STATISTICS IN MEDICINE, VOL. 16, 2311–2325 (1997)

HIERARCHICAL POLYTOMOUS REGRESSION MODELS WITH APPLICATIONS TO HEALTH SERVICES RESEARCH

MICHAEL J. DANIELS[1]* AND CONSTANTINE GATSONIS[2]

[1] *Department of Statistics, Iowa State University, U.S.A.*
[2] *Center for Statistical Sciences, Brown University, U.S.A.*

SUMMARY

The analysis of variations is an important area of interest in health services and outcomes research and has two main goals: to identify and quantify variability across units, such as geographic regions or health care providers, in terms of procedure utilization and outcomes, and to explore the links between process, such as regional or hospital practice patterns, and outcomes, such as patient mortality and functional status. Hierarchical regression models are well suited for this type of analysis. In this paper we formulate a hierarchical polytomous regression model and apply it to the analysis of variations in the utilization of alternative cardiac procedures in a national cohort of elderly Medicare patients who had an acute myocardial infarction during 1987. The model is designed to accommodate clustered multinomial data with covariate vectors available on individual cases and on clusters. We present a Bayesian approach to fitting and checking the model using simulated values from the posterior distribution of the parameters. The simulation algorithms are based on Gibbs sampling in combination with Metropolis steps. Using the hierarchical polytomous regression model, we examine how the rates of cardiac procedures depend on patient-level characteristics, including age, gender and race, and whether there exist interstate differences and regional patterns in the use of these procedures. © 1997 by John Wiley & Sons, Ltd.

Statist. Med., **16**, 2311–2325 (1997)
No. of Figures: 4 No. of Tables: 5 No. of References: 43

1. INTRODUCTION

The analysis of variations in the practice of medicine is an important area of interest in health services and outcomes research.[1–6] As a starting point, the analysis of variations seeks to determine whether comparable patients receive similar treatment across geographic regions or care providers and whether they experience similar endpoints. If there is evidence of differences, the analysis then focuses on what patient, regional, or provider characteristics contribute significantly to these differences. In addition, the analysis examines the link between outcomes, such as patient mortality, morbidity, or functioning and process, such as regional or provider specific practice patterns.

Our case study for this paper involves data on cardiac procedure utilization for elderly patients who had an acute myocardial infarction (AMI) in 1987. The data set was derived from records

* Correspondence to: Michael J. Daniels, Department of Statistics, Iowa State University, Snedecor Hall, Ames, IA 50011, U.S.A.

CCC 0277–6715/97/202311–15$17.15

Received July 1996
Revised December 1996

collected regularly by the Health Care Financing Administration (HCFA) and includes over 200,000 patients in the United States, covered by Medicare and having a principal diagnosis of AMI. Patients were classified by place of residence in one of the 51 states (including the District of Columbia). The goal of the analysis was to examine variations in the patterns of utilization of percutaneous transluminal coronary angioplasty (PTCA), coronary artery bypass graft (CABG) and coronary angiography (catheterization) performed during a 90 day period beginning with the initial hospital admission for an AMI.[7] Variations both within and between states are of interest in this data set, which has a natural hierarchical structure.

Hierarchical regression models are ideally suited to the analysis of data clustered along multi-level, nested structures, such as our AMI data set. Models of this type have been used in several areas of research including the analysis of educational data,[8] epidemiologic data,[9] and data on health care utilization and patient outcomes.[5,6,10–12] In this paper we generalize the hierarchical logistic regression model of Wong and Mason[13,14] (see also Gatsonis *et al.*[6]) to the analysis of polytomous responses and apply it to the analysis of the AMI data. We use the polytomous generalization of the logistic link function and employ the Gibbs sampler[15] for model fitting and model checking. Regression models for binary data have also been discussed by Albert and Chib,[16] and have been used in various applications, including prediction of hospital malpractice claims[17] and modelling household purchase behaviour.[18] Models for polytomous data with a probit link but without cluster-level covariates have been implemented by McCulloch and Rossi using Markov chain Monte Carlo techniques.[19] In related work, Goldstein[20] discussed estimation in multilevel polytomous models using linear approximations and Natarajan,[21] in models for multinomial response data. Mor *et al.*[22] applied non-hierarchical polytomous models in the context of discharges of nursing home residents to hospital, home, or cemetery.

The hierarchical polytomous regression model makes it possible to take into account correlations in the data, and to estimate case and cluster-level covariate effects and variance components simultaneously. The model allows for pooling information across clusters to derive more precise estimates of cluster-specific and cluster-level parameters, and estimates of the variance of the effect parameters taking into account variation in the scale matrix of the cluster-specific parameters. In addition, the polytomous generalization of the logistic link allows for convenient interpretation of coefficients in terms of odds ratios.

In Section 2 we present the hierarchical polytomous regression model, and in Section 3 we describe in more detail the AMI data set. MCMC methods to facilitate model checking are described in Sections 4. The analysis of the cardiac data set and details on model fitting appear in Section 5, and Section 6 discusses alternative models. We summarize our findings and conclusions in Section 7.

2. HIERARCHICAL POLYTOMOUS REGRESSION MODEL

The general structure of the proposed hierarchical regression model for clustered multinomial data is as follows (the specifics for the AMI data set are illustrated in Section 3).

2.1. Level I (within-cluster variation)

We assume that responses Y_{ij} follow a multinomial distribution with sample size, $n_{ij} = 1$, and parameter vector $\vec{p}_{ij}$. At the first level, the probability that individual ij falls into category

k follows a polytomous regression model

$$\log\left(\frac{p_{ijk}}{p_{ij1}}\right) = X_{ijk}\beta_{ik} + Z_{ijk}\alpha_k \tag{1}$$

for $i = 1, \ldots, N$ clusters, $j = 1, \ldots, J_i$ cases within the ith cluster, $k = 2, \ldots, K$ categories and $p_{ij1} = 1 - p_{ij2} - \cdots - p_{ijK}$. Here, we are modelling the logarithm of the probability ratio of case ij being in category k versus the reference category 1, as a linear function of case-level covariates (X_{ijk}, Z_{ijk}). The coefficients of X_{ijk}, of dimension $p \times 1$, are specific to the cluster and category and the coefficients of Z_{ijk}, of dimensions $q \times 1$, are fixed over all clusters, but specific to the category.

