Novel Bayesian Methods for Disease Mapping: An Application to Chronic Obstructive Pulmonary Disease

by

Jie Liu

A Thesis

Submitted to the Faculty

of

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the

Degree of Master of Science

in

Applied Statistics

by

May 2002

APPROVED:

Dr. Balgobin Nandram, Thesis Advisor

Dr. Homer Walker, Department Head

Abstract

Mapping of mortality rates has been a valuable public health tool. We describe novel Bayesian methods for constructing maps which do not depend on a post stratification of the estimated rates. We also construct posterior modal maps rather than posterior mean maps. Our methods are illustrated using mortality data from chronic obstructive pulmonary diseases (COPD) in the continental United States.

Poisson regression models have attracted much attention in the scientific community for their superiority in modeling rare events (including mortality counts from COPD). Christiansen and Morris (JASA 1997) described a hierarchical Bayesian model for heterogeneous Poisson counts under the exchangeability assumption. We extend this model to include latent classes (groups of similar Poisson rates unknown to an investigator).

Also, it is standard practice to construct maps using quantiles (e.g., quintiles) of the estimated mortality rates. For example, based on quintiles, the mortality rates are cut into 5 equal size groups, each containing 20% of the data, and a different color is applied to each of them on the map. A potential problem is that, this method assumes an equal number of data in each group, but this is often not the case. The latent class model produces a method to construct maps without using quantiles, providing a more natural representation of the colors.

Typically, for rare events, the posterior densities of the rates are skewed, making the posterior mean map inappropriate and inaccurate. Thus, although it is standard practice to present the posterior mean maps, we also develop a method to provide the joint posterior modal map (i.e., the map with the highest posterior probability over the ensemble).

For the COPD data, collected 1988-1992 over 798 health service areas, we use Markov chain Monte Carlo methods to fit the model, and an output analysis is used to construct the new maps.

Acknowledgments

First of all, gratitude beyond words is due to my advisor, Dr. Balgobin Nandram, for his great help, guidance, patience and understanding. It is an exciting and rewarding experience to work with such a knowledgeable and kind professor, to catch up with his endless ideas and creativity. Professor Nandram also provided funding for my trips to the National Center of Health Statistics (NCHS) and financial support Summer 2001.

Thanks are due to Dr. Jai W. Choi of the NCHS for his encouragement and assistance throughout the whole project. I am also grateful to Jimmie Givens for discussions about disease mapping and Linda Pickle for providing the data.

Thanks also to my classmates, Mike Johnson and Tatiana Doytchinova, for kindly providing the latex template.

I presented a talk on this thesis in the Department of Mathematical Sciences, Worcester Polytechnic Institute, the sixteenth New England Statistics Symposium at Yale University and the National Center for Health Statistics, Hyattsville, Maryland. I am grateful for their comments also.

A special thank you to my parents, for their forever love and support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Recently, there has been increased interest in estimating mortality rates for small geographical areas. Mapping small area death rates is a valuable public health tool, which may be used to generate etiologic hypotheses and identify high rate areas where intervention or treatment programs may be profitable. Also it has always been of interest to know how and where to allocate limited resources, especially for local and federal government. For example, if we know a particular disease occurrs in some areas more often than in others, we might want to provide better medical facilities and services in these areas. Furthermore, if we can find some potential risk factors which show a statistically significant relation with the occurrence of a disease, we might be able to implement some prevention program much more efficiently.

In this chapter we mainly discuss our motivation in conducting this study, Chronic Obstructive Pulmonary Disease (COPD) and its potential risk factors, the recent related research works and their main contributions, and finally, a plan of this thesis.

1.1 Statement of the Problems

Pickle, Mungiole, Jones, and White (1996) have presented the maps of the leading causes of death in the United States for 1988-1992. Our study is on one of the disease categories studied in this recent Atlas, i.e., chronic obstructive pulmonary disease, International Classification of Diseases 490-496 (World Health Organization 1977), for white males that includes asthma, bronchitis, and emphysema. Recently, the mortality rate of COPD has risen dramatically, making COPD the fourth leading cause of death in the United States, with rates highest for white males.

Table 3.1 and table 1.2 show some recent statistics of COPD for the United States, Figures are from "Monthly Vital Statistics Report", Vol. 49, No. 8 and "Vital and Health Statistics" Series 10, No. 200, the National Center of Health Statistics.

Table 1.1: COPD deaths count

Characteristics	**Statistics**	**Year reported**
Deaths Annually	124,181	1999
Age-Adjusted Death Rate	45.8 deaths/100,000 people	1999
Cause of Death Rank	4	1999

Table 1.2: COPD Cases report

Type of Disease	**Statistics**	**Year reported**
Cases of Bronchitis Reported Annually	14 .2 million	1996
Cases of Emphysema Reported Annually	2 million	1996
Cases of Asthma Reported Annually	14.5 million	1996

So far a lot of work has been done on the analysis of COPD data, among which the mapping of small area mortality rates turned out to be a highly effective tool in such analyses. Figure 1.1 is a map of observed COPD mortality rates for white males age 65 +, map units are the 798 HSA's (health service area).

However, we found two questionable points in the standard practice of disease mapping:

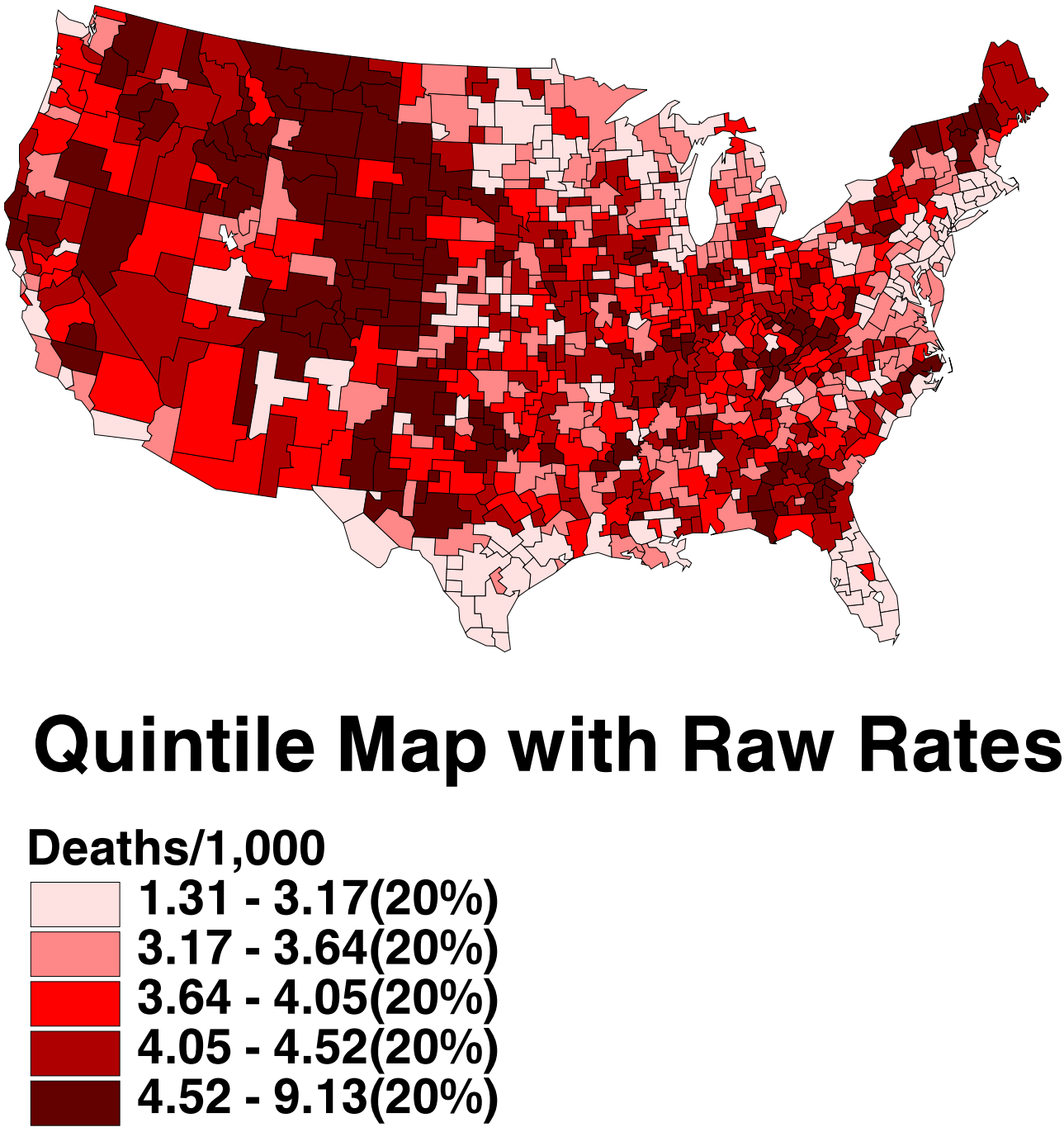

Figure 1.1: Map using observed rates for age 65+

(1) Conventionally, means of the estimated rates are presented on the map.

(2) Colors are assigned by a post-grouping after the rates are obtained.

In a lot of cases, presenting estimated means of the rates is a natural and effective practice. However, in the analysis on rare events (e.g., mortality rates of COPD), this often turns out to be misleading. Because the distributions of such rare events are usually highly skewed. This can produce a large difference between the means and the modes, presenting means will fail to give us accurate information.

Also, when maps are constructed using quantiles, we cut the rates into groups of equal size, each containing the same number of areas, and a different color is applied to each group to represent the group difference. This method assume an equal distribution of the rates among groups, whenever this assumption fail to be satisfied, the map becomes misleading.

In this thesis, we try to come up with better methods which avoid the problems mentioned above, and to provide more accurate analysis on dispersed data, especially, for the analysis on COPD mortality rates.

Christiansen and Morris (1997) describe a hierarchical Bayesian model for heterogeneous Poisson counts under the exchangeability assumption. We extend this model to include latent classes (groups of similar Poisson rates unknown to an investigator). The inclusion of latent classes is to better model the heterogeneous subpopulation (i.e., classes) structure in some data (as seen in COPD data), thus, instead of using quantiles, produce a method to construct maps based on the latent classes, providing a more natural representation of the colors.

We also develop a method to provide the joint posterior modal map, i.e., the map with the highest posterior probability over the ensemble. For data with a highly

skewed distribution (Gamma in our case), the mean is sometimes far away from the mode, becoming a less informative parameter, whereas the mode always stands for the point where the highest probability density lies. In such cases, mapping posterior modes provide us more information about rates than posterior means do.

1.2 Description of the Data

The dataset comes from the National Center of Health Statistics (NCHS). It contains the deaths number, population, and a set of potential explanatory covariates for the 798 HSAs in the contiguous 48 states. Data are collected during 1988-1992 with 10 age groups being identified. In our study, we focus on the age groups which contain age 65+ (65 and older). This group is of particular interest because COPD occurred much more often, and it is justifiable to investigate causes of death of our retirees.

In our study we also tried to link the mortality rate to the potential explanatory variables (covariates). These covariates include smoking history, population density, elevation, annual rainfall level, summer rainfall level, average income level and college student ratio.

Previous studies show that:

- In general, hospitalization for respiratory disease is greater among smokers and persons with lower income and education levels (Morris and Munasinghe 1994).
- In addition, hospitalization for childhood asthma is also related to urbanization and second-hand smoke (Burr, Anderson, Austin, Harkins, Kaur, Strachan, and Warner 1999; McConnochie, Russo, McBride, Szilagyi, Brooks, and Roghmann

1999). These factors may serve as a surrogate for high levels of indoor pollutants in poor housing (Bates 1989).

- Extreme climatic conditions are known to aggravate existing asthma and bronchitis (Bates 1989), as is living at high altitudes because of the reduced oxygen supply (Schoene 1999).

- Repeated exposure to particulate matter and other air pollutants, primarily from traffic exhaust and coal-burning power plants, can aggravate existing lung conditions and can even cause death among the severely ill (English, Neutra, Scalf, Sullivan, Waller, and Zhu 1999; Sunyer, Schwartz, Tobias, Macfarlane, Garcia, and Anto 2000).

- In particular, small airborne particles such as SO_2 found in urban air pollution can be deposited deep in the lungs, causing severe pulmonary effects (Sunyer et al. 2000; Schwartz and Neas 2000).

- Aerosolized toxins and viruses can be inhaled in dusty environments, causing pulmonary effects (Centers for Disease Control 1998).

Because data are not directly available at the county or HSA level for all of these risk factors, we used several proxy variables in the analysis.

Lung cancer mortality rates among white men during 1998-1992 are used as a surrogate for smoking rates (Pickle et al. 1996).

We include rainfall as a climatic factor that may affect lung function at the extremes (both very dry and very wet conditions). For each recording station (a particular place where measurement is taken), the measure of annual rainfall (AR) is the median rainfall over 1973-1992. The measure of AR for an HSA is the median of the values for all stations within the HSA.

Population density is included as a possible proxy for urban pollution and older, more crowded housing. Elevation is available directly.

The EPA reports actual measurements and predicted levels of airborne pollutants, including SO_2 and total suspended particulates, the most relevant for COPD (Environmental Protection Agency 1997; Nizich, Pope, and the Pechan-Avanti Group 1998). However, not every HSA has a monitoring station, and the predicted levels are of questionable quality at the local level. Thus, we did not include these environmental variables in our model.

Finally, nine potential risk factors are included in our study, a list of these nine covariates together with their explanations is given in table 1.3:

Table 1.3: Covariates explanation

Covariates	**Explanations**
wmlung	**white male lung cancer mortality rate a surrogate of smoking history**
sqrtpopd	**square root of population density**
sqrtelev	**square root of elevation**
pcinc	**income per capital**
pctcoll	**college students percentage**
pctpov	**below poverty line percentage**
srain	**summer rainfall level**
arain	**annul rainfall level**
rainpc	**per capital rainfall level**

We tried to regress the mortality rate on these covariates. First we looked at the correlation matrix of the nine covariates, and found there is no serious multicollinearity problem. We also tried include some interactions in the model, but they all turned out to be non-significant, so finally we fit a linear regression model without any interaction, with the normal assumptions basically satisfied.

The fitted results suggest that, among these nine covariates, only four of them are statistically significant. The results are given in table 1.4

Table 1.4: Parameter Estimates of Regression Model (SAS output)

Variable	DF	Parameter Estimate	Standard Error	t Value	P-value
Intercept	1	-5.92840	0.10677	-55.53	$< .0001$
wmlung	1	0.00803	0.00065697	12.22	$< .0001$
sqrtpopd	1	-0.00372	0.00103	-3.61	0.0003
sqrtelev	1	0.00443	0.00060025	7.38	$< .0001$
pcinc	1	-0.00000499	0.00000526	-0.95	0.3436
pctcoll	1	0.00238	0.00562	0.42	0.6715
pctpov	1	-0.00521	0.00363	-1.44	0.1512
srain	1	-0.00996	0.00698	-1.43	0.1539
arain	1	-0.00234	0.00088730	-2.64	0.0084
rainpc	1	0.17117	0.11668	1.47	0.1428

From table 1.4 it is easy to see only wmlung (White male lung cancer rate), sqrtpopd (square root of Population density), sqrtelev (square root of elevation) and arain (Annual rainfall level) are significant at a 5% significant level. So we include only these four covariates as regressors in our linear regression model, in which mortality rate is the response variable. The test statistics by fitting this model are given in table 1.5, and a residual plot is in Figure 1.2.

Table 1.5: Parameter Estimates of Regression Model (SAS output)

Variable	DF	Parameter Estimate	Standard Error	t Value	P-value
Intercept	1	-6.03472	0.04811	-125.44	$< .0001$
wmlung	1	0.00788	0.00060869	12.94	$< .0001$
sqrtpopd	1	-0.00435	0.00082186	-5.29	$< .0001$
sqrtelev	1	0.00468	0.00057727	8.11	$< .0001$
arain	1	-0.00271	0.00074502	-3.64	0.0003

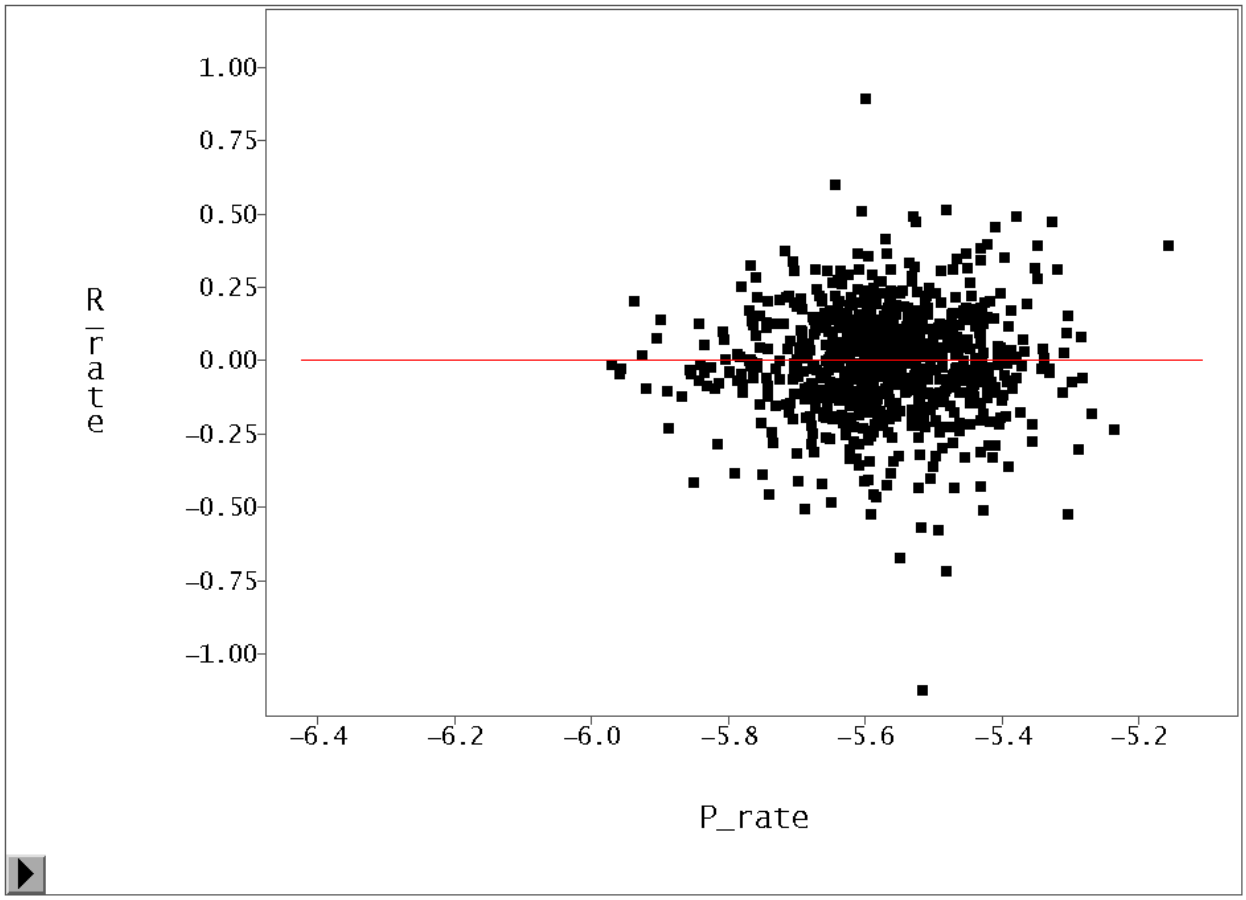

Figure 1.2: A plot of residual versus predicted response

The residual plot Figure 1.2 shows a problem of heteroscadesticity, suggestting that it maybe better to fit a weighted least square model.

