Discharge Rates of Medicare Stroke Patients to Skilled Nursing Facilities: Bayesian Logistic Regression With Unobserved Heterogeneity

Michael J. Kahn; Adrian E. Raftery

Journal of the American Statistical Association, Vol. 91, No. 433 (Mar., 1996), 29-41.

Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28199603%2991%3A433%3C29%3ADROMSP%3E2.0.CO%3B2-J

Discharge Rates of Medicare Stroke Patients to Skilled Nursing Facilities: Bayesian Logistic Regression With Unobserved Heterogeneity

Michael J. KAHN and Adrian E. RAFTERY

We determine factors, both hospital-specific and market area–specific, associated with hospitals' propensities for discharging Medicare stroke patients to skilled nursing facilities (SNF's) in California and Florida. Logistic regression is generalized to the case of a betabinomial, hierarchical model, in which covariate information is included in the hyperparameters of the second-stage beta distribution. It is found that the posterior mean of the proportion discharged to SNF is approximately a weighted average (i.e., shrinkage estimator) of the logistic regression estimator and the observed rate. We develop fully Bayesian inference that takes into account uncertainty about the hyperparameters, and we find that this also allows us to test for overdispersion in a natural way. The number of observed zeros (i.e., hospitals that sent no stroke patients to a SNF) is excessive compared to the number expected from a standard logistic regression model and is fit better by the hierarchical betabinomial model. The factors associated with discharge to SNF differ between California and Florida. In California the case-mix index and percent Medicaid admissions of the hospital, as well as the per capita income for the area and whether there is a rehabilitation facility in the area, are associated with discharge rates to SNF's. In Florida, whether there is a rehabilitation facility in the area is the only factor that exhibits association with discharge rates to SNF's.

KEY WORDS: Betabinomial model; Empirical Bayes procedure; Laplace approximation; Locally uniform prior distribution; Overdispersion.

1. INTRODUCTION

1.1 The Problem

In 1983 the Health Care Financing Administration (HCFA) introduced a prospective payment system (see, for example, Neu and Harrison 1988) for Medicare hospital reimbursement. Before advent of the prospective payment system, a retrospective reimbursement system was in place. In the retrospective system, each Medicare patient's medical costs were tracked by the hospital, and the bill was sent to HCFA. The prospective payment system is an attempt to curb costs without sacrificing quality of care. After introducing the prospective payment system, HCFA found that use of posthospital facilities increased dramatically, and so did HCFA's outlays to such facilities. HCFA is interested in better understanding the relationship between hospital characteristics, including characteristics of the market in which the hospital operates, and the hospital's rate of discharge of Medicare patients to one type of posthospital care: skilled nursing facilities (SNF's). Classically, such investigations begin with logistic regression methods. This allows an estimate for a particular hospital's "true" rate to be determined from information concerning all of the hospitals. (We consider a hospital's "true" rate to be its rate in the near future; in the following year, say.)

Michael J. Kahn is Research Associate, Cancer Center Statistics, Mayo Clinic, Rochester, MN 55905. Adrian E. Raftery is Professor of Statistics and Sociology, Department of Statistics, University of Washington, Seattle, WA 98195-4322. This work was done while Kahn was Assistant Professor of Mathematics, St. Olaf College. This research was supported in part by National Science Foundation Grant DMS-89-0211, Office of Naval Research Contract N00014-J-91-1074, National Institute of Child Health and Human Development Grant 5R01HD26330-02, the RAND Corporation, and St. Olaf College. The authors thank Dr. C. R. Neu for introducing them to the problem and Neal Thomas and James Hodges for much insight into the problem. They are very grateful to Charles Geyer, Michael Newton, Michael Perlman, Trevellore Raghunathan, the editor, and associate editor, and two anonymous referees for extremely helpful conversations and comments that greatly improved this article.

Figure 1 shows the total number of discharges (n_j) of stroke patients for urban hospitals in California and Florida in a 12-month period in 1984–1985. Although there are a few hospitals with more than 50 stroke discharges, nearly one-half of the California hospitals and one-fourth of the Florida hospitals had fewer than 10 discharges. Clearly, the greater the number of discharges, the better the information about that hospital's "true" rate.

This leads to the consideration of a shrinkage estimator, in which the estimate for a particular hospital is a weighted average of an ensemble estimator that draws from all the data (e.g., logistic regression) and an estimator that uses only the information from that particular hospital. If a particular hospital's observed rate contains little information about its "true" rate (e.g., if n_j is small), then more weight should be given to the ensemble estimator. Conversely, when the information in a particular hospital's observed rate is high, less weight should be given to the ensemble estimator. This is precisely how empirical Bayes methods work, leading us to consider empirical Bayes models for these data.

Exploration of these data reveals extrabinomial variation in the form of an excessive number of observed discharge rates of zero (i.e., hospitals that discharged no stroke patients to SNF's), even when we look only at hospitals with at least 20 stroke discharges (see Kahn 1990 and Figures 5 and 6 in Section 4.1). This raises the question of whether some hospitals have policies not to send patients to SNF's or whether this large number of zeros is a result of unobserved heterogeneity, with many hospitals simply having low "true" rates of discharge.

Journal of the American Statistical Association
March 1996, Vol. 91, No. 433, Applications and Case Studies

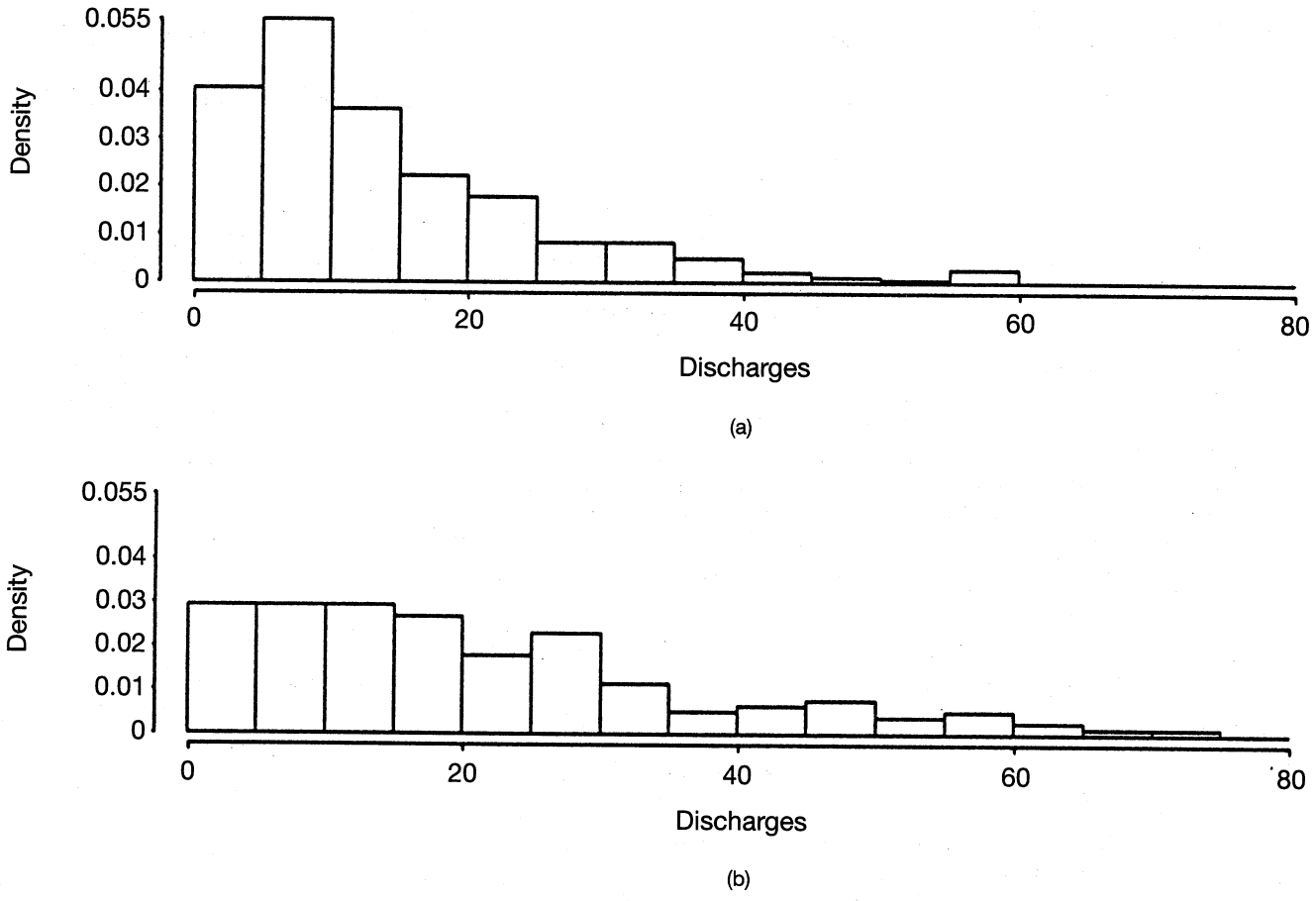

Figure 1. Histograms of the Number of Discharges of Medicare Stroke Patients, July 1984–June 1985, From Urban Hospitals in California (a) and Florida (b). About one-half of the 360 California hospitals had fewer than 10 discharges, compared to one-fourth of the 157 Florida hospitals.

For comparative purposes, we begin by discussing logistic regression and quasi-likelihood methods. Next, we describe an empirical Bayes extension of these models. We specify the betabinomial-logit model, find moments, and show how to obtain maximum likelihood estimates (MLE's) for the regression parameters and their corresponding standard errors. We then show that the parameter estimates from the betabinomial-logit scheme will be nearly the same as the logistic regression estimates and that their standard errors are close to those obtained through the quasi-likelihood method. Further, we show that the empirical Bayes estimate of the "true" rate will be approximately a convex combination (weighted average) of the logistic regression estimate and the observed rate. In fact, the weights are determined by the correlation between the hospital's observed rate and the hospital's "true" rate; see Section 2.3.3. Finally, we develop a fully Bayesian way of incorporating uncertainty about the hyperparameters and apply it to the hospital data.

