Hierarchical Poisson model

- Count data are often modeled using a Poisson model.
- If $y \sim$ Poisson(μ) then $E(y) = var(y) = \mu$.
- When counts are assumed exchangeable given μ and the rates μ can also be assumed to be exchangeable, a Gamma population model for the rates is often chosen.
- The hierarchical model is then

$$\begin{aligned} y_i &\sim \text{Poisson}(\mu_i) \\ \mu_i &\sim \text{Gamma}(\alpha, \beta). \end{aligned}$$

- Priors for the hyperparameters are often taken to be Gamma (or exponential):

$$\begin{aligned} \alpha &\sim \text{Gamma}(a, b) \\ \beta &\sim \text{Gamma}(c, d), \end{aligned}$$

  with (a, b, c, d) known.

Hierarchical Poisson model (cont'd)

- The joint posterior distribution is

$$p(\mu, \alpha, \beta | y) \;\propto\; \Pi_i \mu_i^{y_i} \exp\{-\mu_i\} \mu_i^{\alpha-1} \exp\{-\mu_i \beta\} \\ \alpha^{a-1} \exp\{-\alpha b\} \beta^{c-1} \exp\{-\beta d\}$$

- To carry out Gibbs sampling we need to find the full conditional distributions.
- Conditional for μ_i is

$$p(\mu_i | \text{ all}) \propto \mu_i^{y_i+\alpha-1} \exp\{-\mu_i(\beta + 1)\},$$

which is proportional to a Gamma with parameters $(y_i + \alpha, \beta + 1)$.

- The full conditional for α is

$$p(\alpha | \text{ all}) \propto \Pi_i \mu_i^{\alpha-1} \alpha^{a-1} \exp\{\alpha b\}.$$

- The conditional for α does not have a standard form.

Hierarchical Poisson model (cont'd)

- For β:

$$\begin{aligned} p(\beta|\text{ all}) &\propto \Pi_i \exp\{-\beta\mu_i\}\beta^{c-1}\exp\{-\beta d\} \\ &\propto \beta^{c-1}\exp\{-\beta(\sum_i \mu_i + d)\}, \end{aligned}$$

  which is proportional to a Gamma with parameters $(c, \Sigma_i \mu_i + d)$.

- Computation:
  - Given α, β, draw each μ_i from the corresponding Gamma conditional.
  - Draw α using a Metropolis step or rejection sampling or inverse cdf method.
  - Draw β from the Gamma conditional.

- See Italian marriages example.

Poisson regression

- When rates are not exchangeable, we need to incorporate covariates into the model. Often we are interested in the association between one or more covariate and the outcome.
- It is possible (but not easy) to incorporate covariates into the Poisson-Gamma model.
- Christiansen and Morris (1997, JASA) propose the following model:
- Sampling distribution, where e_i is a known exposure:

  $$y_i|\lambda_i \sim \text{ Poisson}(\lambda_i e_i).$$

  Under model, $E(y_i/e_i) = \lambda_i$.
- Population distribution for the rates:

  $$\lambda_i|\alpha \sim \text{ Gamma}(\zeta, \zeta/\mu_i),$$

  with $\log(\mu_i) = x_i'\beta$, and $\alpha = (\beta_0, \beta_1, ..., \beta_{k-1}, \zeta)$.
- ζ is thought of as an unboserved prior count.

Hierarchical Poisson model (cont'd)

- Under population model,

$$
\begin{aligned}
E(\lambda_i) &= \frac{\zeta}{\zeta/\mu_i} \\
&= \mu_i \\
\mathrm{CV}^2(\lambda_i) &= \frac{\mu_i^2}{\zeta}\frac{1}{\mu_i^2} \\
&= \frac{1}{\zeta}
\end{aligned}
$$

- For $k = 0$, μ_i is known. For $k = 1$, μ_i are exchangeable. For $k \geq 2$, μ_i are (unconditionally) nonexchangeable.
- In all cases, standardized rates λ_i/μ_i are Gamma(ζ, ζ), are exchangeable, and have expectation 1.
- The covariates can include random effects.