This result is also consistent with the work done by Nandram, Sedransk and Pickle (2000), who used a sophisticated hierarchical Bayesian model. Their results are in Table 1.6

Table 1.6: Posterior Means, Standard Deviations and 95% Credible Intervals for the Regression Coefficients for the Covariates (age group 55+)

Variable	Post mean	Post S.T.D.	95% Credible Interval
Covariates for old age classes			
White male lung cancer rate	3.73	.30	(3.15, 4.32)
Population density	-5.58	.20	(-6.00, -5.21)
Elevation	.80	.23	(.34, 1.23)
Annual rainfall	-2.21	.40	(-2.93, -1.39)

For computational convenience, we also make a transformation on two of the covariates, in the end the four covariates in our study are:

(a) White male lung cancer rate per 1000 population

(b) Square root of population density/10^4

(c) Square root of elevation/10^4

(d) Annual rainfall/10^2

Figure 1.3 shows the distributions of these four covariates in the country. Later we link these maps to the mortality rate maps to make statistical inference.

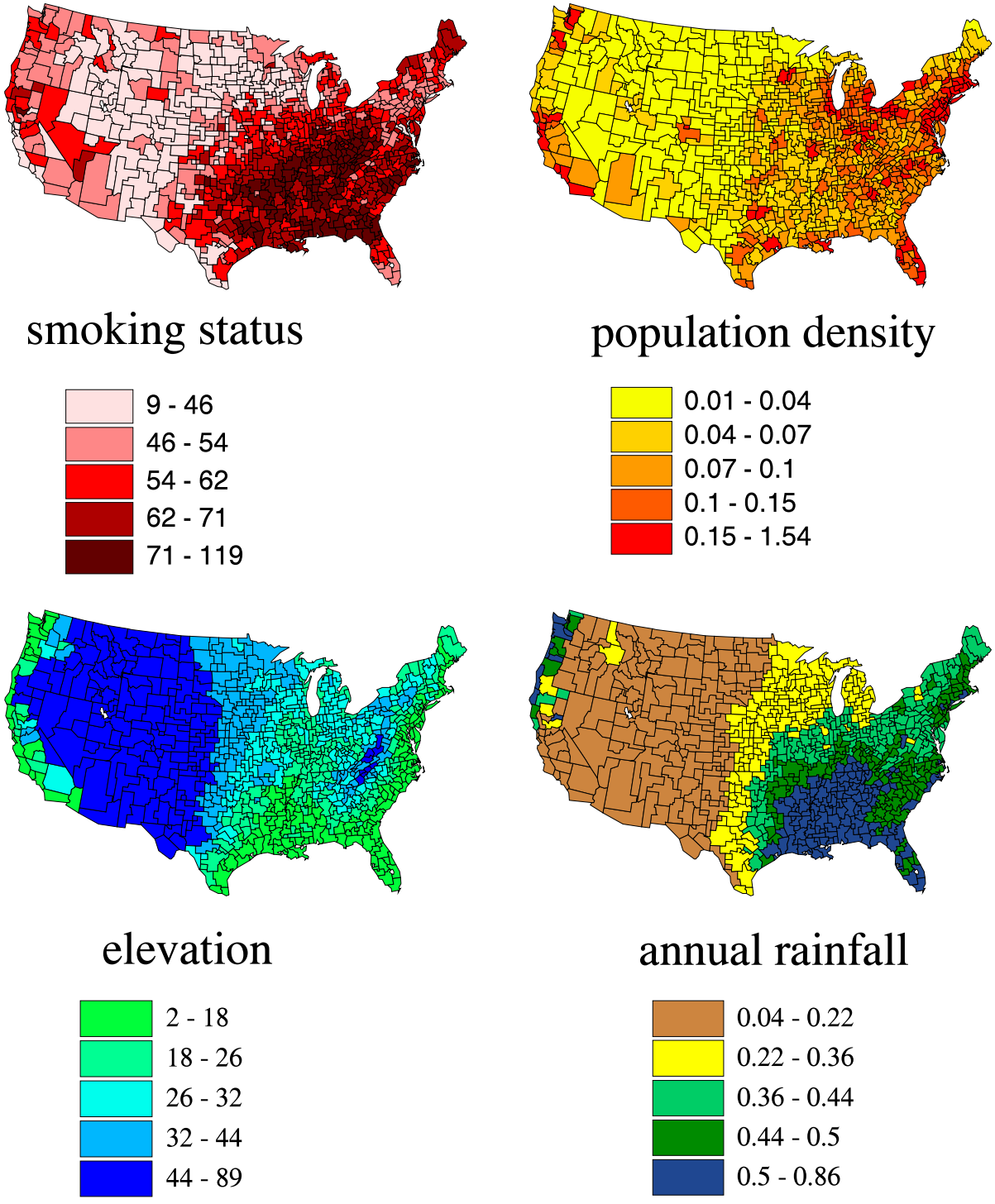

Figure 1.3: Maps For Risk Factors

1.3 Thesis Overview

The goal of this thesis is to propose improved statistical models and methods for analysis of the dispersed COPD mortality rates.

In Chapter 2, we describe the literature. First we describe the Poisson regression model and the latent class model. Then we describe the discrete mixture model, which has been fitted using the EM algorithm.

The theoretical part of this thesis is in Chapter 3, we set up a hierarchical Bayesian latent class model for Poisson regression. Then a Bayesian analysis is conducted to make inference about the mortality rates of the 798 HSAs. The inclusion of regression is to link the mortality rates with the potential risk factors. The latent classes are obtained through a prior density which is a mixture of Gamma densities that include the covariates. A typical problem in a mixture model is the non-identifiability which we eliminate by incorporating an order restriction. We apply Markov chain Monte Carlo method in the computation, and the Metropolis-Hastings sampler is the main algorithm used.

In Chapter 4, we describe how well the latent class model fits and how to choose a reasonable number of classes for the model.

In Chapter 5, in an output analysis, we show how to obtain the new legends for the maps and, how to obtain the posterior modal maps. Our model and these two methods together are our three contributions to disease mapping.

Chapter 6, we make our final remarks based on the data analysis and map comparison. Some future issues are considered.

Chapter 2

Review of Literature

Models and methods of analysis on rate data are abundant, Walker et al. (1996), Christiansen and Morris (1997), Nandram et al. (1999), Nandram (2000), A.F.Militino et al.(2001) and so on. In this chapter, we describe three newly developed models which are important for our work.

2.1 Hierarchical Poisson Regression Model

Christiansen and Morris (1997) proposed a new approach, called Poisson regression interactive multilevel modeling (PRIMM). One of our contributions in the thesis is to extend the Poisson regression model to include latent classes. In this section we give a brief description of their model.

(1) The Descriptive Model

Level 1 of the descriptive model specifies the distribution of the observed data vector, $\mathbf{d} = (d_1, \ldots, d_k)'$, given the individual parameters λ_i. Level 2 specifies gamma distributions for λ_i, given the hyperparameters $\alpha \equiv (\beta, \zeta)$.

*L*evel 1: Individual Model.

The observed counts $(d_1, \ldots, d_k)$ have independent Poisson distributions with expected values $E(d_i) = n_i\lambda_i, i = 1, \ldots, \ell$, for known exposures $n_i > 0$. Given $\lambda = (\lambda_1, \ldots, \lambda_\ell)'$.

$$d_i|\lambda_i \sim \text{Poisson}(n_i\lambda_i), \quad \text{independently } i = 1, \ldots, k. \tag{2.1}$$

The observed rates, $\mathbf{r}$, with $r_i \equiv d_i/n_i$ have expectations $E(r_i) = \lambda_i$.

*L*evel 2: Structural Model.

The individual Poisson parameters $(\lambda_1, \ldots, \lambda_\ell)$ follow conjugate gamma distributions for $i = 1, \ldots, \ell$ independently, given the unknown hyperparameter vector $\alpha = (\beta_0, \ldots, \beta_{k-1}, \zeta)'$, where k is the number of unknown regression coefficients. Thus

$$\lambda_i|\alpha \overset{\text{ind}}{\sim} \text{Gamma}(\zeta, \zeta/\mu_i), \;\; \log(\mu_i) = \mathbf{x_i'}\beta \tag{2.2}$$

The log link (i.e., the natural link for the Poisson distribution) is assumed for the structural means, so $\log(\mu_i) = \mathbf{x_i'}\beta$ for fixed covariates $\mathbf{x_i} = (x_{i0}, \ldots, x_{i,k-1})', \beta = (\beta_0, \ldots, \beta_{k-1})'$, with the intercept term represented by setting $x_{i0} = 1$. Here $\zeta > 0$ corresponds to an unobserved prior count, not necessarily an integer. The squared coefficient of variation of $\lambda_i, 1/\zeta$, is the same for all components in (2.2). The structural model (2.2) allows μ_i to be known ($k = 0$) and allows exchangeable (without covariates, $k = 1$) and nonexchangeable (if $k \geq 2$) distributions for λ_i. In all cases standardized rates, $\lambda_i/\mu_i \sim \Gamma(\zeta, \zeta)$ are exchangeable, with unit expectation.

*L*evel 3: Distribution on the Structural Parameters.

The joint prior density for the $k+1$-dimensional hyperparameter vector is

$$h(\beta, \zeta) = z_0/(\zeta + z_0)^2 \tag{2.3}$$

$$\text{Equivalently,} \quad B_0 \equiv \zeta/(\zeta + z_0) \sim \text{uniform}(0, 1). \tag{2.4}$$

The choice of an improper prior distribution for β is standard, which can provide good repeated sampling properties to the resulting rules. They choose a proper prior distribution like (2.4) for ζ because the maximum likelihood estimate of ζ can occur at infinity (Aragon, Eberly 1992 and Morris 1997). When that happens, maximum likelihood estimates of the shrinkages $B_i = \zeta/(\zeta + n_i\mu_i)$ occur at the boundary point $B_i = 1$, interfering with further assessments and especially with evaluating precision. The uniform distribution for B_0 is relatively uninformative, giving both ζ and $1/\zeta$ infinite expectations. The constant z_0 is the median of ζ, so that small values of z_0 are less informative and encourage less shrinkage a priori. A conservative choice, one that mainly lets the data speak, would choose z_0 small enough so that z_0 is less than the median of ζ.

(2) The Inferential Model

The marginal distributions of the observed data d_i are independent negative binomials, these being gamma mixtures of Poisson distributions. To fix notation, the negative binomial distribution, $NB(\zeta, p) \equiv NB[\mu = \zeta p/(1-p), \mu + \mu^2/\zeta]$, for $\zeta > 0$, $0 < p < 1$ has density function

$$\frac{\Gamma(\zeta + d)}{\Gamma(\zeta)d!} p^d (1-p)^\zeta, \quad d = 0, 1, 2.... \tag{2.5}$$

As used here, (2.5) arises as the density of a mixture of Poisson(Λ) distributions on d if $\Lambda \sim \Gamma(\zeta, 1) \times (p/(1-p))$ is the mixing distribution. For fixed μ, large values of ζ make the negative binomial approximately Poisson.

*L*evel 1: Marginal Model for the Observations.

The observed counts d_i, given the hyperparameters β and ζ, have independent marginal distribution for $i = 1, \ldots, \ell$

$$d_i|\beta, \zeta \sim NB(\zeta, 1 - B_i) \tag{2.6}$$

and

$$B_i \equiv \zeta/(\zeta + n_i\mu_i) \tag{2.7}$$

with $0 < B_i < 1$. For large ζ or, equivalently, for "shrinkage factors" B_i near unity, the negative binomial distribution (1.6) is approximately the Poisson distribution. Smaller values of ζ indicate extra variation beyond the Poisson variation in the true rate λ_i.

*L*evel 2: Conditional Model for the Individual Parameters.

The conditional distributions (1.8) for the unobserved Poisson rate parameters, given (β, ζ) and the data, are gamma distributions with means and variances that are linear in the data.

$$\lambda_i|\text{data}, \beta, \zeta \overset{\text{ind}}{\sim} \text{Gamma}(d_i + \zeta, n_i + \zeta/\mu_i) \tag{2.8}$$

$i = 1, \ldots, \ell$, independently, where the posterior means λ_i^* and variances $(\sigma_i^*)^2$, given the hyperparameters, are denoted by

$$\lambda_i^* \equiv E(\lambda_i|\text{data}, \beta, \zeta) = (1 - B_i)r_i + B_i\mu_i \tag{2.9}$$

and

$$(\sigma_i^*)^2 \equiv \mathrm{Var}(\lambda_i|\mathrm{data}, \beta, \zeta) = \lambda_i^*(1 - B_i)/n_i \tag{2.10}$$

Linearity of the conditional mean (2.9) holds because the gamma distribution is conjugate to the Poisson likelihood. The gamma distribution choice for the structural model (2.2) is conservative (i.e., it provides minimax estimators for mean squared error) among all distributions with the same first two moments for this problem (Morris 1983).

Results from this method show that, the use of Gamma prior density with regression is appropriate, and the whole hierarchical model also works well for analysis of dispersed rates. Compared to the methods earlier, PRIMM has the following advantages. First, PRIMM provides better coverage probabilities and smaller risk than many other published methods. Second, in most cases, the nominal operating characteristics of PRIMM hold up better than the most commonly used alternative methods. Third, PRIMM provides point and interval estimates for both individual and the structural parameters, unlike some of the other methods, so it is widely used for a lot of statistical applications.

2.2 Latent Class Model

One of our major contributions in this thesis is to include the latent classes into the Poisson regression model. Thus, in this section we describe the reasons and advantages to use latent class model.

The structure of many datasets is too complex to be represented by a single parametric model (e.g. Poisson or normal) especially if they exhibit multimodality, outlying subsets of observations, and so on (McLachian and Basford, 1998; Lavine and

West, 1992; Leonard et al., 1994). Nonparametric analysis is one way of circumventing the problems raised by the complexity of observed structures. A Bayesian approach to a completely nonparametric analysis is via Dirichlet process prior. Another method which retains the framework of parametric densities, but approximates the underlying sampling density as a finite mixture of g latent classes is given as follows.

$$d_i|\underset{\sim}{p}, \underset{\sim}{\alpha} \sim \sum_{j=1}^{g} p_j f_j(d_i|\alpha_j),\ i = 1, \ldots, n$$

where $f_j(d_i|\alpha_j)$ is a given parametric density such as the normal, exponential or Poisson. Such a model is appropriate for a heterogeneous population containing subgroups whose sizes are proportional to p_j. The parameters α_j might, for example, relate to means and variances μ_j, V_j of normal subpopulations, or means μ_j of Poisson subgroups. Means can in turn be related to covariates $x_1, \ldots, x_n$ via class-specific regression parameters β_j. For example, for a mixture of Poisson regressions one might have

$$d_i \sim \sum_{j=1}^{g} p_j \text{Poisson}(\mu_{ij})$$
$$\log(\mu_{ij}) = \beta_j z_i.$$

As noted by Dempster et al. (1997) a mixture model can be expressed alternatively in terms of the original data and augmented data, where the latter consists of an indicator variable Z_i of group membership among the g groups, and we assume that Z_i is known. If w_{ij} are dummy indicators equaling one when $Z_i = j$ and zero otherwise, then the likelihood of the complete data (observed and augmented data combined) is

$$\prod_{j=1}^{g} \prod_{i=1}^{n} p_j^{w_{ij}} [f_j(d_i|\alpha_j)]^{w_{ij}}.$$

In practice of course the w_{ij} are unknown. However the complete data representation means that a mixture model can be represented hierarchically, analogous to the random effects models, with at the first level the mixture density parameters α_j, then the (unobserved) class indicators

$$w_{ij} \sim g(w_{ij}|\alpha_j),$$

and finally the observed data

$$d_i \sim f_{Z_i}(d_i|\alpha_j, w_{ij}).$$

The expected posterior probabilities of class membership for individuals, i.e., $\varphi_{ij} = Pr(w_{ij} = 1)$, have the form

$$\varphi_{ij} = p_j f_j(d_i|\alpha_j) / \sum_{k=1}^{g} p_k f_k(d_i|\alpha_k).$$

Posterior means for each case are then weighted averages of the subpopulation means with weights φ_{ij}. So in the Poisson case without covariates the expected posterior mean for case i is $\sum_{j=1}^{g} \varphi_{ij}\mu_j$ and in the regression case is $\sum_{j=1}^{g} \varphi_{ij} e^{\beta_j z_i}$.

The goal in the latent mixture analysis without covariates, is often the same as in random effects approaches to heterogeneity: to provide a smoothed estimate (e.g. of disease morbidity) which results from "borrowing strength" over the ensemble (Bohning, 2000). In the regression case, the goal might be to identify subpopulations with different economic choice patterns or health behaviors.

2.3 Discrete Poisson Mixture Models

Conventional approaches for estimating rates in disease mapping or mortality studies are based on Poisson regression. Frequently, overdispersion is present and this

extra variability is modeled by introducing random effects. A.F.Militino, M.D.Ugarte and C.B.Dean (2001) introduced the use of discrete Poisson mixture models for identifying high rates, assuming the population arises from a discrete mixture of Poisson distributions.

A.F.Militino et al. is not the first to use discrete mixture model. Earlier, several models were used to describe the spatial distribution of rates. The models are mixed Poisson models, which incorporate random region-specific effects, which may represent environmental factors, and sometimes exhibit spatial correlation. Conditional on random region-effects R_i, representing region-specific relative rates, the observed numbers of cases, $Y_i, i = 1, \ldots, N$, are assumed to have independent Poisson distributions with mean $E_i R_i$; here E_i is the expected number of cases in the i^{th} region based on rates from a standard population.

Typically, the distribution of the random effects is assumed to be normal. Different choices have been explored in other contexts of generalized linear mixed modeling when random effects are assumed independent; for example, the gamma and the inverse-Gaussian distributions. Schlattmann and Bohning (1993) illustrated the use of a non-parametric mixing distribution for mapping rates. This yields a discrete mixture model and a key feature of this approach is that it provides a more robust method for including random effects than the parametric approaches described above.

A.F.Militino, M.D.Ugarte and C.B.Dean then compared the use of discrete mixture models with independent random effects for mapping problems, with normal mixtures which incorporate spatial autocorrelation. A focus of the comparison is how well the discrete and normal mixtures identify high rates. The key finding of their work is to show that the discrete mixture model is indeed more efficient than other available models in locating regions with high rates. This is one of the objectives in modeling mortality data for disease mapping.

The computation is done using the EM algorithm. It is pertinent for us to make a brief description of the computation. Let $\mathbf{Y} = (Y_1, \ldots, Y_N)^T$ be the vector of observations corresponding to the COPD mortality counts for the N contiguous regions. Associated with each region are random effects which model heterogeneity. The non-parametric mixture model assumes that these region-specific random effects have a discrete probability distribution taking g values $R_1, \ldots, R_g$ with probabilities $p_1, \ldots, p_g$, respectively. Each of these g components of the mixture represents a class containing a proportion p_j from the population with mortality ratio R_j where $\sum_{j=1}^{g} p_j = 1$. If the population in the i^{th} region is denoted by n_i, the marginal distribution of the counts Y_i is given by

$$Y_i \overset{iid}{\sim} \sum_{j=1}^{g} p_j \text{Poisson}(E_i R_i).$$

where $E_i = n_i m$ and m is the overall mean infant mortality rate. Hence, conditional on the region-specific random effects, the mortality counts are assumed to be distributed as independent Poisson.

