

**Statistical Models for Multi-level Data  
2004 Summer Epi-Biostat Institute**

**Participant Homework - Answer Key**

**Module I: Statistical Background on Multi-level Models**

1. The data from the "Alcohol Dependence" crossover-trial example given in class actually pertains to a 2x2 crossover trial of Cerebrovascular disease examined in the text: *Analysis of Longitudinal Data*, (Diggle, Heagerty, Liang & Zeger, 2002. pg. 148-150 & pg. 180-181) and reproduced below.

Data from a 2 x 2 crossover trial on cerebrovascular deficiency adapted from Jones and Kenward (1989, p. 90), where treatments A and B are active drug and placebo, respectively; the outcome indicates whether an electrocardiogram was judged abnormal (0) or normal (1).

| Group | Responses |       |       |       | Total | Period |    |
|-------|-----------|-------|-------|-------|-------|--------|----|
|       | (1,1)     | (0,1) | (1,0) | (0,0) |       | 1      | 2  |
| AB    | 22        | 0     | 6     | 6     | 34    | 28     | 22 |
| BA    | 18        | 4     | 2     | 9     | 33    | 20     | 22 |

In SAS's program editor, open and run the program: "**Problem 1 xover trial.sas**" found on the class website to reproduce the results in DHLZ (2002). Interpret the regression coefficients and their confidence intervals for the period and treatment effects for both the marginal (GEE) and conditional (Random Effect) analyses. (Note: the results are the same as in the Module 1 class notes so you can use them and look over the program later if you're short on time.)

**Solution:**

From either the SAS output or the Module 1 class notes we have the following table of marginal and conditional estimates and their standard errors.

| Parameter            | Marginal |      | Conditional |      |
|----------------------|----------|------|-------------|------|
|                      | Estimate | S.E. | Estimate    | S.E. |
| Intercept            | 0.67     | 0.29 | 2.27        | 1.28 |
| Time Period (2 vs 1) | -0.30    | 0.23 | -1.06       | 0.85 |
| Treatment (A vs B)   | 0.57     | 0.23 | 1.89        | 0.97 |
| Association          | 3.56     | 0.81 | 5.06        | 2.13 |

From these we find Odds Ratios and Confidence Intervals that describe treatment and period effects by exponentiating.

| Effect               | Marginal   |              | Conditional |               |
|----------------------|------------|--------------|-------------|---------------|
|                      | Odds Ratio | 95% C.I.     | Odds Ratio  | 95% C.I.      |
| Time Period (2 vs 1) | 0.74       | (0.47, 1.18) | 0.35        | (0.06, 1.90)  |
| Treatment (A vs B)   | 1.77       | (1.11, 2.82) | 6.64        | (0.96, 45.85) |

### Marginal Effect Interpretations:

There does not appear to be a large difference between time periods, as the Odds Ratio of 0.74 is close to 1, (i.e. no effect) and the Confidence Interval of (0.47, 1.18) substantially overlaps 1. We do estimate a statistically significant treatment effect, as the overall Odds of a normal electrocardiogram are almost twice as high (OR=1.77) comparing responses on the active drug to responses on placebo and the Confidence Interval for this effect ranges from 1.11 to 2.82.

### Conditional Effect Interpretations:

Again, there does not appear to be a large difference between time periods for a specific subject, as the Confidence Interval for the Odds Ratio (0.06, 1.90) substantially overlaps 1. There may be a treatment effect for a given subject, as the overall Odds of a normal electrocardiogram are estimated to be six or seven times as high (OR=6.64) comparing the response for a specific subject on the active drug to that same subject's response on placebo. The Confidence Interval for this effect ranges from 0.96 to 45.85, which does cover 1, (which some may take to indicate a possibility of no actual effect), but note that we have used  $\exp(1.89 \pm 2*0.97)$  to form the 95% C.I. instead of the usual  $\exp(1.89 \pm 1.96*0.97)$  to be conservative. More important is the increased uncertainty in estimating the subject-specific effect.