2.2. Level II (between-cluster variation)

At the second level, we separate the variation between clusters into a systematic and a random component. We model the systematic component through regression equations linking the cluster specific parameters to cluster level covariates. We model the random component using distributional assumptions on the error term. In particular, we assume that the β_i follow a multivariate t-distribution with location parameter $W_i\gamma$, *scale parameter* (matrix) D, and degrees of freedom v:

$$\beta_i \sim t_v(W_i\gamma, D) \tag{2}$$

where $\beta_i = (\beta_{i2}', \ldots, \beta_{iK}')'$, $\gamma = (\gamma_2', \ldots, \gamma_K')'$, $W_i = \text{diag}(G_i', \ldots, G_i')$, $G_i = U_i \otimes I_{p \times p}$, and U_i is a vector of cluster level covariates.

We also place a vague proper prior on the fixed effects, $\alpha = (\alpha_2', \ldots, \alpha_K')'$, by assuming

$$\alpha \sim \text{N}(\alpha^*, A_\alpha) \tag{3}$$

with α^* an estimate of α and A_α a diagonal matrix with large diagonal elements or chosen as an estimate of the variance of a unit information prior.[2,3]

2.3. Level III

At the final stage, we assume vague proper priors on the cluster-level parameter, the corresponding scale matrix and the degrees of freedom. Specifically

$$\gamma \sim \text{N}(\gamma^*, A_\gamma) \tag{4}$$

where γ^* is an estimate of γ and A_γ is a diagonal matrix with large diagonal elements or chosen as a unit information prior

$$D^{-1} \sim W_{p(k-1)}(B^{-1}, 3) \tag{5}$$

where $W_{p(k-1)}(B^{-1}, 3)$ is a Wishart distribution with 3 degrees of freedom. The Wishart is parameterized such that $E[D] = B$. Thus, we can also choose the hyperparameter B as an estimate of D, and we assume the degrees of freedom to be uniform over the interval from 3 to 100, with 100 corresponding to a normal distribution and 3 corresponding to the smallest degrees of freedom necessary for the prior distribution on the β_i to have finite mean and variance. This last point is important for obtaining finite estimates of the posterior mean and variance of the β_i for the clusters that have a flat likelihood function associated with them.

3. EXAMPLE

We illustrate the use of the hierarchical polytomous regression model with an analysis of data from the Health Care Financing Administration collected on a cohort of patients hospitalized for an acute myocardial infarction (AMI) in 1987. The entire data set comprised more than 280,000 patients, who were discharged alive or dead from a hospital with a principal diagnosis of AMI in 1987. Of these patients, we excluded about 70,000 for the following reasons: atypical patients (for example, end stage renal disease); incomplete data; miscoding of AMI diagnosis; AMI in previous year, or incomplete insurance coverage. The final cohort consisted of 218,427 individuals.[7]

In this analysis, we focus attention on cardiac procedures patients received during a 90 day period from the date of admission for AMI. Patients were classified as having undergone no procedure, only coronary angiography, or coronary angiography and a revascularization procedure. The latter could be coronary artery bypass graft (CABG) or percutaneous transluminal coronary angioplasty (PTCA). Previous studies had reported substantial differences in the use of cardiac procedures by gender, race and age.[7,24–27] The effects of gender and race were shown to vary across states and geographic regions of the country.[5,6,28] In addition, there was evidence that on-site availability of a procedure was related to overall use of the procedure.[6,29]

In view of these earlier findings, our analysis had two major goals: (i) to assess how the rates of cardiac procedures depend on patient-level characteristics, including age, gender and race; (ii) to investigate the presence of interstate differences and regional patterns in the use of these procedures. Since previous studies had shown that the effect of race and gender may vary considerably between states while the effect of age did not, we chose to focus our analysis on interstate variation in the effects of gender and race. Table II describes nationwide procedure use and patient characteristics in our cohort. Table I presents procedure use by state.

We coded race as a binary variable (black versus other) and included a patient-level index of comorbidity with non-cardiac conditions (see reference 6 for details on the construction of this index). Since previous studies had shown both interstate and broad regional (such as Northeast versus South versus West versus Midwest) patterns in the utilization of cardiac procedures, we used census region as a proxy for geographic region in our model. Based on an examination of the effect of age using the entire data set, we decided to express age as a piecewise linear spline function that covered three intervals, 65–74, 75–84, and 85 + , and allowed the slope to vary by interval. In the notation of the previous section, we consider models with covariates:

(i) $X_{ijk} = (1, \text{gender, race})$ to express state-specific effects;
(ii) $Z_{ijk} = (\text{age, comorbidity index})$ associated with fixed effects across states;
(iii) $U_i = (I(\text{West}), I(\text{Midwest}), I(\text{South}), I(\text{Northeast}), \text{angiography availability})$ a state-level covariate to express census region, that is, $I(\text{region}) = 1$ if the ith state is located in that region and an index of the availability of angiography to the state's residents.[6]

4. MODEL CHECKING

We can construct a global check of the model using the posterior predictive distribution of the data by comparing a goodness-of-fit statistic for the observed data to one created with a set of predicted data.[30] In particular, we can use the statistic

$$C^{(1)} = \sum_{i=1}^{N} \sum_{j=1}^{L_i} \sum_{t=1}^{K} \frac{(\tilde{y}_{ij(t)} - \tilde{n}_{ij}\hat{p}_{ij(t)})^2}{\tilde{n}_{ij}\hat{p}_{ij(t)}} \tag{6}$$