Estimation would be trivial if we know which class each observation belongs to. In this case, the likelihood function would have the simple multinomial form.

$$\prod_{i=1}^{\ell} \left\{ \left[p_1 \frac{(E_i R_1)^{y_i} e^{(-E_i R_1)}}{y_i!} \right]^{z_{i1}} \left[p_2 \frac{(E_i R_2)^{y_i} e^{(-E_i R_2)}}{y_i!} \right]^{z_{i2}}, \ldots, \left[p_g \frac{(E_i R_g)^{y_i} e^{(-E_i R_g)}}{y_i!} \right]^{z_{ig}} \right\}$$

where z_{ij}, called latent variable, taking value 1 when observation i belongs to class j for $j = 1, \ldots, g, i = 1, \ldots, N$, take 0 otherwise.

The EM algorithm takes advantage of this simplicity and creates a two-stage iterative estimation scheme where, in the first stage, a 'guess' is made regarding the grouping of the classes. Then estimation based on such grouping is conducted. It

has been applied to a wide variety of mixture problems. Dempster et al. (1997) interpreted the mixture density estimation problem as an estimation problem involving incomplete data. In our situation, it is the z_{ij}'s which are regarded as the 'missing' data and the above likelihood function is considered to correspond to the 'complete' data. They also related estimation for the mixture density problem to a broader class of statistical problems and showed that the EM algorithm for the mixture density problems is essentially a specialization of a more general algorithm for calculating maximum likelihood estimates from incomplete data. The EM algorithm has been found in most instances to have the advantages of reliable global convergence, low cost per iteration and ease of programming. Although it has been criticized because its convergence can be slow, especially when there is a high proportion of missing data, it works very well in many instances, and in particular, in the case considered in the paper of A.F.Militino et al (2001).

The iterative scheme of the EM algorithm has two steps. Step 1 calculate the conditional expectation of z_{ij} given Y_i, then step 2 maximizes the conditional distribution of the complete likelihood after substituting $E(z_{ij}|Y_i)$ for the 'missing' data z_{ij}. This yields solutions

$$\hat{m} = \frac{\sum_{i=1}^{N} Y_i}{\sum_{i=1}^{N} n_i \sum_{j=1}^{g} E(z_{ij}|Y_i) R_j}, \quad \hat{R_j} = \frac{\sum_{i=1}^{N} E(z_{ij}|Y_i) Y_i}{\sum_{i=1}^{N} E(z_{ij}|Y_i) E_i}, \quad \hat{p_j} = \frac{\sum_{i=1}^{N} E(z_{ij}|Y_i)}{N}.$$

The iterative procedure starts with an initial value for m, R_j, p_j and z. The number of components is considered fixed. Step 1 and 2 above are repeated until a specified convergence criterion is reached.

Chapter 3

A Hierarchical Latent Class Regression Model

In this chapter, we describe a hierarchical Bayesian latent class model with g latent classes. We first consider the case when $g = 1$, essentially the Christiansen and Morris model. Then this model is extended to include $g > 1$ latent classes. A related question is how many latent classes there are (i.e., the value of g). We do not need to determine what g is because typically disease maps are usually represented with 4-6 colors, so we choose g to be 4, 5 or 6.

3.1 The Regression Model for a Single Class

3.1.1 The Model

Some previous work has been done with the assumption of a single class, the estimated mortality rates are then cut into equal quantiles. For example, with quintiles each contain 20% of the total data. Then mapping is done based on this grouping. The following model explains how this works in a hierarchical Bayesian analysis.

The observed death numbers for 798 HSAs $(d_1, \ldots, d_\ell)$ have independent Poisson distributions with expected values $E(d_i) = n_i\lambda_i, i = 1, \ldots, \ell$, for known populations $n_i > 0$. Given $\lambda = (\lambda_1, \ldots, \lambda_\ell)'$.

$$d_i|\lambda_i \overset{iid}{\sim} \text{Poisson}(n_i\lambda_i), \quad \text{independently } i = 1, \ldots, \ell. \tag{3.1}$$

The observed rates, $\mathbf{r}$, with $r_i \equiv d_i/n_i$ have expectations $E(r_i) = \lambda_i$. The unobservable parameters $(\lambda_1, \ldots, \lambda_\ell)$ follow conjugate gamma distributions for $i = 1, \ldots, \ell$ independently, given the unknown hyper-parameter vector $\underset{\sim}{\alpha} = (\alpha, \underset{\sim}{\beta})'$. Here $\underset{\sim}{\beta} = \beta_0, \ldots, \beta_k$, denote the potential risk factors which are selected in Chapter 1, where we recall in Table 3.1

Table 3.1: Potential Risk Factors

Covariates $\underset{\sim}{x}$	**Coefficients** $\underset{\sim}{\beta}$	**Corresponding Factors**
$x_0(\equiv 1)$	β_0	Intercept
x_1	β_1	Smoking History
x_2	β_2	Population Density
x_3	β_3	Elevation
x_4	β_4	Annual Rainfall

We try to link the above covariates to the parameter λ_i, which stands for the mortality rate. Fit the covariates in the regression model (with mortality rate λ_i as the response variable) we get the prior distribution,

$$\lambda_i|\alpha, \underset{\sim}{\beta} \overset{iid}{\sim} \text{Gamma}(\alpha, \alpha e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}}) \tag{3.2}$$

where $T \sim \text{Gamma}(\alpha, \lambda)$ means $f_T(t) = \frac{\lambda^\alpha t^{\alpha-1}e^{-\lambda t}}{\Gamma(\alpha)}$. That is,

$$P(\underset{\sim}{\lambda}|\alpha, \underset{\sim}{\beta}) = \prod_{i=1}^{\ell} \frac{(\alpha e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}})^\alpha \lambda_i^{\alpha-1} e^{-\alpha e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}}\lambda_i}}{\Gamma(\alpha)}.$$

Now for the hyper-parameters, α and $\underset{\sim}{\beta}$, we use the following proper prior distribution,

$$P(\alpha) = \frac{1}{(1+\alpha)^2}, \qquad \alpha \geq 0.$$

$$\underset{\sim}{\beta} \sim \text{MN}(\mu_{\underset{\sim}{\beta}}, \Delta_0)$$

where $\mu_{\underset{\sim}{\beta}}$ and Δ_0 are two constants to be specified by the user. We fit a weighted least square regression model and specify $\mu_{\underset{\sim}{\beta}}$ and Δ_0 as below. The computational details can be found in the Appendix A.

$$\mu_{\underset{\sim}{\beta}} = \underset{\sim}{b} = (X'W^{-1}X)^{-1}(X'W^{-1}Y)$$

$$\Delta_0 = (X'W^{-1}X)^{-1}\frac{(Y-Xb)'W^{-1}(Y-Xb)}{n-2}$$

where Y, X and W are the reponse vector, covariate matrix and weight matrix respectively.

3.1.2 Markov Chain Monte Carlo Computation

After setting up the model and all the priors, we can proceed using Markov chain Monte Carlo (MCMC) method to make inferences about the parameters in the model. The particular MCMC method used here is called Metropolis-Hastings sampler. It works by drawing sample from the conditional distribution. After a large number of iterations, the sample will converge to the joint posterior distribution.

Now let us start by finding the conditional distributions for the parameters; we also need a start value for each parameter in the Gibbs sampling iterations.

We use $\frac{d_i}{n_i}$ as the starting value for λ_i, which is the observed mortality rate for HSA i. As for the starts for α and $\underset{\sim}{\beta}$, we use Nelder-Mead algorithm to find the optimal values which maximize the density function, these values will be used as the start values.

First, conditional posterior distributions are proportional to the joint posterior distributions

$$\pi(\underset{\sim}{\lambda}, \alpha, \underset{\sim}{\beta}|\text{data}) \propto \pi(\underset{\sim}{\lambda}, \alpha, \underset{\sim}{\beta}, \text{data})$$

i.e.

$$\pi(\underset{\sim}{\lambda}, \alpha, \underset{\sim}{\beta}|\text{data}) \propto \prod_{i=1}^{\ell} \frac{(n_i\lambda_i)^{d_i} e^{-n_i\lambda_i}}{(d_i)!}$$

$$\times \prod_{i=1}^{\ell} \frac{(\alpha e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}})^{\alpha} \lambda_i^{\alpha-1} e^{-\alpha e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}_{\lambda_i}}}}{\Gamma(\alpha)} \times \frac{1}{(1+\alpha)^2}$$

$$\times \frac{1}{\sqrt{(2\pi)^p \det(\Delta_0)}} e^{-\frac{1}{2}(\underset{\sim}{\beta}-\mu_{\underset{\sim}{\beta}})'\Delta_0^{-1}(\underset{\sim}{\beta}-\mu_{\underset{\sim}{\beta}})} \tag{3.3}$$

where $\mu_{\underset{\sim}{\beta}} = \underset{\sim}{b} = (X'W^{-1}X)^{-1}(X'W^{-1}Y)$ and $\Delta_0 = (X'W^{-1}X)^{-1}\frac{(Y-Xb)'W^{-1}(Y-Xb)}{n-2}$

Then from the joint posterior distributions, we can find the conditional posterior distributions for each parameter. First, for λ_i

$$\lambda_i|\alpha, \underset{\sim}{\beta}, \text{data} \sim \text{Gamma}(\alpha + d_i, n_i + \alpha e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}}). \tag{3.4}$$

For α and $\underset{\sim}{\beta}$, the distribution function is too complex, so here we use the Metropolis-Hastings approximation method. First propose an approximate multivariate-normal distribution density for α and $\underset{\sim}{\beta}$, then we draw a sample from the proposed distribution, and decide to keep it or discard it by Metropolis-Hastings procedure.

The proposed multivariate-normal distribution for vector $(\alpha, \underset{\sim}{\beta})'$ is (detail computation can be found in Appendix B):

$$\begin{pmatrix} \alpha \\ \underset{\sim}{\beta} \end{pmatrix} \sim \text{MN}\left\{ \begin{bmatrix} \hat{\alpha} \\ \underset{\sim}{\hat{\beta}} \end{bmatrix}, \begin{bmatrix} \sigma_\alpha^2 & \underset{\sim}{\nu}' \\ \underset{\sim}{\nu} & \Delta_\beta \end{bmatrix} \right\} \tag{3.5}$$

In which $\hat{\alpha}$ and $\hat{\underset{\sim}{\beta}}$ are the solutions from Nelder-Mead algorithm.

Now we draw a random sample of $\alpha, \underset{\sim}{\beta}$ from (3.5).

First, draw a random $\underset{\sim}{\beta}^*$ from

$$\mathrm{N}(\hat{\underset{\sim}{\beta}}, \Delta_\beta),$$

then, draw a random α^* from Gamma distribution with mean and variance calculated by:

$$\mu_\alpha = \hat{\alpha} + \nu' \Delta_\beta^{-1} (\underset{\sim}{\beta}^* - \hat{\underset{\sim}{\beta}}),$$

$$\mathrm{Var}(\alpha) = \sigma_\alpha^2 - \nu' \Delta_\beta^{-1} \nu.$$

Now we use Metropolis-Hastings method to decide take or throw the new sample drawn.

Let's use the notation $f(\alpha, \underset{\sim}{\beta})$ stand for the complex conditional posterior distribution of $\alpha, \underset{\sim}{\beta}$, and $f_0(\alpha, \underset{\sim}{\beta})$ stand for the proposal density we used, i.e., Gamma multiplied by Multivariate-Normal. The sample we just drawn is denoted by $(\alpha_{n+1}, \underset{\sim}{\beta}_{n+1})$, then we compute

$$\varphi(n) = \frac{f(\alpha_n, \underset{\sim}{\beta}_n)}{f_0(\alpha_n, \underset{\sim}{\beta}_n)},$$

$$\varphi(n+1) = \frac{f(\alpha_{n+1}, \underset{\sim}{\beta}_{n+1})}{f_0(\alpha_{n+1}, \underset{\sim}{\beta}_{n+1})},$$

and

$$a_{n,n+1} = \min\left\{ \frac{\varphi(n+1)}{\varphi(n)}, 1 \right\}.$$

Draw random number μ from $U(0,1)$ distribution. Decision rule is

$$\Rightarrow \text{take} \quad (\alpha_{n+1}, \underset{\sim}{\beta}_{n+1}) \quad \text{if} \quad \mu < a_{n,n+1}$$

$$\Rightarrow \text{retake} \quad (\alpha_n, \underset{\sim}{\beta}_n) \qquad \text{if} \quad \mu \geq a_{n,n+1}.$$

3.2 The Regression Model for Many Classes

The model used in the previous section is appropriate when the observations are from a single population (i.e., one class). As described in Chapter 2, in reality this is not always the case. There is often a random class-specific effect which make the population structure too complex to be modeled by a simple distribution. In such cases, we should use the mixture models.

The key idea in our model is introduced in this section, which combine the discrete Poisson mixture model with regression model, and use MCMC methods to estimate the parameters and make other inferences. Some details of implementing this idea is also given in this section.

3.2.1 The Model

As before, the observed death numbers for 798 HSAs $(d_1, \ldots, d_\ell)$ have independent Poisson distributions with expected values $E(d_i) = n_i\lambda_i, i = 1, \ldots, \ell$, for known populations $n_i > 0$. Let $\lambda = (\lambda_1, \ldots, \lambda_\ell)'$. Now the population contains a set of classes, and each class contains a proportion of the whole data.

$$d_i|\lambda_i \sim \text{Poisson}(n_i\lambda_i) \quad \text{independently} \;\; i = 1, \ldots, \ell. \tag{3.6}$$

We use a mixture prior for the λ_i in the following form,

$$\lambda_i|\underset{\sim}{p}, \underset{\sim}{\mu}, \alpha, \underset{\sim}{\beta} \overset{ind}{\sim} \sum_{j=1}^{g} p_j \text{Gamma}(\alpha, \alpha\frac{1-\mu_j}{\mu_j}e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}}). \tag{3.7}$$

That is,

$$P(\underset{\sim}{\lambda}|\underset{\sim}{p}, \underset{\sim}{\mu}, \alpha, \underset{\sim}{\beta}) = \prod_{i=1}^{\ell}\left\{\sum_{j=1}^{g} p_j \frac{(\alpha\frac{1-\mu_j}{\mu_j}e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}})^\alpha \lambda_i^{\alpha-1} e^{-\alpha\frac{1-\mu_j}{\mu_j}e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}}\lambda_i}}{\Gamma(\alpha)}\right\}.$$

We take the shrinkage prior as the prior density for α

$$\pi(\alpha) = \frac{1}{(1+\alpha)^2},$$

and we use a multivariate normal density as the prior for $\underset{\sim}{\beta}$

$$\underset{\sim}{\beta} \sim \text{MN}(\underset{\sim}{\mu}^{(0)}, c\underset{\sim}{\Delta}^{(0)}).$$

where c is a variance inflation factor. We take $c = 10000$ in our analysis to obtain a proper diffuse prior density for $\underset{\sim}{\beta}$.

Here $\underset{\sim}{\mu}^{(0)}$ and $\underset{\sim}{\Delta}^{(0)}$ are specified as:

$$\underset{\sim}{\mu}^{(0)} = \underset{\sim}{b} = (X'W^{-1}X)^{-1}(X'W^{-1}Y),$$

$$\underset{\sim}{\Delta}^{(0)} = (X'W^{-1}X)^{-1}\frac{(Y-Xb)'W^{-1}(Y-Xb)}{n-2},$$

Y, X and W are respectively the reponse vector, covariate matrix and weight matrix in the regression model. Detailed computations can be found in Appendix A.

The prior for $\underset{\sim}{p}$ is Dirichlet$(1, \ldots, 1)$; this non-informative proper prior density is used since there is no information about $\underset{\sim}{p}$.

A typical problem of the latent class model is nonidentifiability. To maintain identifiability, we incorporate the order restriction.

$$\frac{1}{2} \equiv \mu_1 < \mu_2 < \ldots < \mu_g < 1.$$

By constraining the μ_k between μ_{k-1} and μ_{k+1}, it avoids the possible switching between two μ_k's, thus eliminate the nonidentifiability problem.

When $k = 1$, $\mu_1 = \frac{1}{2}$, the prior for λ_i becomes $\Gamma(\alpha, \alpha e^{-\underset{\sim}{\beta}})$, which is exactly the model by Christiansen and Morris (1997), we described in Section 3.2.

3.2.2 MCMC Computation

Metropolis-Hastings sampler is still used here, the procedures are similar with the case of $(g = 1)$.

Joint Posterior Distribution

The new parameter z_{ij} is introduced as a latent variable, to simply the computation. It takes value 1 when observation i belongs to class j for $j = 1, \ldots, g, i = 1, \ldots, N$, take 0 otherwise. Then the joint posterior distribution is:

$$\pi(\underset{\sim}{\lambda}, \alpha, \underset{\sim}{\beta}, \underset{\sim}{\mu}, \underset{\sim}{p}, \phi|\text{data}) \propto \prod_{i=1}^{\ell} \frac{(n_i\lambda_i)^{d_i} e^{-n_i\lambda_i}}{(d_i)!}$$

$$\times \prod_{i=1}^{\ell}\prod_{j=1}^{g} \left\{ p_j \frac{(\alpha\frac{1-\mu_j}{\mu_j} e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}})^{\alpha} \lambda_i^{\alpha-1} e^{-\alpha\frac{1-\mu_j}{\mu_j} e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}}\lambda_i}}{\Gamma(\alpha)} \right\}^{z_{ij}}$$

$$\times \frac{1}{(1+\alpha)^2} \times \frac{1}{\sqrt{(2\pi)^p \det(\underset{\sim}{\Delta}^{(0)})}} e^{-\frac{1}{2}(\underset{\sim}{\beta}-\underset{\sim}{\mu}^{(0)})'\underset{\sim}{\Delta}^{(0)-1}(\underset{\sim}{\beta}-\underset{\sim}{\mu}^{(0)})} \tag{3.8}$$

where $\underset{\sim}{\mu}^{(0)} = \underset{\sim}{b} = (X'W^{-1}X)^{-1}(X'W^{-1}Y)$ and $\underset{\sim}{\Delta}^{(0)} = (X'W^{-1}X)^{-1}\frac{(Y-Xb)'W^{-1}(Y-Xb)}{n-2}$.

Note that $\Sigma_{j=1}^{g} p_j = 1$, and the inclusion of latent variable make the computation a lot more convenient, as noted from the EM algorithm.

Below we calculate the conditional posterior distributions to run the Metropolis-Hastings sampler. For λ_i

$$\lambda_i|\alpha, \underset{\sim}{\beta}, \underset{\sim}{\mu}, \underset{\sim}{p}, \underset{\sim}{z}, \text{data} \sim \text{Gamma}(\alpha + d_i, n_i + \alpha e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}} \sum_{j=1}^{g} (\frac{1-\mu_j}{\mu_j} z_{ij}))$$

For $\underset{\sim}{p}$

$$p_j|\alpha, \underset{\sim}{\beta}, \underset{\sim}{\mu}, \underset{\sim}{\lambda}, \underset{\sim}{z}, \text{data} \sim \text{Dirichlet}(\sum_{i=1}^{\ell} z_{i1} + 1, \ldots, \sum_{i=1}^{\ell} z_{ig} + 1)$$

For $z_{\underset{\sim}{i}}$

$$z_{\underset{\sim}{i}}|\alpha, \underset{\sim}{\beta}, \underset{\sim}{\mu}, p, \underset{\sim}{\lambda}, \text{data} \sim \text{Multinomial}[1, q]$$

where

$$q_j = \frac{p_j(\frac{1-\mu_j}{\mu_j})^\alpha e^{-\alpha\frac{1-\mu_j}{\mu_j}e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}}\lambda_i}}{\sum_{j=1}^g [p_j(\frac{1-\mu_j}{\mu_j})^\alpha e^{-\alpha\frac{1-\mu_j}{\mu_j}e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}}\lambda_i}]}, \quad \sum_{j=1}^g q_j = 1.$$

To draw a z_{ij}, we just draw a random number μ from $U(0,1)$, then compare with the accumulated probability.

