A Hierarchical Bayes Approach to a Study of Hospital Variation in Surgical Procedures

Patrick J. Farrell
School of Mathematics and Statistics
Carleton University

Susan Groshen
Department of Preventive Medicine
School of Medicine, University of Southern California

Brenda MacGibbon
Département de Mathématiques
Université du Québec À Montréal

and

Thomas J. Tomberlin
Department of Decision Sciences and Management Information Systems
Concordia University

May, 2002

Les Cahiers du GERAD

G-2002-22

Abstract

There is increased interest in rating types of hospitals or geographical regions containing hospitals on the basis of their performance in the provision of certain types of medical services or procedures. Another important aspect of such a study could be the identification of exceptional hospitals or regions. The outcome variable of interest is often a binary one representing the type of procedure used or a successful outcome of a service. Because of the hierarchical nature of the data (for example, patients, doctors, hospitals, hospital type within hospitals, regions or small areas, etc.), a hierarchical model should be used. Such studies lend themselves very easily to a multiple logistic regression model with mixed effects; that is, patient information such as age, gender and hospital information such as caseload are considered as fixed effects while hospital and/or region are modeled as random effects. The hierarchical Bayes approach proposed here also allows for the standardization of the random effects which permits the use of normal probability plots for the detection of outliers and exceptional cases. For a large well-defined hospital population, Simons et al (1997) recently reported statistically significant differences in surgical choices in the treatment of rectal cancer seemingly due to hospital type and caseload within hospital type. Their data are used here to illustrate the utility of hierarchical Bayes techniques for parameter estimation and outlier detection in a logistic regression model with random effects in such a study.

Résumé

Il y a un intérêt croissant dans l'évaluation des types d'hôpitaux ou de régions géographiques contenant des hôpitaux sur la base de leur performance pour fournir certains types de procédures ou services médicaux. Un autre important aspect de telles études peut être l'identification d'hôpitaux ou régions exceptionnels. Les variables d'intérêt sont souvent binaires, représentant le type de procédure utilisée ou les résultats réussis d'un service. À cause de la nature hiérarchique des données (par exemple, patients, docteurs, hôpitaux, types d'hôpitaux, régions ou petits territoires, etc.) un modèle hiérarchique doit être utilisé. De telles études conduisent très facilement à des modèles de régression logistique multiple avec des effets mixes, c'est-à-dire que l'information sur les patients telles que l'age, le sexe et l'information de l'hôpital telle que l'occupation de l'hôpital sont considérés comme des effets fixes, alors que les hôpitaux et/ou les régions sont modelés comme des effets aléatoires. L'approche Bayes hiérarchique proposée ici permet aussi la standardisation des effets aléatoires qui permettent l'utilisation des courbes de probabilités normales pour la détection des valeurs aberrantes et des cas exceptionnels. Pour une population grande et bien définie d'hôpitaux, les données de Simons et al (1997) sont utilisées pour illustrer l'utilité des techniques Bayes hiérarchique pour les paramètres d'estimation et les détections de valeurs aberrantes dans un modèle de régression logistique avec des effets aléatoires dans une telle étude.

1. Introduction

A problem of high priority in medical research is the analysis of health services data that have a hierarchical structure. There is increased interest in rating types of hospitals or geographical regions containing hospitals on the basis of their performance in the provision of certain types of medical services or surgical procedures. The process of comparing the quality of care or the use of services and cost with established standards is called profiling (Normand et al 1997). A study is often organized and analyzed in order to assist in medical decision making *vis a vis* the procedure or service.

According to Normand et al (1997) the most important aspect of profiling is to develop a method of evaluating the providers of medical care. In order to do this successfully, Daniels et al (1997) suggest that the analysis of variation is important in the study of the type of service provided or of other outcomes. They indicate that a very important question to answer is whether comparable patients receive similar treatments or procedures across the medical care providers and/or hospitals and suggest that hierarchical regression models are an excellent choice for analysis of such data that have many levels and a nested structure.

Many generalizations of the hierarchical logistic model originally proposed by Wong and Mason (1989, 1991) exist. In particular, we cite Daniels et al (1997) Malec et al (1997) and Albert and Chib (1993), who generalized the method to allow for the analysis of polychotomous responses. Using another generalized linear model, Christiansen et al (1996, 1997) were very successful in modeling patient mortality rates in heart transplant patients using a two-level Poisson regression model. For an overview of the use of hierarchical generalized linear models in the study of health care, see Daniels et al (1999).

However, as the outcome variable of interest is often a binary one representing the type of procedure used or a successful outcome of a service, we will concentrate on this type of outcome here. Because of the hierarchical nature of the data (for example, patients, doctors, hospitals, hospital type within hospitals, regions or small areas, etc.), a hierarchical model should be used. A goal of such an analysis would be to reduce variation in the estimation of the individual parameters while properly accounting for it.

Such studies lend themselves very easily to a multiple logistic regression model with mixed effects; that is, patient information such as age and gender, and hospital information such as caseload are considered as fixed effects while hospital and/or region are modeled as random effects. One of the difficulties of such an analysis is the numerical computation of model parameter estimates. One possible approach is an empirical Bayes procedure using the EM algorithm (Dempster et al 1977). As it is not feasible in the case of logistic regression to obtain a closed form expression for the posterior distribution of the parameters, the approximation often used to avoid the intractable numerical integration is that originally proposed by Laird (1978). This method has been used successfully for logistic regression with random intercepts by Stiratelli, Laird and Ware

(1984) and for small area estimation of proportions by Dempster et al (1980), MacGibbon et al (1989), Farrell et al (1994, 1997).