Table I. State summary data for 1987 AMI cohort

State	Region	Number of AMI's	Cath	CABG	PTCA	Cath avail
AK	W	93	0·097	0·054	0·011	0·280
AL	S	4100	0·161	0·103	0·082	0·551
AR	S	2833	0·101	0·092	0·048	0·444
AZ	W	2410	0·158	0·100	0·087	0·670
CA	W	14411	0·116	0·093	0·084	0·530
CO	W	1601	0·134	0·084	0·075	0·533
CT	NE	3204	0·121	0·088	0·042	0·655
DC	S	313	0·166	0·042	0·058	0·383
DE	S	694	0·098	0·081	0·022	0·372
FL	S	13188	0·126	0·082	0·060	0·512
GA	S	4820	0·139	0·076	0·039	0·557
HI	W	335	0·128	0·128	0·051	0·573
IA	MW	3305	0·098	0·076	0·092	0·400
ID	W	873	0·134	0·102	0·058	0·285
IL	MW	10226	0·116	0·084	0·057	0·523
IN	MW	4962	0·136	0·075	0·072	0·534
KS	MW	2460	0·146	0·063	0·099	0·332
KY	S	4031	0·112	0·069	0·029	0·498
LA	S	3343	0·142	0·074	0·080	0·535
MA	NE	7378	0·072	0·059	0·028	0·271
MD	S	3928	0·091	0·068	0·036	0·461
ME	NE	1600	0·072	0·053	0·030	0·292
MI	MW	8235	0·107	0·079	0·053	0·494
MN	MW	2989	0·083	0·077	0·048	0·300
MO	MW	5681	0·124	0·090	0·086	0·553
MS	S	2469	0·101	0·073	0·040	0·355
MT	W	703	0·138	0·090	0·166	0·407
NC	S	6452	0·108	0·069	0·051	0·344
ND	MW	782	0·139	0·092	0·032	0·389
NE	MW	1777	0·085	0·065	0·057	0·323
NH	NE	1133	0·090	0·062	0·023	0·229
NJ	NE	7455	0·102	0·069	0·031	0·342
NM	W	791	0·133	0·066	0·106	0·267
NV	W	523	0·141	0·120	0·147	0·776
NY	NE	17818	0·078	0·057	0·024	0·376
OH	MW	11174	0·142	0·077	0·047	0·554
OK	S	3375	0·110	0·085	0·058	0·409
OR	W	2006	0·111	0·078	0·070	0·414
PA	NE	15849	0·102	0·066	0·042	0·450
RI	NE	1408	0·067	0·050	0·031	0·342
SC	S	2741	0·123	0·081	0·068	0·460
SD	MW	938	0·058	0·071	0·046	0·196
TN	S	4766	0·134	0·083	0·045	0·508
TX	S	12266	0·129	0·078	0·065	0·519
UT	W	908	0·170	0·126	0·068	0·621
VA	S	4690	0·114	0·073	0·039	0·493
VT	NE	549	0·089	0·086	0·029	0·173
WA	W	3181	0·114	0·101	0·075	0·579
WI	MW	4695	0·122	0·087	0·065	0·417
WV	S	2611	0·103	0·062	0·034	0·330
WY	W	351	0·114	0·123	0·091	0·319

Table II. Nationwide procedure rates and patient characteristics

Procedure		
None	165056	(76%)
Angiography only	24766	(11%)
CABG and angio	16770	(7%)
PTCA and angio	11832	(5%)
Characteristics		
Age		
65–74	111241	(51%)
76–84	81570	(37%)
$\geqslant 85$	25613	(12%)
Gender		
Female	109259	(50%)
Male	109165	(50%)
Race		
Black	11946	(5%)
White	206478	(95%)

where j indexes the number of distinct covariate patterns in the ith cluster, $\hat{p}_{ij(t)}$ denotes an observation from the posterior distribution of the probabilities $p_{ij(t)}$. $\tilde{n}_{ij}$ equals the number of cases with the jth covariate pattern in the ith cluster, and $\tilde{y}_{ij(t)}$ denotes an observation from the posterior predictive distribution of the response, $y_{ij(t)}$.[31] By repeated sampling, we can define a p-value as the number of times $C^{(1)}_{\text{pred}}$ exceeds $C^{(1)}_{\text{obs}}$, where the statistic $C^{(1)}_{\text{pred}}$ is defined as in equation (6) and $C^{(1)}_{\text{obs}}$ is also defined as in equation (6), but with $\tilde{y}_{ij(t)}$ replaced with the observed values. The p-value is a good indicator of whether the variation in the data is consistent with the variation predicted by the model. If the p-value is very large (for example greater than 0·95), the variation of the observed data (measured by $C^{(1)}_{\text{obs}}$), lies in the upper tail of distribution of the predicted variation (measured by $C^{(1)}_{\text{pred}}$), and is an unlikely value under the model specification. A small p-value, near 0 (for example, less than 0·05), implies that the variation in the data is more than the model predicts. One can also compute other statistics to assess model fit using the posterior predictive distribution as discussed in Gelman *et al*., chapter 6.[30]

We can also attempt to assess relative model fit using observations from the posterior distribution of the parameters. The following statistic can be used to assess relative model fit within clusters using the posterior distribution of $\vec{p}_{ij}$ by computing the statistic for each cluster:

$$C^{(2)} = \sum_{i=1}^{N} \sum_{j=1}^{L_i} \sum_{t=1}^{K} \frac{(y_{ij(t)} - \tilde{n}_{ij}\hat{p}_{ij(t)})^2}{\tilde{n}_{ij}\hat{p}_{ij(t)}} \tag{7}$$

with $\hat{p}_{ij(t)}$ denoting the estimated posterior probability that a case with the jth covariate pattern in the ith cluster is in category t and the other quantities defined as above.[31] This statistic will only indicate relative fit, that is, do some clusters fit worse or better than others.

Model selection using Bayes factors shows some promise for these models through the use of the Savage–Dickey density ratio;[32] for example, testing whether regional effects, γ are non-zero or particular fixed effects α are zero. Note, that to compute Bayes factors for testing the regional effects, we need to place the unit information prior (or any proper prior) on γ or α, respectively.