1.2 Background

There is an enormous literature concerning regression-like methods in which the response variable is dichotomous or, if grouped, binomial. The best-known and most widely used of these methods is logistic regression (see, for example, Cox 1970), which assumes that there is no unobserved heterogeneity. Here we develop regression methods for binomial data with unobserved heterogeneity. We first describe a model that synthesizes logistic regression and the betabinomial distribution to represent overdispersion. We then show how standard empirical Bayes methods can be used for estimation. Such standard methods underestimate variability, however, because they do not take uncertainty about the hyperparameters into account. We develop fully Bayesian inferential methods that remedy this.

The betabinomial distribution has found numerous applications in toxicology, particularly in studies of teratogenesis or carcinogenesis (see, e.g., Haseman and Kupper 1979). The betabinomial distribution has also been useful in studies of consumer purchasing (Chatfield and Goodhardt 1970), advertising campaigns (Danaher 1988), animal litters (Healy 1972), the incidence of disease (Griffiths 1973), radiation (Prentice 1986), and prediction of fires in New York City (Carter and Rolph 1973).

Kleinman (1973) and Tamura and Young (1986, 1987) investigated the properties of maximum likelihood estimation in the betabinomial model and derived alternatives to moment estimators. The reason that moment estimators were ever considered was the difficulty of finding the MLE without sufficient computing power. Today this is of less concern. Maximum likelihood estimation for the mean parameter, α in the parameterization (2) in Section 2.3.1, is stable. The MLE of γ tends to be unstable for large values of γ, but Albert (1988) showed that for moderate values of n, the MLE is stable for γ less than about 3,000. In the Medi-

care data analyzed in this article, there were no problems of instability; the MLE's of γ are all less than 50.

Williams (1975) was the first to consider the problem of unequal n_j for the binomial counts and to test the equivalent of a betabinomial-analysis of variance (ANOVA) model. That is, Williams allowed γ and α to depend on which treatment the subject received, say (γ_i, α_i) for $i = 1, \ldots, K$ treatments (see also Crowder 1978). Williams (1982) generalized the betabinomial-ANOVA model to the case where one has some other, possibly continuous, concomitant information. He developed a quasi-likelihood method and did not discuss maximum likelihood estimation of the betabinomial-logit model (Sec. 2.3). Instead he discussed an iteratively reweighted least squares method for finding quasi-likelihood estimates. Other work in this vein has been done by Stiratelli, Laird, and Ware (1984), Anderson and Aitkin (1985), and Im and Gianola (1988). These authors used different sorts of logit-normal ANOVA models that might be of interest in our problem if we had patient-level data. (Though patient-level data are now available, we did not have access to them at the time of this study.)

Williams (1982) also discussed maximum likelihood estimation when the distribution of p_j is logistic-normal (see Aitchison and Shen 1980). The most striking difference between the logistic-normal-binomial model and the betabinomial-logit model is that in the betabinomial-logit model we have $\text{logit}(\text{E}(p_j)) = \mathbf{x}_j^T\boldsymbol{\beta}$, whereas in the logistic-normal-binomial model $\text{E}(\text{logit}(p_j)) = \mathbf{x}_j^T\boldsymbol{\beta}$. Thus all concomitant information is added together on the logistic scale in the logistic-normal-binomial model. One problem with the logistic-normal-binomial model is that the intractability of the marginal distribution of Y_j makes it difficult to check the fit of the model, forcing one to use ad hoc methods for finding the MLE of the hyperparameters. Williams (1982) presented a reasonable method for estimating the logistic-normal-binomial model, but we prefer having the marginal distribution of Y_j with which to work, and we prefer the conjugacy properties of the beta distribution. Albert (1988) discussed another way of fitting hierarchical generalized linear models, including the betabinomial-logit model. An even more sophisticated model might allow β to vary between metropolitan statistical areas (MSA's). An example of this in the logistic regression case was given by Wong and Mason (1985). Follman and Lambert (1989) also provided a hierarchical logistic model that is especially useful for errors-in-variables logistic regression and for generalizing the linear-on-the-logit-scale specification of logistic regression. But their model does not facilitate predictive inference or evaluation of hospital performance in a natural way. The betabinomial-logit model does allow for predictive inference and hospital evaluation.

The model described in this article is motivated by, and applied to, the analysis of rates at which urban hospitals discharge Medicare stroke patients to nursing homes. We wish to understand the relationship between various hospital characteristics and the discharge rates to nursing homes, as well as how the rates vary between hospitals. Yet, due to extrabinomial variation, primarily manifesting itself in an excess of hospitals that sent no stroke patients to a SNF, standard logistic regression models do not fit well. The betabinomial scheme that we present accounts for overdispersion in the data. Quasi-likelihood (QL) methods are an alternative attempt to account for extrabinomial variation, but, whereas QL simply assumes parametric forms for the first two moments, the methods discussed here are based on a fully specified stochastic model. Using a fully specified model allows us to test for overdispersion and also to assess the fit of the model, particularly the extent to which it accounts for the excess zeros. It also allows us to determine whether the excess zeros are better explained as "true zeros" or as a by-product of routine overdispersion. These additional goals would be difficult to achieve using a quasi-likelihood approach.

Our method is easily implemented and gives point estimates that are often nearly identical to those given by logistic regression (Sec. 2.3), although the standard errors of the point estimates are typically larger than those from logistic regression. Logistic regression and the proposed betabinomial-logit models are the same when each observation is Bernoulli; that is, when each observation is either a zero or a 1.

2. METHODOLOGIES

2.1 Logistic Regression

The logistic regression model is

$$Y_j|\boldsymbol{\beta} \overset{\text{indep}}{\sim} \text{Bin}(n_j, p_j(\mathbf{x}_j;\boldsymbol{\beta})), \qquad j = 1, \ldots, N, \quad (1)$$

where $\mathbf{x}_j \in \mathbb{R}^m$ is a known vector of covariates with the first component equal to 1, $\boldsymbol{\beta} \in \mathbb{R}^m$ is a vector of regression parameters where the first component corresponds to an intercept, and $\text{E}(Y_j|\mathbf{x}_j) = p_j(\mathbf{x}_j;\boldsymbol{\beta}) = (1 + \exp(-\mathbf{x}_j^T\boldsymbol{\beta}))^{-1} \in (0,1)$ or, equivalently, $\text{logit}(p_j) = \log[p_j/(1-p_j)] = \mathbf{x}_j^T\boldsymbol{\beta}$. The estimating equations for the MLE of $\boldsymbol{\beta}$ are given by the usual linear, exponential family results. Although these equations look much like the standard, linear-model equations, in which case $\text{E}(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$, in the logistic regression case we have a nonlinear system of equations to solve for $\boldsymbol{\beta}$. This is easily done using the Newton–Raphson method (see, e.g., Cox 1970).

2.2 Quasi-Likelihood

For a detailed discussion of quasi-likelihood methods and accompanying references, see the text of McCullagh and Nelder (1983). The quasi-likelihood approach makes parametric assumptions about the first and second moments but not about the distributional form of the Y_j. In our case the assumptions are that $\text{E}(Y_j) = n_j p_j(\mathbf{x}_j;\boldsymbol{\beta})$, $\mathbf{V}(Y_j) = \sigma^2 n_j p_j(\mathbf{x}_j;\boldsymbol{\beta})(1-p_j(\mathbf{x}_j;\boldsymbol{\beta}))$, and that the Y_j are independent, where σ^2 is an over(under)dispersion parameter when $\sigma^2 > 1 (< 1)$. Logistic regression and quasi-likelihood have identical first and second moments when $\sigma^2 = 1$. With these assumptions, the quasi-likelihood estimating equations for $\boldsymbol{\beta}$ are the same as in the standard logistic regression case, but the standard errors of the $\hat{\boldsymbol{\beta}}$ are inflated or deflated by a factor of σ. Following McCullagh and Nelder, we use the consistent (as $N \to \infty$; i.e., as the number of hospitals gets large) estimator of σ^2,

$$\tilde{\sigma}^2 = \frac{1}{N-m} \sum_{j=1}^{N} \frac{(y_j - n_j\hat{p}_j)^2}{n_j\hat{p}_j(1-\hat{p}_j)}.$$

2.3 The Betabinomial-Logit Model

2.3.1 The Betabinomial Distribution. In this section we discuss the basic properties of the betabinomial distribution. This distribution arises naturally in a hierarchical or Bayesian context in the following manner. Let

$$Y|p \sim \text{Bin}(n,p)$$

and

$$p \sim \text{Beta}(\gamma\alpha, \gamma(1-\alpha)), \tag{2}$$

with $\gamma \in (0,\infty)$ and $\alpha \in (0,1)$. The marginal distribution of Y is then said to have a betabinomial distribution with support $\{0,\ldots,n\}$,

$$\text{P}(Y=y) = \binom{n}{y} \frac{\Gamma(\gamma\alpha+y)\Gamma(\gamma(1-\alpha)+n-y)\Gamma(\gamma)}{\Gamma(\gamma\alpha)\Gamma(\gamma(1-\alpha))\Gamma(\gamma+n)}. \tag{3}$$

This distribution was first discussed by Skellam (1948) (also see Johnson and Kotz 1970). Here Bayes's theorem gives

$$p|Y \sim \text{Beta}(\gamma\alpha+Y, \gamma(1-\alpha)+n-Y). \tag{4}$$

If we have $n=1$ (i.e., Y is Bernoulli(p)), then (3) becomes

$$\text{P}(Y=1) = \alpha = 1 - \text{P}(Y=0).$$

That is, Y remains Bernoulli with success probability equal to the mean of the distribution for p.

