.609    0.7804
Number of obs: 295, groups:  id, 59

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.03259    0.15269   6.763 1.35e-11 ***
X1          -0.02387    0.21067  -0.113   0.9098
X2           0.11080    0.04689   2.363   0.0181 *
X1:X2       -0.10368    0.06505  -1.594   0.1110
...
```

- Model that excludes a main effect for treatment (i.e. at baseline):

```
>
> fit.RI.1 <- glmer(count ~ X2 + X1X2 + (1 | id), offset=log(weeks),
+                   family=poisson, data=seizure, nAGQ=25)
>
> summary(fit.RI.1)
...
     AIC      BIC   logLik deviance df.resid
   968.1    982.8   -480.0    960.1      291
...
Random effects:
 Groups Name        Variance Std.Dev.
 id     (Intercept) 0.6091   0.7804
Number of obs: 295, groups:  id, 59

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.02006    0.10528   9.689   <2e-16 ***
X2           0.11143    0.04657   2.393   0.0167 *
X1X2        -0.10486    0.06423  -1.633   0.1026
...
```

- Now let's consider random intercepts/slopes:

$$\begin{aligned} Y_{ki} |\ \boldsymbol{X}_{ki}, t_{ki}, \gamma_{0k} &\sim \text{Poisson}(\mu_{ki}) \\ \log \mu_{ki} &= \beta_0^* + \beta_1^* X_{ki,1} + \beta_2^* X_{ki,2} + \beta_3^* X_{ki,1} X_{ki,2} \\ &\quad + \log t_{ki} + \gamma_{0k} + X_{ki,2}\gamma_{1k} \\ \boldsymbol{\gamma}_k &\sim \text{MVN}(\mathbf{0},\ \boldsymbol{G}(\boldsymbol{\alpha})) \end{aligned}$$

- When thinking about what the $(\gamma_{0k},\ \gamma_{1k})$ represent, it may be useful to rewrite the mean-model specification as:

$$\log \mu_{ki} = \beta_{0k}^* + \beta_1^* X_{ki,1} + \beta_{2k}^* X_{ki,2} + \beta_3^* X_{ki,1} X_{ki,2} + \log t_{ki}$$

- The random intercepts, γ_{0k}, serve to characterize *heterogeneity in the pre-randomization seizure rates under the placebo regimen* across patients

- The random slopes, γ_{1k}, serve to characterize *heterogeneity in the pre-post differences in seizure rates under the placebo regimen* across patients

```
>
> fit.RIS.0 <- glmer(count ~ X1 * X2 + (X2 | id), offset=log(weeks),
+                    family=poisson, data=seizure)
>
> summary(fit.RIS.0)
...
Random effects:
 Groups Name        Variance Std.Dev. Corr
 id     (Intercept) 0.4999   0.7070
        X2          0.2319   0.4815   0.17
...
Fixed effects:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.071299   0.140267   7.638 2.21e-14 ***
X1           0.049481   0.192717   0.257   0.7974
X2          -0.002394   0.109092  -0.022   0.9825
X1:X2       -0.307159   0.150452  -2.042   0.0412 *
...
```

- ⋆ suggests fairly substantial variation in the cluster-specific pre-post differences across patients
- ⋆ treatment effect is substantially bigger and now statistically significant!

- Formally evaluate the contribution of the random slopes:

```
> ##
> fit.RI.0  <- glmer(count ~ X1 * X2 + (1 | id), offset=log(weeks),
+                    family=poisson, data=seizure)
>
> ## LRT based on a mixture of chi^2 distributions
> ##
> lrt <- abs(as.numeric(2 * (logLik(fit.RIS.0) - logLik(fit.RI.0))))
> mix.chiSq <- c(rchisq(5e5, df=1), rchisq(5e5, df=2))
> round(c(lrt, mean(mix.chiSq > lrt)), 3)
[1] 172.019   0.000
```

- Strong evidence that the random intercepts/slopes model yields a better fit to the observed data than the random intercepts model

- Recall, however, that there appear to be two 'outlier' patients:
  - ⋆ subject 227 (placebo) had a large increase between weeks 4 and 6
  - ⋆ subject 207 (progabide) had consistently high seizure counts

- Investigate whether these patients may have undue influence on the variation in the random slopes
  - ⋆ restrict analyses to remaining 57 patients

```
> ##
> seizureSub   <- seizure[!is.element(seizure$id, c(207, 227)),]
> ##
> fit.RI.0Sub  <- glmer(count ~ X1 * X2 + (1 | id), offset=log(weeks),
+                       family=poisson, data=seizureSub)
> fit.RIS.0Sub <- glmer(count ~ X1 * X2 + (X2 | id), offset=log(weeks),
+                       family=poisson, data=seizureSub)
> ##
> lrt <- abs(as.numeric(2 * (logLik(fit.RIS.0Sub) - logLik(fit.RI.0Sub))))
> mix.chiSq <- c(rchisq(5e5, df=1), rchisq(5e5, df=2))
> round(c(lrt, mean(mix.chiSq > lrt)), 3)
[1] 79.5  0.0
```

  - ⋆ random intercepts/slopes model still provides a significantly better fit

- See that, while reduced, there is still fairly substantial variation in the cluster-specific pre-post differences across patients