However, when the number of random effects is large, this computational approach is no longer feasible and we turn to a hierarchical Bayes approach with the Gibbs sampler, originally proposed by Geman and Geman (1984). An excellent discussion of this method can be found in Gelfand and Smith (1990). Another potential computational difficulty occurs when the number of cases in certain hospitals is very small or when the random effects distribution is highly skewed. In such cases an adaptive rejection Metropolis sampling scheme as proposed by Gilks, Best and Tan (1995) can be appended to the Gibbs sampling algorithm.

Regression diagnostics, analogous to those introduced by Pregibon (1981) in the context of ordinary logistic regression, including normal probability plots within the hierarchical Bayes framework, using methods developed by Farrell et al (1994) in an empirical Bayes setting can also be used here to identify possible outliers.

Simons et al (1997) recently report statistically significant differences in surgical choices in the treatment of rectal cancer seemingly due to hospital type and caseload within hospital type. Their data are used here to illustrate the utility of the hierarchical Bayes approach described above for parameter estimation in a logistic regression model with hospital modeled as a random effect in such a study.

The paper is organized as follows. Section 2 presents the hierarchical model with fixed and random effects that will be used to analyze this type of data. Section 3 describes how to obtain hierarchical Bayes point and interval estimates for such a model using adaptive rejection Metropolis sampling within the Gibbs sampler as originally proposed by Gilks, Best and Tan (1995). Regression diagnostics are also discussed here. In section 4 the data from Simons et al (1997) are analyzed and discussion of the approach is given in section 5.

2. The hierarchical Bayes model with fixed and random effects for the analysis of dichotomous choice data

We restrict our attention here to the analysis of data leading to a dichotomous choice and for convenience we assume that the random effect variable is labeled as hospital, although it could equally well be a region, etc. More precisely let $Y_{ij} = 1$ if the j-th patient treated at hospital i receives one option, and let $Y_{ij} = 0$ in the case that the patient received any other option. The random variable Y_{ij} follows a Bernoulli distribution with parameter π_{ij}, reflecting the probability that the j-th patient treated at hospital i receives the first option. To study the effects of various covariates on the choice of

surgery for rectal cancer, one possibility is to make use of a simple logistic regression model:

$$Y_{ij} \sim Bernoulli\ (\pi_{ij}),$$

$$logit\ (\pi_{ij}) = ln\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = x_{ij}'\beta, \tag{1}$$

where x_{ij} is a vector, augmented by the constant one, of covariates associated with the *j*-th patient at the *i*-th hospital while β is an associated parameter vector which contains a constant term β_0. Maximum likelihood estimation of the parameters in (1) is straightforward and can be accomplished using any statistical software package.

However, this model may not describe the data well if there are covariates other than those in x_{ij} that influence the variation in rates from one hospital to another. If this is the case, the following random effects logistic regression model may be more appropriate:

$$Y_{ij} \sim Bernoulli\ (\pi_{ij}),$$

$$logit\ (\pi_{ij}) = ln\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = x_{ij}'\beta + \delta_i, \tag{2}$$

$$\delta_i \sim Normal\ (0,\ \tau^2).$$

The quantity δ_i is a random effect associated with the *i*-th hospital. Thus, the model assumes that the hospitals included in the data set are a sample of all possible hospitals. The purpose of the random effect terms in the model is to account for the influence of unobserved covariates on the hospital-to-hospital variation in the proportion of patients receiving a particular option. We assume that these effects are normally distributed with an unknown variance τ^2. Estimation of the parameters in (2) can be accomplished via a hierarchical Bayes approach. This requires the specification of a prior distribution for τ^2. We consider a diffuse version of an inverse gamma distribution for the random effect variance in what follows.

3. Hierarchical Bayes estimation and regression diagnostics for the mixed effects model.

To develop hierarchical Bayes estimates for the parameters in the hierarchical model given by (2) requires posterior distributions of the model parameters. However, it is only possible to know these distributions up to a constant of proportionality (see Gilks, Best, and Tan 1995); specifically the posterior distribution for any given parameter is proportional to the product of all terms in the model that contain it. Therefore, for Model (2), if Y and δ are vectors containing Y_{ij} and δ_i respectively, then

$$f(\beta_0 \mid Y, \beta_1, \ldots, \beta_m, \delta, \tau^2) \;\alpha\; \prod_{ij} \pi_{ij}^{y_{ij}} (1-\pi_{ij})^{1-y_{ij}},$$

$$f(\beta_u \mid Y, \beta_0, \beta_1, \ldots, \beta_{u-1}, \beta_{u+1}, \ldots, \beta_m, \delta, \tau^2) \;\alpha\; \prod_{ij} \pi_{ij}^{y_{ij}} (1-\pi_{ij})^{1-y_{ij}},$$

$$f(\delta_i \mid Y, \beta, \delta_1, \ldots, \delta_{i-1}, \delta_{i+1}, \ldots, \delta_n, \tau^2) \;\alpha\; \prod_{ij} \pi_{ij}^{y_{ij}} (1-\pi_{ij})^{1-y_{ij}} \exp\left(-\frac{1}{2}\sum_i \frac{\delta_i^2}{\tau^2}\right),$$

$$f(\tau^2 \mid Y, \beta, \delta) \;\alpha\; \frac{1}{\tau^{n+2}} \exp\left(-\frac{1}{2}\sum_i \frac{\delta_i^2}{\tau^2}\right),$$

where *u* refers to the *u*-th covariate, *m* is the number of covariates, and *n* is the number of sampled hospitals.

