Bayesian Hierarchical Models to Assess The Impact of Managed-Care Strategies in Length of Hospital Stays

Dalene Stangl and Gabriel Huerta
Institute of Statistics and Decision Sciences
Sanford Institute of Public Policy
Duke University
Box 90251, Durham NC, 27708, U.S.A.

Abstract

Hierarchical models provide a useful framework for the complexities encountered in policy-relevant research in which the impact of social programs is being assessed. Such complexities include multi-site data, censored data and over-dispersion. In this paper, Bayesian inference through Markov Chain Monte Carlo methods is used for the analysis of a complex hierarchical log-normal model that attempts to determine the impact of public and private-sector managed care strategies aimed at limiting length of hospital stays. Parameters in this model account for existing variability across hospital/strategy and also for institution-dependent effects due to program implementation. We emphasize on prior elicitation and sensitivity analysis with respect to prior beliefs. All the calculations for the posterior and predictive distributions of relevance, were obtained using the software BUGS.

Key words: Hierarchical log-normal model, Prior elicitation, Gibbs sampling, Predictive distribution.

1 Introduction

In the past decade the health-care system has undergone a metamorphosis. No state, county, or hospital has gone unchanged by efforts to 'manage care'. Sky-rocketing public-sector costs and tightened government spending has forced states to take dramatic action to contain costs. This has resulted in a boon to an area of research labeled as "Health-Services Research". Health-services research focuses on determining the most cost-effective ways of delivering health-care treatment. Most often this determination relies on naturalistic-observation and field-study designs rather than randomized controlled clinical trials. This paper fits into the category of "Health-Services Research" in that it assesses the impact of a managed-care strategy. The outcome variable of interest is length-of-stay in the hospital. Study observations are neither randomized nor is there much experimental control over the implementation of the alternative service strategies. In determining the impact of the managed care strategy, a Bayesian hierarchical generalized linear model is used to combine data across hospitals and assess the impact on hospital length-of-stay. Section 2 of the

paper presents the available data, introduces the hierarchical log-normal model that will be used to model the data, and describes the prior distributions. Section 3 presents in terms of posterior distributions for key parameters, predictive distributions for future observations. Section 4 will provide conclusions and discussion of these results.

1.1 The Managed-Care Strategy

Since November of 1986, all elective Medicaid-reimbursed in-patient admissions in North Carolina, including those for psychiatric services, have been subject to review by Medical Review of North Carolina, an independent peer-review organization. In August of 1990, however, the North Carolina Division of Medical Assistance sub-contracted with a private managed-care firm to implement a more stringent utilization review (UR) program covering both elective and emergency admissions to selected hospitals. This UR program has two management components. The first component is a formal precertification process conducted prior to admission for elective admissions and within 2 days of admission for emergency admissions. Conducted by an interdisciplinary team, the intent of the precertification process is to ensure that children's psychiatric admissions meet Medicaid's admission criteria (42-CFR 441.152) and to ensure that this fact is recorded in the medical record. These criteria are that (1) outpatient services available in the community are inadequate to meet the needs of the child, (2) the child has a psychiatric condition requiring services under the direction of a physician and (3) the services rendered must be reasonably expected to lead to an improvement in the child's condition or prevent further regression so that hospitalization will no longer be necessary. The second management component is a length-of-stay review designed to ensure that children are not being kept longer than clinically necessary. While a concurrent review process was used for the first 12 months of the program, starting August, 1991, lengths-of-stay are monitored using a retrospective review process. Medical records of all patients with stays of longer than 30 days are evaluated, while records of all patients with stays under 30 days are reviewed on a sample basis. If the retrospective review suggests that a child's length-of-stay was excessive, all Medicaid reimbursements for the excess days are recouped from the in-patient provider. Initially, the UR program was designed to cover all admissions to any specialty public or private psychiatric hospital and 4 general hospitals. In April of 1991, 14 additional general hospitals came under the review process. Altogether the hospitals covered by the UR program account for over 95% of all admissions. Elective admissions to all remaining hospitals continue to be covered by the Medical Review of North Carolina peer review process.

2 Methods

North Carolina has 35 hospitals that admitted children for mental-health or substance-abuse reasons in the period under study. Only hospitals admitting at least two children in both the pre and post implementation periods of the UR program (I=33) were included in the analysis. Data was provided by the Division of Medical Assistance, the State's Medicaid office, via the Division of MH/DD/SAS. Hospitalizations separated by less than 7 days were treated as a single admission.

Of central interest in this analysis is to describe how the distribution of length-of-stay changes with the implementation of the UR program. Because hospitals entered the program at two time points the pre/post dates differ for the two groups of hospitals. For hospitals beginning the UR program in August, 1990, the time periods classified as pre and post included 11/89-6/90 and 11/90-6/91 respectively. For the group of hospitals beginning the program in April, 1991, the pre and post periods included in the analysis were 7/90-2/91 and 7/91-2/92. The months included in the pre and post time periods were held constant to control for seasonal fluctuations. Hospital stays beginning during the pre period but extending beyond the start of the program were censored at the starting date of the program.

To begin, the distribution of length-of-stay for each hospital was compared to several distributions including the log-normal, exponential, and Weibull. The log-normal provided the closest fit to the Kaplan-Meier empirical distributions, and hence is used throughout this paper.

2.1 Hierarchical Log-Normal Model

The statistical model chosen for length-of-stay needs to account for variability across the 35 hospitals, as well as for the general effect of the UR process and the possibility that the effect of the UR process differs across these 35 hospitals. Additionally, the effect of several individual and hospital-level characteristics were of concern. At the individual level, four covariates were considered. First, number of psychiatric diagnoses was used a proxy for illness severity. Second, because the unavailability of community-based services was a condition for hospitalization and continued hospitalization, a variable that assessed this availability was included. Lastly, gender and age were controlled for in the model. At the hospital level, a single covariate was included. Hospitals were classified into three types: state-funded hospitals, privately-funded specialty-psychiatric hospitals, and privately-funded general hospitals.

Before considering the covariates the following simple model was examined.

$$log(T_{ijk})|\mu_{ij}, \sigma^2_{ij} \sim N(\mu_{ij}, \sigma^2_{ij}) I_{(log(C_{ijk}),\infty)}(log(T_{ijk}))$$

$$i = 1, \cdots, I$$
$$j = 1, 2$$
$$k = 1, \cdots, n_{ij}$$

where T_{ijk} denotes length-of-stay which follows a truncated log-normal distribution with location parameter μ_{ij} and scale parameter σ_{ij}. $I_A(\cdot)$ is the indicator function over the set A. The subscript i denotes hospital, j denotes pre versus post implementation of the UR program, and k denotes observations within each hospital by pre/post subgroup. When an observation is censored, $C_{ijk} = T_{ijk}$ otherwise $C_{ijk} = 0$ so C_{ijk} are censoring indicator variables. For convention, if $C_{ijk} = 0$ then $log(C_{ijk}) = -\infty$. Note that both means and variances are allowed to vary across hospitals and across pre versus post implementation.