2.3.2 Approximations. For the case where n is any positive integer, consider Stirling's approximation for $\Gamma(x)$:

$$\Gamma(x) \approx \sqrt{2\pi}x^{x-1/2}e^{-x}.$$

This yields the following approximation for the mass function, (3):

$$\text{P}(Y=y) \approx \binom{n}{y} \alpha^y(1-\alpha)^{n-y}c(y,\gamma,\alpha), \tag{5}$$

where

$$c(y,\gamma,\alpha) = \left(\frac{\gamma}{\gamma+n}\right)^{1/2} \left(\frac{\gamma\alpha+y}{\gamma\alpha+n\alpha}\right)^{\gamma\alpha+y-1/2} \times \left(\frac{\gamma(1-\alpha)+n-y}{\gamma(1-\alpha)+n(1-\alpha)}\right)^{\gamma(1-\alpha)+n-y-1/2}.$$

As γ gets large, $c(y,\gamma,\alpha)$ approaches 1, so that the betabinomial mass function approaches that of a Bin(n,α). Also, because $\text{E}(Y) = n\alpha$, the second and third factors in $c(y,\gamma,\alpha)$ will, on average, be near 1, giving

$$c(y,\gamma,\alpha) \approx \left(\frac{\gamma}{\gamma+n}\right)^{1/2}.$$

2.3.3 Moments. The mean and variance of a betabinomially distributed random variable are

$$\begin{aligned}\text{E}(Y) &= \text{E}(\text{E}(Y|p)) = \text{E}(np) = n\alpha \\ \text{var}(Y) &= \text{var}(\text{E}(Y|p)) + \text{E}(\text{var}(Y|p)) \\ &= n\alpha(1-\alpha)\left(\frac{\gamma+n}{\gamma+1}\right). \end{aligned} \tag{6}$$

Model (2) specifies a joint distribution for Y and p, with

$$\text{E}(Yp) = \text{E}(\text{E}(Yp|p)) = \text{E}(np^2) = \frac{n\alpha(1+\gamma\alpha)}{1+\gamma},$$

so that

$$r \equiv \text{corr}(Y,p) = \sqrt{\frac{n}{n+\gamma}}.$$

From (4), we see that

$$\begin{aligned}\text{E}(p|Y,\gamma,\alpha) &= \frac{\gamma}{\gamma+n}\alpha + \frac{n}{\gamma+n}\left(\frac{Y}{n}\right) \\ &= (1-r^2)\alpha + r^2\left(\frac{Y}{n}\right). \end{aligned} \tag{7}$$

Thus the posterior mean is a weighted average of α, the mean of p, and Y/n, the mean of the data. The weights are determined completely by r, in such a way that the greater the correlation between the observable, Y, and the unobservable "true" rate, p, the more influence Y/n has in estimating p. When Y and p are uncorrelated, the data do not influence the estimate of p at all. (In our case this would only happen approximately if γ were much larger then n.) This says that the "regression line" of p on Y/n has intercept α, slope r^2, and passes through the point (α,α). This emphasizes the empirical Bayes focus of implicitly trying to determine the unknown correlation between Y_j and p_j and drawing regression-like inference for p_j from Y (in the normal theory context, see Stigler 1990).

2.3.4 The Likelihood Function. Suppose that we have a set of data to be modeled by the hierarchical scheme described earlier. Let

$$Y_j|p_j \overset{\text{indep}}{\sim} \text{Bin}(n_j,p_j)$$

and

$$p_j \overset{\text{indep}}{\sim} \text{Beta}(\gamma\alpha_j, \gamma(1-\alpha_j)). \tag{8}$$

We allow for covariates as did Albert (1988). Let

$$\alpha_j(\mathbf{x}_j;\boldsymbol{\beta}) \equiv (1+\exp(-\mathbf{x}_j^T\boldsymbol{\beta}))^{-1} = \text{logit}^{-1}(\mathbf{x}_j^T\boldsymbol{\beta}) \in (0,1)$$

and $\gamma \in (0,\infty)$. Then $\text{E}(Y_j|\mathbf{x}_j,\boldsymbol{\beta}) = n_j\alpha_j = n_j(1+\exp(-\mathbf{x}_j^T\boldsymbol{\beta}))^{-1}$ as in the logistic regression model (1). But from (3),

$$\text{var}(Y_j|\mathbf{x}_j,\gamma,\boldsymbol{\beta}) = n_j\alpha_j(\mathbf{x}_j;\boldsymbol{\beta})(1-\alpha_j(\mathbf{x}_j;\boldsymbol{\beta}))\left(\frac{\gamma+n_j}{\gamma+1}\right),$$

whereas in the logistic regression case (1), we have

$$\text{var}(Y_j|\mathbf{x}_j,\boldsymbol{\beta}) = n_jp_j(\mathbf{x}_j;\boldsymbol{\beta})(1-p_j(\mathbf{x}_j;\boldsymbol{\beta})).$$

Table 1. Logistic Regression, Quasi-Likelihood, and Betabinomial-Logit Results for the California Hospital Data

Variable	Logistic	S.E.	t	QL S.E.	t	BB-Logit	S.E.	t
γ	*	*	*	*	*	41.206	12.52	3.29
intercept	−.735	.112	−6.54	.129	−5.71	−.737	.131	−5.62
bedsupply	.068	.038	1.80	.043	1.58	.078	.045	1.72
cmi	−.134	.039	−3.42	.045	−2.99	−.129	.046	−2.81
%MDadm	−.137	.046	−3.00	.052	−2.62	−.116	.052	−2.22
income	.236	.041	5.81	.047	5.08	.235	.049	4.84
rehabs	−.667	.123	−5.44	.141	−4.75	−.674	.144	−4.67
Log-likelihood (model)		−616.373					−606.015	
Log-likelihood (intercept)		−651.140					−627.427	

NOTE: * = Not Applicable; t = t-Ratio.

As noted earlier, these are equivalent when $n_j = 1$ or $\gamma = \infty$.

The variance in the QL model (see Sec. 2.2) is given by

$$\text{var}(Y_j|\mathbf{x}_j, \boldsymbol{\beta}) = n_j p_j(\mathbf{x}_j;\boldsymbol{\beta})(1 - p_j(\mathbf{x}_j;\boldsymbol{\beta}))\sigma^2.$$

We see from this that the betabinomial-logit model generalizes the QL variance specification in the sense that the overdispersion parameter, $(\gamma + n_j)/(\gamma + 1)$, depends on j through n_j. At the same time, the betabinomial-logit model is more restrictive in the sense that in the QL model $\sigma^2 > 0$, but in the betabinomial-logit model $(\gamma+n_j)/(\gamma+1) \in (1, n_j)$. Hence we cannot model underdispersion with the betabinomial-logit model. Although QL may be a sensible method for fitting overdispersion in some situations, in the logistic regression case it does not yield an actual stochastic model. It is not clear how one might generate random variates from a QL model, but presumably one can use any distribution with the right mean, variance, and support for the Y_j. There are a number of such distributions, each having different characteristics beyond the first two moments. QL may be fine if one is interested solely in how the $\{\mathbf{x}_j\}$ relate to the Y_j; that is, if one's sole interest is $\boldsymbol{\beta}$. But if interest is in predictive inference, then one must have a stochastic model capable of generating the data. The betabinomial-logit model is a proper, stochastic model allowing for extrabinomial variation from which data can be generated.

Further, the betabinomial-logit and logistic regression models can be checked and compared by way of goodness-of-fit tests, by exact or approximate Bayes factors or by likelihood ratio tests. Also, these likelihood-based statistics give us a test of overdispersion in a natural way. The goodness-of-fit test described by Kahn (1990) calculates, based on the fitted models, the expected number of hospitals with no discharges to SNF, 1 discharge to SNF, and so on, and is useful for diagnostic purposes in both the betabinomial-logit and the logistic regression models.

The most important difference between the binomial and betabinomial models is that in the latter we are modeling the $\{p_j\}$ as stochastic variables. This says that the "true" rates of all the hospitals that have covariate information $\mathbf{x}_j$ can be modeled as coming from a beta$(\gamma\alpha_j, \gamma(1-\alpha_j))$ distribution. Thus the data are more variable than in the logistic regression model.

Using (3), we obtain the likelihood for γ and $\boldsymbol{\beta}$ as

$$L(\gamma, \boldsymbol{\beta}) = \prod_{j=1}^{N} \left[\binom{n_j}{y_j} \frac{\left[\prod_{k=0}^{y_j-1}(\gamma\alpha_j + k)\right]\left[\prod_{k=0}^{n_j-y_j-1}(\gamma(1-\alpha_j)+k)\right]}{\left[\prod_{k=0}^{n_j-1}(\gamma+k)\right]} \right] \quad (9)$$

with $\prod_{k=0}^{-1}(a+k) \equiv 1$. Hence the log-likelihood of the jth observation is

$$l_j(\gamma, \boldsymbol{\beta}) = c_j + w(\gamma\alpha_j, y_j) + w(\gamma(1-\alpha_j), n_j - y_j - 1) - w(\gamma, n_j - 1),$$

where $w(a, m) = \sum_{k=0}^{m-1} \log(a+k)$ and $w(a, 0) \equiv 0$. The score function is given by

$$\frac{\partial l_j}{\delta\gamma} = \alpha_j d(\gamma\alpha_j, y_j) + (1-\alpha_j)\, d(\gamma(1-\alpha_j), n_j - y_j) - d(\gamma, n_j)$$

and

$$\frac{\delta l_j}{\delta\beta_i} = \gamma x_{ij}\alpha_j(1-\alpha_j)\{d(\gamma\alpha_j, y_j) - d(\gamma(1-\alpha_j), n_j - y_j)\},$$

Table 2. Logistic Regression, Quasi-Likelihood, and Betabinomial-Logit Results for the Florida Hospital Data

Variable	Logistic	S.E.	t	QL S.E.	t	BB-Logit	S.E.	t
γ	*	*	*	*	*	40.289	14.201	2.84
intercept	−1.432	.077	−18.51	.093	−15.43	−1.435	.099	−14.53
bedsupply	.095	.071	1.34	.085	1.12	.057	.091	.63
cmi	−.009	.055	−.16	.066	−.13	.013	.069	.19
%MDadm	−.184	.081	−2.26	.098	−1.88	−.186	.101	−1.84
income	.129	.057	2.25	.069	1.88	.136	.074	1.35
rehabs	−.217	.096	−2.26	.115	−1.88	−.267	.125	−2.14
Log-likelihood (model)		−298.575					−289.269	
Log-likelihood (intercept)		−309.044					−296.480	

NOTE: * = Not Applicable; t = t-Ratio.