Cell	**Cumulative probability**
c_1	q_1
c_2	$q_1 + q_2$
c_3	$q_1 + q_2 + q_3$
.	
.	
c_{g-1}	$q_1 + q_2 +, \ldots, q_{g-1}$
c_g	1

If the random number μ lies in the interval of $(c_{k-1} < \mu < c_k)$, then let $\phi_{ik} = 1$, others all equal to 0, for all the $i = 1, \ldots, \ell$.

For $\underset{\sim}{\mu}$, we use the grid method under the order constraint to draw the sample. For example, when we draw μ_3 (when $k \geq 3$), since $\frac{1}{2} \equiv \mu_1 < \mu_2 < \ldots < \mu_k < 1$, so we have restriction

$$\mu_2 < \mu_3 < \mu_4 (\text{or } 1, \text{if } k = 3).$$

The conditional posterior density for μ_3 is:

$$P(\mu_3|\text{all other parameters, data}) = \prod_{i=1}^{\ell} \left\{ \frac{(\alpha\frac{1-\mu_3}{\mu_3}e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}})^\alpha \lambda_i^\alpha e^{-(\alpha\frac{1-\mu_3}{\mu_3}e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}})\lambda_i}}{\Gamma(\alpha)} \right\}^{z_{i3}}$$

Then, we cut interval (μ_2, μ_4) into n grids, $n \approx 100$ in our case. For all grids, we calculate the posterior densities (use the value of the mid-point on each grid)

$\pi_1, \ldots, \pi_{100}$.

Similar to drawing z's, we first calculate

$$q_h = \frac{\pi_h}{\sum_{h'=1}^{100} \pi_{h'}}, \quad \text{thus get } \sum_{h=1}^{100} q_h = 1.$$

Then we calculate the cumulative probability of these q_h's.

Cell	**Cumulative probability**
c_1	q_1
c_2	$q_1 + q_2$
c_3	$q_1 + q_2 + q_3$
.	
.	
c_{99}	$q_1 + q_2 +, \ldots, q_{99}$
c_{100}	1

To draw a sample of μ_3, first draw a random number x from $U(0,1)$, then compare with the cumulative probability. If the random number x lies in the interval of $(c_{h-1} < x < c_h)$, then pick the corresponding mid-point on the h^{th} grid as a sample μ_3.

For α and $\underset{\sim}{\beta}$, we still use the Metropolis-Hastings method, and still use multivariate-normal distribution as the proposal distribution. (Computation details can be found in the Appendix B).

Therefore the vector $(\alpha, \underset{\sim}{\beta})'$ have the following multivariate-normal distribution

$$\begin{pmatrix} \alpha \\ \underset{\sim}{\beta} \end{pmatrix} \sim \text{MN} \left\{ \begin{bmatrix} \hat{\alpha} \\ \hat{\underset{\sim}{\beta}} \end{bmatrix}, \begin{bmatrix} \sigma_\alpha^2 & \underset{\sim}{\nu}' \\ \underset{\sim}{\nu} & \Delta_\beta \end{bmatrix} \right\} \tag{3.9}$$

In which $\hat{\alpha}$ and $\hat{\underset{\sim}{\beta}}$ are the solutions from Nelder-Mead algorithm, $\begin{bmatrix} \sigma_\alpha^2 & \underset{\sim}{\nu}' \\ \underset{\sim}{\nu} & \Delta_\beta \end{bmatrix}$ is the inverse of the negative Hessian matrix.

Now we draw a random sample of $\alpha, \underset{\sim}{\beta}$ from (3.9)

First, draw a random $\underset{\sim}{\beta}^*$ from

$$\mathrm{N}(\hat{\underset{\sim}{\beta}}, \Delta_\beta)$$

Then, compute the mean and variance of α^*, μ_{α^*}, $\mathrm{Var}(\alpha^*)$.

$$\mu_{\alpha^*} = \hat{\alpha} + \nu' \Delta_\beta^{-1} (\underset{\sim}{\beta}^* - \hat{\underset{\sim}{\beta}}),$$

and

$$\mathrm{Var}(\alpha^*) = \sigma_\alpha^2 - \nu' \Delta_\beta^{-1} \nu$$

We draw α^* from a Gamma density with mean μ_{α^*} and variance $\mathrm{Var}(\alpha^*)$.

Now we use Metropolis-Hastings method to decide take or throw the new sample drawn.

Let's use the notation $f(\alpha, \underset{\sim}{\beta})$ stand for the complex conditional posterior distribution of $\alpha, \underset{\sim}{\beta}$, and $f_0(\alpha, \underset{\sim}{\beta})$ stand for the proposal joint density of α and $\underset{\sim}{\beta}$, i.e., Gamma multiplied by Multivariate-Normal. The sample we just drawn is denoted by $(\alpha_{n+1}, \underset{\sim}{\beta}_{n+1})$, then we compute

$$\varphi(n) = \frac{f(\alpha_n, \underset{\sim}{\beta}_n)}{f_0(\alpha_n, \underset{\sim}{\beta}_n)},$$

$$\varphi(n+1) = \frac{f(\alpha_{n+1}, \underset{\sim}{\beta}_{n+1})}{f_0(\alpha_{n+1}, \underset{\sim}{\beta}_{n+1})},$$

and

$$a_{n,n+1} = \min\left\{ \frac{\varphi(n+1)}{\varphi(n)}, 1 \right\}.$$

Draw random number μ from $U(0,1)$ distribution. Decision rule is

$$\Rightarrow \text{take} \quad (\alpha_{n+1}, \underset{\sim}{\beta}_{n+1}) \quad \text{if} \quad \mu < a_{n,n+1}$$

$$\Rightarrow \text{retake} \quad (\alpha_n, \underset{\sim}{\beta}_n) \qquad \text{if} \quad \mu \geq a_{n,n+1}.$$

3.2.3 Performance of the Metropolis-Hastings algorithm

We will use $M = 1000$ iterations for analysis. Let $\Omega = (\underset{\sim}{p}, \underset{\sim}{\mu}, \alpha, \underset{\sim}{\beta})$ as the parameter vector, so we need a sample of $\Omega^{(h)}, h = 1, \ldots, 1000$. To get this sample, we let the Metropolis-Hastings sampler run 5500 iterations, and "burn in" the first 500 iterations because it is unlikely to get convergence in a such a short time. Then, we pick every 5^{th} from the remaining 5000. Thus, make the $\Omega^{(h)}$ independent, so in the end we have a sample of 1000 observations.

A further check on the jumping rate of the Metropolis-Hastings sampler shows, this method runs well for all the models, as shown in table 3.2.

Table 3.2: Jumping Rate of the Metropolis-Hastings sampler by the number of classes

g	**Jumping rate**
1	0.4912727
2	0.4900000
3	0.4994546
4	0.4929091
5	0.5027273
6	0.4883636
7	0.4927273
8	0.5012727
9	0.5092728
10	0.4987273

From this table we see the jumping rates are all centered with 0.5, which means the Metropolis-Hastings sampler is running well, sampling with a rejection rate of about 0.5.

The sample we get are basically independent, which can be seen from the autocorrelation table 3.3. (We only include a small part of all the autocorrelations computed, these are for $g = 1$, for some randomly selected rates).

Table 3.3: Autocorrelation Study

Autocorrelation	**Std. Error**
0.1670	0.0315
0.0822	0.0315
0.0411	0.0315
0.0288	0.0315
0.0394	0.0315
0.0170	0.0315
0.0584	0.0314
0.0635	0.0314
0.0596	0.0316
0.1281	0.0316
0.0051	0.0315

Because all the autocorrelations and standard errors are small enough, we treat the sample values as independent.

3.2.4 Sensitivity Test

In the Metropolis-Hastings sampler, when there is no information available, we usually specify the magnitude of the prior variance based on our experience. In such cases, we should test the sensitivity of the model under different specified prior variance.

The way we specify this variance is, first calculate a rough estimate, then multiply it by a constant, called variance inflation factor. Now we will change this factor and compare the result each time, to see how robust the model is.

The test is done for two choices of g, one and five, and the test results are shown in Table 3.4.

Table 3.4: Mean and Standard Deviation of the 798 Posterior Means and Standard Deviations of the Variance Inflation Factor for Two Choices of number of classes

g	**Variable**	**Variance Factor**	**Mean**	**Std Dev**
1	Mean	1000	3.871	0.609
		10000	3.871	0.609
		100000	3.871	0.609
	Std Dev	1000	0.288	0.109
		10000	0.288	0.108
		100000	0.288	0.108
5	Mean	1000	3.873	0.618
		10000	3.873	0.617
		100000	3.873	0.618
	Std Dev	1000	0.291	0.116
		10000	0.291	0.116
		100000	0.292	0.117

Note: All the figures in the table are multiplied by 10^3, then rounded to three significant digits, originally they are different under more significant digits.

It can be seen that the model is pretty stable, no matter what value we take, the mean and standard deviation of the posterior means (of mortality rates) are virtually unchanged.

Chapter 4

Assessment of the Model

In this chapter, we first check how well the model fits by looking at the residuals. We assess each model by using two methods, deviances and marginal likelihood.

4.1 Analysis of Residual

To get residuals of the mortality rates $\underset{\sim}{\lambda} = (\lambda_1, \ldots, \lambda_{798})'$, we start from the death number $\underset{\sim}{d} = (d_1, \ldots, d_{798})'$. Since $\lambda_i = \frac{d_i}{n_i}$, here death number $d_i \sim \text{Poisson}(n_i\lambda_i), i = 1, \ldots, 798$, population n_i is a constant, so we have:

$$E(\lambda_i) = E(\frac{d_i}{n_i}) = \frac{E(d_i)}{n_i} \tag{4.1}$$

and

$$Var(\lambda_i) = Var(\frac{d_i}{n_i}) = \frac{Var(d_i)}{n_i^2}.$$

Also, for standard deviation $std(\lambda_i)$

$$std(\lambda_i) = \sqrt{Var(\lambda_i)} = \sqrt{\frac{Var(d_i)}{n_i^2}} = \frac{std(d_i)}{n_i}. \tag{4.2}$$

Here we use a cross-validation method to compute the mean of d_i given the mortality rates of all the other HSA's, we denote $\underset{\sim}{\lambda}_{(i)}$ the vector of mortality rates, whose elements are all the λ_j except λ_i.

To calculate the mean of d_i given $\underset{\sim}{d}_{(i)}$, we use a weighted sample average:

$$E(d_i|\underset{\sim}{d}_{(i)}) = E_{\Omega|\underset{\sim}{d}_{(i)}}[E(d_i|\underset{\sim}{d}_{(i)}, \Omega)]$$

$$= n_i E(d_i|\underset{\sim}{d}_{(i)}) = n_i \sum_{h=1}^{M} \omega_i^{(h)} \lambda_i^{(h)} \tag{4.3}$$

where $\lambda_i^{(h)}$ is the h^{th} sample of λ_i, M is the sample size, here $M = 1000$, and

$$\omega_i^{(h)} = \frac{[(n_i\lambda_i^{(h)})^{d_i} e^{-n_i\lambda_i^{(h)}}]^{-1}}{\sum_{h=1}^{M}[(n_i\lambda_i^{(h)})^{d_i} e^{-n_i\lambda_i^{(h)}}]^{-1}}.$$

In the same way we can calculate

$$\mathrm{Var}(d_i|\underset{\sim}{d}_{(i)}) = E_{\Omega|\underset{\sim}{d}_{(i)}}[\mathrm{Var}(d_i|\underset{\sim}{d}_{(i)}, \Omega)] + \mathrm{Var}_{\Omega|\underset{\sim}{d}_{(i)}}[E(d_i|\underset{\sim}{d}_{(i)}, \Omega)]$$

$$= n_i E(\lambda_i|\underset{\sim}{d}_{(i)}) + n_i^2 \mathrm{Var}(\lambda_i|\underset{\sim}{d}_{(i)}) \tag{4.4}$$

where

$$\mathrm{Var}(\lambda_i|\underset{\sim}{d}_{(i)}) = E[\lambda_i - E(\lambda_i|\underset{\sim}{d}_{(i)})]^2$$
$$= \sum_{h=1}^{M} \omega_i^{(h)} [\lambda_i^{(h)} - E(\lambda_i^{(h)}|\underset{\sim}{d}_{(i)})]^2.$$

We use $E(d_i|\underset{\sim}{d}_{(i)})$ and $Var(d_i|\underset{\sim}{d}_{(i)})$ to approximate $E(d_i)$ and $Var(d_i)$, and therefore we can also get $E(\lambda_i)$ and $std(\lambda_i)$ by plug in (4.1) and (4.2).

For residuals, first define:

$$ares_i = r_i - E(r_i|\underset{\sim}{r}_{(i)})$$
$$sres_i = \frac{r_i - E(r_i|\underset{\sim}{r}_{(i)})}{std(r_i|\underset{\sim}{r}_{(i)})}.$$

Here $ares_i$ is the regular residuals and $sres_i$ is the standardized residuals. It is easy to prove

$$E(r_i|\underset{\sim}{r}_{(i)}) = \frac{E(d_i|\underset{\sim}{d}_{(i)})}{n_i}$$

$$std(r_i|\underset{\sim}{r}_{(i)}) = \frac{std(d_i|\underset{\sim}{d}_{(i)})}{n_i}.$$

Figure 4.1 is a plot of residual versus standard deviation, for $g = 1, \ldots, 6$. We add bands of ± 2 standard deviations in the plot, where most of the residuals should fall inside if the model fits well.

Figure 4.2 is a plot of the standardized residual versus predicted value, for $g = 1, \ldots, 6$. The band in the plot is (-2, +2), most of the standardized residuals should fall in this band if the model fits well.

Figure 4.3 is a box plot, instead of plot all the mortality rates in a single box, we divide them into 12 groups by region, and plot in each region. This provide a more insightful view of the residuals.

All the plots suggest a good fit, with most of the residuals fall inside the bands, and all the box plots have a normal like distribution around 0 (for $g = 1, 4, 5, 6$). Outliers exist, but not too many. All these are signs of a good fit.

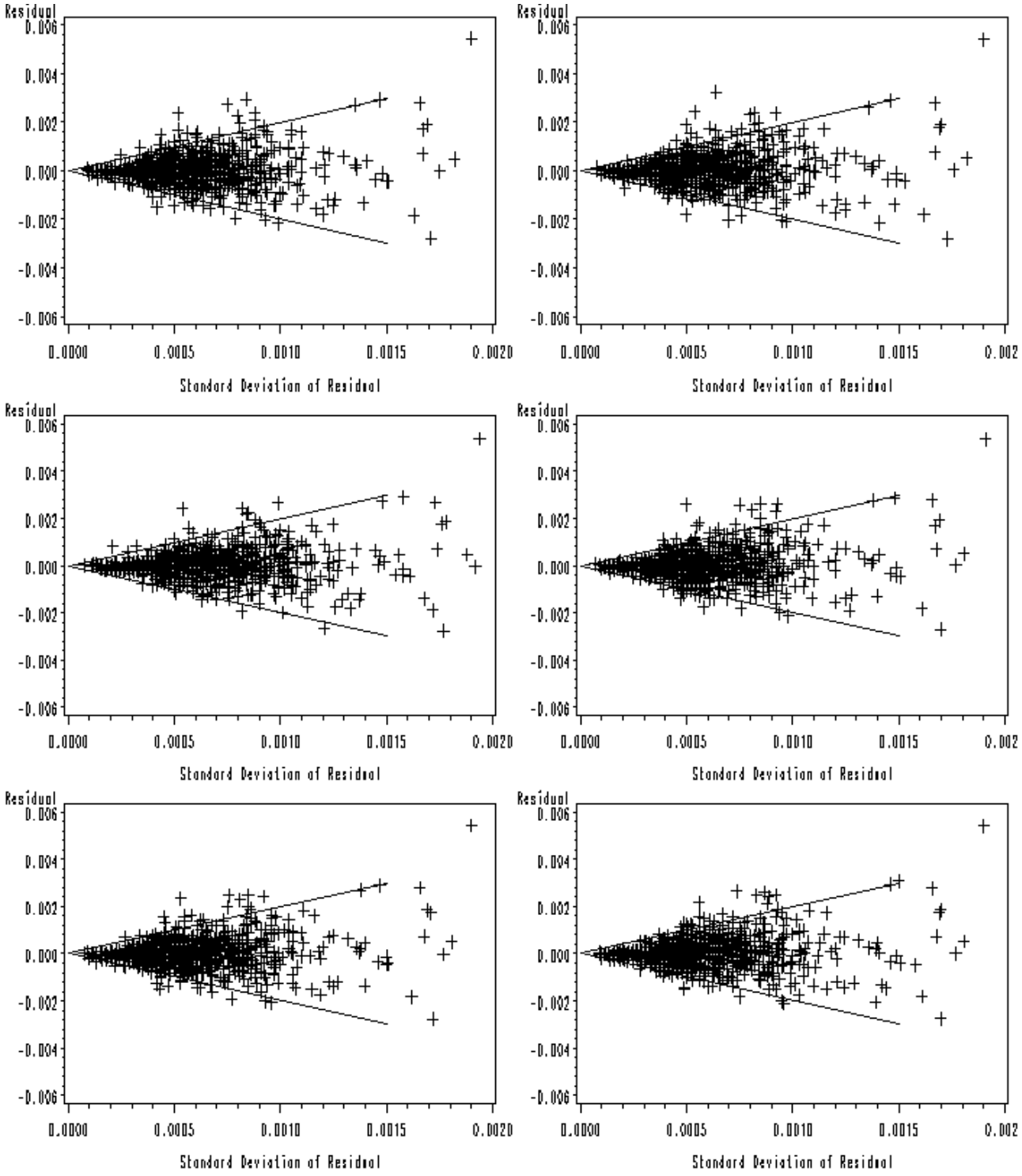

Figure 4.1: Residual Against Standard Deviation With Bands at Two Standard Deviations