```
>
> ##
> summary(fit.RIS.0Sub)
...
Random effects:
 Groups Name        Variance Std.Dev. Corr
 id     (Intercept) 0.4479   0.6693
        X2          0.1943   0.4408   0.00
...
Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.03850    0.13618   7.626 2.42e-14 ***
X1           0.02001    0.18685   0.107    0.915
X2          -0.03224    0.10529  -0.306    0.759
X1:X2       -0.29916    0.14490  -2.065    0.039 *
...
```

- Also see that there is no real impact on the conclusions that one draws regarding the treatment effect

- Finally, returning to the results based on the random intercepts/slopes model using the full data:

```
>
> summary(fit.RIS.0)
...
Fixed effects:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.071299   0.140267   7.638 2.21e-14 ***
X1           0.049481   0.192717   0.257   0.7974
X2          -0.002394   0.109092  -0.022   0.9825
X1:X2       -0.307159   0.150452  -2.042   0.0412 *
...
>
> exp(-0.307159)
[1] 0.7355336
```

Q: Interpretation of $\exp\{-0.307\} = 0.74$?

Marginal vs. conditional parameters

- Throughout the discussion of mixed effects models, we've seen that parameters correspond to cluster-specific contrasts which should be distinguished from the population-averaged contrasts that one estimates via GEE

- In a linear mixed model we saw that the induced marginal model is also linear
  - ⋆ assuming, at least, that the mean of the random effect distribution is zero

- Consequently, even though the two types of contrasts are different they are equivalent numerically for linear models
  - ⋆ if you learn about one then you learn about the other
  - ⋆ one can interpret results from a conditional model as pertaining to marginal contrasts
  - ⋆ analogous to learning about the mean or median for a symmetric distribution

Q: Does the same phenomenon apply to GLMMs? Are the induced marginal mean model and marginal covariance structure readily interpretable?

$$\begin{aligned}
\mathsf{E}[Y_{ki}] &= \mathsf{E}_{\boldsymbol{\gamma}}[\mathsf{E}_{\boldsymbol{Y}}[Y_{ki}|\ \boldsymbol{\gamma}_k]] \\
&= \mathsf{E}_{\boldsymbol{\gamma}}[g^{-1}(\boldsymbol{X}_{ki}\boldsymbol{\beta}^* + \boldsymbol{Z}_{ki}\boldsymbol{\gamma}_k)] \\
&\stackrel{?}{=} \boldsymbol{X}_{ki}\boldsymbol{\beta}^*
\end{aligned}$$

\int \frac{exp(X\beta + Z\gamma)}{1 +exp(X\beta + Z\gamma) } MVN(0, G)d\gamma

Only identity link can do it!

$$\begin{aligned}
\mathsf{V}[Y_{ki}] &= \mathsf{V}_{\boldsymbol{\gamma}}[\mathsf{E}_{\boldsymbol{Y}}[Y_{ki}|\ \boldsymbol{\gamma}_k]] \ + \ \mathsf{E}_{\boldsymbol{\gamma}}[\mathsf{V}_{\boldsymbol{Y}}[Y_{ki}|\ \boldsymbol{\gamma}_k]] \\
&= \mathsf{V}_{\boldsymbol{\gamma}}[g^{-1}(\boldsymbol{X}_{ki}\boldsymbol{\beta}^* + \boldsymbol{Z}_{ki}\boldsymbol{\gamma}_k)] \ + \ \phi\mathsf{E}_{\boldsymbol{\gamma}}[V(\mu_{ki})] \\
&\stackrel{?}{=} \boldsymbol{Z}_{ki}\boldsymbol{G}(\boldsymbol{\alpha})\boldsymbol{Z}_{ki}^T \ + \ \sigma^2
\end{aligned}$$

$$\begin{aligned}
\mathsf{Cov}[Y_{ki}, Y_{kj}] &= \mathsf{Cov}_{\boldsymbol{\gamma}}[\mathsf{E}_{\boldsymbol{Y}}[Y_{ki}|\ \boldsymbol{\gamma}_k],\ \mathsf{E}_{\boldsymbol{Y}}[Y_{kj}|\ \boldsymbol{\gamma}_k]] \ + \ \mathsf{E}_{\boldsymbol{\gamma}}[\mathsf{Cov}[Y_{ki}, Y_{kj}|\boldsymbol{\gamma}_k]] \\
&= \mathsf{Cov}_{\boldsymbol{\gamma}}[g^{-1}(\boldsymbol{X}_{ki}\boldsymbol{\beta}^* + \boldsymbol{Z}_{ki}\boldsymbol{\gamma}_k),\ g^{-1}(\boldsymbol{X}_{kj}\boldsymbol{\beta}^* + \boldsymbol{Z}_{kj}\boldsymbol{\gamma}_k)] \\