Hierarchical Poisson model (cont'd)

- To complete specification of model, we need priors on α.
- Christensen and Morris (1997) suggest:
  - β and ζ independent a priori.
  - Non-informative prior on β's associated to 'fixed' effects.
  - For ζ a proper prior of the form:

    $$p(\zeta|y_0) \propto \frac{y_0}{(\zeta + y_0)^2},$$

    where y_0 is the prior guess for the median of ζ.
- Small values of y_0 (for example, $y_0 < \hat{\zeta}$ and $\hat{\zeta}$ the MLE of ζ) provide less information.

Poisson regression

- When the rates cannot be assumed to be exchangeable, it is common to choose a generalized linear model of the form:

$$\begin{aligned} p(y|\beta) &\propto \Pi_i \exp\{-\lambda_i\}\lambda_i^{y_i} \\ &\propto \Pi_i \exp\{-\exp(\eta_i)\}[\exp(\eta_i)]^{y_i}, \end{aligned}$$

  for $\eta_i = x_i'\beta$ and $\log(\lambda_i) = \eta_i$.

- The vector of covariates can include one or more random effects to accommodate additional dispersion (see epylepsy example).

- The second-level distribution for the β's will typically be flat (if covariate is a 'fixed' effect) or normal

$$\beta_j \sim \text{ Normal}(\beta_{j0}, \sigma^2_{\beta_j})$$

  if jth covariate is a random effect. The variance $\sigma^2_{\beta_j}$ represents the between 'batch' variability.

Epilepsy example

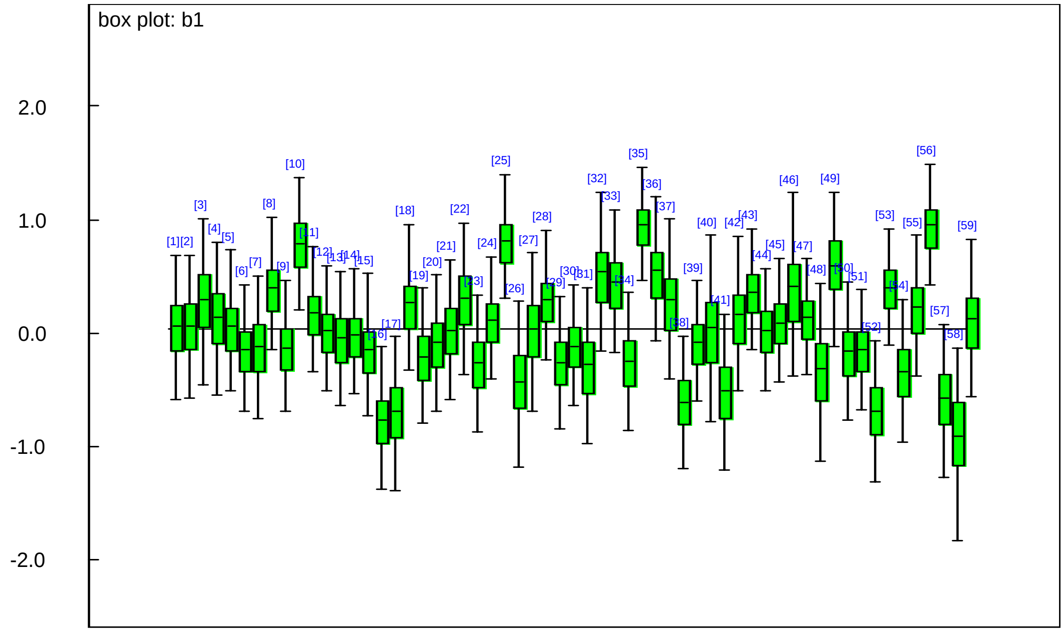
- From Breslow and Clayton, 1993, JASA.
- Fifty nine epilectic patients in a clinical trial were randomized to a new drug: T = 1 is the drug and T = 0 is the placebo.
- Covariates included:
  - Baseline data: number of seizures during eight weeks preceding trial
  - Age in years.
- Outcomes: number of seizures during the two weeks preceding each of four clinical visits.
- Data suggest that number of seizures was significantly lower prior to fourth visit, so an indicator was used for V4 versus the others.
- Two random effects in the model:
  - A patient-level effect to introduce between patient variability.