Furthermore we are considering that

$$\mu_{ij} = h_i + p_{j(i)}$$

with the restriction that $p_{1(i)} = 0$. Hence the location parameter is just the hospital effect, h_i, for the pre-implementation period, plus $p_{2(i)}$, a deviation from h_i due to the UR program. This deviation is expected to be negative for most hospitals implying a reduction in the length-of-stay. To model the variability in hospital and program effects we add a second stage to the hierarchy.

$$\begin{aligned} h_i &\sim N(\mu_h, \sigma^2_h) \\ p_{2(i)} &\sim N(\mu_p, \sigma^2_p) \\ \sigma_{ij} &\sim IG(\alpha_j, \beta_j) \end{aligned}$$

It is worth noting that the adopted parameterization of our model follows closely the *hierarchical centering* ideas introduced by Gelfand, A.E., et. al. (1995) and that can often enable improve algorithm convergence. Here, the centering is applied separately to h_i and $p_{j(i)}$ which define the overall mean μ_{ij}. Finally prior distributions for stage II parameters were added at the third level of the hierarchy. The choice of these priors will be considered in detail in the next section.

2.2 Prior Specification

Prior specification is a challenging task in Bayesian analysis. The logarithmic data transformation used here added an extra step of complexity to the elicitation procedure. Priors were elicited in the original units

(days) and then adjusted to the log scale. The entire prior specification is

$$
\begin{aligned}
Stage I: \quad log(T_{ijk}) &\sim N(h_i + p_{j(i)}, \sigma^2_{ij}) I_{(log(C_{ijk}),\infty)}(log(T_{ijk})) \\
Stage II: \quad h_i &\sim N(\mu_h, \sigma^2_h) \\
p_{1(i)} &\equiv 0 \\
p_{2(i)} &\sim N(\mu_p, \sigma^2_p) \\
\sigma^2_{ij} &\sim IG(\alpha_j, \beta_j) \\
Stage III: \quad \mu_h &\sim N(4.09, .18^2) \\
\sigma^2_h &\sim IG(3.85, .35) \\
\mu_p &\sim N(-.29, .12^2) \\
\sigma^2_p &\sim IG(2.23, .07) \\
\alpha_j &\sim U(4, 6) \\
\beta_j &\sim Ga(3, 1)
\end{aligned}
$$

The elicitation will be explained using days, and the log conversions will be presented in parentheses where necessary. Distributions for h_i, $p_{j(i)}$, and σ^2_{ij} were defined in the second stage of the hierarchical model. All were set up as conjugate. The prior for h_i, the hospitals' pre-UR average length-of-stay, is a normal distribution with mean μ_h and standard deviation σ_h denoted by $N(\mu_h, \sigma^2_h)$. The prior for $p_{2(i)}$, the hospitals' UR-program effect on average length-of-stay, is a normal distribution with mean μ_p and standard deviation σ_p, denoted by $N(\mu_p, \sigma^2_p)$, while $p_{1(i)} \equiv 0$ for all values of i. The prior for σ^2_{ij}, the variability in length-of-stay within hospitals pre and post implementation, is an inverse gamma distribution with shape parameter α_j, scale parameter β_j and denoted by $IG(\alpha_j, \beta_j)$.

Priors in the third stage of the hierarchy, were set up around the belief that the average of h_i, μ_h, would be around 60 days (4.09 on the log scale) with almost all of the probability being between 30 (3.4) and 90 (4.5) days. If we let the range 3.4 to 4.5 represent ± 3 standard deviations, then the prior for μ_h can be set up as a normal distribution with a mean of 4.09 and standard deviation of .18, $N(4.09, .18^2)$. The prior for the UR-program effect, μ_p, was built around the belief that the UR program would result in an average reduction of average length-of-hospital stay of about 15 days. A reduction in average length-of-stay from 60 (4.09) to 45 (3.81) days would be an effect of -15 in days or -.28 in log(days). The uncertainty in the UR-program effect must reflect the range of potential average reductions of 0 to 30 days. A reduction of 0 days corresponds to a difference in log(days) pre versus post of 0, and a reduction of 30 days corresponds to a difference in log(days) pre versus post of -.69. If we let the range of 0 to -.69 represent ± 3 standard

deviations, then the prior for μ_p can be represented by a normal distribution with a mean of -.28 and a standard deviation of .12, $N(-.28, .12^2)$. This set of priors gives small positive weight even to values more extreme than what was plausible, but yet are clearly informative proper priors.

Priors for σ_h^2 and σ_p^2 are both inverse gamma distributions. The parameter σ_h^2 represents the variance in pre-UR average length-of-stay across hospitals. It was believed that the pre-UR average length-of-stay across hospitals would vary between 15 (2.7) and 120 (4.8) days. If the range of 2.7 to 4.8 represents ± 3 standard deviations, then the prior on σ_h^2 should have a mean of $.35^2$. The minimum plausible variability across hospitals was 45 to 75 days and the maximum plausible variability was 5 to 180 days. Converting these estimates to the log scale and letting the difference between the minimum and maximum represent ± 3 standard deviations yields a prior standard deviation for σ_h^2 of .09 which corresponds to an Inverse Gamma distribution of shape parameter 3.85, scale parameter .35 and denoted by $IG(3.85, .35)$.

It was believed that the UR program could result in a change in average length-of-stay for each hospital between +5 and -45 days. An increase was plausible because the program would effect length-of-stay in two ways. The program potentially could eliminate short stays by diverting the least severe cases from hospitalization all together. This would work to increase the average length-of-stay. On the other end of the distribution the program would also be working to minimize long lengths-of-stay, hence reducing average length-of-stay. For a hospital with a pre-UR average length-of-stay of 60 days (4.09), an increase of 5 days would result in a post-UR length-of-stay of 65 days (4.17), hence a change of 5 in days and .08 in log(days). For a hospital with a pre-UR average length-of-stay of 60 days (4.09), a reduction of 45 days would result in a post-UR length-of-stay of 15 days (2.7), hence a change of 45 in days or 1.39 in log(days). If the range .08 to 1.39 represents ± 3 standard deviations, this supports a prior on σ_p^2 with a mean of $.24^2$. The minimum plausible variability across hospitals in program effect was a reduction of 10 to 20 days and the maximum plausible variability was an increase of 20 to a decrease of 59 days. (We did not allow for the possibility that the program would cause a hospital to close its doors completely.) Converting these estimates to the log scale and letting the difference between the minimum and maximum represent ± 3 standard deviations yields a prior standard deviation for σ_p^2 of .12 which leads to a prior $IG(2.23, .07)$.

