Published in final edited form as:

Biometrics. 2010 June ; 66(2): 613–620. doi:10.1111/j.1541-0420.2009.01284.x.

Covariate Adjustment and Ranking Methods to Identify Regions with High and Low Mortality Rates

Huilin Li*, Barry I. Graubard, and Mitchell H. Gail

Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Executive Plaza South, Room 8032, Bethesda, MD 20892-7244

SUMMARY

Identifying regions with the highest and lowest mortality rates and producing the corresponding color-coded maps help epidemiologists identify promising areas for analytic etiologic studies. Based on a two-stage Poisson-Gamma model with covariates, we use information on known risk factors, such as smoking prevalence, to adjust mortality rates and reveal residual variation in relative risks that may reflect previously masked etiologic associations. In addition to covariate adjustment, we study rankings based on standardized mortality ratios, empirical Bayes estimates, and a posterior percentile ranking method and indicate circumstances that warrant the more complex procedures in order to obtain a high probability of correctly classifying the regions with the upper 100 γ % and lower 100 γ % of relative risks for $\gamma = 0.05, 0.1$ and 0.2 . We also give analytic approximations to the probabilities of correctly classifying regions in the upper 100 γ % of relative risks for these three ranking methods. Using data on mortality from heart disease, we found that adjustment for smoking prevalence has an important impact on which regions are classified as high and low risk. With such a common disease, all three ranking methods performed comparably. However, for diseases with smaller event counts, such as cancers, and wide variation in event counts among regions, empirical Bayes and posterior percentile ranking methods outperform ranking based on standardized mortality ratios.

Keywords

Incidence rate maps; Mortality rate maps; Poisson-Gamma model; Ranking procedures; Standardized incidence ratio; Standardized mortality ratio

1. Introduction

Epidemiologists study maps of disease incidence rates and mortality rates to find clues to disease etiology. Regions with the highest and lowest rates are of particular interest, because analytic epidemiologic studies can be conducted in such regions to identify etiologic factors. Standardized incidence ratios (SIRs) or standardized mortality ratios (SMRs), namely the ratio of the number of deaths, O_i , in region i divided by the expected deaths, E_i , are often ranked, and the corresponding regions are color coded according to decile or other quantile of the rank. One often focuses on the regions with the highest risk (Devesa et al., 1999).

Two important issues arise in identifying regions with the highest or lowest mortality rates. (Hereafter we refer to mortality rates, but the same ideas apply to incidence rates). One issue is whether the variation in SMRs is partly explained by known factors such as smoking prevalence. One would like to adjust for such factors to identify regions with the highest or

*lih5@nih.gov.

lowest residual mortality rates where one might find new etiologic factors apart from e.g. smoking. A second issue is what procedures to use for ranking. The SMR is unstable in areas with small values of E_i , and a two-stage Poisson-Gamma model (Clayton and Kaldor, 1987; Pennello, Devesa, and Gail, 1999) can be adapted to produce smoother Bayes or empirical Bayes (EB) estimates for ranking the regions. Although EB estimates produce good estimates of the ranks overall (Laird and Louis, 1989), they are not optimal for identifying the regions with the upper 100 γ % of rates (Lin et al., 2006). For that purpose, Lin et al. (2006) recommended ranking the posterior probability that a region's rank will be in the top 100 γ % of ranks, which we call fully Bayesian posterior percentile ranking (B-PPR). In this paper, we propose an empirically approximated posterior percentile ranking procedure (PPR) that implicitly assumes vague priors on certain parameters, and we compare PPR with B-PPR. We use the abbreviation PPR for ranking either in the upper 100 γ % of ranks or lower 100 γ % of ranks and the abbreviations PPHR and PPLR for the upper 100 γ % and lower 100 γ % of ranks respectively.