&\stackrel{?}{=} \boldsymbol{Z}_{ki}\boldsymbol{G}(\boldsymbol{\alpha})\boldsymbol{Z}_{kj}^T
\end{aligned}$$

- Unfortunately the result does not hold more generally
  - ⋆ marginal and conditional contrasts are, in general, not equivalent numerically

- Summary of parameter interpretations for the intercept (i.e. β_0^*) and slope parameters (i.e. β_1^*, ..., β_p^*) from various common model specifications:

Response - link	Coefficient	Fitted GLMM	
		Random intercepts	Random intercepts/slopes
Continuous - identity	Intercept	M/C	M/C
	Slope	M/C	M/C
Count - log	Intercept	C	C
	Slope	M/C	C
Binary - logit	Intercept	C	C
	Slope	C	C

- Towards examining the actual numerical differences, consider a simple random intercepts logistic regression model:

$$\text{logit P}(Y_{ki} = 1|\ X_{ki}, \gamma_{0k}) \ = \ \beta_0^* \ + \ \beta_1^* X_{ki} \ + \ \gamma_{0k}$$

with $\gamma_{0k} \sim$ Normal(0, σ_γ^2)

  - ⋆ conditional or cluster-specific rate of the response

- The induced marginal or population rate can be obtained by integrating (i.e. averaging) over the distribution of γ_{0k}:

$$\text{P}(Y_{ki} = 1|\ X_{ki}) \ = \ \int \text{P}(Y_{ki} = 1|\ X_{ki}, \gamma_{0k}) f(\gamma_{0k}|\ \sigma_\gamma^2) \partial \gamma_{0k}$$

where $f(\cdot)$ is the Normal density function

  - ⋆ e.g. via Gauss-Hermite quadrature

- As a numerical example, suppose, X_{ki} is binary and $(\beta_0^*,\ \beta_1^*)$ = (-2, 0.4)
  - ⋆ conditional odds ratio is $\exp\{0.4\} = 1.5$

- If $\sigma_\gamma^2 = 2$, the induced population rates are:

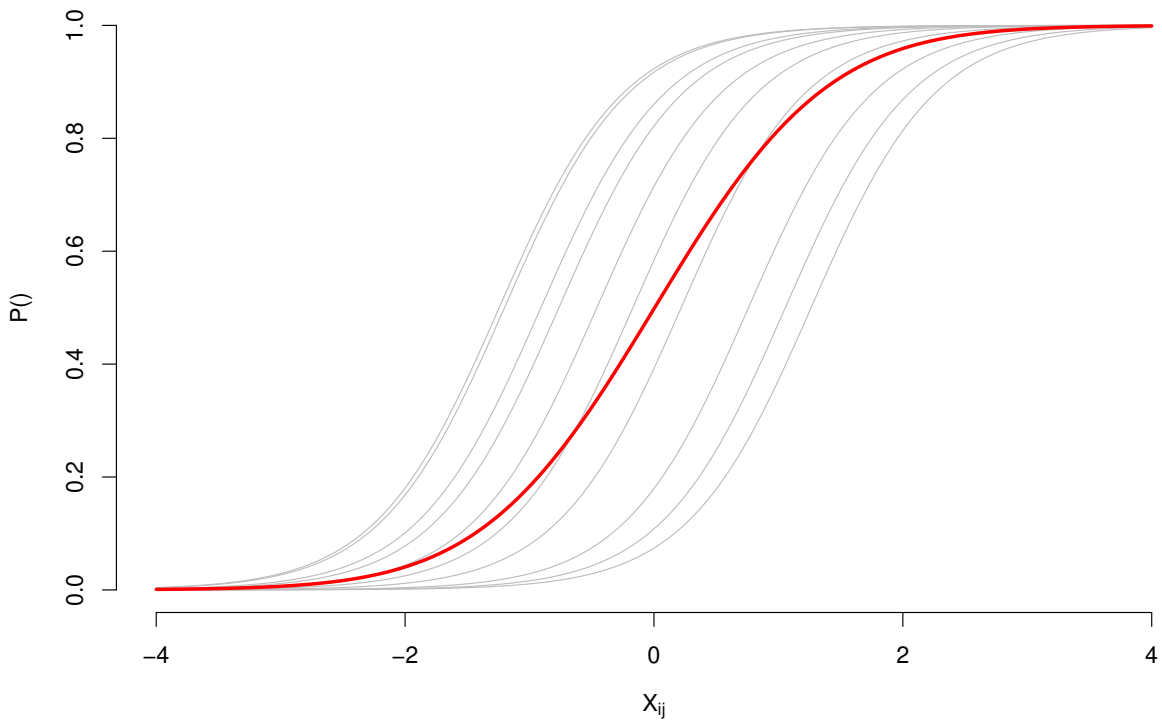
$$\begin{aligned} \mathsf{P}(Y_{ki} = 1|\ X_{ki} = 0) &= 0.18 \\ \mathsf{P}(Y_{ki} = 1|\ X_{ki} = 1) &= 0.23 \end{aligned}$$

  which can be used to compute the induced marginal odds ratio:

$$\frac{\mathsf{P}(Y_{ki} = 1|\ X_{ki} = 1)/\mathsf{P}(Y_{ki} = 0|\ X_{ki} = 1)}{\mathsf{P}(Y_{ki} = 1|\ X_{ki} = 0)/\mathsf{P}(Y_{ki} = 0|\ X_{ki} = 0)} = 1.36 \neq 1.5$$

- See that the marginal odds ratio is attenuated relative to the conditional odds ratio

- We can also see the attenuation visually:



  - ⋆ each of the grey lines has the same slope (i.e. the same conditional parameter in the mixed effects model)
  - ⋆ see that the population slope is shallower (i.e. attenuated)
  - ⋆ a function of the non-linearity of the odds ratio

- More formally, Zeger et al (1988) show that if

$$\text{logit P}(Y_{ki} = 1 | \ \boldsymbol{X}_{ki}, \gamma_{0k}) \ = \ \boldsymbol{X}_{ki}\boldsymbol{\beta}^* \ + \ \gamma_{0k}$$

and $\gamma_{0k} \sim$ Normal(0, σ^2_γ) then the induced marginal model is

$$\text{logit P}(Y_{ki} = 1 | \ X_{ki}) \ \approx \ \boldsymbol{X}_{ki}\boldsymbol{\beta}^* \ \times \ (1 + c^2\sigma^2_\gamma)^{-1/2}$$

where $c \ = \ 16\sqrt{3}/15\pi$ so that

Conditional parameter

$$\boldsymbol{\beta} \ \approx \ \boldsymbol{\beta}^* \ \times \ (1 + c^2\sigma^2_\gamma)^{-1/2}$$

Marginal coefficient

  - ⋆ adjustment factor is positive and less than 1.0, so the marginal parameter is attenuated relative to the conditional parameter

- Furthermore, Neuhaus et al (1991) showed that if V$[\gamma_{0k}] > 0$ then
  - (1) $|\beta_j| < |\beta^*_j| \ \forall \ j = 1, \ldots, p$
  - (2) equality holds iff $\beta^*_j = 0$
  - (3) discrepancy increases as V$[\gamma_{0k}]$ increases

- Returning to the ICHS data, let's compare a random intercepts GLMM to a marginal model with an exchangeable correlation structure:

```
>
> library(lme4)
> fit.RI.25 <- glmer(infection ~ xerop + age + (1 | id),
+                    data=ichs, family=binomial, nAGQ=25)
>
> library(geepack)
> fit1.pack <- geeglm(infection ~ xerop + age,
+                     id=id, data=ichs,
+                     family=binomial, scale.fix=TRUE,
+                     corstr="exchangeable")
>
> round(cbind(summary(fit.RI.25)$coef[,c(1,2,4)],
+             summary(fit1.pack)$coef[,c(1,2,4)]), 3)
            Estimate Std. Error Pr(>|z|) Estimate Std.err Pr(>|W|)
(Intercept)   -2.621      0.171    0.000   -2.370   0.117     0.00
xerop          0.549      0.479    0.251    0.589   0.449     0.19
age           -0.027      0.007    0.000   -0.025   0.005     0.00
```

- Some differences for the intercept and the slope for xerophthalmia, but the substantive conclusions don't change

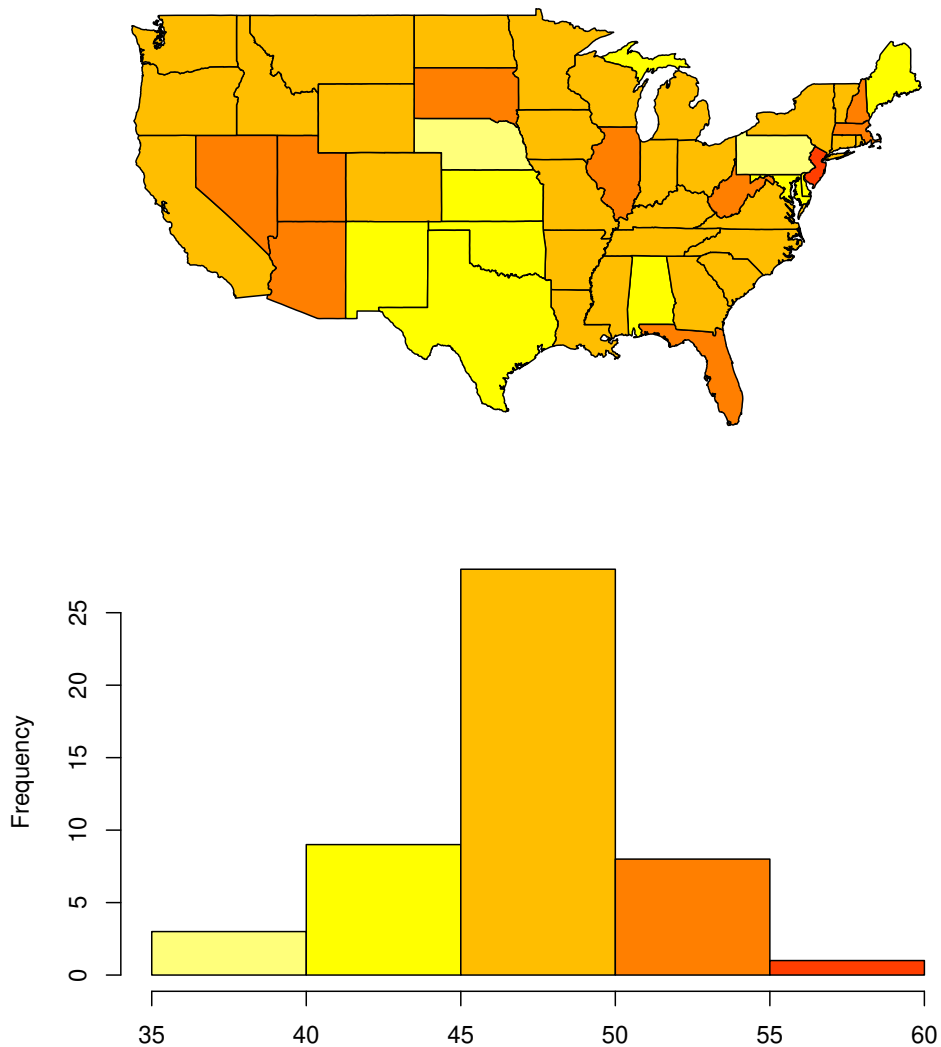
CMS data

- As a final exercise, let's consider an example in which the random effects play a prominent role in the science
- Return to the CMS data on outcomes among patients diagnosed with pancreatic cancer
  - ⋆ focus attention on patients diagnosed during a hospitalization between 2000-2009 at an age of 65 years or older and successfully discharged
  - ⋆ dataset restricted to hospitals with at least 50 admissions
- Take the scientific goal to be characterization of variation in 'quality of care' as measured by 90-day mortality at the level of the state
  - ⋆ focus on the contiguous U.S.
  - ⋆ ignore clustering at the level of the hospital
  - ⋆ covariates adjust for patient case-mix