**2. In slides 50 and 53 of the module 1 presentation, results of a conditional and marginal logistic multilevel model are presented. Focus on Model 1 in the tables (the models with the simplest association structures). Interpret the regression coefficients for gender and for %protestant in each of these two models.**

### Solution:

**Marginal Model:** Slide 53 gives the estimates for the Marginalized Multilevel Model 1:

The Gender comparison, adjusting for Year, Working Class, Religion, & District is given as

$\text{log-odds}(\text{Favoring Abortion} \mid \text{Female}, \mathbf{X})$

$-\text{log-odds}(\text{Favoring Abortion} \mid \text{Male}, \mathbf{X}) = -0.349$

with a standard error of 0.205. Thus, the Odds of Favoring Abortion are estimated to be  $\exp(-0.349) = 0.71$ , or 30% smaller for Females compared to Males, after controlling for Year, Working Class, Religion, and District. However, the 95% Confidence Interval for the OR is (0.47, 1.1), indicating that this Gender difference is not *statistically* significant.

The %-Protestant comparison, adjusting for Year, Working Class, Religion, & District is given as

$\text{log-odds}(\text{Favoring Abortion} \mid \% \text{Protestant} + 1, \mathbf{X})$

$-\text{log-odds}(\text{Favoring Abortion} \mid \% \text{Protestant}, \mathbf{X}) = 0.799$

with a standard error of 0.479. Thus, the Odds of Favoring Abortion are estimated to be  $\exp(0.799) = 2.22$ , or twice as high for Districts that have protestant percentages that differ by 1 percent, after controlling for Year, Working Class, Religion, and District. Again, the 95% Confidence Interval for the OR, (0.85, 5.80) substantially covers 1, indicating we do not find a statistical difference between Districts that have protestant percentages that differ by 1 percent.

**Conditional Model:** Slide 50 gives the estimates for the Conditional Multilevel Model 1:

Recall that the conditional model 1 incorporates latent effects at the person and district level, Hence, the results pertain to a specific person within a specific district. Suppose that the specific person is a female, the Gender comparison would then attempt to examine what her odds of favoring abortion would be if she was in fact a male, (also adjusting for Year, Working Class, Religion, & District). Consider the  $j^{\text{th}}$  female in the  $i^{\text{th}}$  district who has a subject-specific latent effect  $b_{1ij}$  and a district specific latent effect  $b_{2ij}$ .

$$\log\text{-odds}(\text{Favoring Abortion} \mid \text{Female}_{ij}, b_{1ij}, b_{2ij}, \mathbf{X})$$

$$- \log\text{-odds}(\text{Favoring Abortion} \mid \text{Male}_{ij}, b_{1ij}, b_{2ij}, \mathbf{X}) = -0.6$$

with a standard error of 0.358. Thus, the Odds of Favoring Abortion are estimated to be  $\exp(-0.6) = 0.55$ , or 50% smaller if this Female somehow became a Male. Of course, if all of the females do change into males, this abortion issue may be a moot point.

Similar issues in interpretation exist for the % Protestant comparison with the conditional model.

**3. Download the winbugs software on your laptop or visit the School computer lab to visit the BUGS website and watch the movie demo:**

<http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml>

<http://www.statslab.cam.ac.uk/~krice/winbugsthemovie.html>

**Solution:**

Great Movie! The popcorn was a bit salty....

Hopefully you were able to try out a couple of the examples located under the help menu.

## Module II: A two-stage model example: The DZAPS study

Below find a table of maximum likelihood estimates of the log relative risk (percent increase per 10 micrograms per cubic meter) and their statistical standard errors for 6 cities from the hypothetical DZAPS study (note: the data are not real).

| City     | Log RR<br>$\hat{\beta}_c$ | Statistical<br>std error<br>$\sqrt{vc}$ | Total Var<br>(TVc) | 1/TVc  | wc    | $\theta_c$ | RR.EB<br>$\tilde{\beta}_c$ | se(RR.EB) |
|----------|---------------------------|---|--------------------|--------|-------|------------|----------------------------|-----------|
| LA       | 0.30                      | 0.10                                    | 0.074              | 13.541 | 0.260 | 0.865      | 0.31                       | 0.09      |
| NYC      | 0.50                      | 0.12                                    | 0.078              | 12.780 | 0.246 | 0.816      | 0.48                       | 0.11      |
| Chi      | 0.40                      | 0.15                                    | 0.086              | 11.581 | 0.222 | 0.739      | 0.39                       | 0.13      |
| Dal      | 0.00                      | 0.30                                    | 0.154              | 6.500  | 0.125 | 0.415      | 0.22                       | 0.19      |
| Hou      | 1.00                      | 0.40                                    | 0.224              | 4.467  | 0.086 | 0.285      | 0.55                       | 0.21      |
| SD       | -0.10                     | 0.50                                    | 0.314              | 3.186  | 0.061 | 0.203      | 0.27                       | 0.23      |
| Over-all | 0.369                     | 0.139                                   |                    |        |       |            |                            |           |

1. Use the estimates above and their standard errors to estimate the natural variance in the true log relative risks across these 6 cities. Follow the calculations made in the lecture for module 2.