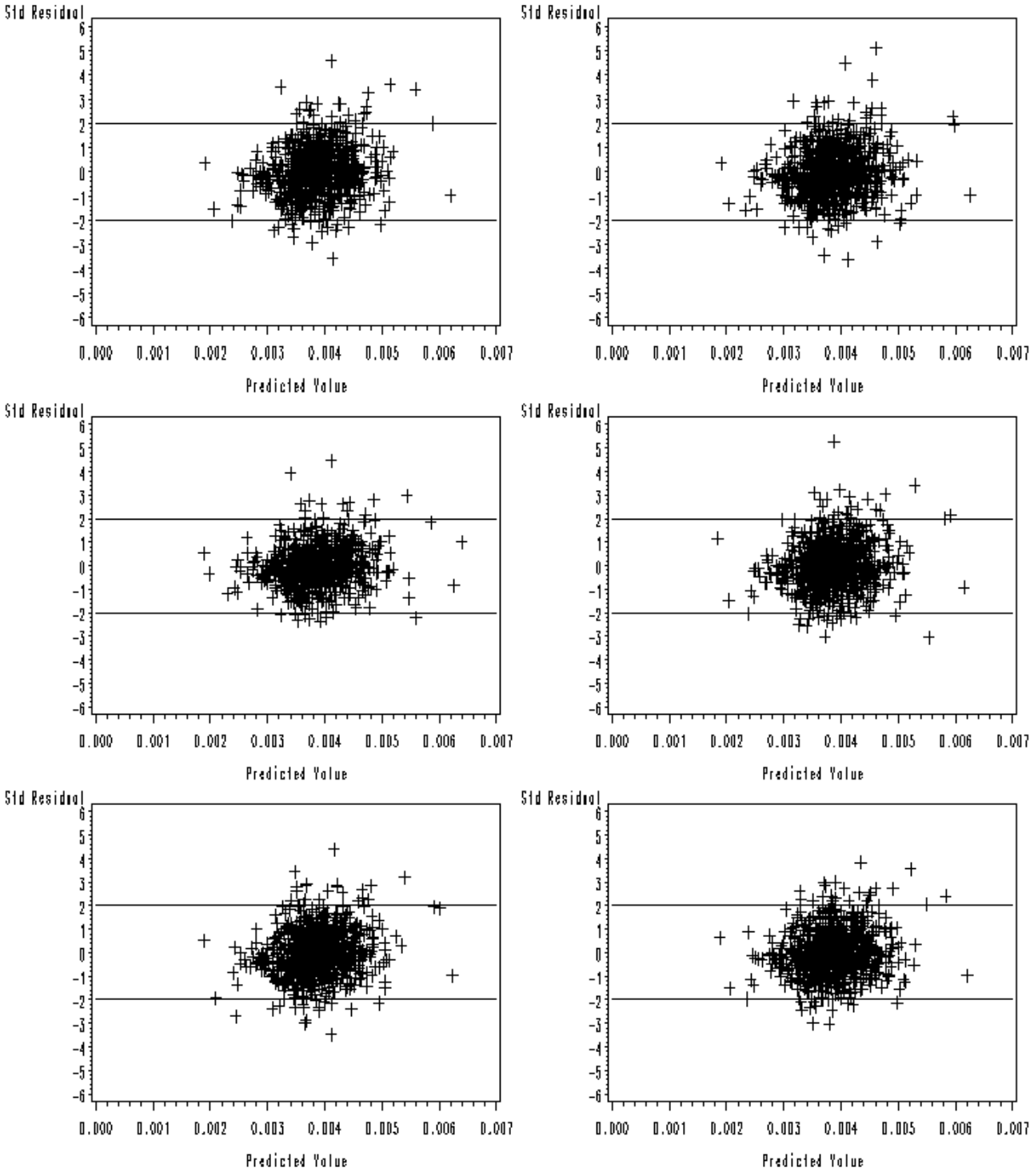

Figure 4.2: Standardized Residual against Predicted Value

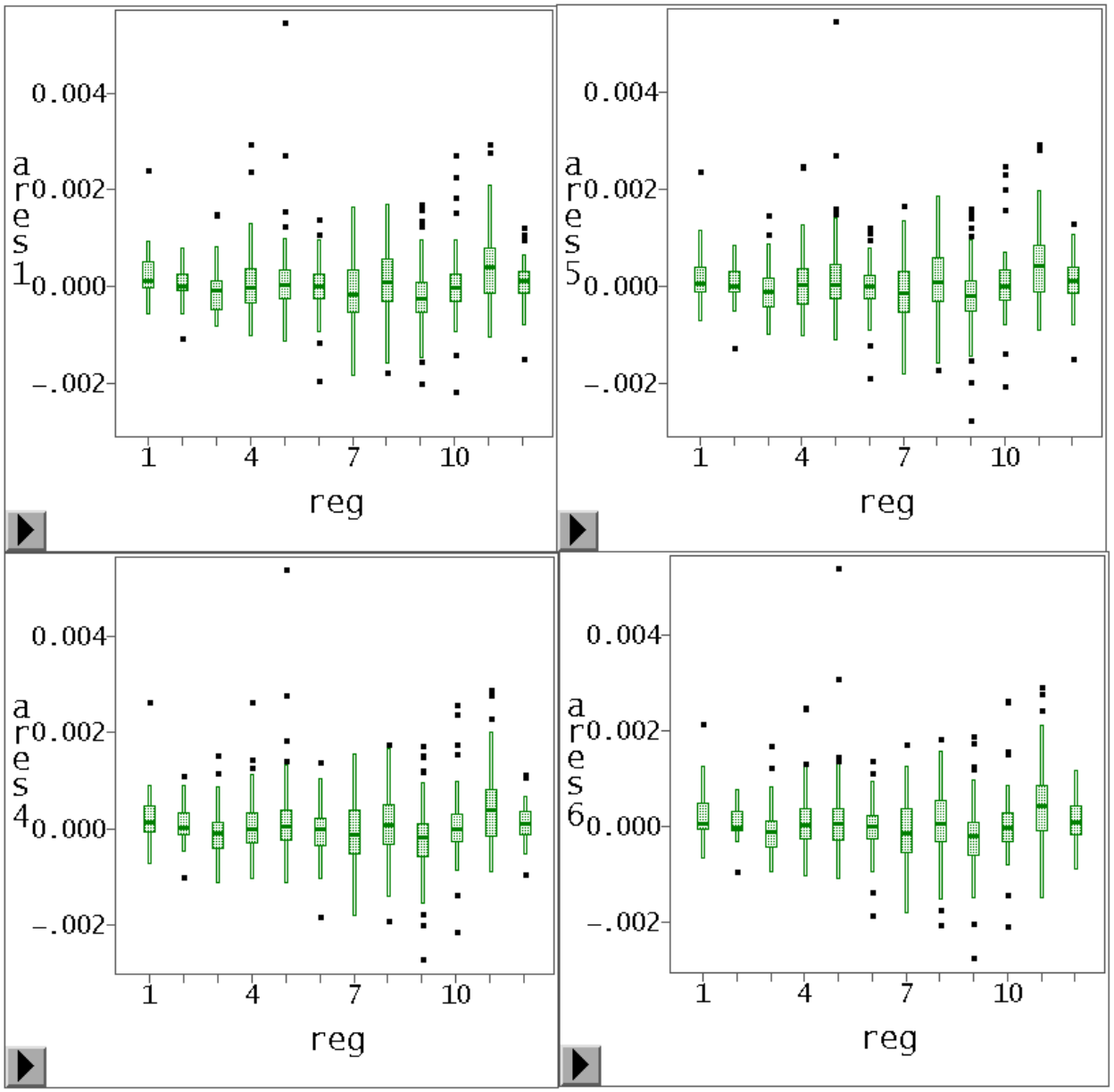

Figure 4.3: Box Plots of Residuals by 12 Regions of the US

4.2 Analysis of Deviance and Marginal likelihood

So far we have done setting the model and checking the model fit, but actually we are always using several models instead of one, since we choose different g in our model, which gives us different models. Now we want to compare the models under different g's, this is a process of model selection. But in our particular case, the most important problem is to provide improved maps.

4.2.1 The Deviance Measure

To assess the goodness of a model, we first use the Gelfand and Ghosh (1997) deviance in Poisson distribution.

Deviance is defined as:

$$\begin{aligned} D_k(M) &= \mathrm{P}_k(M) + \mathrm{G}_k(M) \\ &= 2\sum_{i=1}^{\ell} \omega_i \left\{ t_i^{(M)} - t(\mu_i^{(M)}) \right\} \\ +2(k+1)\sum_{i=1}^{\ell} \omega_i &\left\{ \frac{t(\mu_i^{(M)}) + kt(y_{i,obs})}{k+1} - t(\frac{\mu_i^{(M)} + ky_{i,obs}}{k+1}) \right\}. \end{aligned} \tag{4.5}$$

Here M refers to the model, k is a constant to be specified (we use 100), $y_{i,obs}$ is the observed data value y_i, $t(x)$ is a function defined as

$$t(x) \equiv x \log x - x$$

and

$$t_i^{(M)} = E[t(y_i)|y_{i,obs}, M], \quad \mu_i^{(M)} = E[y_i|y_{i,obs}, M].$$

Here $\omega_i \equiv 1$ for Poisson distribution.

To calculate the $t_i^{(M)}$, we start from

$$t_i^{(M)} = E[t(y_i)|y_{i,obs}, M] = \sum_{y=0}^{\infty} [y \log y - y] \frac{(n_i\lambda_i)^y e^{-n_i\lambda_i}}{y!}.$$

Notice that

$$y \log y - y|_{y=0} = 0.$$

So we get

$$t_i^{(M)} = E[t(y_i)|y_{i,obs}, M] = \sum_{y=1}^{\infty} [y \log y - y] \frac{(n_i\lambda_i)^y e^{-n_i\lambda_i}}{y!}$$

$$= \sum_{y=1}^{\infty} y \log y \frac{(n_i\lambda_i)^y e^{-n_i\lambda_i}}{y!} - n_i\lambda_i.$$

Let $y - 1 = \zeta$, then

$$\overset{y-1=\zeta}{=} \sum_{\zeta=0}^{\infty} (\zeta + 1) \log (\zeta + 1) \frac{(n_i\lambda_i)^{(\zeta+1)} e^{-n_i\lambda_i}}{(\zeta + 1)!} - n_i\lambda_i$$

$$= n_i\lambda_i \sum_{\zeta=0}^{\infty} \log (\zeta + 1) \frac{(n_i\lambda_i)^{\zeta} e^{-n_i\lambda_i}}{\zeta!} - n_i\lambda_i$$

$$= n_i\lambda_i \left\{ \sum_{\zeta=0}^{\infty} \log (\zeta + 1) \frac{(n_i\lambda_i)^{\zeta} e^{-n_i\lambda_i}}{\zeta!} - 1 \right\}$$

For Poisson distribution with parameter $n_i\lambda_i$,

$$\frac{P(y = s + 1)}{P(y = s)} = \frac{(n_i\lambda_i)}{s + 1}$$

This is an increasing function of s, we get

$$\text{when} \quad s + 1 \leq n_i\lambda_i, \quad \frac{P(y = s + 1)}{P(y = s)} \geq 1, \quad \text{i.e.}, P(y = s + 1) \geq P(y = s)$$

$$\text{when} \quad s + 1 \geq n_i\lambda_i, \quad \frac{P(y = s + 1)}{P(y = s)} \leq 1, \quad \text{i.e.}, P(y = s + 1) \leq P(y = s)$$

so the mode lies on the point where $s = [n_i\lambda_i - 1]$, $[n_i\lambda_i - 1]$ is the nearest integer to $n_i\lambda_i - 1$.

In practice, we do not need to take ζ from 0 to ∞. We only care about a limited interval around the mode, the interval we use here is:

$$\{\text{Max}(0, mode - 10std), mode + 10std\}.$$

To calculate $t_i^{(M)}$, we just sum up $[y \log y - y]\frac{(n_i\lambda_i)^y e^{-n_i\lambda_i}}{y!}$ on interval $(\text{Max}(0, mode - 10std), mode + 10std)$. That is,

$$t_i^{(M)} \approx \sum_{y=a}^{b} [y \log y - y]\frac{(n_i\lambda_i)^y e^{-n_i\lambda_i}}{y!}.$$

Here $mode = \text{Max}(0, [n_i\lambda_i - 1])$, std is the standard deviation of ζ, $std \approx \sqrt{n_i\lambda_i}$. $a = \text{Max}(0, mode - 10std)$ and $b = mode + 10std$.

To calculate $\mu_i^{(M)}$, we use the sample mean as an approximation, $\mu_i^{(M)} \approx n_i\hat{\lambda}_i$. Here $\hat{\lambda}_i$ is the sample average of the 1000 iterates, $\sum_{h=1}^{1000} \lambda_i^{(h)}$.

After we calculate $\mu_i^{(M)}$ and $t_i^{(M)}$, plug these in the formula (4.5) give us the deviance.

4.2.2 Computing Marginal Likelihood

Another important measurement to assess the model is to compute marginal likelihood. For different g, we can easily get the joint posterior distribution of all the parameters and data. From this joint posterior density, if we can integrate all the parameters out, then we will get the marginal likelihood function of data given model, comparing this function value under different g will give us the best model.

The equation below explain this idea.

$$f_g(\underset{\sim}{d}) = \int f_g(\underset{\sim}{d}|\Omega)\pi(\Omega)d\Omega, \text{ for some } g$$

Here Ω is the vector of all model parameters, because $\underset{\sim}{\lambda}$ can be integrated out, so for $g = 1$, $\Omega = (\alpha, \underset{\sim}{\beta})$; for $g > 1$, $\Omega = (\alpha, \underset{\sim}{\beta}, \underset{\sim}{\mu}, p)$.

A problem is, the joint posterior density is usually very complicated, it is extremely hard to integrate all the parameters out. So we use Monte Carlo integration to solve this problem.

First construct the importance function $f_g^*(\underset{\sim}{d}, \Omega)$ by looking at the joint posterior density $f_g(\underset{\sim}{d}, \Omega)$, $f_g^*(\underset{\sim}{d}, \Omega)$ is a probability density function (integratable over Ω) which satisfy $f_g^*(\underset{\sim}{d}, \Omega) \approx f_g(\underset{\sim}{d}, \Omega)$ on the whole range of Ω, then we make a transformation

$$f_g(\underset{\sim}{d}) = \int f_g(\underset{\sim}{d}, \Omega) d\Omega = \int \frac{f_g(\underset{\sim}{d}, \Omega)}{f_g^*(\underset{\sim}{d}, \Omega)} f_g^*(\underset{\sim}{d}, \Omega) d\Omega$$

let

$$G_g(\underset{\sim}{d}, \Omega) = \frac{f_g(\underset{\sim}{d}, \Omega)}{f_g^*(\underset{\sim}{d}, \Omega)}$$

then the problem turned out to be finding $E(G_g(\underset{\sim}{d}, \Omega))$. Here

$$f_g(\Omega|\underset{\sim}{d}) = \frac{1}{\mathrm{D}_g(1, \ldots, 1)} \times (g-1)! \times \mathrm{P}(\underset{\sim}{\mu})$$

$$\times \frac{1}{\sqrt{(2\pi)^p \det(\Delta_0)}} e^{-\frac{1}{2}(\underset{\sim}{\beta} - \mu_{\underset{\sim}{\beta}})' \Delta_0^{-1} (\underset{\sim}{\beta} - \mu_{\underset{\sim}{\beta}})} \times \frac{1}{(1+\alpha)^2},$$

and

$$f_g^*(\underset{\sim}{d}, \Omega) = \mathrm{Dirichlet}_{\underset{\sim}{p}}(\hat{z}_1, \ldots, \hat{z}_g) \times \mathrm{P}(\underset{\sim}{\mu}) \times \mathrm{h}(\alpha, \underset{\sim}{\beta}|\underset{\sim}{d}).$$

Here $\mathrm{h}(\alpha, \underset{\sim}{\beta}|\underset{\sim}{d})$ is the proposal density in the Metropolis-Hastings step,

$$\begin{pmatrix} \alpha \\ \underset{\sim}{\beta} \end{pmatrix} \sim \mathrm{MN} \left\{ \begin{bmatrix} \hat{\alpha} \\ \hat{\underset{\sim}{\beta}} \end{bmatrix}, \begin{bmatrix} \sigma_\alpha^2 & \nu' \\ \nu & \Delta_\beta \end{bmatrix} \right\}$$

We draw α and $\underset{\sim}{\beta}$ from Gamma and multivariate-normal just as in the Metropolis-Hastings sampler step. Then plug these sample values into the above formula to evaluate $G_g(\underset{\sim}{d}, \Omega^{(h)})$. Finally, we estimate $E(G_g(\underset{\sim}{d}, \Omega))$ by

$$\widehat{E(G_g(\underset{\sim}{d}, \Omega))} = M^{-1} \sum_{h=1}^{M} G_g(\underset{\sim}{d}, \Omega^{(h)}),$$

$\Omega^{(h)}$ is the sample parameter vector, $h = 1, \ldots, M$, M is the sample size. We first use the results from Metropolis-Hasting sampler to construct the importance function $f_g^*(\underset{\sim}{d}, \Omega)$. Then draw a sample $\hat{\Omega}^{(1)}$, plug it into $G_g(\underset{\sim}{d}, \Omega)$ and get $G_g(\underset{\sim}{d}, \Omega^{(1)}$. Draw

another sample $\hat{\Omega}^{(2)}$ and repeat this process until reach a big enough sample size M (here we use 10000), we will get a good approximation.

Table 4.1 show the results from computing deviance and marginal likelihood, we choose up to $g = 8$ different models.

Table 4.1: Results From Model Selection Methods

g	$P_k(M)$	$G_k(M)$	Deviance	log Marginal Likelihood	N.S.E
1	1212.504	329.9967	1542.501	-3580.421	0.0029
2	1202.819	324.3637	1527.182	-3582.652	0.1222
3	1209.665	324.3541	1534.019	-3584.219	0.2219
4	1212.547	317.5014	1530.048	-3586.351	0.5064
5	1303.804	222.6683	1526.472	-3596.543	0.6426
6	1241.213	320.3492	1561.562	-3596.032	0.9976
7	1258.806	272.7744	1531.580	-3626.447	0.9451
8	1236.289	315.3004	1551.589	-3632.781	0.7983

The model selection criteria is, for deviance, choose the model which produce the minimum deviance, for marginal likelihood, choose the model which produce the maximum marginal likelihood.

Based on what the two measurements show, it is not clear what value of g to choose. The marginal likelihood indicate that $g = 1$ is fine, but it is not very different from $g = 2, 3$, or 4. The deviance indicates that $g = 5$ is the best, but it is not very different from $g = 2, 3$, or 4, even though the values at $g = 2, 3, 4, 5$ are smaller than the others. However, it is encouraging that for the likelihood $g = 1$ is not chosen. Therefore, for comparison, we consider $g = 1, 4, 5, 6$.

Chapter 5

Methods for Improved Maps

In this chapter we introduce the method of using posterior modes in mapping, for the purpose of comparison, the standard method (using posterior means) is also described.
Finally, we develop a new method in constructing map legends, which has some advantages over the standard method.