```
> ##
> load("CMS.RData")
> CMS <- CMS[!is.element(CMS$state, c("AK",
+                                     "Guam",
+                                     "HI",
+                                     "PR",
+                                     "SAIPAN OR NORTHERN MARIANAS",
+                                     "VI")),]
> CMS <- CMS[(CMS$year > 1999),]
>
> CMS[1:5,]
  hospID hospVol year state female age  race admission deyo LOS  discharge T1 T2
1      1     228 2008    IL      1  72 Other        ER    1  14     1.Home  8 NA
2      1     228 2000    IL      1  65 White        ER    1   3    9.Other NA 45
3      1     228 2005    IL      0  77 White        ER    1  10 2.HomeCare  6 NA
4      1     228 2009    IL      1  67 White        ER    1  10  3.SNF/ICF NA 16
5      1     228 2000    IL      1  78 White     Other    1   4  3.SNF/ICF NA 19
> dim(CMS)
[1] 120789     13
> length(unique(CMS$state)) ## 48 states and DC
[1] 49
> length(unique(CMS$hospID))
[1] 1024
```

```
>
> ## State-specific mortality rates
> ##
> CMS$T2[is.na(CMS$T2)] <- 999
> CMS$D.30 <- as.numeric(CMS$T2 <= 30)
> CMS$D.90 <- as.numeric(CMS$T2 <= 90)
> rateD.30 <- tapply(CMS$D.30, list(CMS$state), FUN=mean) * 100
> rateD.90 <- tapply(CMS$D.90, list(CMS$state), FUN=mean) * 100
>
> ##
> tab.Rate <- rbind(summary(rateD.30),
+                   summary(rateD.90))
> dimnames(tab.Rate)[[1]] <- c("rateD.30", "rateD.90")
>
> print(tab.Rate)
          Min. 1st Qu. Median  Mean 3rd Qu.  Max.
rateD.30 19.44   24.16  25.43 25.37   26.61 30.79
rateD.90 36.73   45.28  47.20 46.84   49.15 56.95
```

- Visualization of 90-day mortality rates across the contiguous U.S.:



```
>
> ## Code for producing figure
> ##
> library(maps)
> my.colors <- function(n) heat.colors(n)[n:1]
>
> ##
> myCuts <- c(seq(from=35, to=60, by=5), 100)
> nCats  <- length(myCuts) - 1
>
> ##
> colorBuckets <- as.numeric(cut(rateD.90, myCuts))
> dataColors   <- my.colors(length(myCuts))[colorBuckets]
>
> ##
> par(mfrow=c(2,1))
> map("state", fill=TRUE, col=dataColors, xlim=c(-125, -65), ylim=c(25, 50))
> hist(rateD.90, nclass=5, col=my.colors(length(myCuts)), xlab="", main="")
```

```
>
> ## Covariate manipulations prior to modeling
> ##  - note: much of this is arbitrary and, in the real world, one
> ##    would want to decide upon these changes with a collaborator
>
> ##
> summary(CMS$age)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  65.00   70.00   76.00   76.48   82.00  102.00
> CMS$age <- (CMS$age - 75) / 10
>
> ##
> cbind(table(CMS$race),
+       round(tapply(CMS$D.90, list(CMS$race), FUN=mean), 2))
       [,1] [,2]
Black  12466 0.50
Other   4765 0.45
White 103558 0.46
> CMS$raceBlack <- 0
> CMS$raceBlack[CMS$race == "Black"] <- 1
```

```
>
> ##
> cbind(table(CMS$admission),
+       round(tapply(CMS$D.90, list(CMS$admission), FUN=mean), 2))
          [,1] [,2]
ER       52062 0.58
Transfer   821 0.55
SNF/ICF    828 0.69
Other    67078 0.38
> CMS$admission <- as.numeric(CMS$admission != "Other")
>
> ##
> rbind(table(CMS$deyo),
+       round(tapply(CMS$D.90, list(CMS$deyo), FUN=mean), 2))
           0         1       2       3      4     5  6 7
[1,] 8617.00 103435.00 6531.00 1755.00 387.00 52.00 11 1
[2,]    0.28      0.47    0.56    0.61   0.68  0.65  1 1
> CMS$deyo[CMS$deyo > 3] <- 3
```

```
>
> ##
> cbind(table(CMS$discharge),
+       round(tapply(CMS$D.90, list(CMS$discharge), FUN=mean), 2))
             [,1] [,2]
1.Home      60634 0.33
2.HomeCare  22861 0.42
3.SNF/ICF   17208 0.60
4.Hospice   15657 0.90
5.Rehab      1278 0.39
6.Inpatient  1830 0.71
7.LTC         670 0.56
8.Swing bed   120 0.47
9.Other       531 0.39
> CMS$discharge[CMS$discharge == "5.Rehab"]     <- "9.Other"
> CMS$discharge[CMS$discharge == "7.LTC"]       <- "9.Other"
> CMS$discharge[CMS$discharge == "8.Swing bed"] <- "9.Other"
```