Finally at third stage, priors for α_j and β_j are defined. Almost all pre-UR lengths-of-stay were expected to be between 1 and 180 days, and post-UR stays between 1 and 120 days. Letting these ranges represent ± 3 standard deviations, the former converts to a standard deviation of approximately .87 and the latter .80 on the log scale. Hence third stage priors were defined that would yield a distribution for σ_{ij}^2 with a mean of $.87^2$. The mean of an inverse gamma distribution is $\frac{\beta}{\alpha-1}$, so priors for α and β were chosen so that prior means yielded $\frac{\beta}{\alpha-1} = .87^2$. A Uniform distribution with support on (4,6) and denoted by $U(4,6)$ was used as the prior for α_j. A Gamma distribution prior with scale parameter equal to 3 and shape equal to 1 ($Ga(3,1)$) was used for β_j.

These prior distributions represent the belief of one researcher (DS). To assess the sensitivity of results to these priors a series of priors was developed ranging from the base prior above to highly diffuse yet proper priors. In addition, priors that modeled progressively smaller UR-program effects of 10, 5, and 0 days, were also compared to the base prior that modeled a UR-program effect of 15 days.

To incorporate further covariates, the μ_{ij} were modeled as

$$\mu_{ij} = h_i + p_{j(i)} + \beta^{'} x.$$

The $\beta^{'} x$ represents the vector of additional covariates. The effect of each covariate was not allowed to vary across the hospitals. The prior distribution for each β was an independent normal centered at 0 with a precision of .01.

3 Results

Posterior and predictive distributions were calculated using BUGS software. BUGS samples from the full conditional distributions to come up with samples from the desired posterior and predictive distributions. Censored observations are fully incorporated because they are part of the model definition and BUGS allows for truncated distributions through the use of indicator functions. The obtained samples serve as the basis of inference about the hospital and program effects. To obtain inferences about the mean length-of-stay for each hospital we need the mean of T_{ijk} under the log-normal distribution. This mean equals

$$exp(h_i + p_{j(i)} + \sigma_{ij}^2/2)$$

when $C_{ijk} = 0$, otherwise to obtain the correct mean this expression needs to be multiplied by the factor

$$\frac{1 - \Phi((log(C_{ijk}) - h_i - p_{j(i)} - {\sigma_{ij}}^2)/\sigma_{ij})}{1 - \Phi((log(C_{ijk}) - h_i - p_{j(i)})/\sigma_{ij})}$$

where Φ denotes the cumulative distribution function of a standard Normal. Samples from the distribution of this mean are obtained by using the above transformation on samples of h_i, $p_{j(i)}$, σ_{ij}^2. Predictive distributions for future observations can be obtained by generating T_{ijk} from the corresponding Log-normal distributions given the sampled parameter values.

3.1 Descriptive Analysis

Figure 1 presents Kaplan-Meier plots for all 35 hospitals combined and the five hospitals with the most admissions. From the plots we see a UR program effect for all hospitals combined, and for 4 of the 5 largest hospitals we see a clear UR program effect. With all hospitals combined the mean length-of-stay declines from 61.5 pre to 49.1 post, a decline of 11.4 days. The median length-of-stay declines by 17 days from 49 days

pre to 32 days post. The heterogeneity between the 5 hospitals with the most admissions is considerable. Four of the 5 hospitals show clear declines in length-of-stay with means declining between 16.7 and 21.4 days and medians declining between 9 and 22 days. Table 1 presents samples sizes, median length-of-stay pre and post implementation of the UR program, difference in medians and type of institution for all 35 hospitals. Tables 2 and 3 summarize information with respect to the related covariables: "Age", "Gender" and "Number of Diagnosis". The distribution of both "Age" and "Diagnosis" appeared to be homogeneous across "Gender". This justifies the use of a general covariate effect. None of the subjects in the sample reported more than 3 cases of diagnosis. Figure 2 presents a plot of hospital number against the percentage of censored observations. The figure shows that the percentages have a large variability across hospitals and particularly for hospital 19 the percentage of censored data is closed to 50%. In addition, the overall percentage of censored observation is 6.5%. Clearly, the estimation of individual program effects can be misleading if censoring is not given appropriate consideration.

3.2 Posterior and Predictive Distributions

Based on 2000 samples from the desired posterior, Figure 3 presents density estimators for the posterior densities of μ_h, μ_p, σ_h^2, and σ_p^2. The posterior distributions for μ_p and σ_p^2 are of most interest as they reflect the UR-program effect and the heterogeneity in the UR-program effect across hospitals. We see that the posterior mode for μ_p is -0.45. The area to the right of 0 represents the posterior probability that the program increased rather than decreased average length-of-stay. This probability is very small. Also with this 2000 samples, 95% highest posterior density regions, posterior medians for mean length of stay, pre and post of an "unobserved" hospital, and difference in posterior medians are presented in Table 4 for several models. The only difference among the first 4 models, numbered 1.1 to 1.4, is the location of the UR-program effect, μ_p. The second set of 4 models numbered in the table as 2.1-2.4, replicates the first 4, but increases the standard deviations of each prior distribution by a factor of 2. The last set of 4 models indexed as 3.1-3.4, replicates the first 4 but increases the standard deviation of each prior distribution by a factor of 4. The table shows that the posterior distribution for μ_p is quite robust across the various prior specifications. The difference between posterior medians for mean length-of-stay pre and post UR-program implementation ranges between 10 and 14 days. Even for model 1.4, a model that placed equal prior mass on positive and negative UR-program effects, the 95% HPD for μ_p is (-0.44,-0.22). To compute this posterior medians for an unobserved hospital, at the model specification we increased I, the number of total institutions by one, and sampled the new additional parameters from stage II of the hierarchical model. Then samples of the mean length-of-stay of a new hospital are obtained and posterior summaries can be produced.

The density estimator for the posterior of σ_p^2 in Figure 3, is clearly centered away from zero, supporting the notion that the UR-program effect was not homogeneous across hospitals. Figure 4 shows more clearly

the heterogeneity in program effect across the hospitals. It displays box-plots for 2000 posterior samples of $p_{2(i)}$, the UR-program effect for each hospital. Here we see that all hospitals show a decline in average length-of-stay, with only a few hospitals (8, 12, and 32) showing posterior medians greater than -.25. Similar plots were drawn for the other 11 prior distributions considered. Differences were minute, and hence are not presented here.