The Natural Variance is calculated as:

$$NV = \text{Variance}(\beta_c) - \text{Average}(vc) = 0.155 - 0.09115 = 0.06385$$

2. Calculate the overall estimate of the log relative risk weighting the individual city estimates by the inverse of their total variances

$$\text{Overall Estimate: } \hat{\alpha} = 0.369$$

3. Calculate the standard error for the overall estimate and make a 95% confidence interval for the true population mean.

$$\text{Standard Error}(\hat{\alpha}) = 0.139$$

$$95\% \text{ C.I.} \cong \hat{\alpha} \pm 2 \times \text{se}(\hat{\alpha}) = 0.369 \pm 2 \times 0.139 = (0.092, 0.647)$$

4. Now complete the table above producing the empirical-Bayes estimate and standard error for each city

Please see table above

5. **Compare the Empirical-Bayes and maximum likelihood estimates for San Diego (SD). Which estimate do you prefer and why? Comment on whether you think air pollution saves lives?**

The Empirical-Bayes estimate of 0.23 is a compromise between San Diego's specific MLE of -0.1 and the overall estimate for the six cities of 0.37. Since San Diego's MLE is fairly imprecise, (the standard error is 0.50) the Empirical-Bayes estimate "borrows" a good deal of information from the other cities and "shrinks" San Diego's estimate towards the overall estimate. It was mentioned in Module I that a useful model combines data with prior information to address the question of interest. Prior information here would be knowledge that Air Pollution is generally bad for people, contrary to San Diego's negative MLE, which would indicate that pollution is a protective factor and "saves lives". Thus, the Empirical-Bayes estimate is preferred since it more likely represents the true effect of Air Pollution.

6. **Fit the two-stage normal-normal model below in Winbugs to re-analyze the NMMAPS 6 cities data using MCMC.**

### Model

```
model{
  for( i in 1 : N)
  {
    p.hat[i] <- (1/se[i])*(1/se[i])
    beta.hat[i] ~dnorm(b[i],p.hat[i])
    b[i] <-alpha + u[i]
    u[i] ~dnorm(0,tau)
  }
  tau ~dgamma(0.001,0.001)
  sigma <- 1 / sqrt(tau)
  alpha ~ dnorm(0.0,1.0E-6)
}
```

**Data** list(N=6,beta.hat = c(0.3,0.5,0.4,0.0,1.0,-0.1), se =c(0.10,0.12,0.15,0.30,0.40,0.50))

**Inits** list(alpha =0,tau=1,u=c(0,0,0,0,0,0))

**Results:** (after 50,000 samples)

| node  | mean | sd   | 2.50% | median | 97.50% |
|-------|------|------|-------|--------|--------|
| alpha | 0.37 | 0.10 | 0.16  | 0.37   | 0.57   |
| b[1]  | 0.34 | 0.08 | 0.17  | 0.34   | 0.50   |
| b[2]  | 0.43 | 0.10 | 0.25  | 0.42   | 0.64   |
| b[3]  | 0.38 | 0.10 | 0.18  | 0.38   | 0.60   |
| b[4]  | 0.31 | 0.16 | -0.09 | 0.34   | 0.58   |
| b[5]  | 0.44 | 0.18 | 0.17  | 0.41   | 0.91   |
| b[6]  | 0.34 | 0.18 | -0.10 | 0.35   | 0.64   |

**Discussion:**

Francesca, what would you like to say here? I was going to discuss how the Fully Bayesianized versions incorporate the uncertainty in the Natural Variance whereas the Empirical-Bayes version above considers it fixed, but the standard errors are all smaller for the Full Bayes version and this will probably confuse them. -- Mike