```
> ## Fit three logistic-Normal random intercept models
> ##
>
> ## Unadjusted
> ##
> fit0 <- glmer(D.90 ~ 1 + (1 | state),
                data=CMS,
                family=binomial)
>
> ## "Partial" adjustment
> ##
> fit1 <- glmer(D.90 ~ female + age + raceBlack + admission + factor(deyo)
                       + (1 | state),
                data=CMS,
                family=binomial)
>
> ## "Full" adjustment
> ##
> fit2 <- glmer(D.90 ~ female + age + raceBlack + admission + factor(deyo)
                       + factor(discharge)
                       + (1 | state),
                data=CMS,
                family=binomial)
```

- Point estimates for β^* and σ_γ:

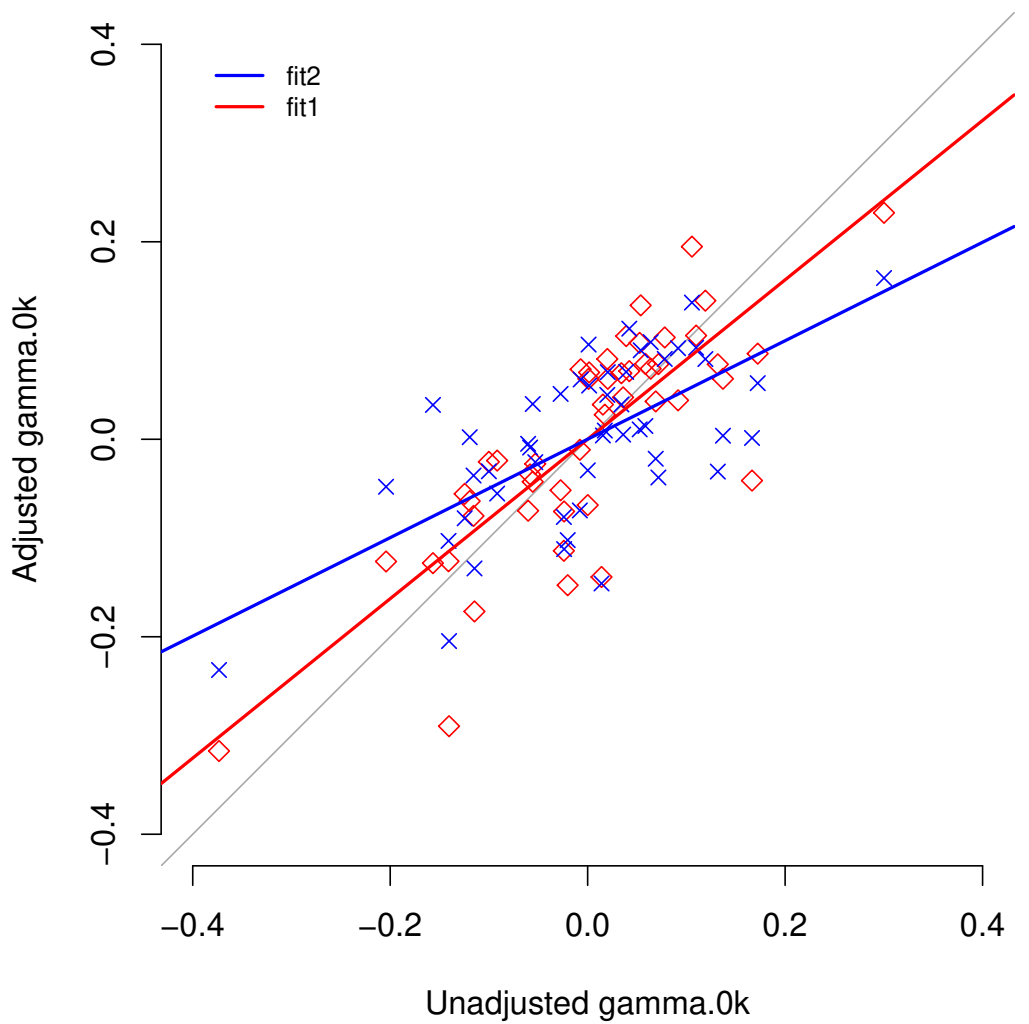
	fit0	fit1	fit2
(Intercept)	-0.128	-1.173	-1.473
female	NA	-0.082	-0.159
age	NA	0.460	0.277
raceBlack	NA	0.059	0.049
admission	NA	0.697	0.554
factor(deyo)1	NA	0.754	0.673
factor(deyo)2	NA	1.105	0.942
factor(deyo)3	NA	1.372	1.130
factor(discharge)2.HomeCare	NA	NA	0.339
factor(discharge)3.SNF/ICF	NA	NA	0.897
factor(discharge)4.Hospice	NA	NA	2.660
factor(discharge)6.Inpatient	NA	NA	1.517
factor(discharge)9.Other	NA	NA	0.366
sigma.gam	0.124	0.125	0.100

random intercept is meaningful, medical care level across states are different

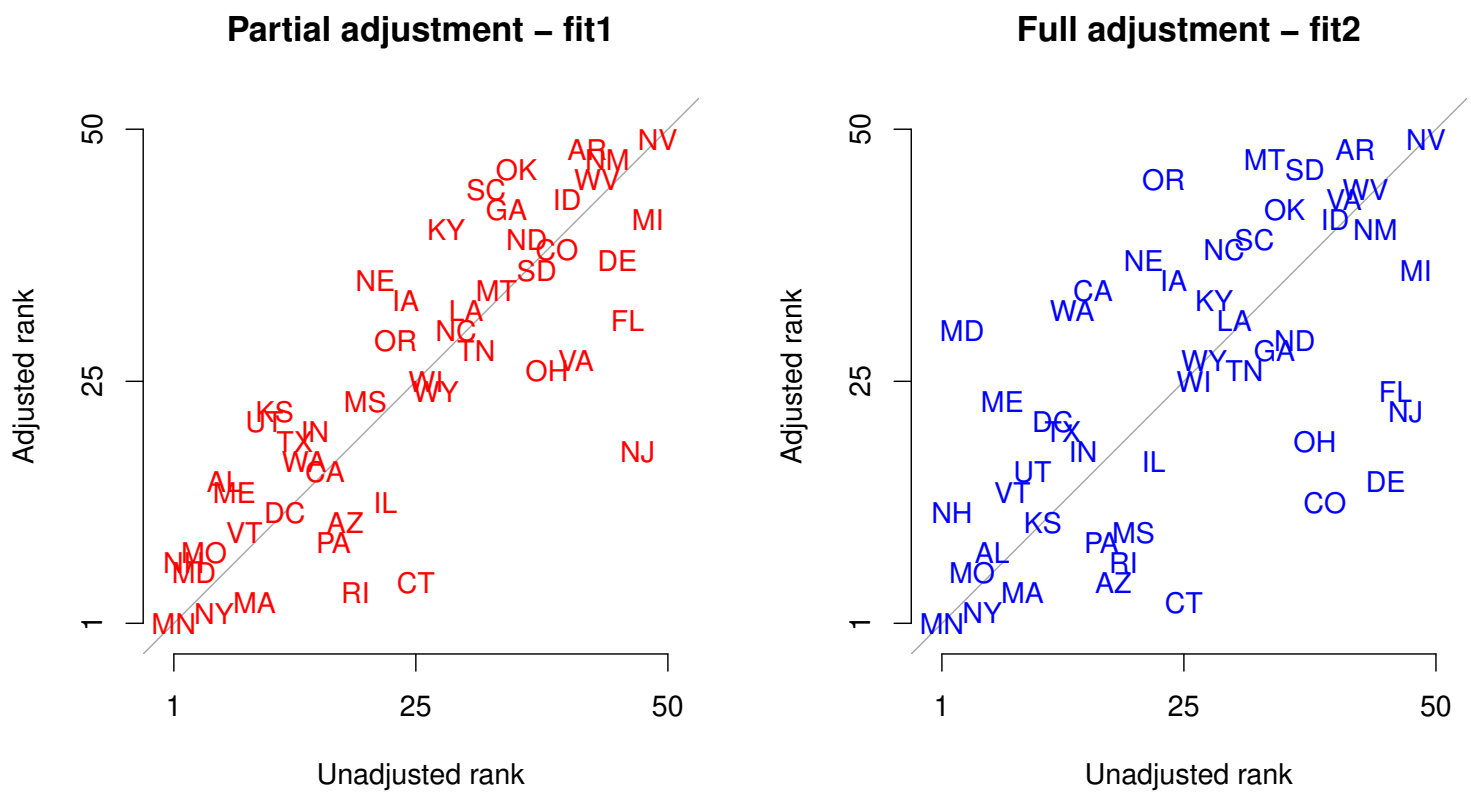
- First set of adjustment factors:
  - ⋆ all highly statistically significant and, in many cases, have strong effects
  - ⋆ inclusion does little, if anything, to explain the dependence structure

- Including discharge location has a fairly large impact on the estimate of σ_γ

- Comparison of the empirical Bayes estimates of γ_{0k}:



- Investigate the impact of case-mix adjustment on the ranks of the empirical Bayes estimates of γ_{0k} heterogeneity across states



- From the full adjustment model:
  - ⋆ CT, AZ, DE, and NJ all 'benefit' from the case-mix adjustment
  - ⋆ NH, MD, OR and MT all 'loose' from the case-mix adjustment

Summary

GLMM: ML
marginal: GEE

multi-level model
hierarchical model