5.1 Disease Mapping by using Posterior Means

A standard method in disease mapping is to use the means of the mortality rates. The procedure is the following:

(a) Calculate the posterior means of the mortality rates in the 798 HSA's

(b) Order the 798 rates and cut them into K groups based on certain criteria

(c) Apply a different color to each group on the map

We obtained a Rao-Blackwellized map for the procedure above for any value of g. Based on the model in Chapter 3, conditional posterior density of the mortality rate

λ_i is written independently for $i = 1, \ldots, \ell$.

$$\pi(\lambda_i|\alpha, \underset{\sim}{\beta}, \underset{\sim}{\mu}, \underset{\sim}{p}, \underset{\sim}{d}) \propto$$

$$\frac{(n_i\lambda_i)^{d_i}e^{-n_i\lambda_i}}{(d_i)!} \times \left\{ \sum_{j=1}^{k} p_j \frac{(\alpha\frac{1-\mu_j}{\mu_j}e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}})^\alpha \lambda_i^{\alpha-1} e^{-\alpha\frac{1-\mu_j}{\mu_j}e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}}\lambda_i}}{\Gamma(\alpha)} \right\} \tag{5.1}$$

$$\pi(\lambda_i|\alpha, \underset{\sim}{\beta}, \underset{\sim}{\mu}, \underset{\sim}{p}, \underset{\sim}{d}) \propto \sum_{j=1}^{k} p_j \frac{(\alpha\frac{1-\mu_j}{\mu_j}e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}})^\alpha \lambda_i^{d_i+\alpha-1} n_i^{d_i} e^{-(n_i+\alpha\frac{1-\mu_j}{\mu_j}e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}})\lambda_i}}{\Gamma(\alpha)}$$

$$\propto \sum_{j=1}^{k} p_j \frac{(\alpha\frac{1-\mu_j}{\mu_j}e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}})^\alpha n_i^{d_i}}{[n_i + \alpha\frac{1-\mu_j}{\mu_j}e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}}]^{\alpha+d_i}} \times \frac{[n_i + \alpha\frac{1-\mu_j}{\mu_j}e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}}]^{\alpha+d_i}\lambda_i^{d_i+\alpha-1}e^{-(n_i+\alpha\frac{1-\mu_j}{\mu_j}e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}})\lambda_i}}{\Gamma(\alpha+d_i)}.$$

Letting

$$\omega_{ij} = \frac{\alpha\frac{1-\mu_j}{\mu_j}e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}}}{[n_i + \alpha\frac{1-\mu_j}{\mu_j}e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}}]},$$

we get

$$\pi(\lambda_i|\alpha, \underset{\sim}{\beta}, \underset{\sim}{\mu}, \underset{\sim}{p}, \underset{\sim}{d}) \propto \sum_{j=1}^{k} p_j\omega_{ij}^\alpha(1-\omega_{ij})^{d_i} \times \frac{[n_i + \alpha\frac{1-\mu_j}{\mu_j}e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}}]^{\alpha+d_i}\lambda_i^{d_i+\alpha-1}e^{-(n_i+\alpha\frac{1-\mu_j}{\mu_j}e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}})\lambda_i}}{\Gamma(\alpha+d_i)}.$$

Now we construct new weights

$$\Lambda_{ij} = \frac{p_j\omega_{ij}^\alpha(1-\omega_{ij})^{d_i}}{\sum_{j=1}^{k} p_j\omega_{ij}^\alpha(1-\omega_{ij})^{d_i}}$$

then

$$\pi(\lambda_i|\alpha, \underset{\sim}{\beta}, \underset{\sim}{\mu}, \underset{\sim}{p}, \underset{\sim}{d}) = \sum_{j=1}^{k} \Lambda_{ij}\text{Gamma}(\alpha+d_i, n_i + \alpha\frac{1-\mu_j}{\mu_j}e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}}). \tag{5.2}$$

From (4.6) we can get the posterior mean of mortality rates

$$E(\lambda_i|\alpha, \underset{\sim}{\beta}, \underset{\sim}{\mu}, \underset{\sim}{p}, \text{data}) = \sum_{j=1}^{k} \Lambda_{ij}\frac{\alpha+d_i}{n_i + \alpha\frac{1-\mu_j}{\mu_j}e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}}}. \tag{5.3}$$

These posterior means will be used in mapping, all the legends are constructed based on them.

5.2 Disease Mapping by using Posterior Mode

Drawing maps by using posterior means has been the standard practice. But for the highly dispersed data like COPD, since the distribution is often skewed, the mean is far away from the mode, so the use of means in mapping can produce biased results.

To fix this problem, we provide an alternative method, which is based on the posterior mode of the mortality rates. This method is actually more robust for different types of data, and the computation is not too expensive. The modes of the joint posterior density of mortality rates $\underset{\sim}{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_{798})$, is indeed the highest posterior density map. The procedure is to obtain the value of the posterior $\pi(\underset{\sim}{\lambda}|\underset{\sim}{d})$ at each of the 1000 iterates obtained from the Metropolis-Hastings sampler. Suppose we order $\pi(\underset{\sim}{\lambda}^{(h)}|\underset{\sim}{d})$, $h = 1, \ldots, M$, we draw the map for $\underset{\sim}{\lambda}^{(h*)}$ where $\underset{\sim}{\lambda}^{(h*)}$ maximize $\pi(\underset{\sim}{\lambda}|\underset{\sim}{d})$ over the M iterations.
The procedure is as follows:

(a) Draw a sample of the model parameters $\Omega^{(1)}, \Omega^{(2)}, \ldots, \Omega^{(M)}$, where $\Omega^{(i)}, i = 1, \ldots, M$ is a vector of all the model parameters $(\alpha, \underset{\sim}{\beta}, \underset{\sim}{\mu}, \underset{\sim}{p})$. As seen earlier, $M = 1000$ is a conventional choice.

(b) Draw $\underset{\sim}{\lambda}^{(1)} = (\lambda_1^{(1)}, \ldots, \lambda_{798}^{(1)})'$ from $\pi(\underset{\sim}{\lambda}|\underset{\sim}{d}, \Omega^{(1)})$, and calculate $\pi(\underset{\sim}{\lambda}^{(1)}|\underset{\sim}{d})$.

$\pi(\underset{\sim}{\lambda}|\underset{\sim}{d}, \Omega^{(1)})$ is the posterior conditional distribution of $\underset{\sim}{\lambda}$ given data and other parameters. Here,

$$\pi(\underset{\sim}{\lambda}^{(1)}|\underset{\sim}{d}) = \int \pi(\underset{\sim}{\lambda}^{(1)}|\underset{\sim}{d}, \Omega)\pi(\Omega|\underset{\sim}{d})\, d\Omega$$

$$\approx M^{-1} \sum_{h=1}^{M} \pi(\underset{\sim}{\lambda}^{(1)}|\underset{\sim}{d}, \Omega^{(h)})$$

where,

$$\pi(\lambda_i|d_i,\Omega)=\sum_{j=1}^{k}\Lambda_{ij}\Gamma(\alpha+d_i,n_i+\alpha\frac{1-\mu_j}{\mu_j}e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}}),$$

$$\Lambda_{ij}=\frac{p_j\omega_{ij}^{\alpha}(1-\omega_{ij})^{d_i}}{\sum_{j=1}^{k}p_j\omega_{ij}^{\alpha}(1-\omega_{ij})^{d_i}},$$

and

$$\omega_{ij}=\frac{\alpha\frac{1-\mu_j}{\mu_j}e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}}}{[n_i+\alpha\frac{1-\mu_j}{\mu_j}e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}}]}.$$

In fact, to make the computation easier, we work on the log scale, i.e., compute the $\log\pi(\underset{\sim}{\lambda}^{(1)}|\underset{\sim}{d})$ instead of $\pi(\underset{\sim}{\lambda}^{(1)}|\underset{\sim}{d})$. Then it turned out to be finding

$$M^{-1}\sum_{h=1}^{M}e^{\left\{\Sigma_{i=1}^{\ell}\log\pi(\underset{\sim}{\lambda}_i^{(1)}|\underset{\sim}{d},\Omega^{(h)})\right\}}$$

Below are the procedures we draw $\underset{\sim}{\lambda}^{(1)}$ from $\pi(\underset{\sim}{\lambda}|\underset{\sim}{d},\Omega^{(1)})$.

For the i^{th} HSA, calculate the weights Λ_{ij} using $\Omega^{(1)}$ and then calculate the accumulated weights of Λ_{ij}, as below

Cumulative Weights	
C_{i1}	Λ_{i1}
C_{i2}	$\Lambda_{i1}+\Lambda_{i2}$
C_{i3}	$\Lambda_{i1}+\Lambda_{i2}+\Lambda_{i3}$
.	
.	
$C_{i(k-1)}$	$\Lambda_{i1}+\Lambda_{i2}+,\ldots,+\Lambda_{i(k-1)}$
C_{ik}	1

Draw a random number μ from uniform distribution $U(0,1)$, if μ lies in the interval of the $(n-1)^{th}$ and n^{th} accumulated weights, i.e, $\Lambda_{i1}+,\ldots,+\Lambda_{i(n-1)}<\mu\leq\Lambda_{i1}+,\ldots,+\Lambda_{in}$ (here $n\leq k$), then we draw λ_i from $\Gamma(\alpha+d_i,n_i+\alpha\frac{1-\mu_n}{\mu_n}e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}})$.

(c) Draw $\underset{\sim}{\lambda}^{(h)}$ from $\pi(\underset{\sim}{\lambda}|\underset{\sim}{d},\Omega^{(h)})$, and calculate $\log\pi(\underset{\sim}{\lambda}^{(h)}|\underset{\sim}{d})$.

(d) Repeat (c) until we get all the 1000 $C^{(h)}=\log\pi(\underset{\sim}{\lambda}^{(h)}|\underset{\sim}{d}),h=1,\ldots,1000$

(e) Order $C^{(h)}$, find the maximum value C^M, the $\underset{\sim}{\lambda}$ correspond to C^M is chosen as the mode, draw the map based on this $\underset{\sim}{\lambda}$.

For our COPD data, we know the distributions of mortality rates are a little skewed, so the modes should not be very close to the means as in the symmetrical distribution cases. In Figure 5.1 and Figure 5.2 we show the difference between the posterior means and modes for the 798 HSA's, the variable to describe this difference is:

$$(\frac{\text{Mean}}{\text{Mode}} - 1) \times 100\%$$

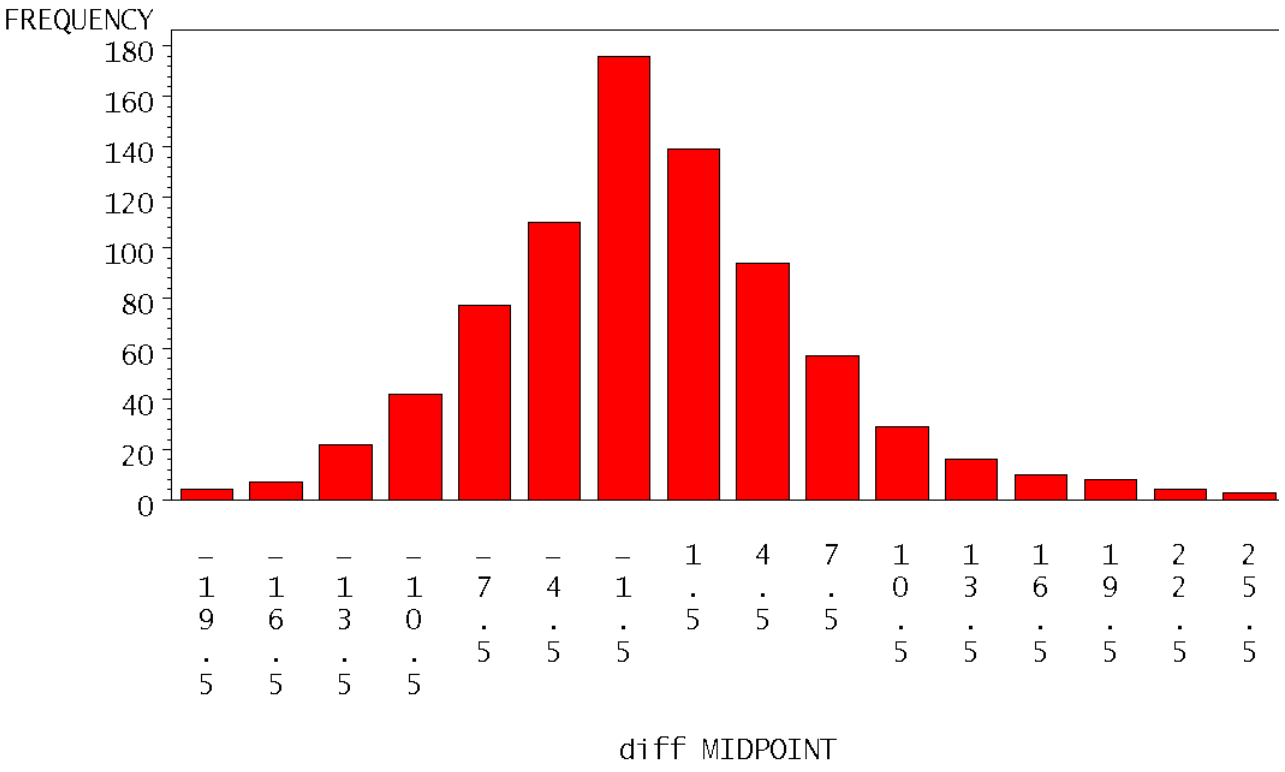

Figure 5.1: Percentage difference of posterior means and modes ($g = 1$)

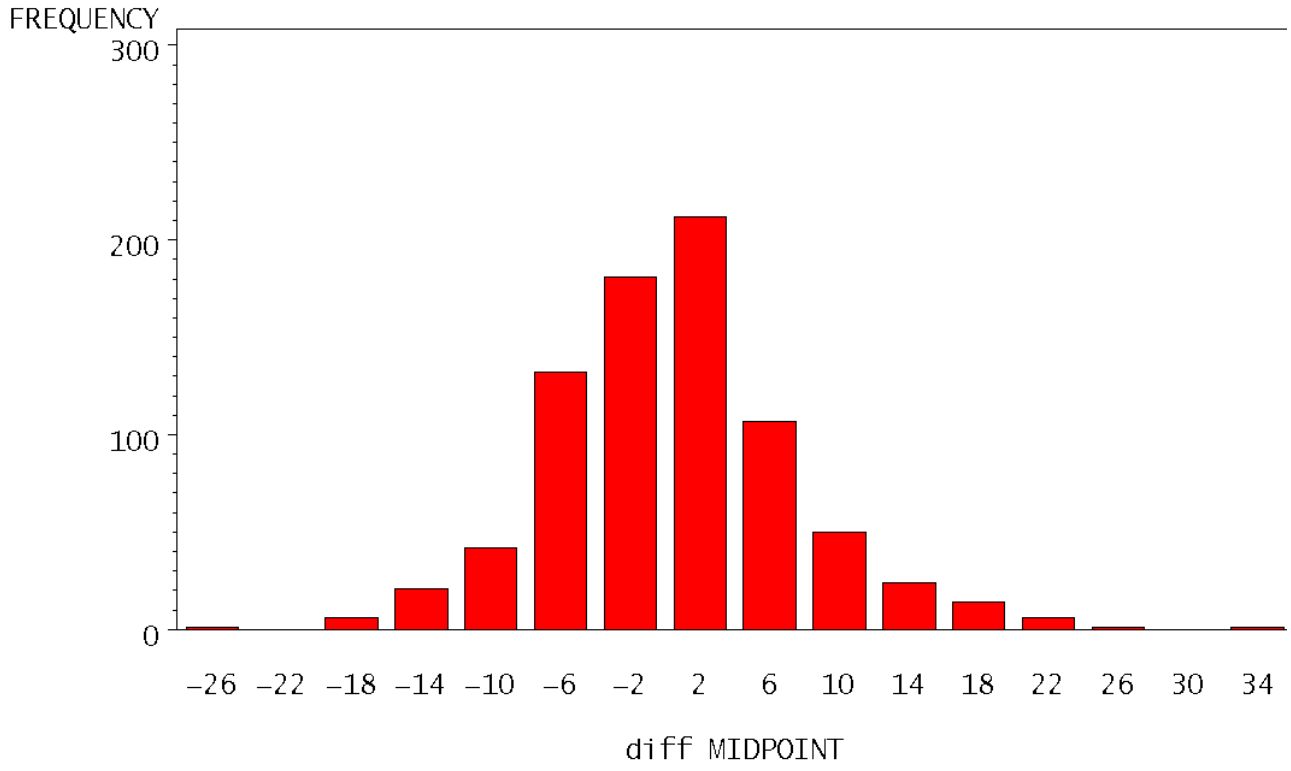

Figure 5.2: Percentage difference of posterior means and modes ($g = 5$)

From these two plots we can see, for both the model without latent class and the model with 5 latent classes, the difference between posterior means and posterior modes are apparent. There are some places where the differences are more than 20%.

5.3 New Legend Construction

The standard practice in legend construction is to use quantile of the posterior mean rates. This method is straightforward and convenient in computation. The computer package Arc view-GIS provide any number of quantiles, for example, quintiles. Then the data are divided into 5 groups each contain 20% of the total observations, ordered from the lowest to highest, a different color is assigned to each of the 5 groups and the map is drawn based on these colors. This method is effective in a lot of cases,

and will still be in use in practice. But a potential problem is, when the data is not evenly distribution in terms of magnitude, this method could provide misleading information.

An alternative method is to cut the data based on its own structure. As described in previous chapters, our model uses some parameters to measure the latent structure in the data. So we can build some "cutting points" from these model parameters.

The model parameters used to measure the latent structure are: $\underset{\sim}{\mu} = (\mu_1, \ldots, \mu_g)'$. They are ordered as:

$$\frac{1}{2} = \mu_1 \leq \mu_2 \leq \mu_3, \ldots, \leq \mu_g < 1.$$

Here $g \geq 2$, when $g = 1$, and we do not have mixture model.

We can see that, for a particular μ_k, it lies in the interval (μ_{k-1}, μ_{k+1}), μ_{k-1} and μ_{k+1} are the boundaries of μ_k, therefore, we can build "cutting points" based on these "boundary values".

Next, we show how to carry out these information about $\underset{\sim}{\mu}$ to the mortality rates $\underset{\sim}{\lambda_i}$. Suppose that we know that the i^{th} HSA belongs to group k, then

$$E(\lambda_i|\Omega, d_i, k) = \frac{\alpha + d_i}{n_i + \alpha\frac{1-\mu_k}{\mu_k}e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}}} \tag{5.4}$$

Transform (4.8), we can get

$$E(\lambda_i|\Omega, d_i, k) = \frac{\alpha + n_i(\frac{d_i}{n_i})}{n_i + \alpha\frac{1-\mu_k}{\mu_k}e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}}} = \frac{\frac{\alpha\frac{1-\mu_k}{\mu_k}e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}}}{\frac{1-\mu_k}{\mu_k}e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}}} + n_i(\frac{d_i}{n_i})}{n_i + \alpha\frac{1-\mu_k}{\mu_k}e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}}}$$

$$= \omega_{ik}\frac{\mu_k}{1-\mu_k}e^{\underset{\sim}{x}_i'\underset{\sim}{\beta}} + (1-\omega_{ik})(\frac{d_i}{n_i}) \tag{5.5}$$

where

$$\omega_{ik} = \frac{\alpha\frac{1-\mu_k}{\mu_k}e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}}}{[n_i + \alpha\frac{1-\mu_k}{\mu_k}e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}}]}.$$

Thus, to obtain the cut points for the map in (4.8), we replace ω_{ik} by arithmetic mean $\overline{\omega_k} = \ell^{-1} \sum_{i=1}^{\ell} \omega_{ik}$, $e^{x_{\sim i}'\beta}_{\sim}$ by their geometric mean $e^{\overline{x'}\beta}_{\sim}$, where $\overline{x'}_{\sim} = \ell^{-1} \sum_{i=1}^{\ell} x_{\sim i}'$, and $\frac{d_i}{n_i}$ by the overall average death rate $\frac{\sum d_i}{\sum n_i}$.

$$\overline{\omega_k} \frac{\mu_k}{1-\mu_k} e^{\overline{x'}\beta}_{\sim} + (1 - \overline{\omega_k})(\frac{\sum d_i}{\sum n_i})$$

Here $k = 2, \ldots, g$.

Table 5.1 lists all the cutting points for the COPD mortality rate under different model settings ($g = 2, \ldots, 6$). The table 5.2 shows the quantiles under three different model settings ($g = 4, 5, 6$). Compare the two tables for same model settings we can see, these breakpoints are different.

Table 5.1: Cutting Points

g	Cutting Points				
2	3.82×10^{-3}				
3	3.52×10^{-3}	3.62×10^{-3}			
4	3.59×10^{-3}	3.71×10^{-3}	4.04×10^{-3}		
5	3.59×10^{-3}	3.69×10^{-3}	3.84×10^{-3}	4.24×10^{-3}	
6	3.53×10^{-3}	3.61×10^{-3}	3.69×10^{-3}	3.82×10^{-3}	4.19×10^{-3}

Table 5.2: Quantiles

g	Cutting Points				
4	3.41×10^{-3}	3.81×10^{-3}	4.27×10^{-3}		
5	3.35×10^{-3}	3.67×10^{-3}	3.99×10^{-3}	4.33×10^{-3}	
6	3.23×10^{-3}	3.58×10^{-3}	3.88×10^{-3}	4.10×10^{-3}	4.47×10^{-3}

Chapter 6

Data Analysis and Map Comparison

In the previous chapters we developed our models, and then obtained parameter estimates and mortality rate estimates through Metropolis-Hastings sampler. Finally the model was checked and new methods in mapping were described. In this chapter, we use the methods in the previous chapters to analyze the COPD mortality rate, and use our new methods to draw maps. A comparison of the maps drawn from traditional method and new method is made.