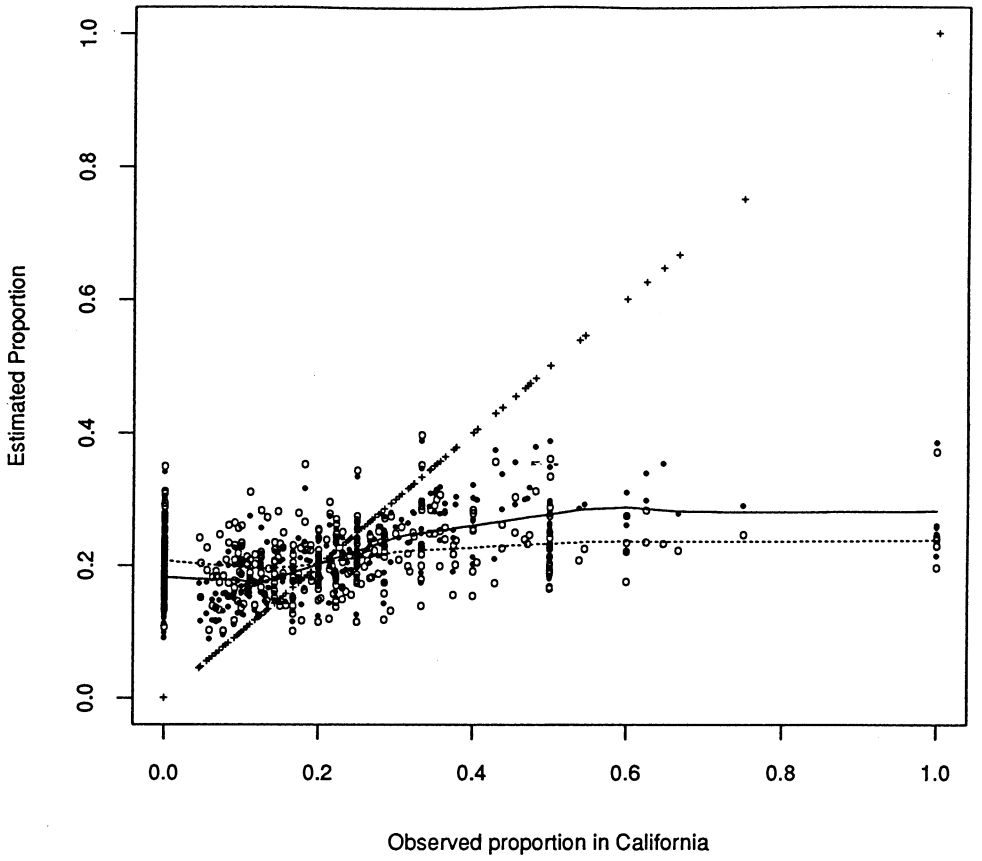

Figure 2. Scatterplot of the Empirical Bayes Point Estimate, in California, of p_j (Symbol "·") and the Logistic Regression Estimate of p_j (Symbol "○") Versus the Individual Estimate of p_j (Symbol "+"). The curves are robust, locally weighted scatterplot smooths (Cleveland 1979). The naive empirical Bayes estimate (solid line) always lies between Y/n and the logistic regression estimate (dotted line).

where

$$d(a,m) = \sum_{k=0}^{m-1} \frac{1}{a+k} \quad \text{and} \quad d(a,0) = 0.$$

The second derivatives are

$$\frac{\delta^2 l_j}{\delta\gamma^2} = \sum_{k=0}^{n_j-1} \frac{1}{(\gamma+k)^2} - \sum_{k=0}^{y_j-1} \frac{\alpha_j}{(\gamma\alpha_j+k)^2} - \sum_{k=0}^{n_j-y_j-1} \frac{(1-\alpha_j)}{(\gamma(1-\alpha_j)+k)^2},$$

$$\frac{\delta^2 l_j}{\delta\gamma\,\delta\beta_i} = x_{ij}\alpha_j(1-\alpha_j)\left\{\sum_{k=0}^{y_j-1} \frac{k}{(\gamma\alpha_j+k)^2} - \sum_{k=0}^{n_j-y_j-1} \frac{k}{(\gamma(1-\alpha_j)+k)^2}\right\},$$

and

$$\frac{\delta^2 l_j}{\delta\beta_i\delta\beta_l} = A_j\left\{\sum_{k=0}^{y_j-1} \frac{k(1-2\alpha_j)-\gamma\alpha_j^2}{(\gamma\alpha_j+k)^2} - \sum_{k=0}^{n_j-y_j-1} \frac{k(1-2\alpha_j)-\gamma(1-\alpha_j)^2}{(\gamma(1-\alpha_j)+k)^2}\right\},$$

where $A_j = \gamma x_{ij} x_{lj} \alpha_j (1-\alpha_j)$. These are used in Newton–Raphson procedures for finding the MLE's of γ and $\boldsymbol{\beta}$. Kahn (1990) showed that MLE's from the betabinomial model are consistent and asymptotically normal. Constrained methods are necessary if $\hat{\gamma} \le 0$, because $\gamma > 0$.

Let $\hat{\beta}^{(\mathrm{BB})}$ be the MLE of $\boldsymbol{\beta}$ from the betabinomial-logit model and let $\hat{\beta}^{(\mathrm{L})}$ be the MLE of $\boldsymbol{\beta}$ from the logistic regression (or quasi-likelihood) model. Then $\hat{\alpha}_j = \mathrm{logit}^{-1}(\mathbf{x}_j^T\hat{\beta}^{(\mathrm{BB})})$ and $\hat{p}_j^{(\mathrm{L})} \equiv \mathrm{logit}^{-1}(\mathbf{x}_j^T\hat{\beta}^{(\mathrm{L})})$. From Section 2.3.2, we know that $\hat{\alpha}_j \approx \hat{p}_j^{(\mathrm{L})}$, so that (7) implies that the empirical Bayes estimator obtained by plugging in $\hat{\gamma}$ and $\hat{\alpha}$ for γ and α in (4) is

$$\hat{p}_j \approx (1-\hat{r}_j^2)\hat{p}_j^{(\mathrm{L})} + \hat{r}_j^2\left(\frac{Y_j}{n_j}\right),$$

a weighted average of the logistic regression estimator and the observed rate at hospital j, with $\hat{r}_j = \sqrt{n_j/(n_j+\hat{\gamma})}$.

3. ACCOUNTING FOR UNCERTAINTY ABOUT THE HYPERPARAMETERS: A FULLY BAYESIAN APPROACH

We now develop a fully Bayesian analysis of the betabinomial-logit model. This requires that a prior distribution be specified for the hyperparameters, $(\gamma, \boldsymbol{\beta})$. To begin, we assume that the scale parameter, γ, is independent a priori of the regression parameter, $\boldsymbol{\beta}$. (This was the motivation for the parameterization given by (2).) Further,

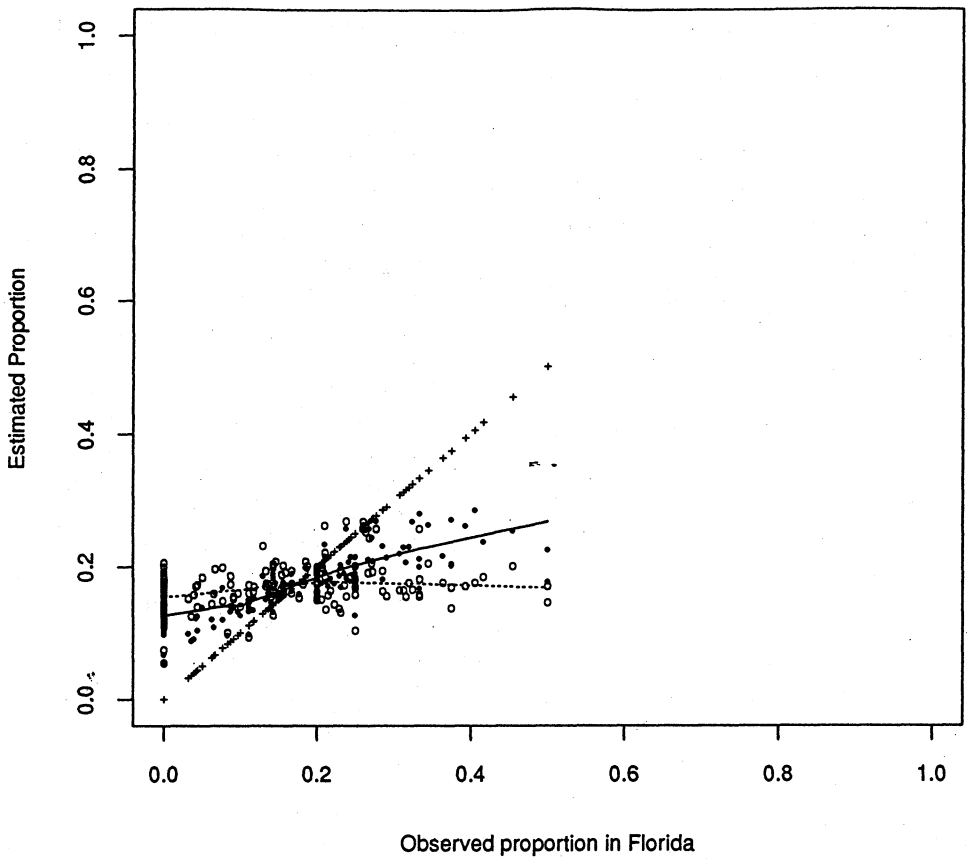

Figure 3. Scatterplot of the Naive Empirical Bayes Point Estimate (Solid Line), in Florida, of p_j (Symbol "·") and the Logistic Regression Estimate (Dotted Line) of p_j (Symbol "○") Versus the Individual Estimate of p_j (Symbol "+"). The curves are robust, locally weighted scatterplot smooths (Cleveland 1979).