Under Gibbs sampling, an initial set of values would be assumed as the estimates for β, δ, and τ^2, say $\hat{\beta}_{\{0\}}$, $\hat{\delta}_{\{0\}}$, and $\hat{\tau}^2_{\{0\}}$. An updated estimate for β_0, say $\hat{\beta}_{0\{1\}}$, is obtained by sampling from the full conditional distribution $f(\beta_0 \mid Y, \hat{\beta}_{1\{0\}}, \ldots, \hat{\beta}_{m\{0\}}, \hat{\delta}_{\{0\}}, \hat{\tau}^2_{\{0\}})$. Sampling from the full conditional distribution $f(\beta_1 \mid Y, \hat{\beta}_{0\{1\}}, \hat{\beta}_{2\{0\}}, \ldots, \hat{\beta}_{m\{0\}}, \hat{\delta}_{\{0\}}, \tau^2_{\{0\}})$ based on $\hat{\beta}_{0\{1\}}$ yields the revised estimate $\hat{\beta}_{1\{1\}}$ for β_1. The completion of a first iteration is realized once the revised estimates $\hat{\beta}_{\{1\}}$, $\hat{\delta}_{\{1\}}$, and $\hat{\tau}^2_{\{1\}}$ are obtained. This procedure of sampling from full conditional distributions using the most up-to-date revised estimates continues until the estimates of each parameter are deemed to have stabilized from one iteration to the next. See Geman and Geman (1984) and Gelfand and Smith (1990) for a general discussion on Gibbs sampling, and Gelman and Rubin (1992) for methods of convergence.

Note that a different full conditional distribution must be sampled every time a new estimate is obtained, regardless of which parameter is being estimated. Since many iterations are usually needed to ensure that estimates for each parameter have stabilized,

efficient methods for constructing full conditional distributions and sampling from them are required. For log-concave distributions, this can be accomplished through adaptive rejection sampling (See Gilks and Wild, 1992). For applications where the full conditional distributions are not log-concave, Gilks, Best, and Tan (1995) propose appending a Hasting-Metropolis algorithm step to the adaptive rejection sampling scheme. They suggest using the resulting adaptive rejection Metropolis sampling scheme within the Gibbs sampling algorithm. We follow this approach here.

Specifically, suppose that the Gibbs sampler has been applied to the full conditional distribution of the parameter θ, $f(\theta \mid Y, \hat{\psi})$, to obtain an updated estimate, say $\hat{\theta}_{CUR}$. Here, $\hat{\psi}$ contains the most recent updated estimates for all other parameters with associated full conditional distributions. For example, one possibility is that $\theta = \beta_0$, $\hat{\psi} = \{\hat{\beta}_{1\{10\}}, \ldots, \hat{\beta}_{m\{10\}}, \hat{\delta}_{\{10\}}, \hat{\tau}^2_{\{10\}}\}$, so that $\hat{\theta}_{CUR} = \hat{\beta}_{0\{11\}}$. In what follows, the various distributions referred to are conditional upon Y and $\hat{\psi}$; however we will suppress the conditioning, writing $f(\theta \mid Y, \hat{\psi})$ as $f(\theta)$, for example. Let $S_K = \{\theta_i; i = 0, 1, \ldots, K + 1\}$ denote a set of values in ascending order for θ at which $f(\theta)$ is to be evaluated, where θ_0 and θ_{K+1} are possibly infinite lower and upper limits. Further, for $1 \le i \le j \le K$, let $L_{ij}(\theta; S_K)$ denote the straight line through the points $[\theta_i, \ln f(\theta_i)]$ and $[\theta_j, \ln f(\theta_j)]$; for other (i, j) assume that $L_{ij}(\theta; S_K)$ is undefined. Under adaptive rejection Metropolis sampling, in order to determine if $\hat{\theta}_{CUR}$ is to be kept or replaced when applying the Gibbs sampler to the full conditional of the next parameter, we proceed as follows:

(1) Sample θ from $g_K(\theta) = \frac{1}{v_K}\exp[h_K(\theta)]$ where $v_K = \int \exp[h_K(\theta)]d\theta$, and $h_K(\theta)$ is a piecewise linear function given by:

$$h_K(\theta) = \max[L_{i,\,i+1}(\theta; S_K), \min\{L_{i-1,\,i}(\theta; S_K), L_{i+1,\,i+2}(\theta; S_K)\}], \qquad \theta_i \le \theta \le \theta_{i+1}.$$

(2) Sample W_1 from a uniform (0, 1) distribution.

(3) If $W_1 > f(\theta) / \exp[h_K(\theta)]$, set $S_{K+1} = S_K \cup \{\theta\}$, ensure that all values for θ in S_{K+1} are arranged in increasing order, increment K, and go back to (1). Otherwise, set $\theta_A = \theta$, and continue.

(4) Sample W_2 from a uniform (0, 1) distribution.

(5) If $W_2 > \min\left[1, \frac{f(\theta_A)\min\{f(\hat{\theta}_{CUR}), \exp[h_K(\hat{\theta}_{CUR})]\}}{f(\hat{\theta}_{CUR})\min\{f(\theta_A), \exp[h_K(\theta_A)]\}}\right]$, then use $\hat{\theta}_{CUR}$ when applying the Gibbs sampler to the next full conditional distribution. Otherwise, use θ_A instead.

When making use of adaptive rejection Metropolis sampling within the Gibbs sampler here, for each parameter $S_K = \{\theta_i; i = 0, 1, \ldots, K + 1\}$ initially comprised six θ_i values

based on the 5^{th}, 30^{th}, 45^{th}, 55^{th}, 70^{th}, and 95^{th} percentiles of $h_K(\theta)$ from the previous Gibbs iteration. This adaptive rejection Metropolis sampling scheme is applied immediately following each time a full conditional distribution is sampled via the Gibbs sampler.

Once the parameter estimates have been obtained, it is important to determine whether or not there are any outliers in the data. Pregibon (1981) was the first to study this in the case of logistic regression with fixed effects only. Analogous to these ideas, in an empirical Bayes approach to logistic regression parameter estimation, Farrell et al (1994) estimated the sampling variance of the estimated random effects as the difference between the prior and posterior variances. This allowed the standardization of these estimated random effects and the use of the normal probability plots for the detection of outliers. We adapt this method to the hierarchical Bayes model described here in order to identify outlying hospitals in the data of Simons et al. (1997).