### Module III: Applications of Multilevel Models to Profiling of Health Care Providers

1. From the Winbugs help menu, copy the “Institutional ranking” example (look under the help menu under Vol I examples)

| Institutional ranking Data |           |              |  |
|----------------------------|-----------|--------------|--|
| Hospital                   | No of ops | No of deaths |  |
| A                          | 47        | 0            |  |
| B                          | 148       | 18           |  |
| C                          | 119       | 8            |  |
| D                          | 810       | 46           |  |
| E                          | 211       | 8            |  |
| F                          | 196       | 13           |  |
| G                          | 148       | 9            |  |
| H                          | 215       | 31           |  |
| I                          | 207       | 14           |  |
| J                          | 97        | 8            |  |
| K                          | 256       | 29           |  |
| L                          | 360       | 24           |  |

We assume that the failure rates across hospitals are *similar* in some way. This is equivalent to specifying a *random effects* model for the true failure probabilities  $p_i$  as follows:

Let  $r_i$  = number of deaths for hospital  $i$

Let  $n_i$  = number of operations (surgeries) performed at hospital  $i$

Assume the conditional model:

$$r_i \sim \text{Binomial}(n_i, p_i)$$

$$\text{logit}(p_i) = \alpha^c + a_i$$

$$a_i \sim N(0, \sigma^2)$$

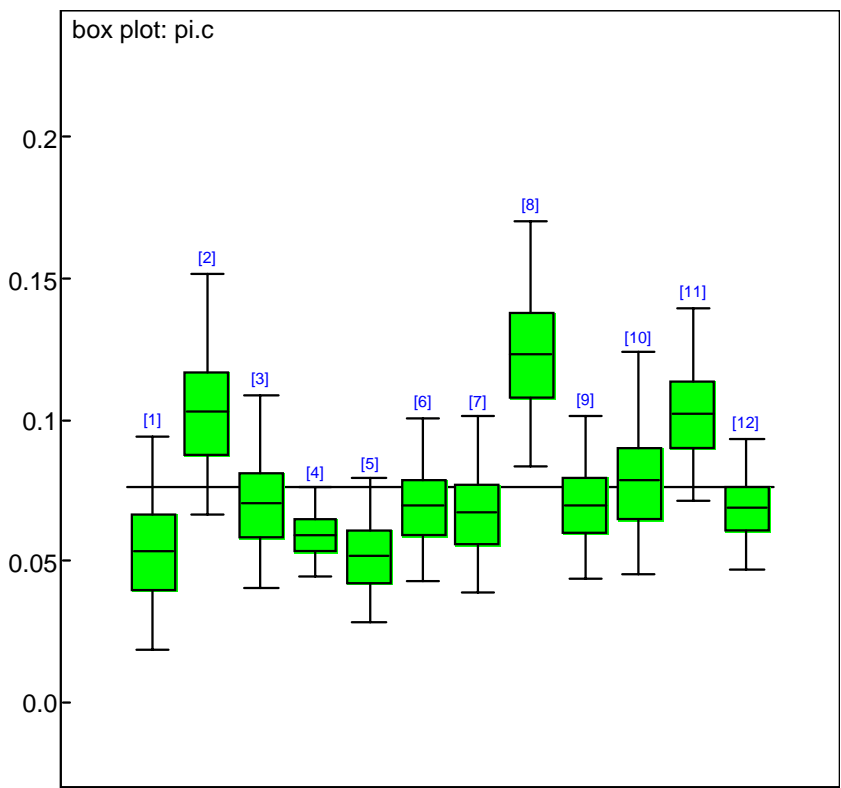
Where:  $p_i$  is the hospital specific probability of death