Another method of model checking is cross-validation implemented, for example, by dropping out one or several clusters at a time and refitting the model to determine how well the reduced data model fits the removed clusters. Various statistics can be computed to determine how well the model predicts the removed clusters.[33]

5. RESULTS

5.1. Gibbs sampler

To simulate the posterior distribution of the parameters, we used the Gibbs sampler.[15] The full conditionals of D and γ have inverse Wishart and Normal forms, respectively, so we can generate these deviates directly (see Appendix). To simulate values from the full conditional distributions for v, we used the random-walk Metropolis algorithm[34,35] and for β_i and α, a modification of the procedure suggested in Gilks *et al.* p. 344.[36] To sample from the random coefficients, β_i, we first assumed a normal approximation to their maximum likelihood estimate (MLE), conditional on the current values of the fixed coefficients, α. In particular, we assumed $\hat{\beta}_i|\alpha \sim \mathrm{N}(\beta_i, \Sigma_i)$, where Σ_i is the inverse of the observed information matrix. Using this assumption, the full conditional of β_i is a multivariate normal distribution, with mean equal to a weighted combination of the MLE and the current value of the prior mean. Formally, $\beta_i|\gamma,\ D, y \sim \mathrm{N}(\beta_i^*, V_i)$ where $\beta_i^* = (D + \Sigma_i)^{-1}(D\hat{\beta}_i + \Sigma_i W_i\gamma)$ and $V_i = (D^{-1} + \Sigma_i^{-1})^{-1}$. To ensure that we sampled from the correct conditional distribution, we included a Metropolis step with $a(\beta_i^{(0)}, \vec{\beta}_i) = \min[(1, \pi(\vec{\beta}_i)/(\hat{\pi}(\vec{\beta}_i))/(\pi(\beta_i^{(0)})/\hat{\pi}(\beta_i^{(0)}))]$, where π denotes the true conditional distribution of the β_i, $\hat{\pi}$ denotes the approximate full conditional distribution of the β_i, $\beta_i^{(0)}$ denotes the value of β_i at the previous iteration, $\vec{\beta}_i$ the candidate value, and $a(\cdot)$ the probability of accepting $\vec{\beta}_i$.[37] Since the MLE of β_i was computed conditionally on the current values of α, we used one Newton step from the MLE of the previous iteration to find the MLE for the current iteration as follows: $\hat{\beta}_i^{\mathrm{new}} = \hat{\beta}_i^{\mathrm{old}} - H^{-1}(\hat{\beta}_i^{\mathrm{old}})g(\hat{\beta}_i^{\mathrm{old}})$, where H denotes the Hessian matrix and g the gradient vector evaluated at the MLE from the previous iteration, $\hat{\beta}_i^{\mathrm{old}}$. We used only one Newton step to avoid slowing down the sampler; with a good initial value, this should be quite accurate. As we have a large sample, a normal approximation to the MLE at the first stage is likely a good approximation.

To sample from the fixed effect parameters, α, we took a similar approach. In particular, to find the conditional MLE of α, we used one Newton step from the previous MLE conditional on the current values of β_i. In addition, as recommended in Zeger and Karim[38] for mixed models, we updated β_i and D several times for each update of α. This updating also accelerated the computations.

We had four strings of the sampler in the final run and used the covariates centred at their population means. The sampler appeared to have converged after about 100 iterations for each of the four strings. We then continued the sampler to 1600 iterations for each string and we sampled every 10th point from each string to minimize autocorrelation. The point estimates reported below are based on posterior sample means, computed from the posterior distribution samples generated by the Gibbs sampler.

5.2. Effects of age and comorbidity

In the final model, we entered age and comorbidity as fixed effects (α). However, we allowed the effects of gender and race to vary across states (β_i). Patient age was a significant predictor of procedure utilization in this cohort. The odds of each procedure relative to no procedure

Table III. Procedure rates* by gender and race

	Female/non-black	Male/non-black	Female/black	Male/black
Angiography only	0·10	0·12	0·09	0·11
CABG	0·06	0·09	0·03	0·05
PTCA	0·04	0·05	0·03	0·03

* Estimates are the median of state posterior means

(p_{ijk}/p_{ij1}, for $k = 2, 3, 4$) decreased with age and the probability or rate of no procedure (p_{ij1}) increased with age (with the other covariates fixed). Focusing on non-black male patients with average comorbidity, as an example, the medians of p_{ij2}, the state-specific rates for only angiography, for patients of age 65, 75, and 85 were 0·23, 0·14, and 0·02, respectively; the corresponding medians for CABG were 0·17, 0·11, and 0·00, and for PTCA, 0·13, 0·06 and 0·01 (not shown in tables or figures). Patients with more comorbid conditions had an increased likelihood of undergoing a procedure.

5.3. Effects of gender and race

Overall, the adjusted procedure rates for blacks were lower than those for non-blacks and the adjusted rates for females were lower than those for males (Table III). Here and elsewhere in this paper, an adjusted procedure rate is the predicted probability of a procedure for a patient with average covariate values. For all these procedures, black females had the lowest rates and white males the highest.

The coefficients for race and for gender (β_i) were quite variable across states (Figure 1, Table IV). For example, state specific black to non-black odds ratios for CABG versus no procedure ($\exp\{\beta_{i33}\}$) ranged from 0·57 to 1·17 with a median of 0·82. Similarly constructed odds ratios for PTCA versus only angiography ($\exp\{\beta_{i43} - \beta_{i23}\}$) ranged from 0·23 to 1·41 with a median of 0·65 (Table IV). Racial differences for PTCA versus no procedure are shown in Figure 1. Regionally, the South appears to have the lowest odds ratios and hence the largest effect of race. Plots of the other four odds ratios of interest showed similar patterns. Note that in Table IV and Figure 1, we report the exponential of the posterior means of the log odds ratio as opposed to the posterior means of the odds ratios. The posterior distribution of the odds ratio is often quite skewed and the posterior mean does not give an accurate measure of the centre of the distribution while the posteriors of the log odds ratios are usually quite symmetric.