- A patients by visit effect to introduce between visit within patient dispersion.

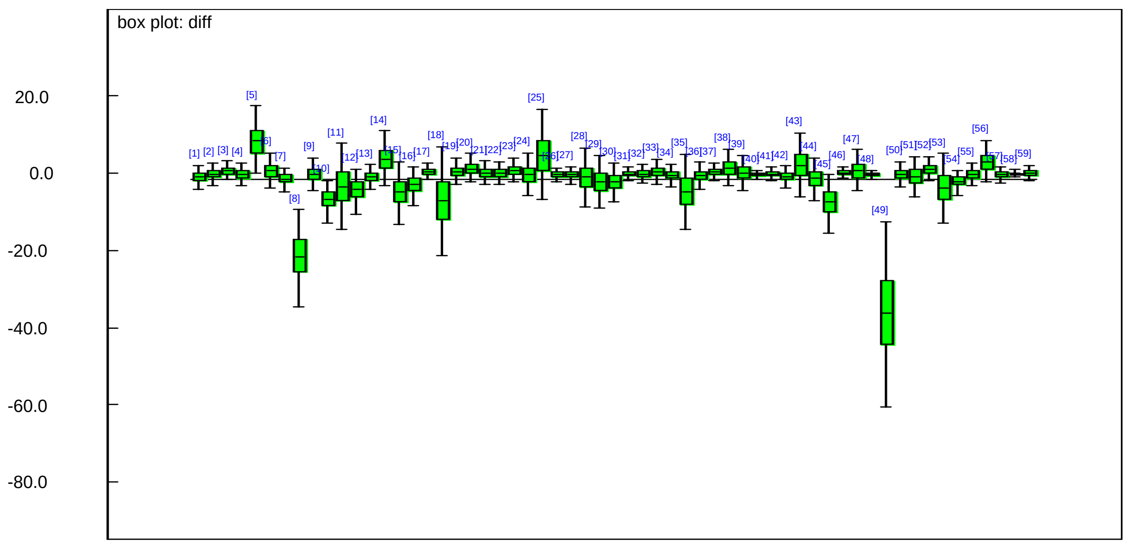
Epilepsy study – Program and results

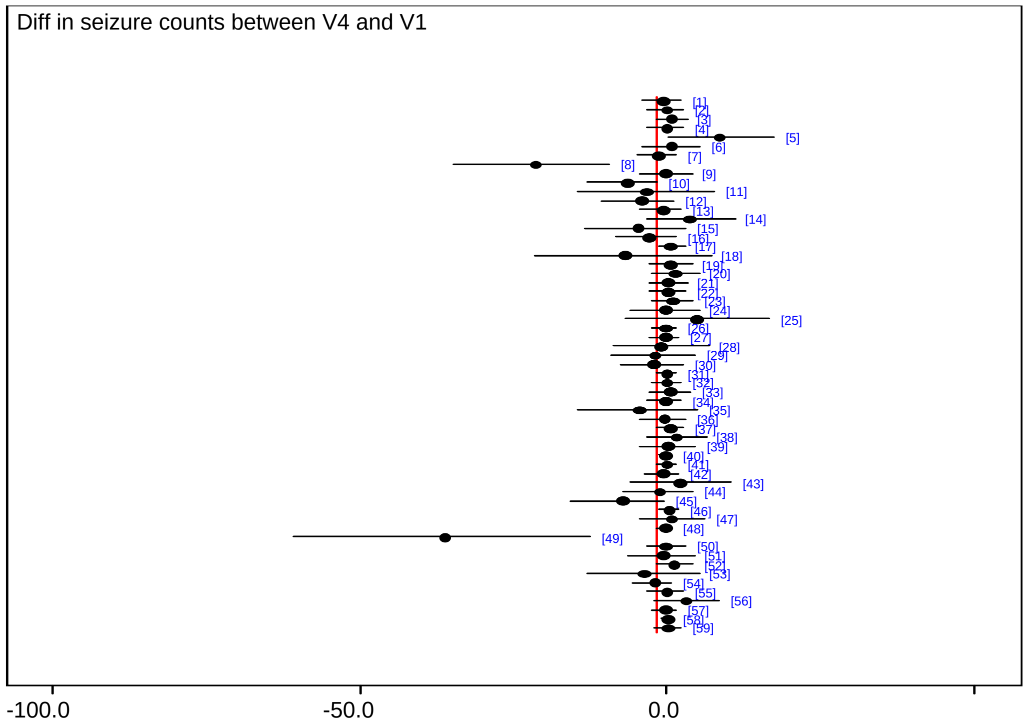
```
model {
                for(j in 1 : N) {
                        for(k in 1 : T) {
                                log(mu[j, k]) <- a0 + alpha.Base * (log.Base4[j] -
log.Base4.bar)
                     + alpha.Trt * (Trt[j] - Trt.bar)
                     + alpha.BT  * (BT[j] - BT.bar)
                     + alpha.Age * (log.Age[j] - log.Age.bar)
                     + alpha.V4  * (V4[k] - V4.bar)
                     + b1[j] + b[j, k]
                                y[j, k] ~ dpois(mu[j, k])
                                b[j, k] ~ dnorm(0.0, tau.b);       # subject*visit random
effects
                        }
                        b1[j]  ~ dnorm(0.0, tau.b1)        # subject random effects
                        BT[j] <- Trt[j] * log.Base4[j]    # interaction
                        log.Base4[j] <- log(Base[j] / 4) log.Age[j] <- log(Age[j])
                        diff[j] <- mu[j,4] – mu[j,1]
                }
        # covariate means:
                log.Age.bar <- mean(log.Age[])