4. Statistical inference for the study of variations in choice between hospitals

Simons et al (1997) examined variation in the surgical treatment of rectal cancer for a large, well-defined patient population and specifically tried to determine if differences exist based on hospital type and surgical caseload. The database of the University of Southern California Cancer Surveillance Program (CSP), a population-based cancer registry in Los Angeles County was used to retrieve information collected on all patients who underwent definitive surgery for rectal cancer in Los Angeles County during the five-year period from 1988 to 1992. Surgical procedures were classified as either sphincter sparing procedures (SSP) or abdominoperineal resection (APR). Information such as patient age, gender, date of surgery, type of surgery, tumor stage, hospital type, status at last follow-up (dead or alive), and date of last follow-up was collected. From the database, hospitals could be classified into 7 categories following established guidelines of the American College of Surgeons (ACOS) Commission on Cancer. The categories were comprehensive cancer centers, teaching hospitals, health maintenance organizations, hospitals with ACOS-approved cancer programs, ACOS-approved hospitals, hospitals not approved by ACOS, and veteran's administration hospitals. Simons et al (1997) assigned each hospital to a unique subgroup based on the predominant characteristics of the hospital. In order to assess the role of caseload volume in the study, they classified hospitals according to whether the number of rectal surgery cases during the five-year period was 25 or less (that is, an average of 5 or fewer cases per year) versus greater than 25, a number initially chosen arbitrarily by Simons et al (1997) because of a break point seen in the distribution of data.

Simons et al (1997) reported results of various comparisons of the above variables using Kruskal-Wallis χ^2 tests and a multiple comparison test. They also used Kaplan-Meier plots and the log-rank test to compare survival differences between the various hospital types. They concluded that variability exists in the choice of procedure for rectal cancer even within the Los Angeles County region. In particular, hospital type and

caseload experience seem to have a significant effect on the choice of surgical procedure and on survival.

Here we model hospitals as random effects and use a hierarchical Bayes model to study this variability. The covariates selected for x_{ij} were chosen by applying a stepwise logistic regression procedure to the data set. Because of the importance attached to controlling for age and to hospital type in the original study, these variables were included in the model regardless of their level of significance. Random effects were not included in the model during this selection procedure. The resulting covariates are summarized in Table 1. Six indicator variables, labeled *HOSTYPE2* through *HOSTYPE7* are included in order to acknowledge the seven different hospital categories given above. Table 2 provides various summary statistics for the different types of hospitals included in the study, while Table 3 summarizes the association of the outcome variable with type of hospital and stage of disease (dichotomized into local versus regional).

Models (1) and (2) were fit to the University of Southern California Cancer Surveillance Program (CSP) database using the above covariates. In fitting Model (2) the procedure employed by Gilks, Best, and Tan (1995) was used. Specifically, the Gibbs sampler was run for 15000 iterations twice, each with a different set of starting values for the parameter estimates. The method of Gelman and Rubin (1992) was used to assess convergence of the Gibbs sampler. To ensure proper convergence only the last 3000 iterations of each of the two runs were used to construct posterior distributions. Specifically, the results over these two sets of 3000 iterations were combined in order to approximate these distributions.

Table 4 reports the results of fitting Model (1), the usual fixed effects logistic regression analysis ignoring hospital-to-hospital variation and Table 5 reports the results where hospitals are treated as random effects. As can be seen, the fixed effect variables that are statistically significant in the first logistic regression such as caseload, sex, stage of the cancer, and *HOSTYPE5* (ACOS-approved or not) remain significant when hospital is considered as a random effect. However, the level of significance is reduced and the confidence intervals are wider.

Figure 1 shows the results of the regression diagnostics and the normal probability plots for the standardized random effects for each hospital type. There are only 3 or 4 hospitals each of the comprehensive cancer centers, teaching hospitals, and veteran's administration hospitals which makes it difficult to detect departures from normality. Standardized random effects for the health maintenance organizations, the hospitals with ACOS-approved Cancer Programs, and the hospitals not approved by ACOS do not exhibit departures from normality, but the plot of the effects for the ACOS-approved hospitals exhibits several interesting features. First, the hypothesis of normality for the random effects can be rejected at the 10% level and one hospital (with standardized random effect = 3.2) is a significant outlier. Moreover, the standardized random effects do not cluster around zero but their average is equal to 0.45. This is exhibited in Figure 2.

After identifying the outlier, we verified that it did not account for all of the statistically significant difference between ACOS-approved hospitals and those of other types by eliminating it from the sample and redoing the analysis. Very little change in the estimators was observed when this was done. Omitting this hospital only raised the P-value slightly. This supports our conclusion that the ACOS-approved hospitals are different from the others and that within this hospital type there is an interesting outlier.

In order to further understand the nature of the intrinsic difference between the outlier and the remaining ACOS-approved hospitals, one and two sample t-tests and χ^2 tests were performed. The outlier had a significantly different caseload (lower) (P-value = 0.0000) age (lower) (P-value = 0.056), but the most significant difference was in the outcome variable where all patients had received sphincter sparing in the outlier hospital versus a proportion of 0.6269 among the other ACOS-approved hospitals. In order to rule out the possibility that the cancer was more localized (versus non-localized) in this hospital, a 2x2 Fisher exact test was performed using a collapsed version of the data in Table 6. A P-value of 0.18 was obtained; we can therefore conclude that the choice of sphincter-sparing surgery in the outlier hospital does not seem to be due to differences in rates of localized versus non-localized cancers.

5. Discussion

The hierarchical Bayes approach to logistic regression with random effects has met with success in analyzing the data from the variations in surgical procedures by hospital study originally considered by Simons et al (1997). It is an extremely flexible method that easily allows multilevel analysis. In addition, the regression diagnostics using standardized versions of the posterior estimates of the random effects can be used to identify outliers as well as different trends in the data.