Results did not change when additional covariates as "Age", "Number of Diagnosis" or "Gender", were added to the model. None of the covariates either at the individual level or at the hospital level added predictive power to the model. That is, the 95 % posterior probability intervals for each of the βs included the value zero. In addition the inclusion of these covariates did not change the posterior distributions of the hospital or UR-program effects.

While these posterior distributions are useful for understanding the impact of the program and the heterogeneity across hospitals, they do not provide the answers in a manner easily understood by non-statisticians. Decision makers want to know how many hospital days are saved at each of the hospitals and for the hospitals in general, and what would we expect if we implemented the program at hospitals not represented in the sample. These questions are more directly addressed by the use of predictive distributions. Predictive distributions provide our updated beliefs about the distribution of the outcome of interest for observations not yet observed. Predictive distributions are in the metric most understandable to the decision maker, length-of-stay in days, rather than the often times foreign metric of the model parameters. Figure 5 presents the predictive distributions for length-of-stay for the 5 hospitals with the most observations as well as for a 6th as yet unobserved hospital. The two curves represent the predictions with/without the UR program and except for hospital 23, a reduction in length-of-stay is noticeable. These distributions are usually compared to Kaplan-Meier curves to assess lack of fit of the model but here such comparisons are avoided. A predictive distribution is built on a combination of prior beliefs and sample information and its performance should not only be evaluated in terms of a data summary such as the Kaplan-Meier estimator. The curves were obtained by smoothing histograms of samples from the predictive distributions.

3.3 MCMC convergence

When BUGS is implemented in practice for complicated hierarchical models convergence of the simulation method can only be obtained after several thousand iterations. In our case, we implemented the MCMC and after a burn-in of 2000 iterations, collected the next 2000 iterations of two chains created with different starting values. Then, we applied the convergence diagnostic described in Gelman, A., et.al. (1992), for which the basic idea consists in estimating the factor by which the scale of the current distribution is reduced, if more iterations of the chain were continued. This *shrink factor* can also be thought as the comparison of a current variance estimate to a within-sequence variance estimate. The median and 97.5% quantile of

the shrink factors for the parameters μ_h, μ_p, σ_h^2 and σ_p^2 appear plotted against the iteration in Figure 6. Essentially we noticed that after 2200 iterations there is evidence that the shrink factor will be below the suggested value of 1.2 for all the cases. Similar results were obtained with other parameters of the model.

4 Conclusions and Discussion

This paper has demonstrated the use of Bayesian hierarchical models to explore the existence of a UR program effect. The analysis clearly shows an effect and gives us predictive distributions that allow to easily incorporate results into a decision-theoretic framework. Although many issues deserve discussion.

Eliciting priors on the log scale resulted simple but has some drawbacks. For example, prior probability assessments in log parameters are not preserved in the original units. Also, properties as symmetry and skewness can be affected due to the transformation. In this sense, the priors used here require a more careful study of what their meaning is in the original scale.

The assumption of independence between hospital means h_i and program effect $p_{2(i)}$ could be in general unrealistic. Perhaps length-of-stay will change dependent on the average length-of-stay by hospital. For the data analyzed here, Figure 7 presents the sample means Pre for each hospital plotted against the mean increase/decrease with the least square for the scatter plot. There are no obvious trends in the picture, in fact, the square of the correlation coefficient is 0.0019. We produce a similar plot but for the logarithm of length-of-stay and did not noted any particular patterns.

On the other hand and based on the model, Hospitals 12, 19 and 20 showed decreases in the length-of-stay pre/post while the exploratory analysis exhibits the contrary. To determine if this was due to the Normal distribution assumpiton, we ran again BUGS using t-distributions with 5 degrees of freedom at the second and third stages and with location and scale parameters so that the same prior means and standard deviations were preserved. For this case, a picture like the one on Figure 4 is presented in Figure 8. This leads essentially to the same conclusions as before, though the variability of the boxes is wider because of the fat tails of the t distribution.

Acknowlegements

The authors are greatful for the valuable comments and suggestions of the two reviewers. This research was supported by .....

Table 1: Data: Sample Sizes and Median Length-of-Stay Pre and Post Implementation of the UR Program.

Hospital	N pre-UR	N post-UR	Median LOS pre-UR	Median LOS post-UR	Difference	Type Of Hospital
1	27	24	31	14	-17	General
2	52	62	73	46	-27	General
3	3	4	6	4	-2	General
4	33	13	61	29	-32	General
5	50	44	39	27	-12	General
6	18	15	36	24	-12	General
7	9	17	28	24	-4	General
8	53	43	24	23	-1	General
9	18	8	13	6	-8	General
10	27	16	29	38	9	General
11	0	6	–	27	–	General
12	29	33	69	123	54	General
13	38	70	42	16	-26	General
14	27	32	21	16	-5	General
15	15	34	71	35	-36	General
16	12	23	30	15	-15	General
17	4	1	15	13	-2	General
18	21	35	57	39	-28	Private
19	7	3	122	150	28	Private
20	62	42	58	77	22	State
21	87	101	37	28	-9	State
22	78	87	47	30	-17	State
23	69	97	48	42	-6	State
24	28	52	81	63	-28	Private
25	25	35	33	32	-1	Private
26	117	153	49	29	-20	Private
27	5	5	51	21	-30	Private
28	19	49	24	27	3	Private
29	39	14	43	16	-27	Private
30	100	82	44	31	-13	Private
31	90	80	52	37	-15	Private
32	49	50	25	23	-2	Private
33	5	15	76	36	-40	Private
34	51	30	45	44	-1	Private
35	5	3	14	16	2	Private
Average					-9.35	
Median					-10.50	
SD					19.03	

Note that 19 of the 35 hospitals admitted more children post-UR, 15 admitted fewer, and 1 had no change. The average change in admissions was 16.6%. The increase in number of children eligible for Medicaid increased by approximately 17% between fiscal years '90-'91 and by approximately 16% between fiscal years '91-'92.

Table 2: Basis statistics for covariable Age for all the sample and by gender.

	Mean	Standard Deviation
All the sample	13.54 yrs.	2.61 yrs.
Boys	13.17 yrs.	2.84 yrs.
Girls	13.91 yrs.	2.24 yrs.

Table 3: Proportions of number of diagnosis for all the sample and by gender.

Number of cases	0	1	2	3
All the sample	0.56	0.26	0.07	0.11
Boys	0.58	0.23	0.07	0.11
Girls	0.52	0.28	0.07	0.12

Table 4: Posterior Medians of Mean Length-of-Stay for an Unobserved Hospital and 95% HPD Regions for UR-Program Effect.