we assume that the scale parameter follows a Jeffreys's invariant (locally uniform) distribution,

$$h(\gamma) \propto \gamma^{-1}.$$

Even though this prior is improper, it is straightforward to show that the resulting posterior is proper. A more challenging problem is the specification of a prior for the regression parameter, $\boldsymbol{\beta}$. We follow the approach taken by Raftery (1988). Suppose that the independent variables have been centered and scaled to have mean zero and variance 1. Then let

$$\boldsymbol{\beta} \sim N((\nu, 0, \ldots, 0)^T, \operatorname{diag}\{\tau^2, \phi^2, \ldots, \phi^2\}).$$

This has the appealing property that it is objective in the sense of Berger and Sellke (1987) (i.e., symmetric and nonincreasing away from the intercept, β_1, relative to the scaled covariates).

This prior then leads to the following posterior for p_j:

$$f(p_j|Y) = C^{-1} \int g(p_j|Y_j, \gamma, \boldsymbol{\beta}) \prod_{i=1}^{N} p(Y_i|\gamma, \boldsymbol{\beta}) h(\gamma, \boldsymbol{\beta})\, d\gamma\, d\boldsymbol{\beta}, \quad (10)$$

where

$$C = \int \prod_{i=1}^{N} p(Y_i|\gamma, \boldsymbol{\beta}) h(\gamma, \boldsymbol{\beta})\, d\gamma\, d\boldsymbol{\beta},$$

$$g(p_j|Y_j, \gamma, \boldsymbol{\beta}) = \frac{p_j^{\gamma\alpha_j + Y_j - 1}(1-p_j)^{\gamma(1-\alpha_j)+n_j-Y_j-1}}{B(\gamma\alpha_j + Y_j, \gamma(1-\alpha_j)+n_j-Y_j)},$$

$$p(Y_i|\gamma, \boldsymbol{\beta}) = \frac{B(\gamma\alpha_i + Y_i, \gamma(1-\alpha_i)+n_i-Y_i)}{B(\gamma\alpha_i, \gamma(1-\alpha_i))},$$

and

$$h(\gamma, \boldsymbol{\beta}) \propto \frac{1}{\gamma}\tau^{-1}\phi^{-(m-1)} \exp\left\{-\frac{1}{2}\left[\left(\frac{\beta_1 - \nu}{\tau}\right)^2 + \left(\frac{\beta_2}{\phi}\right)^2 + \cdots + \left(\frac{\beta_m}{\phi}\right)^2\right]\right\},$$

with $B(a,b) = [\Gamma(a)\Gamma(b)]/[\Gamma(a+b)], p_i \in (0,1), \alpha_i = \operatorname{logit}^{-1}(x_i^T\boldsymbol{\beta}) \in (0,1), \gamma > 0, \boldsymbol{\beta} \in \mathbb{R}^m, \tau > 0$, and $\phi > 0$. There is no simple, closed-form expression for this integral, so one needs an approximation. We choose to use the Laplace approximation for integrals (De Bruijn 1970; Tierney and Kadane 1986; Tierney, Kass, and Kadane 1989), which works as follows. Suppose that $F : \mathbb{R}^k \to \mathbb{R}$ is twice differentiable. Then

$$\int F(\mathbf{x})\, d\mathbf{x} \approx (2\pi)^{k/2}|A|^{1/2}F(\mathbf{x}^*), \quad (11)$$

where $\mathbf{x}^*$ is the value of $\mathbf{x}$ at which $F(\mathbf{x})$ attains its maximum and A is minus the inverse Hessian of $\ln(F(\mathbf{x}))$ evaluated at $\mathbf{x}^*$.

Table 3. Shrinkage Summary for the California Hospital Data

n	Min	Q1	Median	Mean	Q3	Max
360	.41	.70	.80	.78	.89	.98

In our case we have two integrals to approximate: one for the numerator and one for the denominator (i.e., calculation of C). In the numerator we have

$$\begin{aligned}\log(F(\gamma,\boldsymbol{\beta})) &= -\log(B(\gamma\alpha_j + Y_j, \gamma(1-\alpha_j) + n_j - Y_j)) \\ &\quad + (\gamma\alpha_j + Y_j - 1)\log(p_j) \\ &\quad + (\gamma(1-\alpha_j) + n_j - Y_j - 1)\log(1-p_j) \\ &\quad + \sum_{i=1}^{N} S_i(\gamma,\boldsymbol{\beta}), \qquad (12)\end{aligned}$$

where

$$\begin{aligned}S_i(\gamma,\boldsymbol{\beta}) &= \log\binom{n_i}{Y_i} \\ &\quad + \log(B(\gamma\alpha_i + Y_i, \gamma(1-\alpha_i) + n_i - Y_i)) \\ &\quad - \log(B(\gamma\alpha_i, \gamma(1-\alpha_i))) \\ &\quad - \log(\gamma) - \frac{m}{2}\log(2\pi) \\ &\quad - \log(\tau) - (m-1)\log(\phi) \\ &\quad - \left\{\frac{1}{2}\left[\left(\frac{\beta_1-\nu}{\tau}\right)^2 + \left(\frac{\beta_2}{\phi}\right)^2 \right.\right. \\ &\qquad \left.\left. + \cdots + \left(\frac{\beta_m}{\phi}\right)^2\right]\right\}.\end{aligned}$$

In the denominator we have

$$\log(G(\gamma,\boldsymbol{\beta})) = \sum_{i=1}^{N} S_i(\gamma,\boldsymbol{\beta}). \qquad (13)$$

Let γ^* and β^* be the values of γ and $\boldsymbol{\beta}$ that maximize (12) and, similarly, let γ^{**} and β^{**} be the values of γ and $\boldsymbol{\beta}$ that maximize (13). Let A be minus the inverse Hessian of (12) evaluated at (γ^*, β^*) and let B be minus the inverse Hessian of (13) evaluated at $(\gamma^{**}, \beta^{**})$. Then the Laplace approximation to (10) becomes

$$\begin{aligned}f(p_j|\mathbf{Y}) \approx |A|^{1/2}|B|^{-1/2}\exp(&\log(F(\gamma^*,\beta^*)) \\ &- \log(G(\gamma^{**},\beta^{**}))). \qquad (14)\end{aligned}$$

4. APPLICATION: MEDICARE DATA

4.1 Standard Empirical Bayes Analysis

We use data obtained from Dr. C. R. Neu at the RAND Corporation. For this example, we will consider urban (as defined by the U.S. Census Department's Metropolitan Statistical Areas [MSA's] in 1985) hospitals in California and Florida. The data are for Medicare stroke patients discharged from hospitals between July 1, 1984 and June 31, 1985.

Let n_j be the number of stroke patients discharged from hospital j and let Y_j be the number of the n_j who initially went to a nursing home on discharge from hospital j.

The covariates used in this example are as follows:

- cmi: The "case mix index" is the relative costliness of the group of Medicare patients at the hospital. This variable is a proxy for tertiary-care hospitals, or hospitals that are better suited to deal with Medicare patients' more complicated illnesses.
- %MDadm: The percentage of the hospital's admissions that are paid through Medicaid, a state-run health care program for the underprivileged. This variable is a proxy for inner-city, poor area hospitals.
- bedsupply: The ratio of Medicare-approved nursing home beds in the MSA to the number of persons age 65 or older in the MSA.
- income: The per capita income for the MSA.
- rehabs: A dichotomous variable indicating whether there is a secondary-care rehabilitation unit in the MSA (rehabs = 1) or not (rehabs = 0). These facilities are more intensive secondary-care facilities than SNF's.

With the exception of the indicator variable, rehabs, all other variables have been centered to have mean zero and scaled to have standard deviation 1. Also note that the case mix index (cmi) and percent of Medicaid admissions (%MDadm) are hospital-specific variables, whereas the standardized bed supply (bedsupply), per capita income (income), and whether or not there is a rehabilitation facility in the area (rehabs) are MSA-level variables serving as proxies for market characteristics of the area in which the hospital operates.

Table 1 gives results from fitting the logistic regression model (1) to the California data, as well as the quasi-likelihood model ($\hat{\sigma}^2 \approx 1.3$) and the betabinomial-logit model (3). In each case $N_{\mathrm{CA}} = 360$ and $\sum_{j=1}^{N_{\mathrm{CA}}} n_j = 4{,}811$. Table 2 gives results for Florida with $\hat{\sigma}^2 \approx 1.45$, $N_{\mathrm{FL}} = 157$, and $\sum_{j=1}^{N_{\mathrm{FL}}} n_j = 3{,}207$.