5.4. Geographic variation in adjusted rates

Table V and Figure 2 summarize the estimates of adjusted procedure rates by state, that is estimates of the probabilities of each procedure for a patient with covariate values set at the national average. The lowest rates for all three procedures occurred in Northeastern states (Rhode Island and New York), and overall regionally, the Northeast tended to have lower rates than the rest of the country for all three procedures. The three highest rates occurred in the South (Alabama) for only angiography and CABG and in the West (Montana) for PTCA. High on-site availability of angiography was associated with increased procedure utilization as indicated by the significant positive coefficients for angiography availability, associated with the state-specific intercepts (relevant components of γ, not shown in the tables). However, availability was not

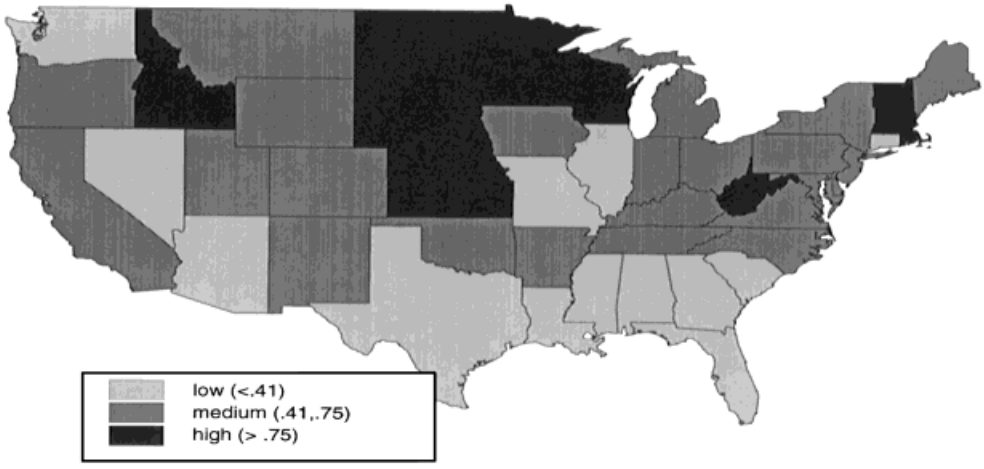

Figure 1. Black/non-black odds ratio estimates for PTCA versus no procedure

Table IV. Black/non-black odds ratio estimates

	Only angiography/ none	CABG/none	PTCA/none	CABG/only angiography	PTCA/only angiography
High	1·17 (Indiana)	0·74 (Indiana)	1·00 (Alaska)	0·87 (Delaware)	1·41 (Alaska)
Low	0·57 (Mississippi)	0·28 (Mississippi)	0·22 (Nevada)	0·32 (Connecticut)	0·23 (Nevada)
Median	0·82	0·44	0·53	0·55	0·65

Table V. Estimated procedure rates for an average patient

Rates	Angiography	CABG	PTCA
High	0·160 (Alabama)	0·101 (Alabama)	0·130 (Montana)
Low	0·063 (Rhode Island)	0·046 (Rhode Island)	0·019 (New York)
Median	0·109	0·071	0·046

significantly related to gender or race differential. This was indicated by the fact that the 95 per cent credible intervals for the relevant components of γ, contained 0.

5.5. Model checking

The p-value for the discrepancy statistic, $C^{(1)}$, was 0·14, indicating that the variation of the observed data does not lie too far in either tail of the distribution of the predicted variation and implies an adequate fit to the data.

To determine whether fit was relatively worse in some states, we computed the goodness-of-fit statistic for each of the states in our model and 'normalized' it by dividing it by the number of patients in the state. Figure 3 shows the 'normalized' values of the statistic for the individual

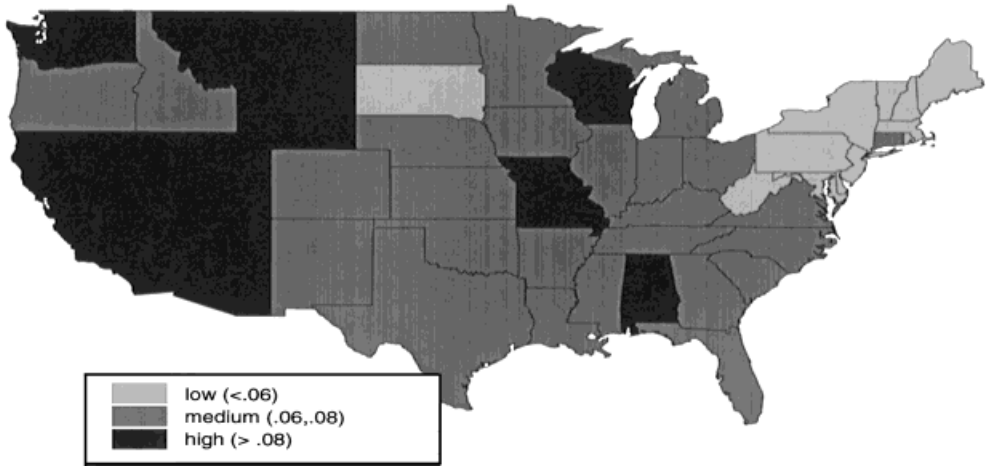

Figure 2. 1987 estimated CABG rates for an average patient

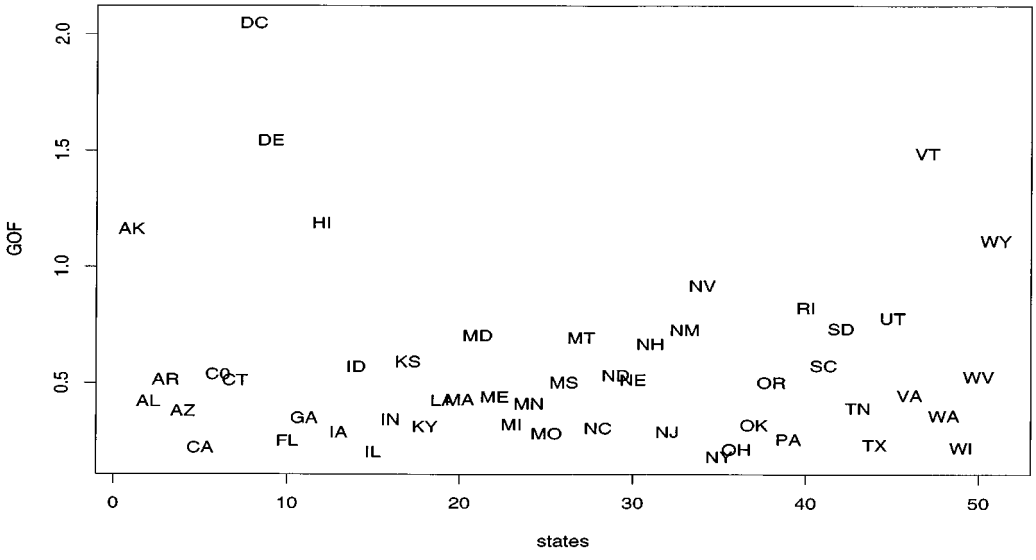

Figure 3. Goodness-of-fit statistic by state

states. Three states, the District of Columbia, Delaware and Vermont, had somewhat large statistics relative to the other states.