Model	Prior mean of μ_p	Median LOS pre-UR Days	Median LOS post-UR Days	Difference	95% HPD for μ_p
1.1	-.29	42	28	14	(-0.51, -0.29)
1.2	-.18	40	27	13	(-0.48, -0.27)
1.3	-.09	40	28	12	(-0.47, -0.26)
1.4	-.00	39	28	11	(-0.44, -0.22)
2*Baseline SDs					
2.1	-.29	39	27	12	(-0.53, -0.29)
2.2	-.18	37	27	10	(-0.52, -0.28)
2.3	-.09	40	27	13	(-0.51, -0.28)
2.4	-.00	39	26	13	(-0.51, -0.28)
4*Baseline SDs					
3.1	-.29	38	27	11	(-0.54, -0.30)
3.2	-.18	40	28	12	(-0.53, -0.29)
3.3	-.09	39	27	12	(-0.53, -0.29)
3.4	-.00	38	25	13	(-0.52, -0.29)

Figure 1: **Kaplan-Meier Plots for all Hospitals Combined and the Five Hospitals with the Most Admissions**

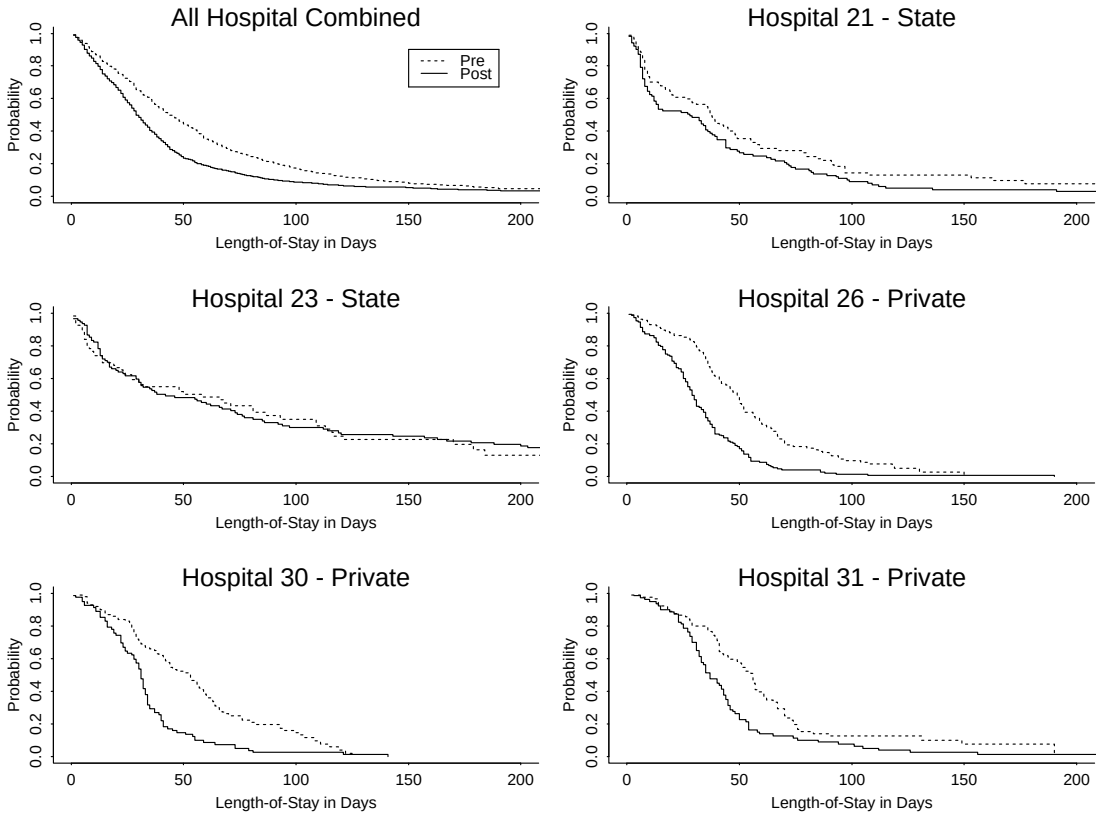


Figure 2: **Proportion of Censored Observations Across Hospitals**

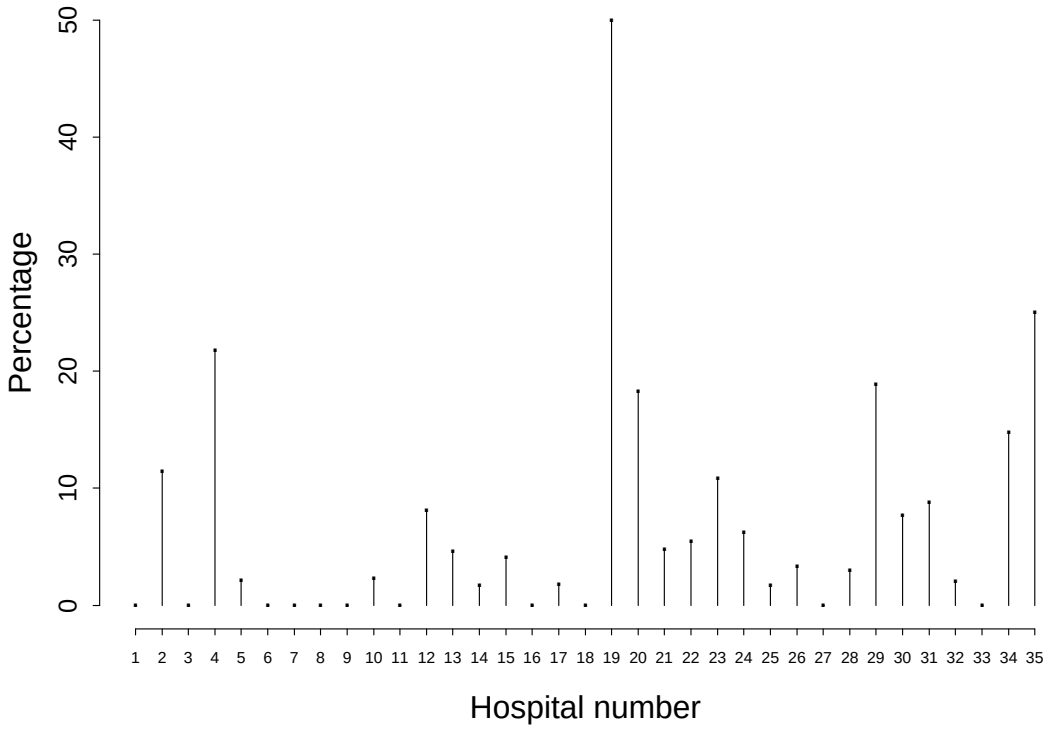


Figure 3: **Posterior Distributions for Stage II Means and Variances**

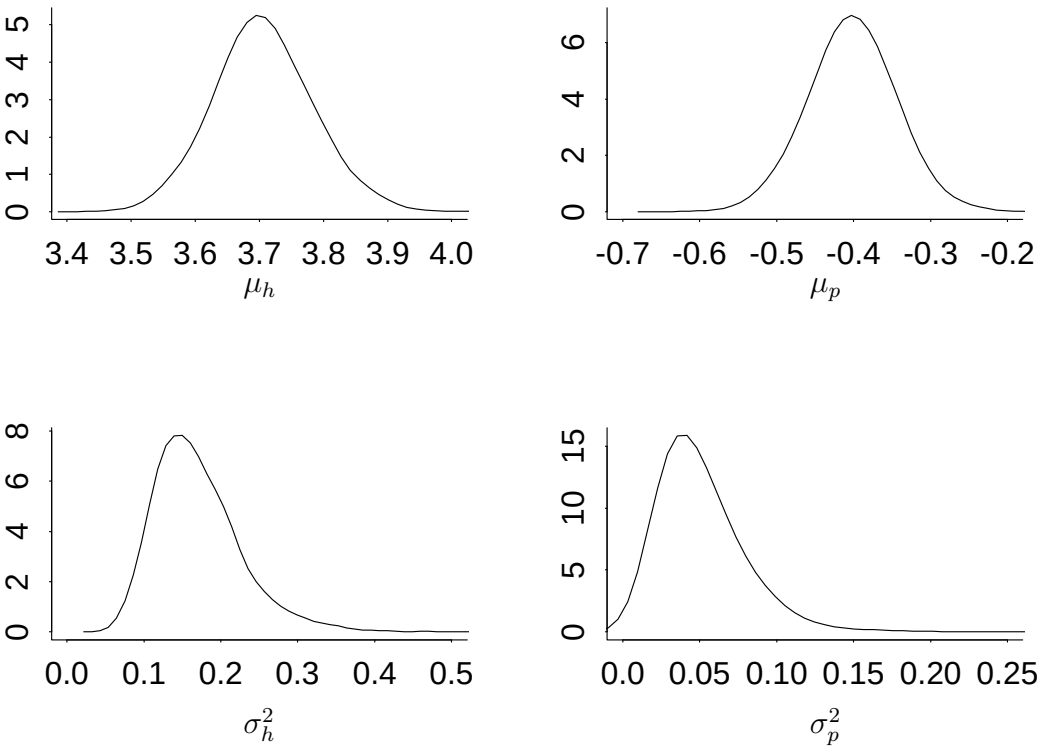


Figure 4: **Heterogeneity of UR-Program Effect Across Hospitals**