Acknowledgements

The authors would like to thank Dr. A. J. Simons of the University of South California School of Medicine for making the data available to us. This research was supported in part by National Cancer Institute Grant CA17054 and the California Public Health Foundation, subcontract 050-F-8709. The ideas and opinions expressed herein are those of the authors, and no endorsement by the State of California, Department of Health Services, or the California Public Health Foundation is intended or should be inferred. The research of the first and third authors was supported by NSERC of Canada.

References

Albert, J.H., and Chib, S. (1993), "Bayesian analysis of binary and polychotomous response data", *J. Amer. Statist. Assoc*., **88**, 669-679.

Christiansen, C.L., and Morris, C.N. (1996), "Fitting and checking a two-level Poisson model: modeling patient mortality rates in heart transplant patients", in Bayesian Statistics, eds. D. Berry and D. Stangl, Marcel Decker, New York, 467-501.

Christiansen, C.L., and Morris, C.N. (1997), "Hierarchical Poisson regression modeling", *J. Amer. Statist. Assoc*., **92**, 618-632.

Cowles, M. K., and Carlin, B. P. (1996), "Markov chain Monte Carlo convergence diagnostics: a comparative review", *J. Amer Statist. Assoc*., **91**, 883-904.

Daniels, M.J, and Gatsonis, C.A. (1997), "Hierarchical polytomous regression models with applications to health services research", *Statistics in Medicine*, **16**, 2311-2325.

Daniels, M.J. and Gatsonis, C.A. (1999), "Hierarchical generalized linear models in the analysis of variations in health care utilization", *J. Amer. Statist. Assoc*., **94**, 29-42.

Dempster, A.P., Laird, N.M, and Rubin, D.B (1977). "Maximum likelihood estimation from incomplete data via the EM algorithm" *J. Royal Statist. Soc*., Series B, **39**, 1-38.

Dempster, A.P., and Tomberlin, T.J. (1980), "The analysis of census undercount from a postenumeration survey", *Proceedings of the Conference on Census Undercount*, Arlington, Va., 88-94.

Farrell, P. J., MacGibbon, B., and Tomberlin, T. J. (1994), "Protection against outliers in empirical Bayes estimation", *Canad. J. Statist*., **22**, 365-376.

Farrell, P. J., MacGibbon, B., and Tomberlin, T. J. (1997), "Empirical Bayes estimators of small area proportions in multistage designs", *Statist. Sinica*, **7**, 1065-1083.

Gelfand, A.E, and Smith, A.F.M. (1990), "Sampling based approches to calculating marginal densities", *J. Amer. Statist. Assoc*., **85**, 398-409.

Gelman, A., and Rubin, D. B. (1992), "Inference from iterative simulations using multiple sequences", *Statist. Sci*., **7**, 457-511.

Geman, S., and Geman, D. (1984), "Stochastic Relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721-741.

Gilks, W.R., Best, N.G. and Tan, K.K. (1995), "Adaptive rejection metropolis sampling within Gibbs sampling", *Appl. Statist.*, **44**, 455-472.

Gilks, W.R, and Wild, P. (1992), "Adaptive rejection sampling for Gibbs sampling", *J. Royal Statist. Soc*., Series C, **41**, 337-348.

Laird, N.M. (1978), "Empirical Bayes methods for two-way contingency tables" *Biometrika*, **65**, 581-590.

MacGibbon, B., and Tomberlin T.J. (1989), "Small area estimates of proportions via empirical Bayes techniques", *Survey Methodology*, **15**, 232-252.

Malec, D., Sedransk, J., Moriarity, C. L., and LeClere, F. B. (1997), "Small area inference for binary variables in the National Health Interview Survey", *J. Amer. Statist. Assoc*., **92**, 815-826.

Normand, S.-L.T., and Glickman, M.E, and Gatsonis, C.A. (1997), " Statistical methods for profiling provides of medical care: issues and applications", *J. Amer. Statist. Assoc*., **92**, 803-814.

Pregibon, D. (1981) "Logistic regression diagnostics", *Ann. Statist*., **9**, 705-724.

Simons, A. J. , Ker, R. , Groshen, S. , Gee, C. , Anthone, G.J. , Ortega, A. E, Vuk asin, P. , Ross R. K. , Beast R. W. (1997), "Variations in treatment of rectal cancer", *Diseases of the Colon and Rectum*, **40**, 641-646.

Stiratelli, R., Laird, N.M., and Ware, J.H. (1984), "Random-effects model for several observations with binary response", *Biometrika*, **40**, 961-971.

Wong, G.Y., and Mason, W.M. (1985), "The hierarchical logistic regression model for multilevel analysis", ", *J. Amer. Statist. Assoc*., **80**, 513-524.

Wong, G.Y., and Mason, W.M. (1991), "Contextually specific effects and other generalizations of the hierarchical linear model for comparative analysis", *J. Amer. Statist. Assoc*., **86**, 487-503.

Table 1.
List of variables included in the analysis.

HOSPITAL RELATED COVARIATES (based on hospital of j^{th} patient)		
Average Annual Number of Surgical Cases for Cancer of the Rectum	*AVGCSLD* =	Quantitative variable
Indicator Variables for Type of Hospital: (Baseline: Comprehensive Cancer Center)	*HOSTYPE2* =	1 if Teaching Hospital
	HOSTYPE3 =	1 if Health Maintenance Organization
	HOSTYPE4 =	1 if hospital with ACOS* approved cancer program
	HOSTYPE5 =	1 if hospital approved by ACOS*
	HOSTYPE6 =	1 if hospital not approved by ACOS*
	HOSTYPE7 =	1 if Veteran's Administration Hospital 0 otherwise
PATIENT RELATED COVARIATES (for j^{th} patient at i^{th} hospital)		
Gender:	*SEX* =	1 if Male, 0 if Female
Age:	*AGE* =	Quantitative variable
Indicator Variables for Stage of Rectal Cancer: (Baseline: Localized Disease)	*STAGE2* =	1 if regional disease – direct extension only
	STAGE3 =	1 if regional disease – involved lymph nodes only
	STAGE4 =	1 if regional disease with both direct extension and involved lymph nodes 0 otherwise
DEPENDENT/OUTCOME VARIABLE (for j^{th} patient at i^{th} hospital)		
Type of Surgery	Y_{ij} =	1 if SSP 0 if APR (requiring colostomy)

* American College of Surgeons (ACOS)

Table 2.
Summary statistics for the different types of hospitals.