6. COMPARISON WITH ALTERNATIVE MODELS

In this section we discuss three alternative models for analysing our data and make comparisons to our approach. The first alternative is a fixed effects polytomous model for the entire country

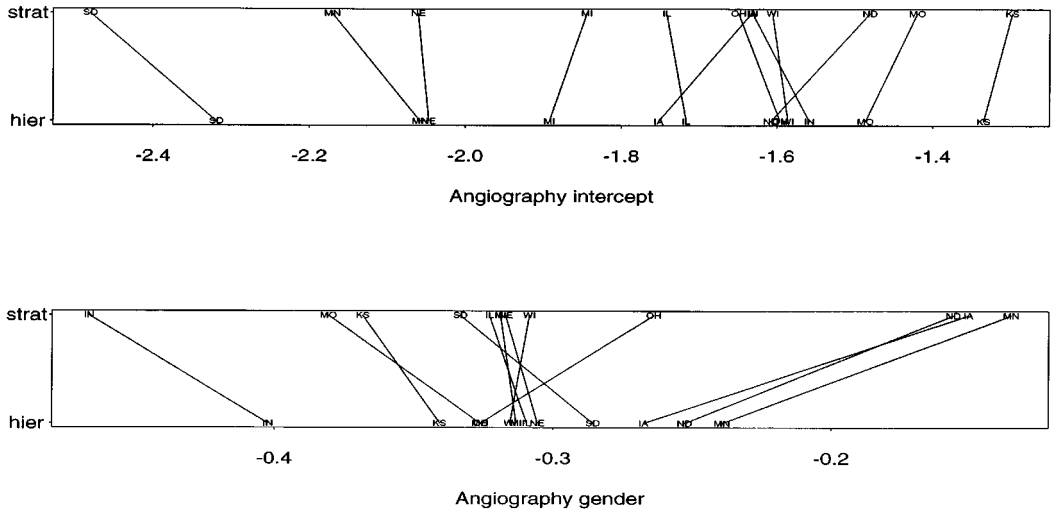

Figure 4. Comparison of estimates of the angiography intercept and gender coefficient derived by fitting a polytomous regression within states (strat) and the hierarchical model (hier)

with dummy variables for each state and interaction effects. This analysis, however, does not account for clustering effects, does not allow state-level covariates to be included and does not borrow strength for estimation of interaction effects for states with small sample sizes.

A second alternative is to fit separate polytomous regression models within each state. A key problem with this stratified approach is that estimates of state specific coefficients can be very imprecise or even impossible to obtain. For example, in the case of our data, we would be unable to estimate at least one of the coefficients in 30 states. In addition, standard errors of the estimates from this approach would be larger than those of our model due to the absence of pooling information across clusters. Figure 4 shows the shrinkage evident in our model versus this model. Note that when covariate effects are well estimated, such as the angiography intercept term, there will be little shrinkage. However, when covariates effects are less well estimated, substantial shrinkage can occur. This will be the case in states with small overall sample sizes, as shown in Figure 4.

A third alternative approach is to use a standard mixed model for polytomous data. The main problem with this model is the failure to account for uncertainty of the covariance matrix in the state-specific effects (variances will be underestimated) and potential biases from the approximation method.[39] This type of model can be fit using the multilevel models package, *MLn*.[20]

7. DISCUSSION

In this paper, we proposed a hierarchical polytomous regression model and discussed approaches for estimation and checking goodness-of-fit. We illustrated the approach by analysing data on the utilization of cardiac procedures on elderly Medicare patients, with emphasis on the heterogeneity in racial and gender effects across states. There appears to be considerable heterogeneity in the effects of race and gender on all three procedures, with black females having the lowest

procedure utilization. This heterogeneity demonstrates the significant effect of between-state variability in addition to the within-state variability accounted for by the polytomous regressions within states. Regionally, Northeastern states tend to have lower adjusted rates for all three procedures.

The posterior simulation approach for model fit can be very computer intensive. We are currently exploring approximations similar to Kass and Steffey[40] that allow for estimation of the variability of the effect parameters taking into account the variation of the scale matrix D. In general, one could use this model most efficiently with data sets of less than about 5000 observations. For models of that size, one can easily and quickly obtain additional model checks by cross-validation and model comparisons with Bayes factors as discussed earlier.

In our model, we assumed a common scale matrix D over all clusters. We can extend the model to allow the scale matrix D to depend on covariates. For example, model fit might improve by allowing a different D matrix for each region. The model-fitting framework described in the paper could easily handle this adjustment as could an expansion of the model through the addition of more levels, for example, patients within hospitals within states.

The programs to fit the hierarchical polytomous model were written in FORTRAN and are currently being streamlined for interested users. In addition, one could use the newest version of BUGS,[41] which allows for multivariate distributions, to fit the hierarchical polytomous model, although for very large data sets, such as the AMI data, the FORTRAN program would probably result in quicker computations.