- GLMMs, fit via ML, and marginal models for dependent data, fit via GEE, represent the two main regression frameworks for cluster-correlated or longitudinal data

Q: Which one should you use?

- Beyond linear models for continuous response data, the important distinction between the two frameworks is in the interpretation and numerical values of the regression coefficients
  - ⋆ marginal vs. conditional with respect to the clustering

- In my opinion this is not a particularly useful distinction to focus on
  - ⋆ we are constantly comparing models with and without certain covariates
  - ⋆ marginal vs. conditional with respect to these covariates

- It's worth noting how certain parts of the literature have focused almost exclusively on specific contrasts:
  - ⋆ e.g. the causal-inference literature ⇒ marginal contrasts
  - ⋆ e.g. the Bayesian correlated data literature ⇒ conditional contrasts

- If the scientific focus is on the clusters themselves, the GLMM framework will be the way forward between the two
  - ⋆ estimation of random effects via empirical Bayes
  - ⋆ although not covered in class, the 'full' Bayesian framework is appealing here because of the capacity to quantify uncertainty

- If the focus is on the regression coefficients, concerns regarding the 'robustness' of results from a GLMM are often cited as a potential problem
  - ⋆ specifically with respect to the choice of random effects distributions
  - ⋆ the literature on this is quite contentious
    - ∗ see references cited in Antonelli et al (*Statistical Science*, 2016)
  - ⋆ my sense is that, in practice, this is not a big concern

- In small-sample settings (i.e. small K), I think there is an argument to be made that GLMMs will often be the way forward
  - ⋆ validity of the sandwich estimator for GEE may be questionable
    - ∗ 'rules of thumb' have been put forward suggesting that K needs to be at least 40
  - ⋆ stability of likelihood-based estimation/inference is appealing, although we should (as ever) still be careful when drawing conclusions based on small samples

Ridge regression Bayesian view: impose normal prior to all coefficient