Type of Hospital	Number of Hospitals	Total Number of Patients	Annual Number of Surgeries for Rectal Cancer (Median & Range)	Percent of Patients that are Male	Median Age (Range)	Percent of Patients with Localized Disease
Comp CC	3	71	6.2 (1.2-6.8)	41%	62 (27-93)	51%
Teaching	3	65	4.2 (1.4-7.4)	49%	57 (20-80)	35%
ACOS-CP	35	1029	6.2 (0.6-24.0)	53%	69 (21-102)	54%
HMO	9	301	6.8 (0.4-11.0)	63%	66 (26-93)	52%
ACOS-App	15	139	1.8 (0.2-4.0)	50%	72 (19-93)	55%
NonACOS-App	56	341	0.8 (0.2-5.0)	53%	70 (29-95)	51%
Veteran's	4	60	3.4 (0.4-4.8)	98%	66 (37-88)	40%
Total	125	2006	2.0 (0.2-24.0)	55%	68 (19-102)	52%

Table 3.
Patients undergoing sphincter-sparing surgical procedure (SSP) by stage of disease and hospital type, number of cases, and patient demographic characteristics.(*)

	Localized Disease		Regional Disease	
	No. of Patients	% SSP	No. of Patients	% SSP
TYPE OF HOSPITAL				
Comp CC	36	61%	35	34%
Teaching	23	52%	42	43%
ACOS-CP	560	67%	469	43%
HMO	156	72%	145	41%
ACOS-approved	76	78%	63	44%
Non ACOS-approved	173	61%	168	42%
Veteran's	24	58%	36	39%
CASELOAD				
≤5 Cases per Year	419	63%	404	42%
> 5 Cases per Year	629	69%	554	43%
GENDER				
Male	555	64%	555	40%
Female	493	70%	403	45%
AGE				
67 Years orLess	518	67%	578	42%
More Than 67 Years	530	67%	380	42%
Total	1048	67%	958	42%

* unadjusted percents (calculated as number of patients undergoing SSP divided by the total number of patients)

Table 4.
Estimates of fixed effects based on Model (1).

Fixed Effect	Estimate	Std Err	Est / Std Err	95% Conf Int
CONSTANT	-0.1505	0.3544	-0.42	(-0.8451, 0.5441)
AVGCSLD	0.0499	0.0113	4.44	(0.0279, 0.0720)
SEX	-0.2381	0.0960	-2.48	(-0.4264, -0.0499)
AGE	0.0057	0.0038	1.49	(-0.0018, 0.0132)
STAGE2	-0.9647	0.1247	-7.74	(-1.2091, -0.7203)
STAGE3	-0.8021	0.1715	-4.68	(-1.1382, -0.4660)
STAGE4	-1.1091	0.1220	-9.09	(-1.3482, -0.8700)
HOSTYPE2	0.1491	0.3567	0.42	(-0.5500, 0.8482)
HOSTYPE3	0.3096	0.2758	1.12	(-0.2310, 0.8502)
HOSTYPE4	0.1984	0.2576	0.77	(-0.3065, 0.7033)
HOSTYPE5	0.7505	0.3102	2.42	(0.1425, 1.3585)
HOSTYPE6	0.3338	0.2756	1.21	(-0.2064, 0.8740)
HOSTYPE7	0.2764	0.3688	0.75	(-0.4465, 0.9993)

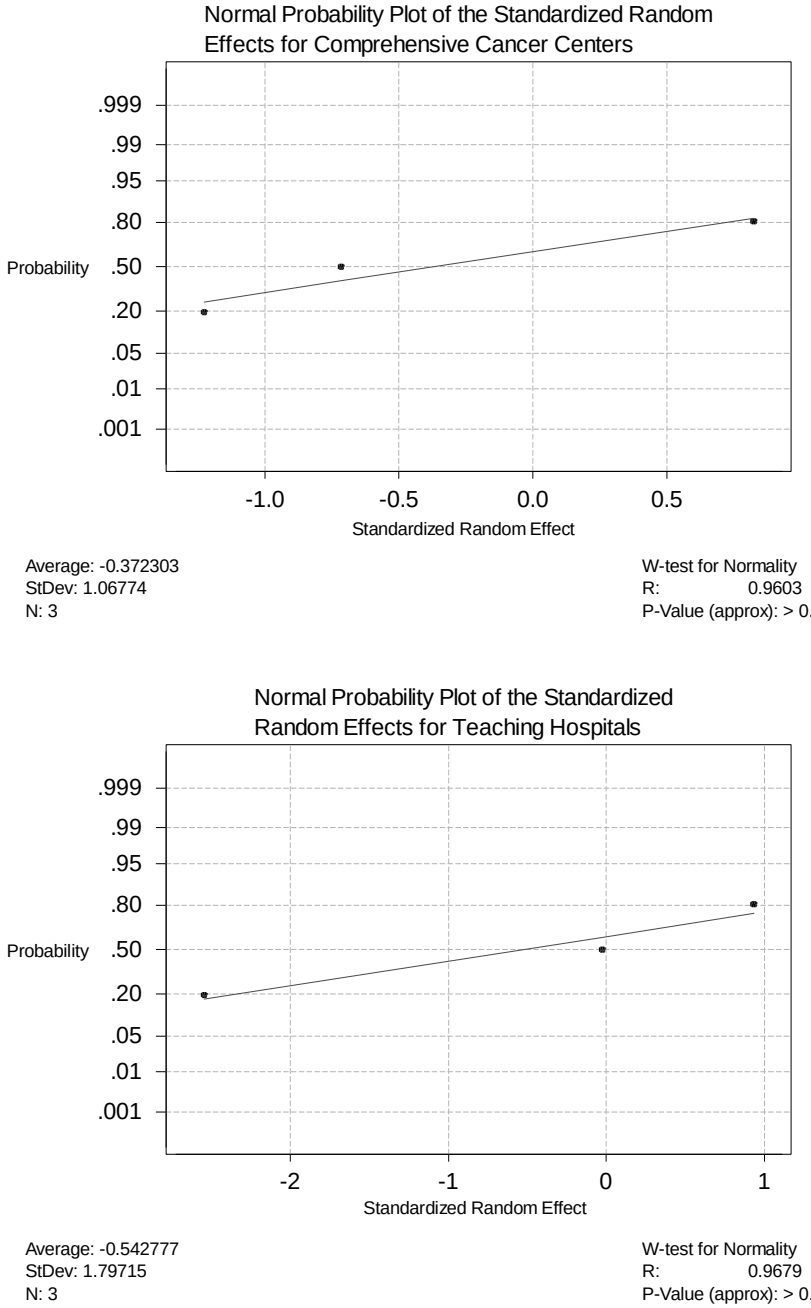
Table 5.
Estimates of fixed effects based on Model (2).