2. Reproduce the example discussed in class on “Institutional ranking”

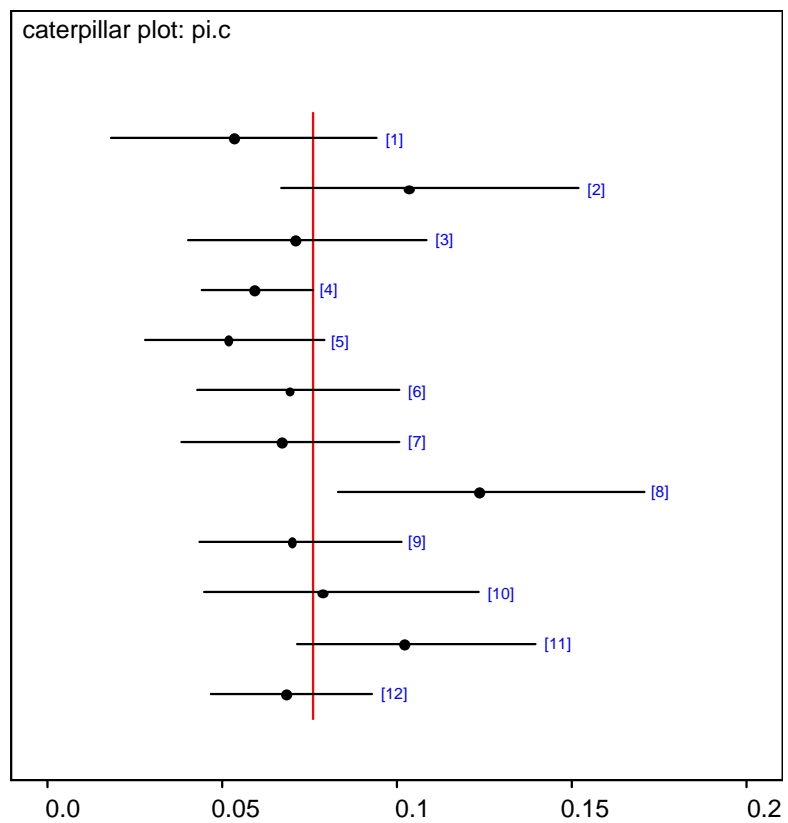
**Results:** (after 100,000 samples)

| Hospital Specific Probabilities of Death (pi.c) |       |       |       |        |        |
|---|-------|-------|-------|--------|--------|
| node  | mean  | sd    | 2.50% | median | 97.50% |
| pi.c[1]   | 0.053 | 0.020 | 0.018 | 0.052  | 0.094  |
| pi.c[2]   | 0.103 | 0.022 | 0.067 | 0.101  | 0.152  |
| pi.c[3]   | 0.071 | 0.018 | 0.040 | 0.069  | 0.109  |
| pi.c[4]   | 0.059 | 0.008 | 0.045 | 0.059  | 0.076  |
| pi.c[5]   | 0.052 | 0.013 | 0.028 | 0.051  | 0.079  |
| pi.c[6]   | 0.069 | 0.015 | 0.043 | 0.068  | 0.101  |
| pi.c[7]   | 0.067 | 0.016 | 0.038 | 0.066  | 0.101  |
| pi.c[8]   | 0.123 | 0.022 | 0.083 | 0.122  | 0.171  |
| pi.c[9]   | 0.070 | 0.015 | 0.044 | 0.069  | 0.101  |
| pi.c[10]  | 0.079 | 0.020 | 0.045 | 0.077  | 0.124  |
| pi.c[11]  | 0.102 | 0.018 | 0.071 | 0.101  | 0.140  |
| pi.c[12]  | 0.069 | 0.012 | 0.047 | 0.068  | 0.093  |