As is to be expected from Section 2.3.2, the point estimates for $\boldsymbol{\beta}$ in the two models are close. Figures 2 and 3 plot each hospital's observed proportion, each hospital's logistic regression estimate, and each hospital's naive empirical Bayes estimate and scatterplot smooths for each (see Cleveland 1979). Further, consider $1 - r_j^2$ to be the amount of shrinkage for each hospital's estimate, $r_j = 1 \Leftrightarrow 1 - r_j^2 = 0$ representing no shrinkage from the individual hospital's estimate of Y_j/n_j and $r_j = 0 \Leftrightarrow 1 - r_j^2 = 1$ being complete shrinkage to the ensemble, logistic regression estimate. In this case Tables 3 and 4 describe the amount of shrinkage for the California and Florida hospitals. Because the average rate for the hospitals is small, about .2, there is a large

Table 4. Shrinkage Summary for the Florida Hospital Data

n	Min	Q1	Median	Mean	Q3	Max
157	.35	.59	.72	.70	.83	.98

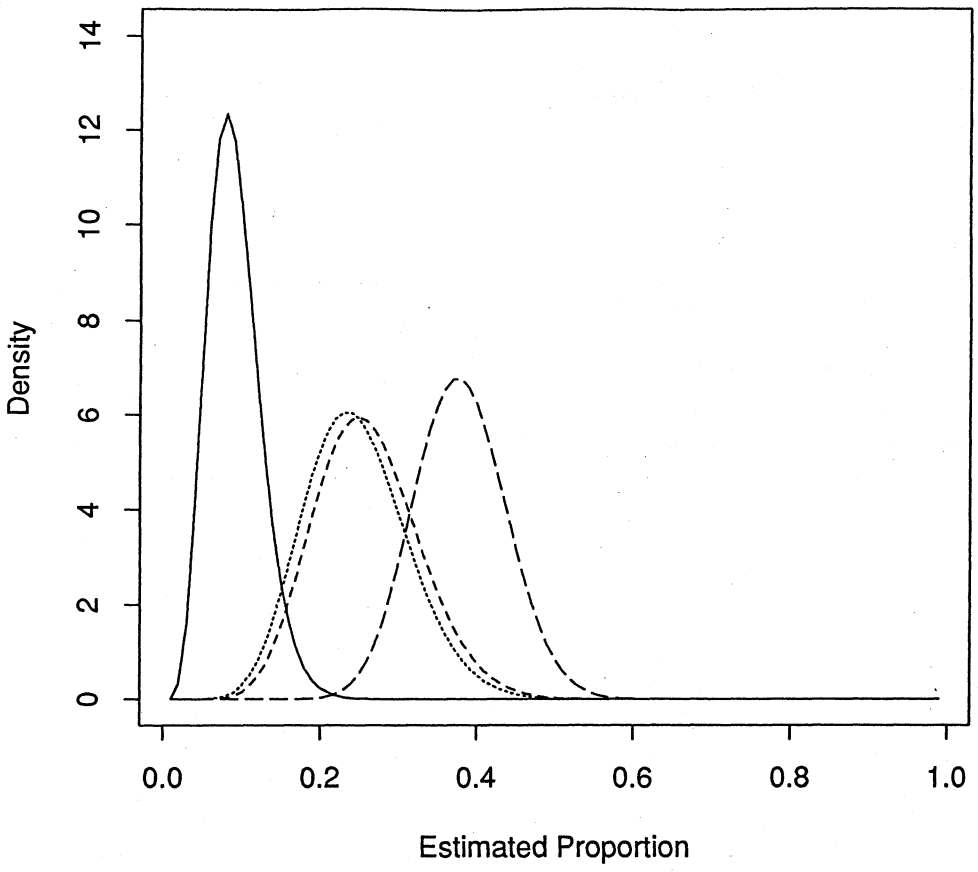

Figure 4. Four Naive Empirical Bayes Beta Densities. The middle two correspond to California hospitals with only one Medicare stroke discharge; one went to a nursing home (short-dashed line), one did not (dotted line). The two densities on the ends correspond to hospitals with relatively large numbers of Medicare stroke discharges. One had 31 discharges, of whom none went to a nursing home (solid line). The other had 27 discharges, of whom 13 went to a nursing home (long-dashed line).

cluster of points close to .2, making it difficult to see the shrinkage in this region. It is easy, though, to see the shrinkage for the California hospitals with observed rates above .3. The smooth of the logistic regression values indicates that the logistic regression estimates vary around the MLE of p_j under the assumption that $p_j = p$ for all j, namely $\sum Y_j / \sum n_j$. The smooth of the empirical Bayes estimates indicates that these estimates vary about a line with a slope between that for the logistic regression case (slope $= 0$) and that for the individual estimation of p_j by Y_j/n_j (slope $= 1$). Also, notice that the hospitals whose observed proportions are 1.0 have empirical Bayes and logistic regression estimates that are almost identical. This is due to the fact that all of these hospitals had a single Medicare stroke discharge who went to a nursing home, so that $\hat{r}_j^2$ is small. Hospitals with large numbers of Medicare stroke patients discharged have an empirical Bayes estimate nearer Y_j/n_j.

Tables 1 and 2 show that bedsupply is not a significant predictor of nursing home use. In California, cmi, %MDadm, and rehabs are significantly negatively associated with nursing home use, whereas income is significantly positively associated with nursing home use. Thus California hospitals in wealthy MSA's without a rehabilitation facility tend to have higher rates of discharge to nursing homes. After adjusting for other relevant factors, a 1 SD increase in percentage of hospital Medicaid admissions in California is associated with a decline of about .12 on the logit scale, which translates into roughly a 10% decline (from .2 to .18) in SNF discharge rate for an otherwise average hospital. The effect of having a rehabilitation facility in the MSA is fairly large, resulting in a decline of about a .67 on the logit scale, which translates into roughly a 45% decline (from .2 to .11) in SNF discharge rate for an otherwise average hospital.

In Florida the picture looks much different; cmi and bedsupply are uncorrelated with nursing home use, whereas %MDadm, income, and rehabs are marginally correlated with nursing-home use, although the signs of the coefficients for these three covariates are the same as they were in California. Figure 4 plots four densities corresponding to hospitals with four different discharge patterns, to illustrate the shrinkage another way. Two of the hospitals had relatively large numbers of Medicare stroke patients discharged but very different proportions going to nursing homes: 0/31 for one hospital, 13/27 for the other. The other two hospitals each had a single Medicare stroke patient discharged, but one hospital sent the patient to a nursing home and the other did not. It is clear from this picture that the two hospitals with a single Medicare stroke discharge have densities that are very similar and near the center of the data. The other two densities have nearly disjoint supports. The hospital with an observed proportion of 0/31 has an empirical Bayes density for the "true" rate of discharging Medicare stroke patients to nursing homes concentrated approximately on (0, .22), whereas the hospital with an observed rate of 13/27 has support concentrated on (.21, .57).

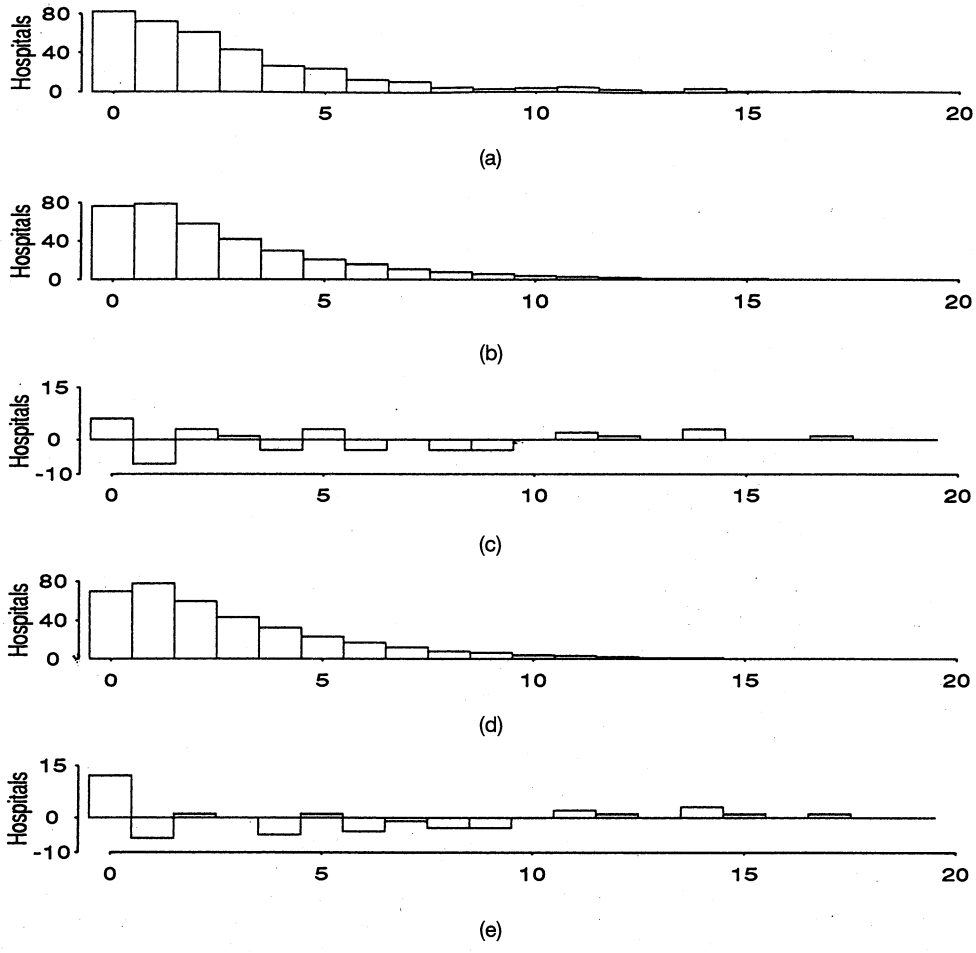

Figure 5. A Comparison of the Observed Number of Discharges per Hospital to the Expected Number Under the Betabinomial-Logit and Logistic Regression Models in California. (a) Observed discharges per hospital in California, (b) expected discharges per hospital under the betabinomial-logit model; (c) observed–expected (betabinomial-logit); (d) expected discharges per hospital under the logistic regression model; (e) observed-expected (logistic).