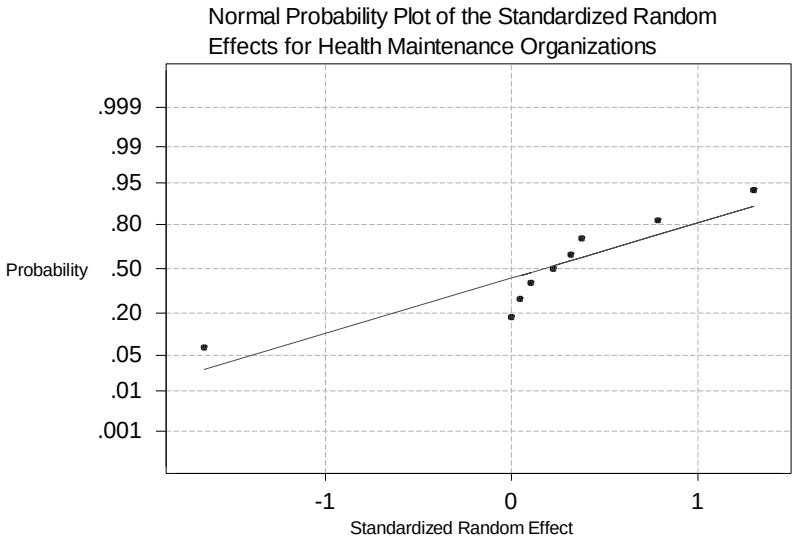
Fixed Effect	Estimate	Std Err	Est / Std Err	95% Conf Int	WidM1 / WidM2
CONSTANT	-0.1509	0.3913	-0.39	(-0.9821, 0.6408)	0.8560
AVGCSLD	0.0544	0.0135	4.03	(0.0252, 0.0769)	0.8529
SEX	-0.2414	0.1053	-2.29	(-0.4165, -0.0354)	0.9878
AGE	0.0057	0.0042	1.37	(-0.0027, 0.0131)	0.9507
STAGE2	-0.9660	0.1367	-7.07	(-1.2586, -0.7051)	0.8832
STAGE3	-0.8033	0.1857	-4.32	(-1.2257, -0.4173)	0.8316
STAGE4	-1.1106	0.1316	-8.44	(-1.3274, -0.7931)	0.8950
HOSTYPE2	0.1508	0.3842	0.39	(-0.6143, 0.8678)	0.9434
HOSTYPE3	0.3088	0.3007	1.03	(-0.3003, 0.9612)	0.8571
HOSTYPE4	0.1999	0.2787	0.72	(-0.3776, 0.7835)	0.8697
HOSTYPE5	0.7508	0.3368	2.23	(0.0834, 1.5379)	0.8360
HOSTYPE6	0.3321	0.2992	1.11	(-0.2798, 0.9737)	0.8618
HOSTYPE7	0.2775	0.4009	0.69	(-0.4930, 1.0208)	0.9551

Table 6.
Number of cases at each stage in all ACOS-approved hospitals and the outlying ACOS-approved hospital.

Stage of Disease	Cases in ACOS-Approved Hospitals	Cases in Outlier of ACOS-Approved Hospitals
Localized	76	7
Regional (Directed Extension Only)	28	1
Regional (Lymph Nodes Only)	12	1
Regional (Directed Extension/Lymph Nodes)	23	0

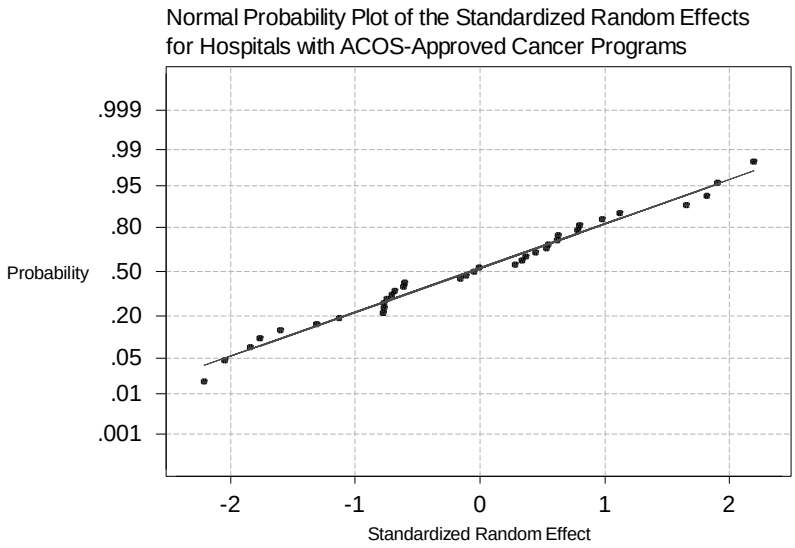
Figure 1.
Normal probability plots of standardized random effects by hospital type.