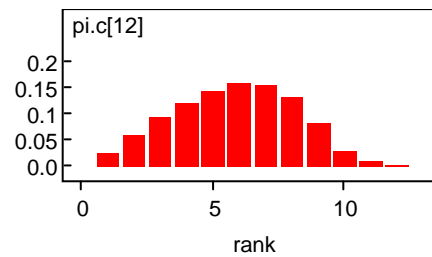
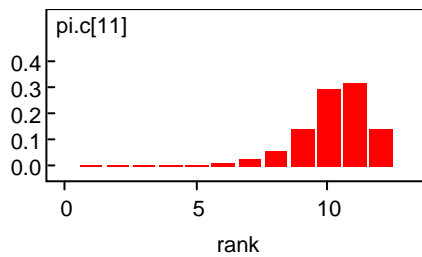
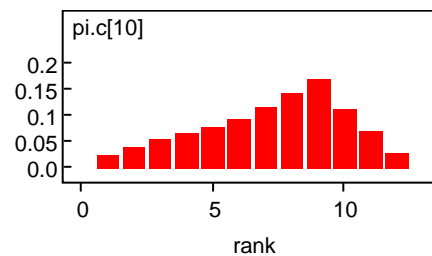
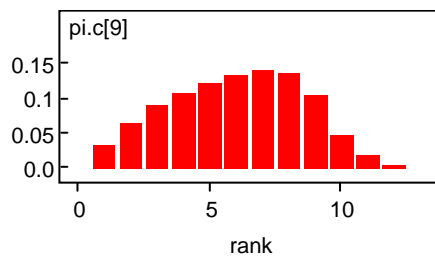
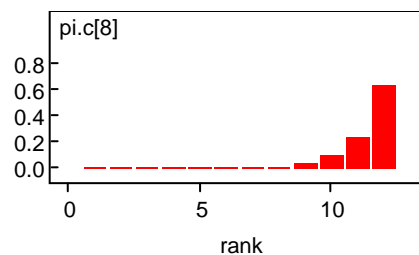
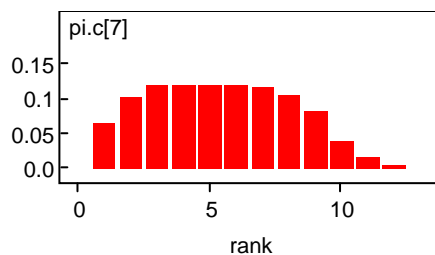
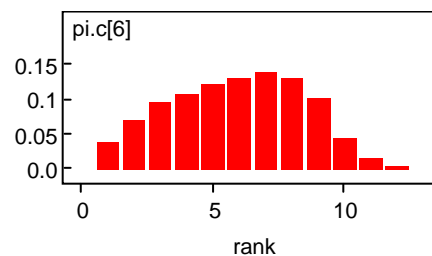
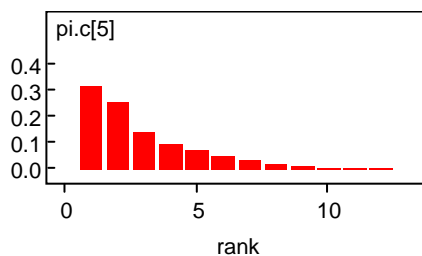
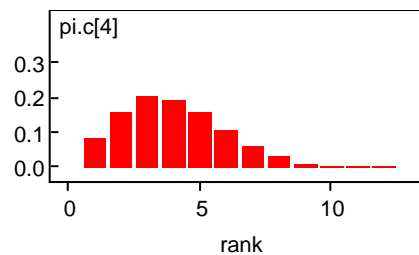
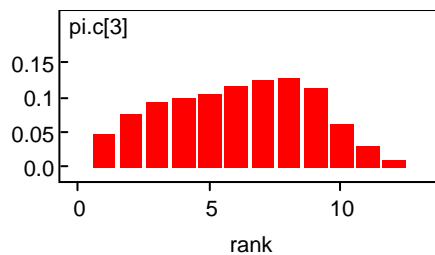
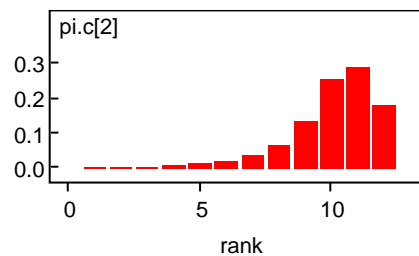
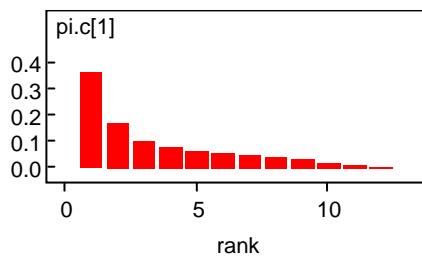
Boxplot Comparisons of Hospital Specific Probabilities of Death



Caterpillar Comparisons of Hospital Specific Probabilities of Death



## Histograms of Rank





### **Results Discussion:**

Hospital 8 has the highest probability of surgical death, ( $\text{Pr}(\text{Death}) \approx 0.12$ ), and has a posterior probability of approximately 0.60 of having the highest probability of surgical death. Hospitals 1 & 5 have the lowest probabilities of surgical death ( $\text{Pr}(\text{Death}) \approx 0.05$ ) with posterior probabilities of being the lowest of approximately 0.35 and 0.30 respectively.

### **3. Discuss whether a marginal model may or may not be appropriate for the analysis of the institutional ranking data.**

A marginal model could also be used for the institutional ranking data. Fixed effects would be included for each hospital and within-hospital correlation could be accounted for using GEE or MMM methods. Inference on ranking would then proceed via the estimates and confidence intervals for each hospital's probability of surgical death. Benefits of using the conditional model include borrowing strength across hospitals, automatic ranking and calculation of uncertainty in the rankings via the MCMC draws. Disadvantages include extrapolating away from the data that was actually observed. The random effects (conditional) model, essentially compares results for a patient in a particular hospital to the results for the patient *if that patient had gone to a different hospital*, but we have no such data.