APPENDIX

Gibbs sampler

The full conditional distributions of the parameters are as follows. Note that we decomposed the multivariate t-distribution on the β_i $[\beta_i|\gamma, D \sim t_v(W_i\gamma, D)]$ into a gamma mixture of normals $[\beta_i|\gamma, D, \tau_i \sim \mathrm{N}(W_i\gamma, D/\tau_i), \tau_i \sim \mathrm{Gamma}\ (v/2, v/2)]$, introducing the additional parameters, τ_i.[16]

$$\gamma|\beta_i, \tau_i, D \sim \mathrm{N}(\psi, \Omega)$$

where $\Omega^{-1} = \sum_{i=1}^{N} W_i'\tau_i D^{-1} W_i + A_\gamma^{-1}$ and $\psi = \Omega(\sum_{i=1}^{N} W_i\tau_i D^{-1}\beta_i + A_\gamma^{-1}\gamma^*)$,

$$D^{-1}|\beta_i, \tau_i, \gamma \sim W(S^{-1}(\gamma), N+3)$$

where $S^{-1}(\gamma) = B + \sum_{i=1}^{N}(\beta_i - W_i\gamma)\tau_i(\beta_i - W_i\gamma)'$, and

$$\tau_i \sim \mathrm{Gamma}\left(\frac{v + p(K-1)}{2}, \frac{v + (\beta_i - W_i\gamma)' D^{-1}(\beta_i - W_i\gamma)}{2}\right)$$

where v denotes the degrees of freedom of the multivariate-t distribution.

The full conditional density of the cluster specific parameters has the following form:

$$\pi(\beta_i|\tau_i, \gamma, D, y) \propto \left(\prod_{j=1}^{J_i}\prod_{k=2}^{K}\left\{\frac{\exp((x_{ij}\beta_{ik} + z_{ij}\alpha_k))}{1 + \sum_{l=2}^{K}\exp((x_{ij}\beta_{il} + z_{ij}\alpha_k))}\right\}^{y_{ijk}}\right)$$
$$\times \exp\left(-\frac{1}{2}(\beta_i - W_i\gamma)'\tau_i D^{-1}(\beta_i - W_i\gamma)\right).$$

The full conditional density of the fixed effect parameters has the following form:

$$\pi(\alpha|\tau_i, \gamma, D, y) \propto \left(\prod_{i=1}^{N} \prod_{j=1}^{J_i} \prod_{k=2}^{K} \left\{ \frac{\exp((x_{ij}\beta_{ik} + z_{ij}\alpha_k))}{1 + \sum_{l=2}^{K} \exp((x_{ij}\beta_{il} + z_{ij}\alpha_k))} \right\}^{y_{ijk}} \right) \times \exp\left(-\frac{1}{2}(\alpha - \alpha^*) A_\alpha^{-1} (\alpha - \alpha^*)' \right).$$

where $y_{ijk} = I(y_{ij} = k)$.

The full conditional density of the degrees of freedom has the following form:

$$\pi(v|\tau_i, \beta_i, \gamma, D) \propto \left(\prod_{i=1}^{n} \frac{\exp(-\tau_i v/2)\tau_i^{v/2}(v/2)^{v/2}}{\Gamma(v/2)} \right) \times I \quad (3 \leqslant v \leqslant 100).$$

Monitoring convergence

Initial values for the sampler can be obtained by fitting polytomous regression models to the data at appropriate levels of aggregation. In particular, we can obtain initial values for β_i, the random coefficients, by fitting individual polytomous regressions within clusters using all the covariates; we can obtain initial values for α, the vector of fixed coefficients, by fitting a polytomous regression to the entire data set.

To determine whether the sampler has converged, we can use a multivariate version of the statistic of Gelman and Rubin[42] which requires the running of m parallel strings of the sampler from overdispersed and underdispersed starting values. In addition, plots of the average of each parameter over parallel strings and the running average can be helpful as well. The combination of these plots and statistics, along with good initial values, should suffice for determining convergence of the sampler.[43]

ACKNOWLEDGEMENTS

This work was supported in part by a Howard Hughes Medical Institute Predoctoral Fellowship. The authors would like to thank the Department of Health Care Policy at Harvard University for the use of computing facilities and for access to the data sets used in this paper.

REFERENCES

1. Diehr, P. 'Small area statistics: Large statistical problems', *American Journal of Public Health*, **74**, 313–314 (1984).
2. Diehr, P., Cain, K., Connell, F. and Volinn, E. 'What is too much variation? The null hypothesis in small area analysis', *Health Services Research*, **24**, 741–771 (1990).
3. Wennberg, J. and Gittelsohn, A. 'Variations in medical care among small areas', *Scientific American*, **246**, 120–134 (1982).
4. Paul-Shaheen, P., Clark, J. D. and Williams, D. 'Small area analysis: A review and analysis of the North American Literature', *Journal of Health Politics, Policy and Law*, **14**, 741–809 (1987).
5. Gatsonis, C. A., Normand, S. L., Liu, C. and Morris, C. 'Geographic variation of procedure utilization: A Hierarchical Model Approach', *Medical Care*, **31**, YS54–YS59 (1993).
6. Gatsonis, C. A., Epstein, A. M., Newhouse, J. P., Normand, S. L. and McNeil, B. J. 'Variations in the utilization of coronary angiography for elderly patients with an acute myocardial infarction: an analysis using hierarchical logistic regression', *Medical Care*, **33**, 625–642 (1995).
7. Udvarhelyi, I. S., Gatsonis, C., Epstein, A. M., Pashos, C. L., Newhouse, J. P. and McNeil, B. J. 'Acute myocardial infarction in the Medicare population: process of care and clinical outcomes', *Journal of the American Medical Association*, **268**, 2530–2536 (1992).