Tables 1 and 2 also show that the standard errors of the components of β using quasi-likelihood are close to those from the betabinomial-logit scheme. This has been the case for a number of other data sets to which we have applied this methodology. Also, in California the log-likelihood of the betabinomial-logit model is greater than that of the logistic regression models by 10.4 points for just one additional parameter. In terms of Bayes information criterion (BIC) (Schwarz 1978), the betabinomial-logit model is also favored, by a margin of about 14.8 points. This corresponds to an approximate Bayes factor of about 1,660:1 in favor of the betabinomial-logit model, which is strong evidence.

In Florida, the betabinomial-logit model yields an observed likelihood larger than that of the logistic regression model by 9.3 points. BIC favors the betabinomial-logit model here by a margin of 13.6 points, corresponding to an approximate Bayes factor of about 880:1. Because the QL model does not specify a stochastic model, neither a likelihood nor a goodness-of-fit statistic can be calculated. Figures 5 and 6 show the expected values for each cell in the betabinomial-logit model and the logistic regression model. The betabinomial-logit model accounts for more of the extrabinomial variability in California. In Florida, the betabinomial-logit model does a better job of explaining the hospitals with observed rates of zero, although the fit is less than perfect, and deserves closer inspection and possibly inclusion of other independent variables.

4.2 Fully Bayesian Posterior Densities

The numerical integrations of Section 3 were performed for the four hospitals in Figure 4. A Newton–Raphson procedure and a minus inverse Hessian routine were written in Fortran to find the maximum values required in (14); these are available from the first author (kahn@mayo.edu). Figure 7 shows both the naive empirical Bayes density from Figure 4 and the approximate posterior density from (14). The fully Bayesian posterior densities are wider than the naive densities, by about 15% as measured by the posterior standard deviation, whereas the posterior means are practically identical. This exhibits how far off one can be when using the naive estimate of variability. It is interesting to note that the inference is highly robust with respect to specification of the prior variances of β. We calculated the posterior distribution for various combinations of τ^2 and ϕ^2 between 1 and 100 and found that it exhibited little variability in this broad range. For this problem with these data, the likelihood dominates the prior.

Despite the large t statistics for some predictors, the posterior densities of the California hospital discharge rates

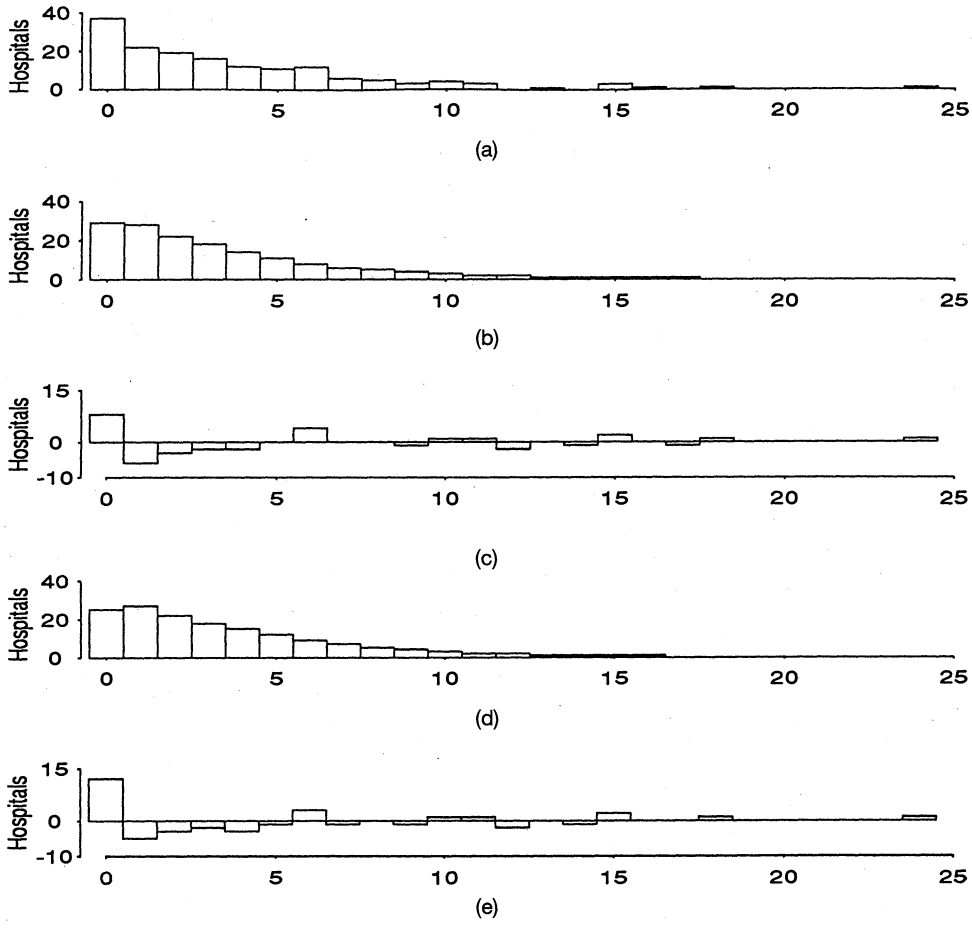

Figure 6. A Comparison of the Observed Number of Discharges per Hospital to the Expected Number Under the Betabinomial-Logit and the Logistic Regression Models in Florida. (a) Observed discharges per hospital in Florida; (b) expected discharges per hospital under the betabinomial-logit model; (c) observed–expected (betabinomial-logit); (d) expected discharges per hospital under the logistic regression model; (e) observed-expected (logistic).

show a high degree of overlap. The two most disparate cases—0/31 and 13/27—have posterior densities that overlap by a small amount (5–7%), and it can be said with probability about 95% that these hospitals have different true rates of initially discharging stroke patients to nursing homes. But no other two densities are as different. Further investigation as to why the hospital with 31 stroke discharges had none going to a nursing home would be in order. For the covariates reported in this article, the 0/31 hospital had Medicaid admission of more than 44% (this is 2.24 standard deviations above the mean percentage of Medicaid admissions), whereas the 13/27 hospital had about 13% Medicaid admissions (very nearly the mean value). Because a high percentage of Medicaid admissions is typically a feature of poorer, inner-city hospitals, and because hospitals with higher percentages of Medicaid admissions tend to have lower rates of stroke patients discharged to SNF's, one might want to investigate the possibility that nursing homes in the area prefer not to deal with this particular hospital.

The betabinomial-logit model better accounts for the excess of observed zeros than the logistic regression model, as is shown by Figures 5 and 6. The betabinomial-logit also accounts for this excess of zeros better than a model using a mixture of a point mass at zero and a logistic regression model (Kahn 1990). Hence the excess of zeros is accounted for by the unobserved heterogeneity of "true" hospital rates of discharge to SNF's. There is no evidence that any hospital has a "true" discharge rate of zero.

5. DISCUSSION

In this article we have developed the betabinomial-logit model as a way of accounting for unobserved heterogeneity in logistic regression. Maximum likelihood estimation for the betabinomial-logit model is straightforward and takes little more computing time than logistic regression. The betabinomial-logit model behaves like a quasi-likelihood method but enjoys the advantages of a true probability model. Logistic regression is also a probability model but does not allow for unobserved heterogeneity. The naive empirical Bayes, betabinomial-logit estimate is (approximately) a shrinkage estimator that lies between the logistic regression estimator and the individual estimate, $\hat{p}^{(\mathrm{NEB})} \approx (1 - r^2)\hat{p}^{(\mathrm{L})} + r^2(Y/n)$, where $r = \mathrm{corr}(Y, p)$.

The fully Bayesian approach is straightforward using the Laplace approximation to obtain a proper posterior distribution for the $\{p_i\}$ from which we can draw inference. This accounts for variability in the estimation of the hyperparameters, which standard empirical Bayes procedures ignore. In the California data we found that the fully Bayesian posterior variances were about 20–30% higher than the standard empirical Bayes posterior variances and were robust to the choice of prior variance for the regression parameters