### **4. Write an abstract for a scientific journal that summarizes the results of the “Institutional ranking” example. Report statistical uncertainty associated with ranking.**

Using mortality data for cardiac surgery on infants, distributions of ranks were created for 12 hospitals according to their hospital-specific probabilities of death. The number of operations ranged from 47 to 810 across hospitals and the number of deaths ranged from 0 to 46. Assuming the mortality probabilities across hospitals were similar, we specified a random effects model to borrow strength across hospitals and account for reliability in the data. Markov Chain Monte Carlo sampling produced posterior distributions of the hospital-specific probabilities of death and of the hospital ranks. These distributions show heavy uncertainty associated with ranking. The estimated probability of surgical death for an 'average infant' among all twelve hospitals was 0.068 (95% Posterior Interval: 0.047, 0.093). Hospital 8 had the highest probability of surgical death, ( $\text{Pr}(\text{Death}) \approx 0.12$ ), and has a posterior probability of approximately 0.60 of ranking the worst of the 12 hospitals. Hospitals 1 & 5 have the lowest probabilities of surgical death ( $\text{Pr}(\text{Death}) \approx 0.05$ ) with posterior probabilities of ranking the best of approximately 0.35 and 0.30 respectively.

### **5. From the case study by Normand and et al JASA 1997, which are the three most “aberrant hospitals”? How is the uncertainty in ranking reported?**

The answer to this depends on your definition of an "aberrant hospital", however, from the 4 ranking methods in Table 4, hospitals 1, 28, and 2 consistently rank among the 9 worst hospitals. Unfortunately, estimates of the uncertainty in these rankings does not appear in the tables or discussion. If the uncertainty in the ranks had also been reported, we would see that trying to pick the three most aberrant hospitals may be of limited value.

## Module IV: Applications of Multilevel Models to Spatial Epidemiology

### 1. From the Winbugs help menu, copy the “Scottish Lip Cancer” example (look under the map menu under examples)

The rates of lip cancer in 56 counties in Scotland have been analysed by Clayton and Kaldor (1987) and Breslow and Clayton (1993). The form of the data includes the observed and expected cases (expected numbers based on the population and its age and sex distribution in the county), a covariate measuring the percentage of the population engaged in agriculture, fishing, or forestry, and the "position" of each county expressed as a list of adjacent counties.

| County | Observed cases<br>$O_i$ | Expected cases<br>$E_i$ | Percentage in agric.<br>$x_i$ | SMR   | Adjacent counties |
|--------|-------------------------|-------------------------|-------------------------------|-------|-------------------|
| 1      | 9                       | 1.4                     | 16                            | 652.2 | 5,9,11,19         |
| 2      | 39                      | 8.7                     | 16                            | 450.3 | 7,10              |
| ...    | ...                     | ...                     | ...                           | ...   | ...               |
| 56     | 0                       | 1.8                     | 10                            | 0.0   | 18,24,30,33,45,55 |

We note that the extreme SMRs (Standardised Mortality Ratios) are based on very few cases.

We may smooth the raw SMRs by fitting a random-effects Poisson model allowing for spatial correlation, using the intrinsic conditional autoregressive (CAR) prior proposed by Besag, York and Mollie (1991). For the lip cancer example, the model may be written as:

$$\begin{aligned} O_i &\sim \text{Poisson}(\mu_i) \\ \log \mu_i &= \log E_i + \alpha_0 + \alpha_1 x_i / 10 + b_i \end{aligned}$$

where  $\alpha_0$  is an intercept term representing the baseline (log) relative risk of disease across the study region,  $x_i$  is the covariate "percentage of the population engaged in agriculture, fishing, or forestry" in district  $i$ , with associated regression coefficient  $\alpha_1$  and  $b_i$  is an area-specific random effect capturing the residual or unexplained (log) relative risk of disease in area  $i$ . We often think of  $b_i$  as representing the effect of latent (unobserved) risk factors.

To allow for spatial dependence between the random effects  $b_i$  in nearby areas, we may assume a CAR prior for these terms.

### 2. Reproduce the statistical analyses performed in class

### 3. For the area with the largest observed SMR, estimate the posterior probability of having the largest relative risk of lip cancer

### 4. For the area with the lowest observed SMR, estimate the posterior probability of having the largest relative risk of lip cancer . Why do you think the rank of the observed SMR and the rank of the smoothed SMR might be different?