8. Bryk, A. S. and Raudenbush, S. W. *Hierarchical Linear Models*: *Applications and Data Analysis Methods*, Sage Publishing, Inc, Newbury Park, 1992.
9. Breslow, N. and Clayton, D. 'Approximate inference in generalized linear models', *Journal of the American Statistical Association*, **88**, 9–25 (1993).
10. Calvin, J. A. and Sedransk, J. 'Bayesian and frequentist predictive inference for the patterns of care studies', *Journal of the American Statistical Association*, **86**, 36–48 (1991).
11. Malec, D. and Sedransk, J. 'Small area inference for binary variables in the National Health Interview Study', *Journal of the American Statistical Association*, (in press), (1997).
12. Normand, S. L., Glickman, M. and Gatsonis, C. 'Statistical methods for profiling providers: issues and applications', Department of Health Care Policy, Harvard Medical School Technical Report, 1994.
13. Wong, G. and Mason, W. 'The hierarchical logistic regression model for multilevel analysis', *Journal of the American Statistical Association*, **80**, 513–524 (1985).
14. Wong, G. and Mason, W. 'Contextually specific effects and other generalization of the hierarchical linear model for comparative analysis'. *Journal of the American Statistical Association*, **86**, 487–503 (1991).
15. Gelfand, A. E. and Smith A. F. M. 'Sampling-based approaches to calculating marginal densities', *Journal of the American Statistical Association*, **85**, 398–409 (1990).
16. Albert, J. H. and Chib, S. 'Bayesian analysis of binary and polychotomous response data', *Journal of the American Statistical Association*, **88**, 669–679 (1993).
17. Gibbons, R. D., Hedeker, D., Charles, S. C. and Frisch, P. 'A random-effects probit model for predicting medical malpractice claims', *Journal of the American Statistical Association*, **89**, 760–767 (1994).
18. Allenby, G. M. and Lenk, P. J. 'Modeling household purchase behavior with logistic normal regression', *Journal of the American Statistical Association*, **89**, 1218–1231 (1994).
19. McCulloch, R. and Rossi, P. E. 'An exact likelihood analysis of the multinomial probit model', *Journal of Econometrics*, **64**, 207–240 (1994).
20. Goldstein, H. *Multilevel Statistical Models*, Edward Arnold, London, 1995.
21. Natarajan, R. 'Estimation in models for multinomial response data: Bayesian and frequentist approaches', Doctoral Thesis, Cornell University, 1995.
22. Mor, V., Intrator, O., Fries, B. E., Phillips, C., Teno, M. D., Hiris, J., Hawes, C. and Morris, J. 'Changes in hospitalization associated with introducing the resident assessment instrument', *Journal of the Geriatric Society*, to appear.
23. Kass, R. E. and Wasserman, L. 'A Reference Bayesian test for Nested Hypotheses and its relationship to the Schwarz Criterion', *Journal of the American Statistical Association*, **90**, 928–934 (1995).
24. Ayanian, J. and Epstein, A. 'Differences in the use of procedures between women and men hospitalized for coronary heart disease', *New England Journal of Medicine*, **325**, 221–225 (1991).
25. Steingart, R. M., Packer, M., Hamm, P., *et al*. 'Sex differences in the management of coronary artery disease', *New England Journal of Medicine*, **325**, 226–230 (1991).
26. Wenneker, M. B. and Epstein, A. M. 'Racial inequalities in patients with ischemic heart disease in Massachusetts', *Journal of the American Medical Association*, **261**, 253–257 (1989).
27. Whittle, J., Conigliaro, J., Good, C. B. and Lofgren, R. P. 'Racial differences in the use of invasive cardiovascular procedures in the department of Veteran Affairs Medical System', *New England Journal of Medicine*, **329**, 621 (1993).
28. Goldberg, K. C., Hartz, A. J., Jacobsen, S. J., Krakauer, H. and Rimm, A. A. 'Racial and community factors influencing coronary artery bypass graft surgery rates for all 1986 Medicare patients', *Journal of the American Medical Association*, **267**, 1473–1477 (1992).
29. Every, N. R., Larson, E. B., Litwin, P. E., Maynard, C., Fihn, S. D., Eisenberg M. S., Hallstrom, A. P., Martin, J. S. and Weaver, W. D. 'The association between on-site cardiac catheterization facilities and the use of coronary angiography after acute myocardial infarction', *New England Journal of Medicine*, **329**, 546–551 (1993).
30. Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. *Bayesian Data Analysis*, Chapman and Hall, London, 1995.
31. Lesaffre, E. and Albert, A. 'Multiple-group logistic regression diagnostics', *Applied Statistics*, **38**, 425–440 (1989).
32. Verdinelli, I. and Wasserman, L. 'Computing Bayes factors using a generalization of the Savage-Dickey density ratio', *Journal of the American Statistical Association*, **90**, 614–618 (1995).

33. Daniels, M. J. 'Hierarchical regression models with applications', Doctoral Thesis, Harvard School of Public Health, 1995.
34. Hastings, W. K. 'Monte Carlo Sampling methods using Markov chains and their applications', *Biometrika*, **57**, 97–109 (1970).
35. Smith, A. F. M. and Roberts, G. O. 'Bayesian computations via the Gibbs sampler and related Markov Chain Monte Carlo Methods', *Journal of the Royal Statistical Society B*, **55**, 3–23 (1993).
36. Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. *Markov Chain Monte Carlo in Practice*, Chapman and Hall, London, 1996.
37. Tierney, L. 'Markov chains for exploring posterior distributions', *Annals of Statistics*, **22**, 1701–1762 (1994).
38. Zeger, S. and Karim, M. 'Generalized linear models with random effects: a Gibbs sampling approach', *Journal of the American Statistical Association*, **86**, 79–86 (1991).
39. Breslow, N. E. and Lin, X. 'Bias correction in generalised linear mixed models with a single component of dispersion', *Biometrika*, **82**, 81–91 (1995).
40. Kass, R. E. and Steffey, D. 'Approximate Bayesian inference for conditionally independent hierarchical models', *Journal of the American Statistical Association*, **84**, 717–726 (1989).
41. Thomas, A., Spiegelhalter, D. J. and Wilks, W. R. 'BUGS: A program to perform Bayesian inference using Gibbs sampling', *in* Bernardo, J. M., Berger, J. O., Dawid, A. P. and Smith, A. F. M. (eds), *Bayesian Statistics 4*, Clarendon Press, Oxford, 1992, pp. 837–842.
42. Gelman, A. and Rubin, D. B. 'Inference from iterative simulation using multiple sequences', *Statistical Science*, **7**, 457–511 (1992).
43. Cowles, M. K. and Carlin, B. P. 'Markov Chain Monte Carlo convergence diagnostics: a comparative review', *Journal of the American Statistical Association*, **91**, 883–904 (1996).