6.1 Covariates Analysis

The Metropolis-Hastings sampler generates a sample of the regression coefficients, along with the standard deviation and 95% confidence interval. We can use these estimate to link the mortality rates with the risk factors, and even do some predictions. Table 6.1 contains the sample summary of these regression coefficients, for $g = 1, \ldots, 6$.

Table 6.1: Sample Summary for Regression Coefficients by the number of classes

g	β_j	Mean	Std. Dev.	N.S.E.	95% C.I.
1	β_0	-6.00506	0.03251	0.00609	(-6.07665, -5.93602)
1	β_1	0.00804	0.00045	0.00009	(0.00710, 0.00896)
1	β_2	-0.44775	0.04349	0.00862	(-0.54584, -0.35104)
1	β_3	0.00449	0.00039	0.00008	(0.00365, 0.00527)
1	β_4	-0.32746	0.05000	0.00951	(-0.43326, -0.23049)
2	β_0	-6.02452	0.03774	0.01751	(-6.10270, -5.94924)
2	β_1	0.00813	0.00046	0.00009	(0.00719, 0.00906)
2	β_2	-0.43798	0.04706	0.00988	(-0.53577, -0.34300)
2	β_3	0.00447	0.00041	0.00009	(0.00367, 0.00525)
2	β_4	-0.33307	0.05300	0.01033	(-0.44594, -0.22464)
3	β_0	-6.04060	0.04991	0.03451	(-6.16549, -5.95579)
3	β_1	0.00812	0.00045	0.00010	(0.00719, 0.00903)
3	β_2	-0.43371	0.05082	0.01211	(-0.54149, -0.32824)
3	β_3	0.00442	0.00041	0.00010	(0.00358, 0.00524)
3	β_4	-0.33716	0.05263	0.01214	(-0.44429, -0.22956)
4	β_0	-6.05964	0.05070	0.03448	(-6.16492, -5.97013)
4	β_1	0.00814	0.00045	0.00008	(0.00723, 0.00904)
4	β_2	-0.43683	0.05063	0.01243	(-0.54419, -0.33586)
4	β_3	0.00441	0.00041	0.00012	(0.00359, 0.00524)
4	β_4	-0.33990	0.05159	0.01065	(-0.43595, -0.23676)
5	β_0	-6.08932	0.06344	0.05192	(-6.22880, -5.98198)
5	β_1	0.00815	0.00048	0.00011	(0.00717, 0.00908)
5	β_2	-0.43464	0.05379	0.01176	(-0.55045, -0.32982)
5	β_3	0.00444	0.00042	0.00010	(0.00357, 0.00523)
5	β_4	-0.33874	0.05525	0.01297	(-0.43966, -0.23060)
6	β_0	-6.11109	0.05995	0.04841	(-6.25431, -6.01598)
6	β_1	0.00816	0.00047	0.00011	(0.00722, 0.00907)
6	β_2	-0.43533	0.05187	0.01032	(-0.54200, -0.33582)
6	β_3	0.00440	0.00042	0.00008	(0.00358, 0.00520)
6	β_4	-0.34017	0.05346	0.01195	(-0.45107, -0.23604)

• All the 95% credible intervals do not contain 0, means they are all significant. This confirm our conclusion in the exploratory data analysis.

• β_1, the coefficient for white male lung cancer mortality rate is positive. This confirms Morris and Munasinghe (1994); those places where more people smoke tend to have a higher COPD mortality rate.

• β_2, the coefficient for population density is negative, this confirms Nandram et al (2000). The possible reason might be, those places with a high population density usually have better medical services, and when there is an emergency, people living in a remote area are more likely to be delayed by the long travel to the nearest hospital.

• β_3, the coefficient for elevation is positive. This confirms extreme climatic conditions aggravate existing asthma and bronchitis (Bates 1989), as is living at high altitudes because of the reduced oxygen supply (Schoene 1999).

• β_4, the coefficient for the annual rainfall level is negative. As claimed before, repeated exposure to particulate matter and other air pollutants, primarily from traffic exhaust and coal-burning power plants, can aggravate existing lung conditions and can even cause death (Neutra et al. 1999; Sunyer et al. 2000). In particular, small airborne particles such as SO_2 found in urban air pollution can be deposited deep in the lungs, causing severe pulmonary effects (Sunyer et al. 2000; Schwartz and Neas 2000). Aerosolized toxins and viruses can be inhaled in dusty environments, causing pulmonary effects (Centers for Disease Control 1998). Rainfall, on the contrary, can lower the density of airborne particles and dust in the air, thus lower the chance of catching a pulmonary disease.

6.2 Map Comparisons

On the next few pages, we compare the maps drawn by our new methods and the maps drawn by using the standard methods.

In the order, the maps are respectively:

[Figure 6.1] Maps drawn with and without using latent class model.
These two maps both present posterior modes of mortality rates. In the map on top, the rates are estimated without using latent class model, but in the map on bottom, the rates are estimated by using our latent class regression model.

[Figure 6.2] Maps presenting posterior means and posterior modes.
There are two maps in this figure, for $g = 5$ (using five latent classes). As labeled, they present the posterior means and modes of the mortality rates respectively, map legends are constructed by our new method.

[Figure 6.3] Maps with legends constructed by standard and new methods.
As labeled, the maps present posterior modes, drawn from model using five latent classes, with legends constructed by both the traditional method and our new method.

[Figure 6.4] Maps by using the current methods and all of our new methods.
In the map on top, the mortality rates are estimated without using latent class model, then the posterior means are presented by quantile grouping method. In the map on the bottom, the mortality rates are estimated by using latent class model($g = 5$), then the posterior modes are presented, they are grouped by using our new legend construction method.

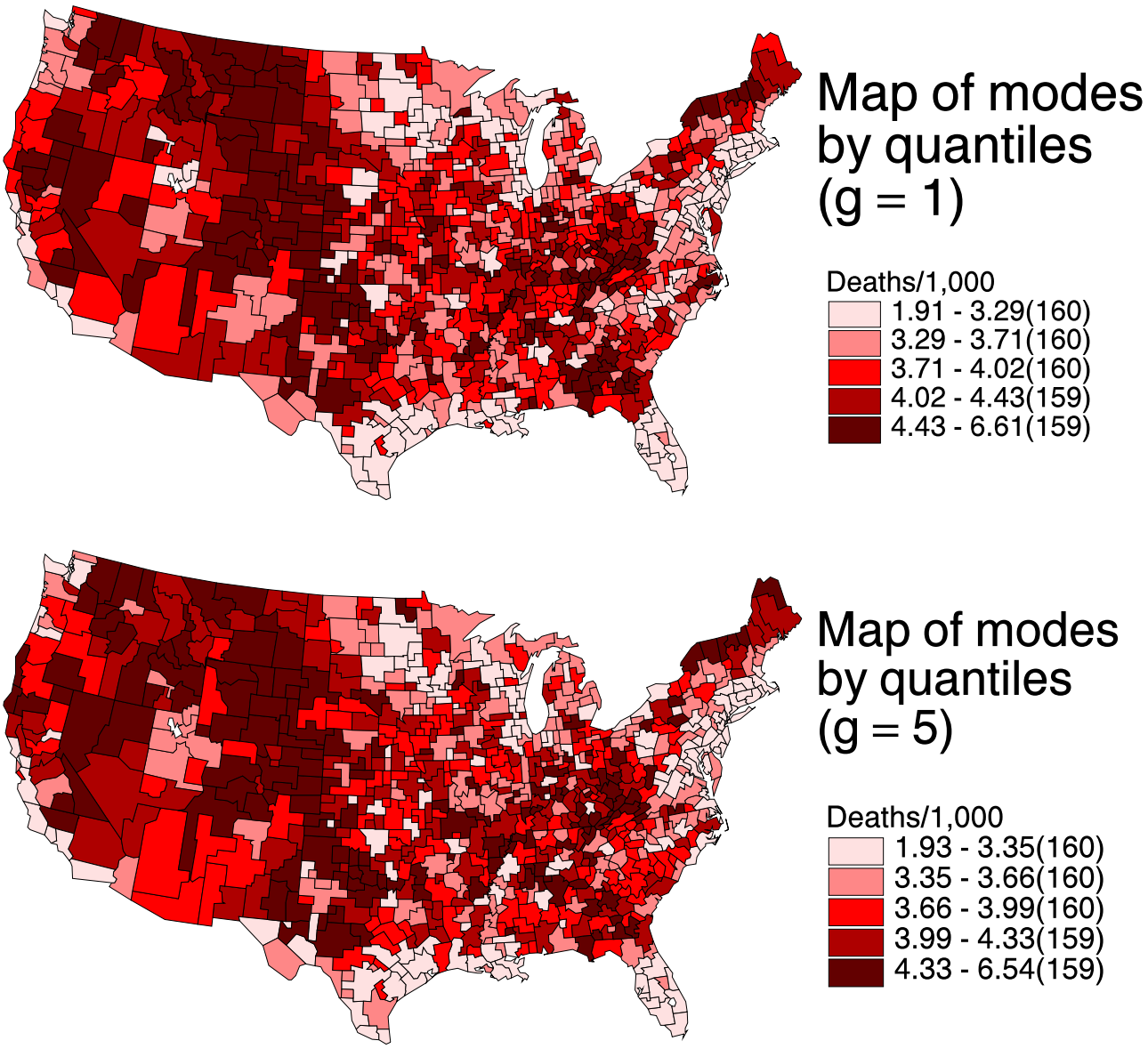

Figure 6.1: Comparison of models with and without latent classes

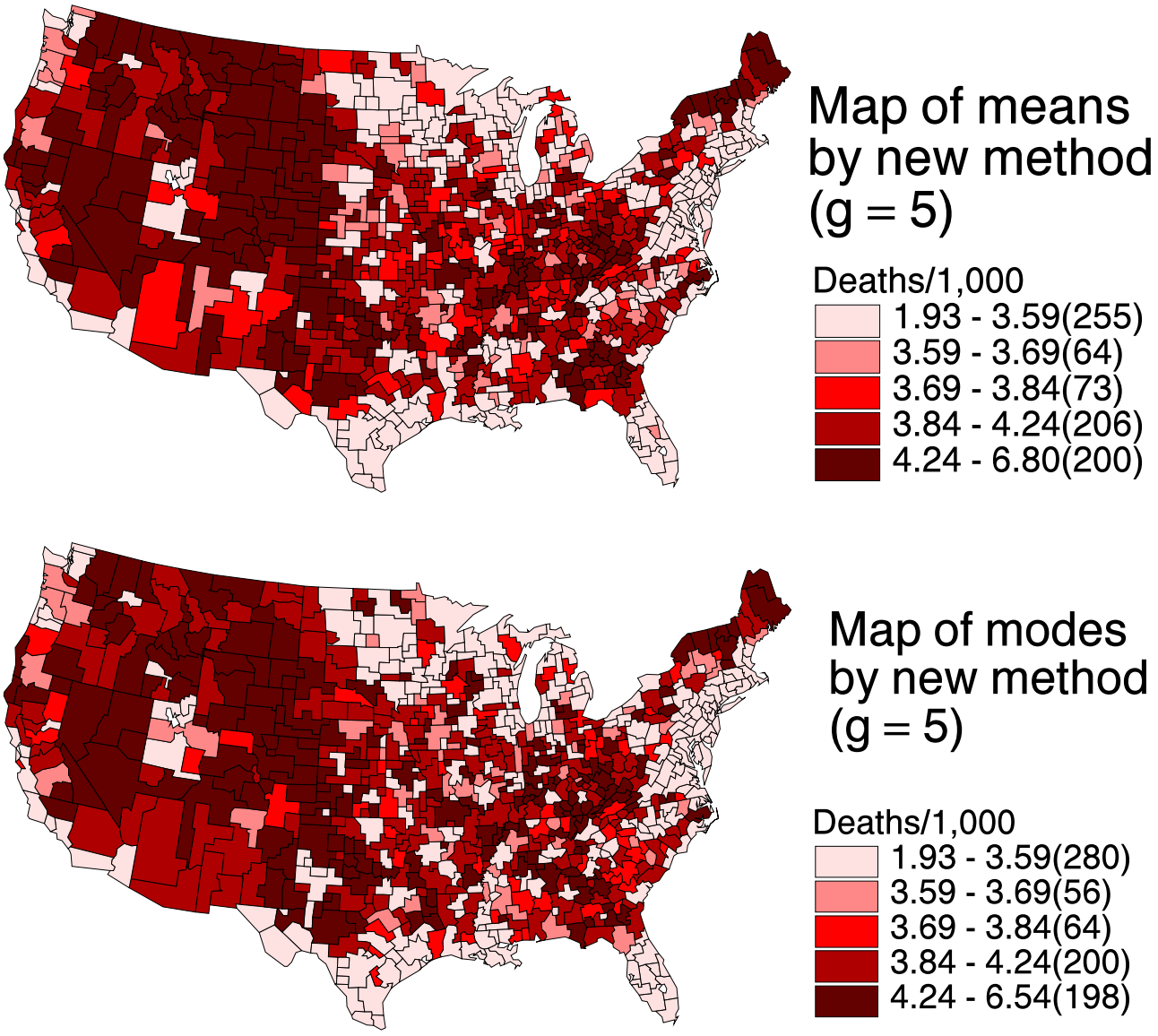

Figure 6.2: Comparison of maps using posterior means and modes

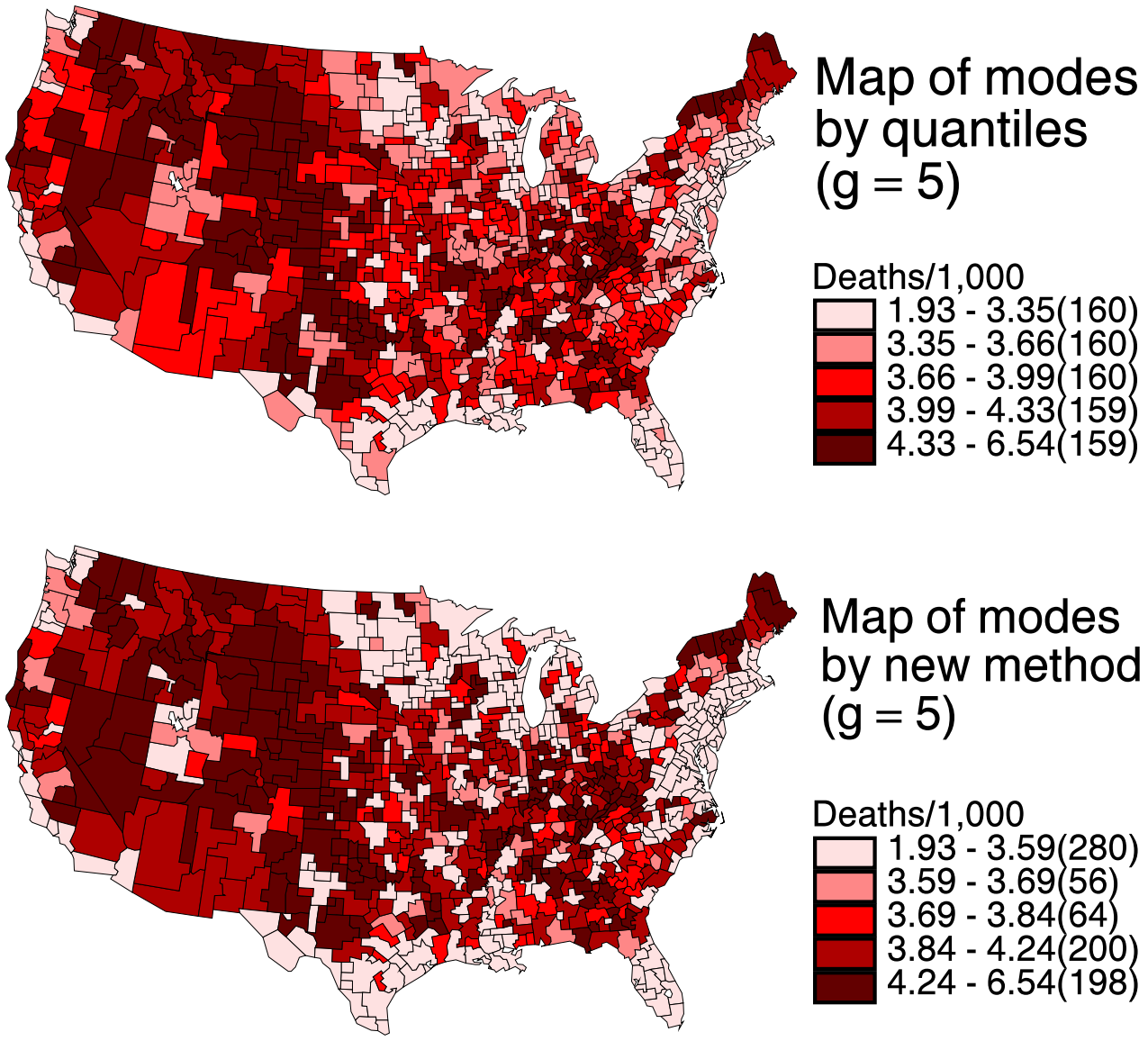

Figure 6.3: Comparison of map legend construction methods

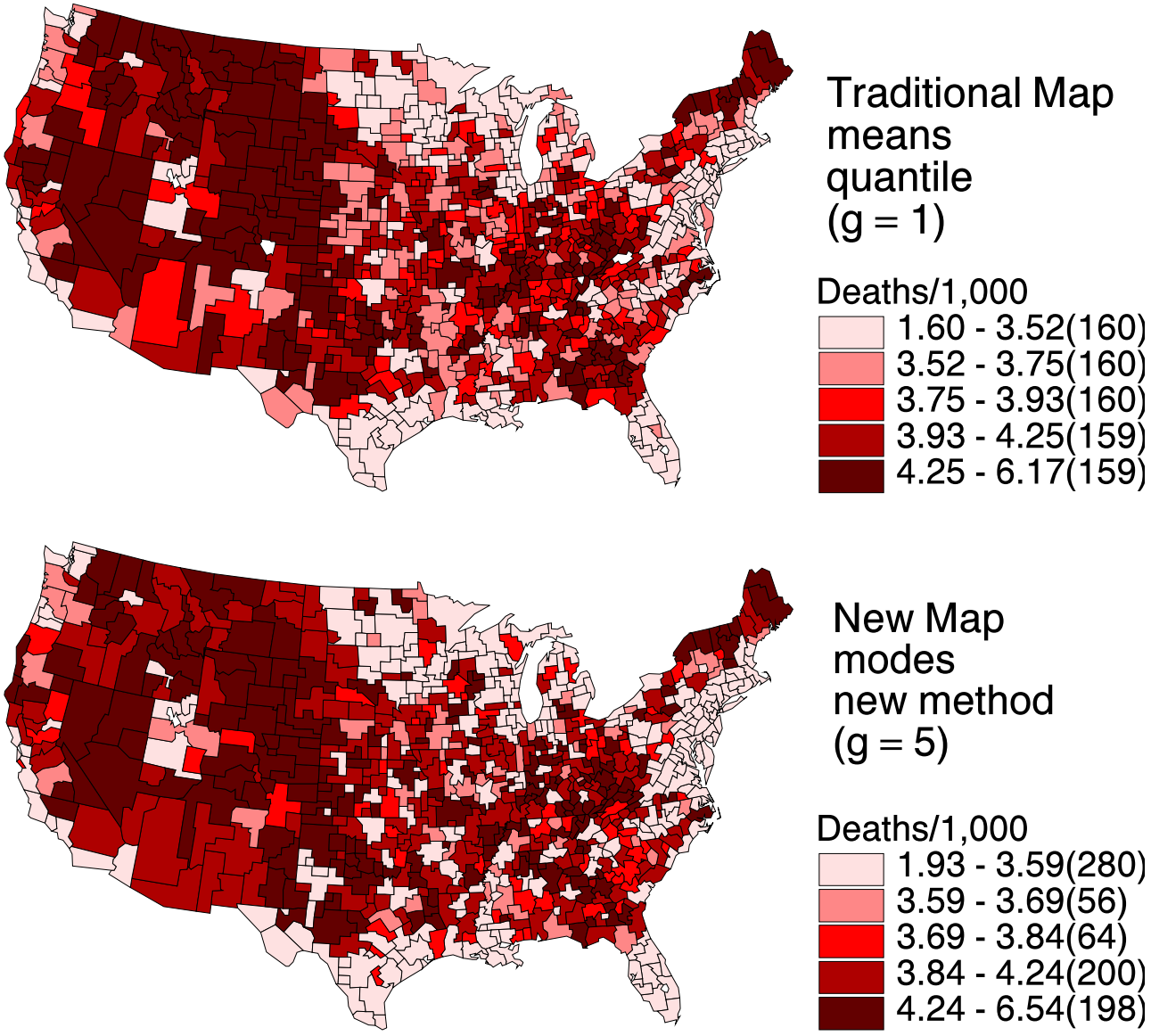

Figure 6.4: Comparison of map legend construction methods

From Figure 6.1, we can see that, in the two maps, even though the general patterns are the same, quite a few HSA's are presented by different colors.