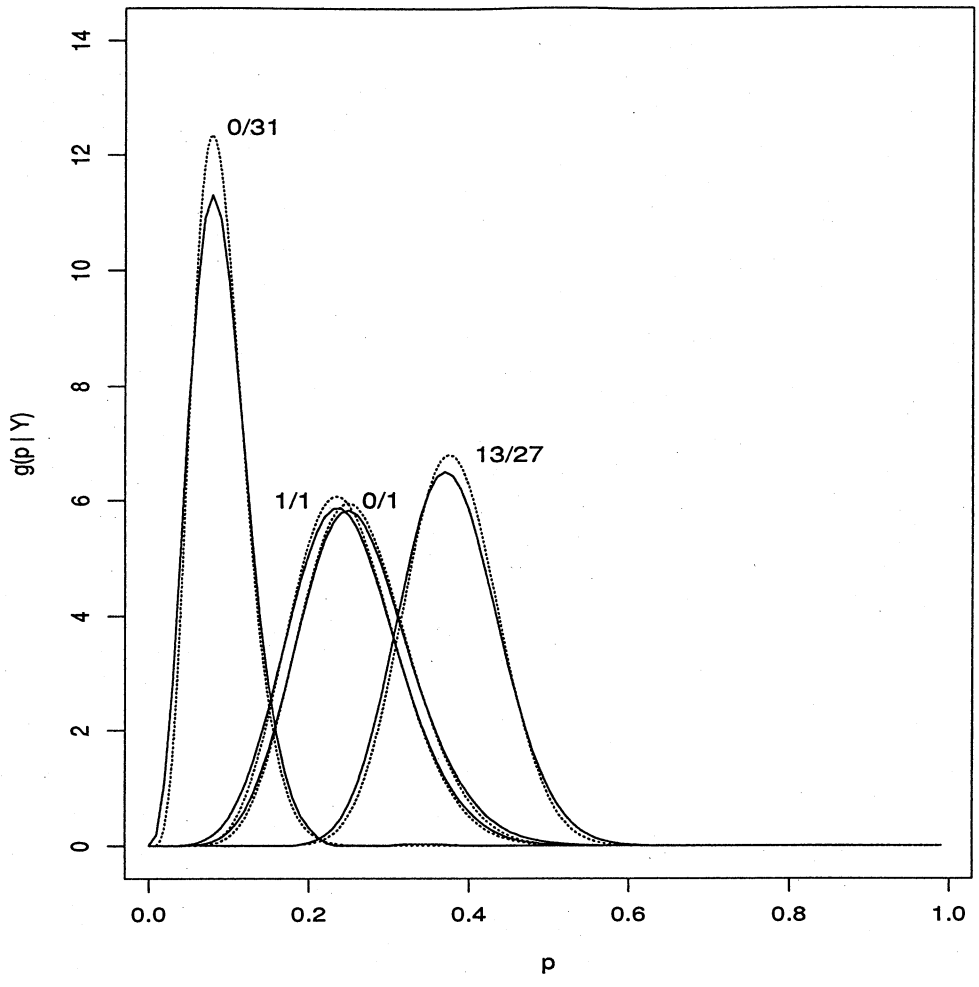

Figure 7. The Naive Empirical Bayes (Dotted Line) Posterior Density Versus the Bayes Empirical Bayes (Solid Line) Density for Four California Hospitals; One With No Stroke Discharges to SNF From a Total of 1 Stroke Discharge, One With 1 Stroke Discharge to SNF From a Total of 1 Stroke Discharge, One With 13 Stroke Discharges to SNF From a Total of 27 Stroke Discharges, and One With No Stroke Discharges to SNF From a Total of 31 Stroke Discharges.

of the betabinomial-logit model. Also, the fully Bayesian posterior mean is nearly the same as the standard empirical Bayes mean, so that the interpretation of the posterior mean as a shrinkage estimator between the logistic regression estimate and the individual estimate remains valid under the fully Bayesian analysis.

For this problem, we found that the hospital-specific and market area–specific factors, associated with hospitals' propensities for discharging Medicare stroke patients to SNF's differ between California and Florida. In California we found that both a hospital's case-mix index and its percentage of Medicaid admissions are negatively correlated with its rate of discharge to a SNF. At the market-area level, the per capita income of the MSA in which the hospital operates is positively correlated with a hospital's rate of discharge to SNF's, whereas the existence of a rehabilitation facility (a more intensive secondary-care facility) in the hospital's MSA is associated with fewer discharges to SNF's. In Florida, the only covariate significantly associated with discharge rates to SNF's is whether there is a rehabilitation facility in the hospital's MSA, with the presence of such a facility associated with a lower discharge rate to a SNF. And finally, in both states the large number of observed zeros is best explained by the unobserved heterogeneity between hospitals' rates. Though there is evidence that some hospitals have small rates of discharging Medicare stroke patients to a SNF, there is no evidence that any hospital in California or Florida has a "true" rate of zero.

Finally, many authors have suggested using the Gibbs sampler for calculating posterior distributions in hierarchical models (Gelfand and Smith 1990) and in particular for the betabinomial distribution (Casella and George 1992). But for the betabinomial-logit model, we have found that the Laplace method is accurate and requires far less computer time than the Gibbs sampler. Though the Laplace approximation requires quite a bit more analytical work (i.e., calculation of gradients and Hessians), the work pays off generously. The Laplace method has also been found to be both more exact and less computationally expensive than the Gibbs sampler for calculating Bayes factors in a Poisson gamma model (Rosenkranz 1992).

[Received August 1992. Revised August 1995.]

REFERENCES

Aitchison, J., and Shen, S. M. (1980), "Logistic-Normal Distributions: Some Properties and Uses," *Biometrika*, 67, 261–272.

Albert, J. H. (1988), "Computational Methods Using a Bayesian Hierarchical Generalized Linear Model," *Journal of the American Statistical Association*, 83, 1037–1044.

Anderson, A., and Aitkin, M. (1985), "Variance Component Models With Binary Response: Interviewer Variability," *Journal of Royal Statistical Society*, Ser. B, 47, 203–210.

Berger, J., and Sellke, T. (1987), "Testing a Point Null Hypothesis: The Irreconcilability of P Values and Evidence," *Journal of the American Statistical Association*, 82, 112–122.

Carter, G., and Rolph, J. (1973), "New York City Fire Alarm Prediction: I. Box-Reported Serious Fires," The NYC RAND Institute, R-1214-NYC.

Cassela, G., and George, E. (1992), "Explaining the Gibbs Sampler," *The American Statistician*, 46, 167–174.

Chatfield, C., and Goodhardt, G. J. (1970), "The Betabinomial Model for Consumer Purchasing Behavior," *Applied Statistics*, 19, 240–250.

Cleveland, W. S. (1979), "Robust Locally Weighted Regression and Smoothing Scatterplots," *Journal of the American Statistical Association*, 74, 829–836.

Cox, D. R. (1970), *The Analysis of Binary Data*, London: Chapman and Hall.

Crowder, M. J. (1978), "Beta-Binomial ANOVA for Proportions," *Applied Statistics*, 27, 34–37.

Danaher, P. J. (1988), "Parameter Estimation and Applications for the Generalisation of the Betabinomial Model," *Australian Journal of Statistics*, 30, 263–275.

De Bruijn, N. G. (1970), *Asymptotic Methods in Analysis*, Amsterdam: North-Holland.

Follman, D., and Lambert, D. (1989), "Generalized Logistic Regression by Nonparametric Mixing," *Journal of the American Statistical Association*, 84, 295–300.

Gelfand, A., and Smith, A. M. F. (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398–409.

Griffiths, D. A. (1973), "Maximum Likelihood Estimation of the Betabinomial Distribution and an Application to the Household Incidence of the Total Number of Cases of a Disease," *Biometrics*, 29, 637–648.

Haseman, J. K., and Kupper, L. L. (1979), "Analysis of Dichotomous Response Data From Certain Toxicological Experiments," *Biometrics*, 35, 281–293.

Healy, M. J. R. (1972), "Animal Litters as Experimental Units," *Applied Statistics*, 21, 155–159.

Im, S., and Gianola, D. (1988), "Mixed Models for Binomial Data, With an Application to Lamb Mortality," *Applied Statistics*, 37, 196–204.

Johnson, N. L., and Kotz, S. (1970), *Distributions in Statistics, Continuous Univariate Distributions*, Vol. 2, Boston: Houghton-Mifflin.

Kahn, M. J. (1990), "Incorporating Covariates Into a Betabinomial Model With Applications to Medicare Policy: A Bayes Empirical Bayes Approach," PhD dissertation, University of Washington, Dept. of Statistics.

Kleinman, J. C. (1973), "Proportions With Extraneous Variation: Single and Independent Samples," *Journal of the American Statistical Association*, 68, 46–54.

McCullagh, P., and Nelder, J. (1989), *Generalized Linear Models* (2nd ed.), New York: Chapman and Hall.

Neu, C. R., and Harrison, S. C. (1988), "Posthospital Care Before and After the Medicare Prospective Payment System," The RAND Corporation, R-3590-HCFA.

Prentice, R. L. (1986), "Binary Regression Using an Extended Betabinomial Distribution With Discussion of Correlation Induced by Covariate Measurement Errors," *Journal of the American Statistical Association*, 81, 321–327.

Raftery, A. E. (1988), "Approximate Bayes Factors for Generalized Linear Models," Technical Report 121, University of Washington, Dept. of Statistics.

Rosenkranz, S. L. (1992), "The Bayes Factor for Model Evaluation in a Hierarchical Poisson Model for Area Counts," Ph.D. dissertation, University of Washington, Dept. of Biostatistics.

Schwarz, G. (1978), "Estimating the Dimension of a Model," *The Annals of Statistics*, 6, 461–464.

Skellam, J. G. (1948), "A Probability Distribution from the Binomial Distribution by Regarding the Probability of Success as Variable Between the Sets of Trials," *Journal of the Royal Statistical Society*, Ser. B, 10, 257–265.

Stigler, S. M. (1990), "A Galtonian Perspective on Shrinkage Estimators," *Statistical Science*, 5, 147–155.

Stiratelli, R., Laird, N., and Ware, J. (1984), "Random-Effects Models for Serial Observations With Binary Response," *Biometrics*, 40, 961–971.

Tamura, R. N., and Young, S. S. (1986), "The Incorporation of Historical Control Information in Tests of Proportions: Simulation Study of Tarone's Procedure," *Biometrics*, 42, 343–349.

——— (1987), "A Stabilized Moment Estimator for the Betabinomial Distribution," *Biometrics*, 43, 813–824.

Tierney, L., and Kadane, J. (1986), "Accurate Approximations for Posterior Moments and Marginal Densities," *Journal of the American Statistical Association*, 81, 82–86.

Tierney, L., Kass, R., and Kadane, J. (1989), "Fully Exponential Laplace Approximations to Expectations and Variances of Nonpositive Functions," *Journal of the American Statistical Association*, 84, 710–716.

Williams, D. A. (1975), "The Analysis of Binary Responses From Toxicological Experiments Involving Reproduction and Teratogenicity," *Biometrics*, 31, 949–952.

——— (1982), "Extra-Binomial Variation in Logistic Linear Models," *Applied Statistics*, 31, 144–148.

Wong, G. Y., and Mason, W. M. (1985), "The Hierarchical Logistic Regression Model for Multilevel Analysis," *Journal of the American Statistical Association*, 80, 513–524.