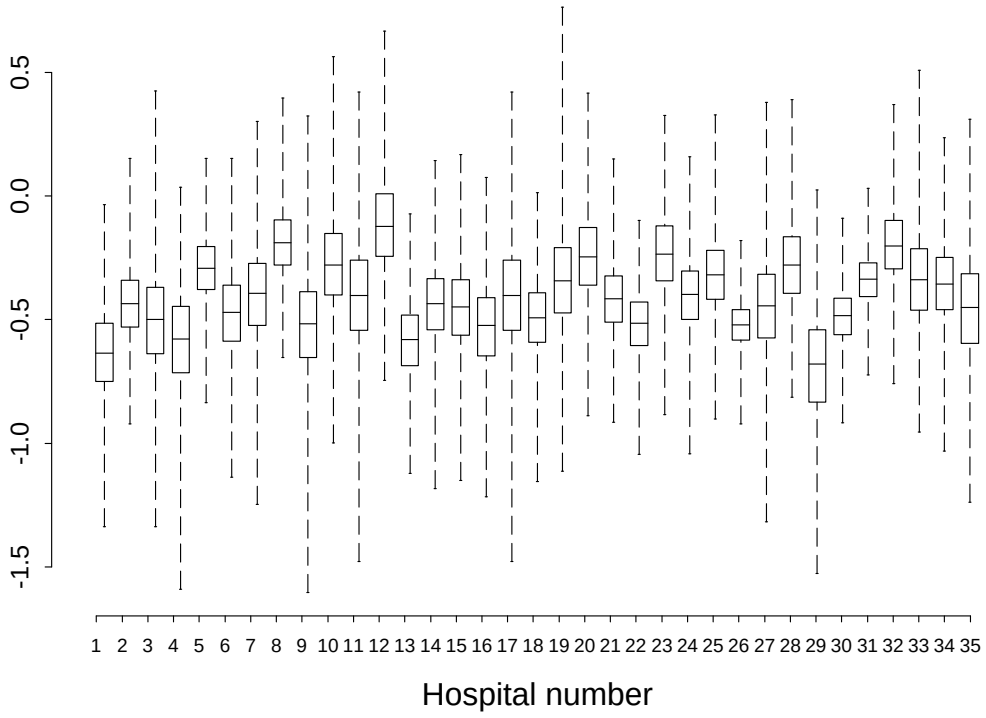


Figure 5: **Predictive Distributions of Length-of-Stay for Future Observations**

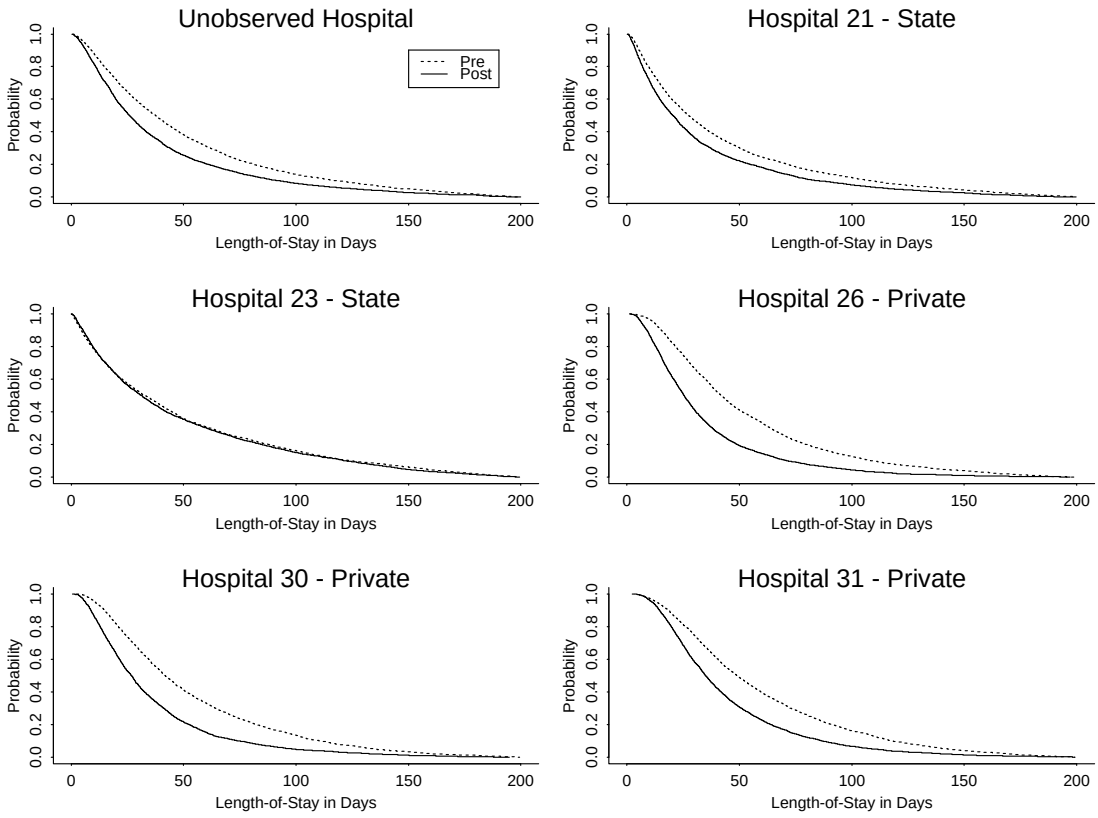


Figure 6: **Gelman and Rubin Diagnostic for Stage II Means and Variances**

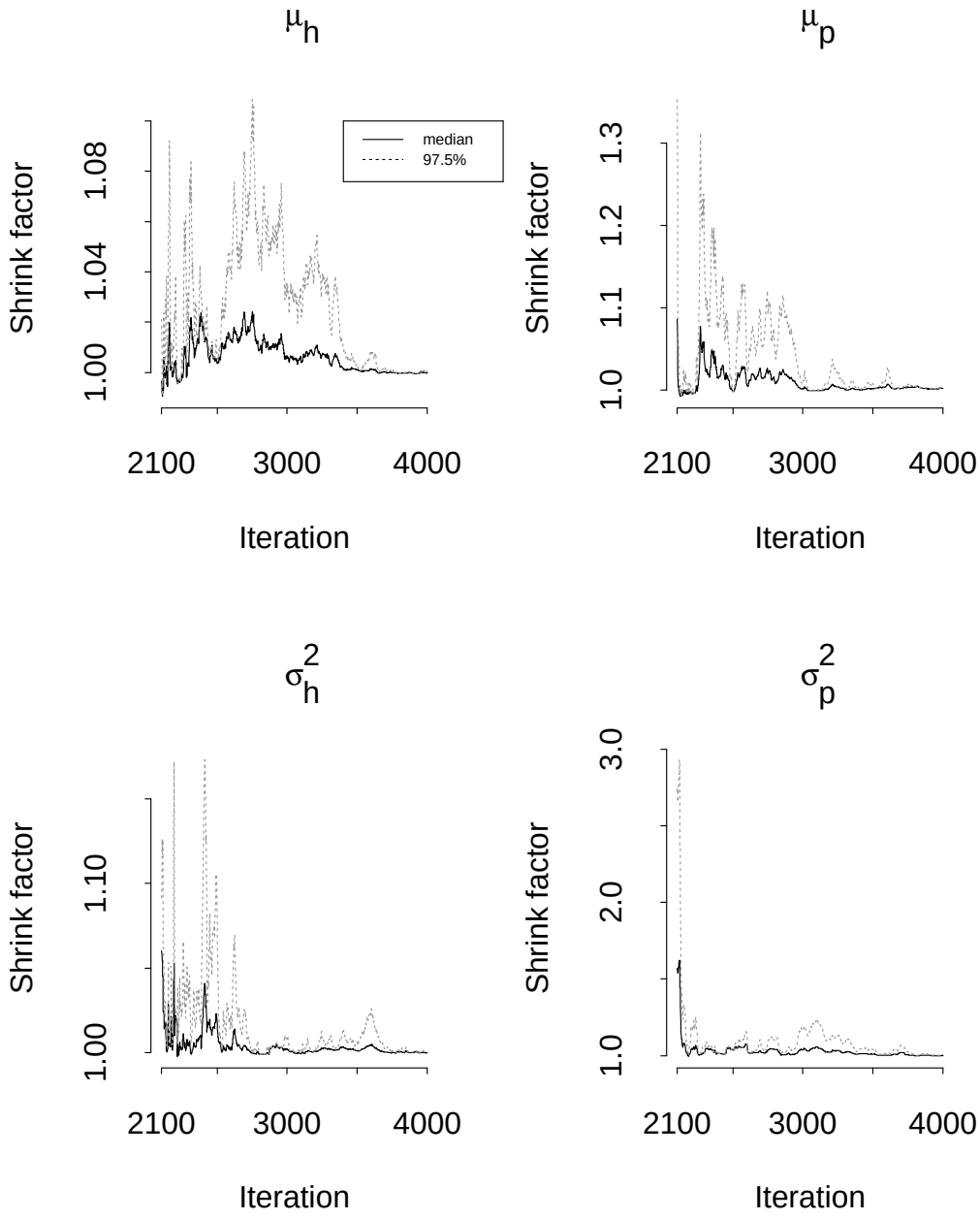


Figure 7: **Means PRE vs. Mean Increase-Decrease Across Hospitals**

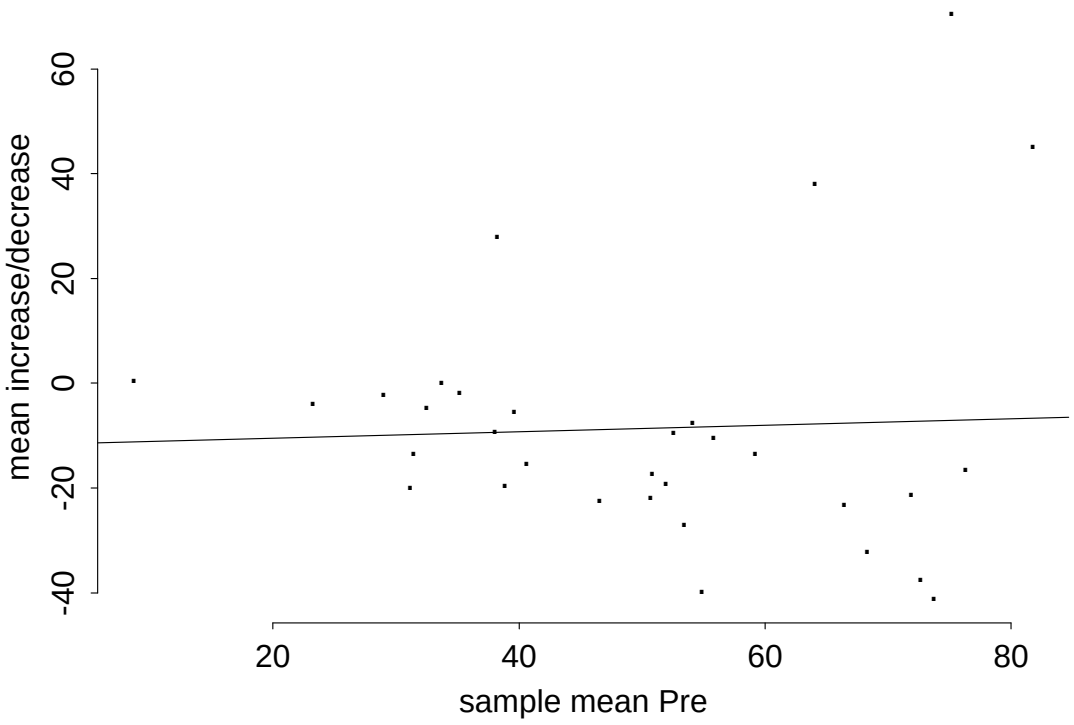
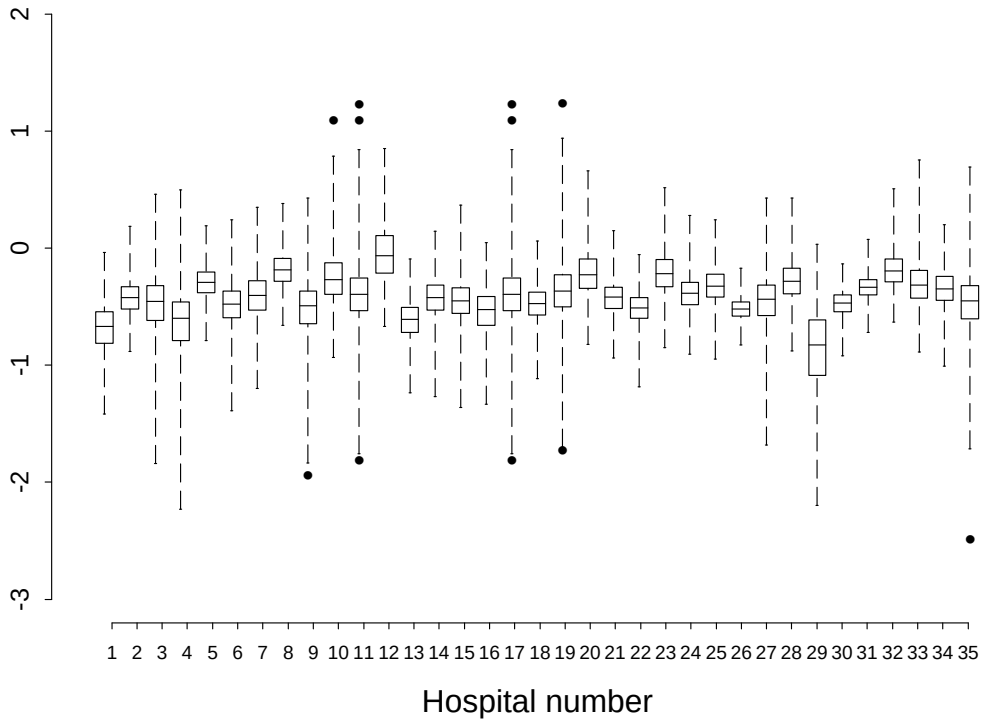


Figure 8: **Heterogeneity of UR-Program Effect Across Hospitals. T distribution**



References:

- Best, N., Cowles, M.K., Vines K. (1995). CODA: Convergence Diagnosis and Output Analysis Software for Gibbs sampling output. Version 0.30. MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge.
- Clayton, D. G. (1991). A Monte Carlo Method for Bayesian Inference in Frailty Models. *Biometrics*, **47**, 467-485.