Here are two examples of big changes across the maps. Henrico, VA and Prince Edward, VA HSA, turned to be much darker; St. Louis, MN and Douglas, WI HSA, turned to be lighter.

From Figure 6.2 we can see there are some differences between the mean map and mode map, even though they are drawn from exactly the same model, and the legends are constructed by the same method.

Here are some examples of big changes across the maps. Mercer, WV and Tazewell, VA HSA, is shown as among the lightest color in mean map, but presented by the darkest color in the mode map. The same happens to Fremont, CO - Custer, CO HSA. The Elko, NV and Lander, NV HSA, on the contrary, changed from darkest to lightest.

In Figure 6.3, the map on the top is constructed by the traditional quantile method, so there are equal number of HSA's in each color. The map at the bottom is drawn by using our new method, the numbers of HSA's in each color are shown beside the legend. By simply looking at these numbers, we can see how different the two maps are!

Here are some examples of the differences between the traditional method and the new method. Nueces, TX and San Patricio, TX HSA, changed from the second lightest color to the most lightest color. Maricopa, AZ and Yavapai, AZ HSA is changed from the third lightest color to the second lightest color.

Figure 6.4 is a comparison of the map drawn by the current standard practices, with the map drawn by using the three new methods provided in this thesis.

6.3 Conclusion and Future Research

The Bayesian hierarchical latent class model is good for modeling rare events such as COPD mortality rates, as shown in this thesis. The conventional method used in map legend construction is based on post-grouping of the estimated rates. That is, rates are first estimated, and then they are grouped by some criterion. There are many different grouping methods, each one is appropriate for a particular type of data. In practice it is hard to decide which method is appropriate. The new method we proposed eliminates this problem by using a grouping that is part of the model. Present posterior modes rather than posterior means is also superior when the rates are skewed, such as COPD mortality rates.

In future, we can do research in the following directions:

(1) Present variation in the map.

(2) Extend the model to include spatial-temporal components.

(3) Apply our model to other data. Although we have studied the COPD data, our model can be used to study any rare event, such as traffic accidents.

Appendix: A,B

Appendix A: Specify $\mu_{\underset{\sim}{\beta}}$ and Δ_0

To specify $\mu_{\underset{\sim}{\beta}}$ and Δ_0, start from the prior density for λ_i, since

$$\lambda_i|\alpha, \underset{\sim}{\beta} \overset{iid}{\sim} \text{Gamma}(\alpha, \alpha e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}}),$$

so we have

$$E(\lambda_i) = \frac{\alpha}{\alpha e^{-\underset{\sim}{x}_i'\underset{\sim}{\beta}}} = e^{\underset{\sim}{x}_i'\underset{\sim}{\beta}}.$$

take a log of both sides of this equation, and let $\lambda_i = e^{\tau_i}$, then we have $E(\tau_i) \approx \underset{\sim}{x}_i'\underset{\sim}{\beta}$. Fit a weighted least square regression on the data, we estimate the parameter coefficients as

$$\underset{\sim}{b} = \hat{\underset{\sim}{\beta}} = (X'W^{-1}X)^{-1}(X'W^{-1}Y),$$

where X is covariates matrix and the Y is the response vector of $\underset{\sim}{\tau}$.

The estimates $\underset{\sim}{b}$ are used in $\underset{\sim}{\beta}$'s prior distribution, taken as the means of $\underset{\sim}{\beta}$, i.e.

$$\mu_{\underset{\sim}{\beta}} = \underset{\sim}{b}.$$

By the same way, we get the covariance matrix by

$$\Delta_0 = (X'W^{-1}X)^{-1}\frac{(Y - Xb)'W^{-1}(Y - Xb)}{n - 2}$$

Appendix B: Proposed multivariate-normal density for parameter vector $(\alpha, \underset{\sim}{\beta})$', for general g.

First, take a logarithm of the real density function

$$\Delta(\alpha, \underset{\sim}{\beta}) = \sum_{i=1}^{\ell} \left\{ \alpha \sum_{j=1}^{g} \phi_{ij} \log(\frac{1-\mu_j}{\mu_j}) + (\alpha - 1) \log \lambda_i - \alpha e^{-\underset{\sim}{x}_i' \underset{\sim}{\beta}} \lambda_i \sum_{j=1}^{g} (\frac{1-\mu_j}{\mu_j} \phi_{ij}) \right\}$$

$$+ \sum_{i=1}^{\ell} \{\alpha[\log \alpha - \underset{\sim}{x}_i' \underset{\sim}{\beta}] - \log \Gamma(\alpha)\} - 2 \log(\alpha + 1)$$

$$-\frac{1}{2} \log((2\pi)^p \det(\Delta_0)) - \frac{1}{2} (\underset{\sim}{\beta} - \mu_{\underset{\sim}{\beta}})' \Delta_0^{-1} (\underset{\sim}{\beta} - \mu_{\underset{\sim}{\beta}}) \tag{6.1}$$

Then, we take the first, second and cross derivatives of $\Delta(\alpha, \underset{\sim}{\beta})$ against α and $\underset{\sim}{\beta}$, form a derivative matrix called Hessian Matrix.

$$\frac{\partial \Delta}{\partial \alpha} = \sum_{i=1}^{\ell} \left\{ \sum_{j=1}^{g} \phi_{ij} \log(\frac{1-\mu_j}{\mu_j}) + \log \lambda_i - e^{-\underset{\sim}{x}_i' \underset{\sim}{\beta}} \lambda_i \sum_{j=1}^{g} (\frac{1-\mu_j}{\mu_j} \phi_{ij}) \right\}$$

$$\sum_{i=1}^{\ell} \{\log \alpha - \underset{\sim}{x}_i' \underset{\sim}{\beta} + 1 - \Phi(\alpha)\} - 2 \frac{1}{1+\alpha} \tag{6.2}$$

$$\frac{\partial^2 \Delta}{\partial \alpha^2} = \sum_{i=1}^{\ell} \left\{ \frac{1}{\alpha} - \Psi(\alpha) \right\} + 2 \frac{1}{(1+\alpha)^2} \tag{6.3}$$

here $\Phi(\alpha)$ and $\Psi(\alpha)$ are the first and second derivative of $\log \Gamma(\alpha)$.

$$\frac{\partial \Delta}{\partial \beta_k} = \sum_{i=1}^{\ell} \left\{ -\alpha x_{k(i)} + \alpha x_{k(i)} e^{-\underset{\sim}{x}_i' \underset{\sim}{\beta}} \lambda_i \sum_{j=1}^{g} (\frac{1-\mu_j}{\mu_j} \phi_{ij}) \right\}$$

$$- \frac{\partial \left\{ \frac{1}{2} (\underset{\sim}{\beta} - \mu_{\underset{\sim}{\beta}})' \Delta_0^{-1} (\underset{\sim}{\beta} - \mu_{\underset{\sim}{\beta}}) \right\}}{\partial \beta_k} \tag{6.4}$$

$$\frac{\partial^2 \Delta}{\partial \beta_{k1} \partial \beta_{k2}} = \sum_{i=1}^{\ell} \left\{ -\alpha x_{k1(i)} x_{k2(i)} e^{-\underset{\sim}{x}_i' \underset{\sim}{\beta}} \lambda_i \sum_{j=1}^{g} \left(\frac{1-\mu_j}{\mu_j} \phi_{ij} \right) \right\} - \frac{1}{{\sigma_{k1k2}}^2} \tag{6.5}$$

$$\frac{\partial^2 \Delta}{\partial \alpha \partial \beta_k} = \sum_{i=1}^{\ell} \left\{ -x_{k(i)} + x_{k(i)} e^{-\underset{\sim}{x}_i' \underset{\sim}{\beta}} \lambda_i \sum_{j=1}^{g} \left(\frac{1-\mu_j}{\mu_j} \phi_{ij} \right) \right\} \tag{6.6}$$

We get the covariance matrix of $(\alpha, \underset{\sim}{\beta})'$ by take the inverse of the Hessian matrix.

$$\Sigma = \begin{bmatrix} \sigma_\alpha^2 & \nu' \\ \nu & \Delta_\beta \end{bmatrix} = \begin{bmatrix} -\frac{\partial^2 \Delta}{\partial \alpha^2} & -\frac{\partial^2 \Delta}{\partial \alpha \partial \beta} \\ -\frac{\partial^2 \Delta}{\partial \alpha \partial \beta} & -\frac{\partial^2 \Delta}{\partial \beta_{k1} \partial \beta_{k2}} \end{bmatrix}^{-1} \tag{6.7}$$

Where the ν is a covariance vector related to $\frac{\partial^2 \Delta}{\partial \alpha \partial \beta}$ and the Δ_β is a covariance matrix related to $\frac{\partial^2 \Delta}{\partial \beta_{k1} \partial \beta_{k2}}$.

A special case: $g = 1$.

When $g = 1$, the number of parameters in the model is reduced greatly, thus the computation is also simplified a lot.

First, take a logarithm of the real density function.

$$\Delta(\alpha, \underset{\sim}{\beta}) = \sum_{i=1}^{\ell} \Big\{ \alpha [\log \alpha - \underset{\sim}{x}_i' \underset{\sim}{\beta}] + (\alpha - 1) \log \lambda_i - \alpha e^{-\underset{\sim}{x}_i' \underset{\sim}{\beta}} \lambda_i - \log \Gamma(\alpha) \Big\}$$

$$-2 \log (\alpha + 1) - \frac{1}{2} \log((2\pi)^p \det(\Delta_0)) - \frac{1}{2} (\underset{\sim}{\beta} - \mu_{\underset{\sim}{\beta}})' \Delta_0^{-1} (\underset{\sim}{\beta} - \mu_{\underset{\sim}{\beta}}) \tag{6.8}$$

Then, we take the first, second and cross derivatives of $\Delta(\alpha, \underset{\sim}{\beta})$ against α and $\underset{\sim}{\beta}$, form a derivative matrix called Hessian Matrix.

$$\frac{\partial \Delta}{\partial \alpha} = \sum_{i=1}^{\ell} \Big\{ \log \alpha - \underset{\sim}{x}_i' \underset{\sim}{\beta} + 1 + \log \lambda_i - e^{-\underset{\sim}{x}_i' \underset{\sim}{\beta}} \lambda_i - \Phi(\alpha) \Big\}$$

$$-2 \frac{1}{1+\alpha} \tag{6.9}$$

$$\frac{\partial^2 \Delta}{\partial \alpha^2} = \sum_{i=1}^{\ell} \left\{ \frac{1}{\alpha} - \Psi(\alpha) \right\} + 2\frac{1}{(1+\alpha)^2} \tag{6.10}$$

here $\Phi(\alpha)$ and $\Psi(\alpha)$ are the first and second derivative of $\log \Gamma(\alpha)$.

$$\frac{\partial \Delta}{\partial \beta_k} = \sum_{i=1}^{\ell} \left\{ -\alpha x_{k(i)} + \alpha x_{k(i)} e^{-\underset{\sim}{x}_i' \underset{\sim}{\beta}} \lambda_i \right\} - \frac{\partial \left\{ \frac{1}{2} (\underset{\sim}{\beta} - \mu_{\underset{\sim}{\beta}})' \Delta_0^{-1} (\underset{\sim}{\beta} - \mu_{\underset{\sim}{\beta}}) \right\}}{\partial \beta_k} \tag{6.11}$$

$$\frac{\partial^2 \Delta}{\partial \beta_{k1} \partial \beta_{k2}} = \sum_{i=1}^{\ell} \left\{ -\alpha x_{k1(i)} x_{k2(i)} e^{-\underset{\sim}{x}_i' \underset{\sim}{\beta}} \lambda_i \right\} - \frac{1}{{\sigma_{k1k2}}^2} \tag{6.12}$$

$$\frac{\partial^2 \Delta}{\partial \alpha \partial \beta_0} = \sum_{i=1}^{\ell} \left\{ -x_{k(i)} + x_{k(i)} \lambda_i e^{-\underset{\sim}{x}_i' \underset{\sim}{\beta}} \right\} \tag{6.13}$$

We get the covariance matrix of $(\alpha, \underset{\sim}{\beta})'$ by take the inverse of the Hessian matrix.

$$\Sigma = \begin{bmatrix} \sigma_\alpha^2 & \nu' \\ \nu & \Delta_\beta \end{bmatrix} = \begin{bmatrix} -\frac{\partial^2 \Delta}{\partial \alpha^2} & -\frac{\partial^2 \Delta}{\partial \alpha \partial \beta} \\ -\frac{\partial^2 \Delta}{\partial \alpha \partial \beta} & -\frac{\partial^2 \Delta}{\partial \beta_{k1} \partial \beta_{k2}} \end{bmatrix}^{-1} \tag{6.14}$$

Where the ν is a covariance vector related to $\frac{\partial^2 \Delta}{\partial \alpha \partial \beta}$ and the Δ_β is a covariance matrix related to $\frac{\partial^2 \Delta}{\partial \beta_{k1} \partial \beta_{k2}}$.

References

Bates, D.V. (1989), *Respiratory Function in Disease*, Philadelphia: W.B.Saunders.

Burr, M.L., Anderson, H.R., Austin, J.B., Harkins, L.S., Kaur, B., Strachan, D.P., and Warner, J.O. (1999), Respiratory Symptoms and Home Environment in Children: A National Survey, *Thorax*, 54, 27-32.

Christiansen, C.L., Morris, C.N. (1997), Hierarchical Poisson Regression Modeling, *Journal of the American Statistical Association*, Vol. 92, 618-631.

English, P., Neutra, R., Scalf, R., Sullivan, M., Waller, L., and Zhu, L. (1999), "Examining Associations Between Childhood Asthma and Traffic Flow Using a Geographic Information System," *Environmental Health Perspectives*, 107, 761-767.

Escobar, M.D., West, M (1998), Computing nonparametric hierarchical models, pp. 1-22 in Dey, Dipak *et al.* (eds), *Practical nonparametric and semiparametric Bayesian statistics*. Lecture Notes in Statistics, 133. New York: Springer.

Gelfand, A., and Ghosh, S. (1998), "Model Choice: A Minimum Posterior Predictive Approach," *Biometrika*, 85, 1-11.

Gelfand, A.E., Sahu, S.K. (1999), Identifiability, Improper Priors, and Gibbs Sampling for Generalized Linear Models, *Journal of the American Statistical Association*, Vol. 94, 247-253.

Lavine, M., West, M. (1992), A Bayesian method for classification and discrimination, *Canadian Journal of Statistics*, Vol. 20, 451-461.

Laird, N. (1982), Empirical Bayes estimates using the nonparametric maximum likelihood estimate for the prior, *Journal of Statistical Computation and Simulation*, Vol. 15, 211-220.

MacEachren, A.M., Brewer, C.A., and Pickle, L.W. (1998), "Visualizing Georeferenced Data: Representing Reliability of Health Statistics," *Environment and Planning A*, 30, 1547-1561.

McConnochie, K.M., Russo, M.J., McBride, J.T., Szilagyi, P.G., Brooks, A.M., and Roghmann, K.J. (1999), "Socioeconomic Variation in Asthma Hospitalization: Excess Utilization or Greater Need?" *Pediatrics*, 103, e75.

Morris, R.D., and Munasinghe, R.L. (1994), "Geographic Variability in Hospital Admission Rates for Respiratory Disease Among the Elderly in the United States" *Chest*, 106, 1172-1181.

Militino, A.F., Ugarte, M.D., Dean, C.B. (1997), *Statistics In Medicine*, Vol. 20, 2035-2049.

Muller, P., Rosner, G.L. (1997), A Bayesian Population Model with Hierarchical Mixture Priors Applied to Blood Count Data, *Journal of the American Statistical Association*, Vol. 92, 1279-1291.

Nandram, B. (2000), Bayesian Generalized Linear Models for Inference about Small Areas, *Generalized Linear Models: A Bayesian Perspective, eds.* D.K., Dey, S.K., Ghosh, and B.K., Mallick, New York: Marcel Dekker, pp. 91-114.

Nandram, B., Sedransk, J., Pickle L.W. (1999), Bayesian Analysis of Mortality Rates for U.S. Health Service Areas, *Sankhya*, B, 61, 145-165.

Nandram, B., Sedransk, J., Pickle L.W. (2000), Bayesian Analysis and Mapping of Mortality Rates for Chronic Obstructive Pulmonary Disease, *Journal of the American Statistical Association*, Vol. 95, 1110-1118.

Nizich, S., Pope, A., and the Pechan-Avanti Group (1998), *National Air Pollutant Emission Trends Update, 1970-1997*, EPA/454-E-98-007. Washington, DC: U.S. Environmental Protection Agency.

Pickle, L.W., Mungiole, M., Jones, G.K., and White, A.A. (1996), *Atlas of United States Mortality*, Hyattsville, MD: National Center for Health Statistics.

Richardson, S., Green, P. (1997), On Bayesian analysis of mixtures with an unknown number of components, *Journal of the Royal Statistical Society*, Series B, 59, 731-792.

Roeder, K., Wasserman, L. (1997), Practical Bayesian Density Estimation Using Mixtures of Normals, *Journal of the American Statistical Association*, Vol. 92, 894-902.

Schwartz, J., and Neas, L.M. (2000), "Fine Particles are More Strongly Associated Than Coarse Particles With Acute Respiratory Health Effects in Scholchildren," *Epidemiology*, 11, 6-10.

Sunyer, J., Schwartz, J., Tobias, A., Macfarlane, D., Garcia, J., and Anto, J.M. (2000), "Patients With Chronic Obstructive Pulmonary Disease are at Increased Risk of Death Associated With Urban Particle Air Pollution: A Case-Crossover Analysis," *American Journal of Epidemiology,* 151, 50-56.

Schlattmann, P., Bohning, D. (1993), Mixture Models and Disease Mapping, *Statistics In Medicine*, Vol. 12, 1943-1950.

Schoene, R.B. (1999), "Lung Disease at High Altitude," *Advances in Experimental Medicine and Biology*, 474, 47-56.