                Trt.bar  <- mean(Trt[])
                BT.bar <- mean(BT[])
                log.Base4.bar <- mean(log.Base4[])
                V4.bar <- mean(V4[])
        # priors:

                a0 ~ dnorm(0.0,1.0E-4)
                alpha.Base ~ dnorm(0.0,1.0E-4)
                alpha.Trt  ~ dnorm(0.0,1.0E-4);
                alpha.BT   ~ dnorm(0.0,1.0E-4)
                alpha.Age  ~ dnorm(0.0,1.0E-4)
                alpha.V4   ~ dnorm(0.0,1.0E-4)
                tau.b1     ~ dgamma(1.0E-3,1.0E-3); sigma.b1 <- 1.0 / sqrt(tau.b1)
                tau.b      ~ dgamma(1.0E-3,1.0E-3); sigma.b  <- 1.0/  sqrt(tau.b)


        # re-calculate intercept on original scale:
                alpha0 <- a0 - alpha.Base * log.Base4.bar - alpha.Trt * Trt.bar
                - alpha.BT * BT.bar - alpha.Age * log.Age.bar - alpha.V4 * V4.bar
        }
```

Individual random effects



Difference in expected seizure counts between fourth and first measurements





Parameter	Mean	Std	2.5th	Median	97.5th
alpha.Age	0.4677	0.3557	-0.2407	0.4744	1.172
alpha.Base	0.8815	0.1459	0.5908	0.8849	1.165
alpha.Trt	-0.9587	0.4557	-1.794	-0.9637	-0.06769
alpha.V4	-0.1013	0.08818	-0.273	-0.09978	0.07268