- Gelfand, A.E., Sahu, S.K., and Carlin, B.P. (1995). Efficient parameterizations for generalized linear mixed models. Bayesian Statistics 5, eds. J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, Oxford University Press.
- Gelfand, A.E. and Smith, A.F.M. (1990). Sampling based approaches to calculation of marginal densities. *Journal of the American Statistical Association*, **85**, 398-409.
- Gelman, A. and Rubin, D.B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, **7**, 457-511.
- Gilks, W.R., Clayton, D.J. , Spiegelhalter, N.G., Best, A.J., McNeil, A.J., Sharples, L. D., and Kirby, A. J. (1993). Modeling Complexity: Applications of Gibbs Sampling in Medicine. *JRSS B*, **55**, 39-52.
- Gray R. (1994) On Institutional Effects in Multicenter Clinical Trials *Biometrics*, **50**, 244-253.
- Kuo, L. and Smith, A.F.M. (1991). Bayesian computations in Survival Models via Gibbs Sampler. Invited Paper NATO Advanced Studies Workshop on Survival Analysis and Related Topics, Columbus, Ohio.
- Morris, C.N. and Normand, S.L. (1992) Hierarchical models for combining information and for meta-analysis. Bayesian Statistics 4, eds. J.M. Bernardo, J.O. Berger, D.V. Lindley, and A.F.M. Smith, Oxford University Press.
- Morris, C., Norton E., and Zhou X. (1994). Parametric Duration Analysis of Nursing Home Usage. Cases Studies in Biometry, eds. N. Lange, L. Ryan, L. Billard, D. Brillinger, L. Conquest and J. Greenhouse, John Wiley and Sons, Inc.
- Smith, A.F. and Roberts, G.O. (1993). Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods. *JRSS B*, **55**, 3-23.
- Spiegelhalter, D., Thomas, A., Best, N., and Gilks, W. (1995). BUGS. Bayesian Inference using Gibbs sampling. Version 0.30. MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge.

- Stangl, D. (1995) Hierarchical Analysis of Continuous Time Survival Models. In Bayesian Biostatistics, ed. D. Berry and D. Stangl, Marcel Dekker.

- Stangl, D. (In Press) Prediction and Decision Making Using Bayesian Hierarchical Models. *Statistics in Medicine*.