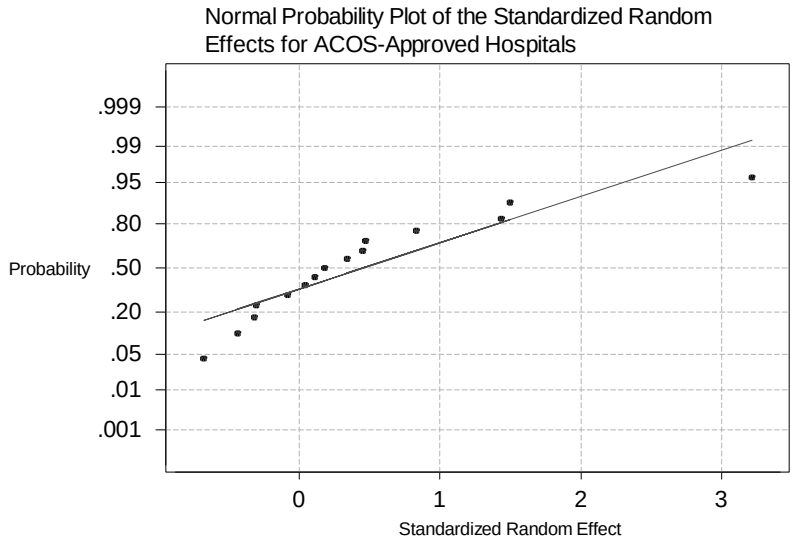
Average: 0.167434
StDev: 0.796833
N: 9

W-test for Normality
R: 0.9028
P-Value (approx): 0.0408



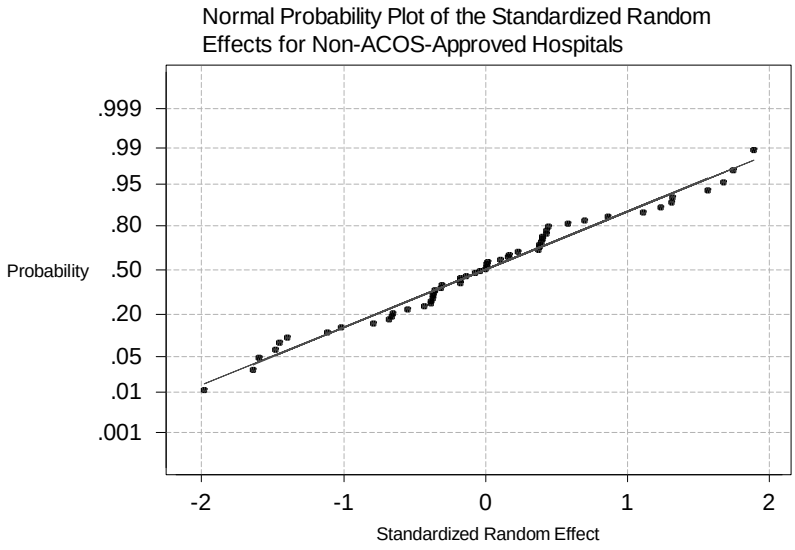
Average: -0.0804439
StDev: 1.14783
N: 35

W-test for Normality
R: 0.9917
P-Value (approx): > 0.1000



Average: 0.451329
StDev: 0.989835
N: 15

W-test for Normality
R: 0.9146
P-Value (approx): 0.0153



Average: -0.0141290
StDev: 0.879800
N: 56

W-test for Normality
R: 0.9897
P-Value (approx): > 0.1000

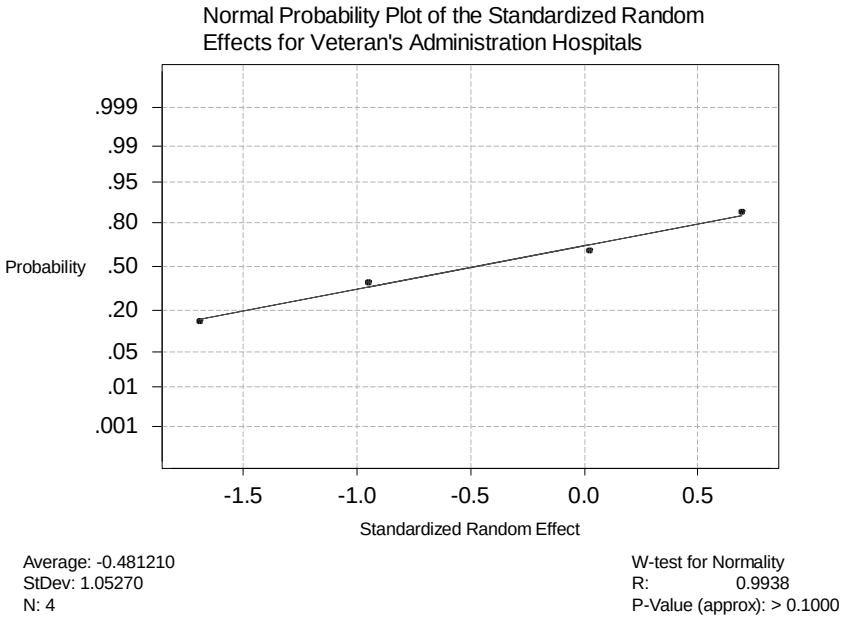


Figure 2.
Simultaneous dotplots of standardized random effects by hospital type.