We describe methods for covariate adjustment for the SMR and for the EB and PPR approaches under the Poisson-Gamma model (Section 2) and compare these procedures in an analysis of heart disease mortality data (Section 3). In this example, we find that covariate adjustment has a greater impact than the choice of ranking method, which motivates simulations and analytic studies to determine under what circumstance the choice of ranking method has a material impact (Section 4). The choice of ranking method is only important when there is wide variation in E_i . In this circumstance, PPR outperforms EB slightly, and both can be noticeably better than SMR if the expected variance of the SMR is large. We present accurate analytic approximations to guide the choice of ranking procedure, both for the Poisson-Gamma and Gaussian-Gaussian models (Section 4). We conclude with a discussion in Section 5.

2. Notation and Methods for Covariate Adjustment and Ranking

Consider I regions, $i = 1, 2, \dots, I$ with J age-groups $j = 1, \dots, J$ in each. Let y_{ij} be the number of person-years exposure and O_{ij} be the observed deaths in region i and age-group j . Note that j might also index a cross-classification on age, race and gender. Suppose we know the smoking prevalence P_{ij} in region i and age-group j . Then the expected numbers of deaths are modeled as

$$E_{ij} = y_{ij} \xi_j \{P_{ij}r + (1 - P_{ij})\}, \quad (1)$$

where $\{\xi_j\}$, the age-specific mortality rates for non-smokers, and r , the relative risk associated with smoking, are to be estimated.

We would like to know the unmeasured random effect θ_i which is assumed to come from a gamma distribution $\Gamma(\alpha, 1/\alpha)$ with mean $\alpha(1/\alpha)=1$ and variance $1/\alpha$, where α is unknown. The quantity θ_i represents the intrinsic relative risk (or frailty) of region i apart from the measured factors encompassed in E_{ij} . Ideally we would be able to rank the $\{\theta_i\}$ and identify those regions with ranks in the upper 100 γ % of ranks and those regions with ranks in the bottom 100 γ % of ranks. Conditional on θ_i , O_{ij} have the Poisson distribution with mean $E_{ij}\theta_i$,

$$O_{ij}|\theta_i \sim \text{Poi}(E_{ij}\theta_i). \quad (2)$$

We assume θ_i is independent of E_{ij} . Thus, unconditionally, $E(O_{ij}) = E_{ij}E(\theta_i) = E_{ij}$. This model for covariates emphasizes that we wish to identify regions with high and low intrinsic θ_i apart from measured factors such as age and smoking prevalence. Clayton and Kaldor

(1987) and Christiansen and Morris (1997) incorporate covariates into the mean of θ_i . If they had used the mixture model (1), their methods would yield the same marginal distribution of O_{ij} as we obtained, namely the negative binomial,

$$O_{ij} \sim \text{NB}(\text{success}=\alpha, \text{prob}=\alpha/(\alpha+E_{ij})), \quad (3)$$

with mean = success(1-prob)/prob and variance = success(1-prob)/prob². The marginal likelihood is the product over regions and age groups of (3), which can be maximized to yield maximum likelihood estimates $\hat{\psi}=\{\hat{\alpha}, \hat{r}$ and $\hat{\xi}_j=\exp(\hat{\delta}_j)\}$. The parameterization in terms of δ_j avoids numerical instabilities. Assuming $\psi=\{\alpha, r$ and $\xi_j\}$ are known, the posterior distribution of θ_i given $O_i=\sum_j O_{ij}$ is

$$\theta_i | O_i \stackrel{\text{iid}}{\sim} \Gamma\{O_i+\alpha, 1/(E_i+\alpha)\}. \quad (4)$$

Thus the EB estimate of θ_i is

$$\hat{\theta}_i^{\text{EB}} = (O_i + \hat{\alpha}) / (\hat{E}_i + \hat{\alpha}), \quad (5)$$

where $\hat{E}_i = \sum_j \hat{E}_{ij}$, and $\hat{E}_{ij}$ is obtained from (1) with $\hat{r}$ and $\hat{\xi}_j$ in place of r and ξ_j .

The fully Bayesian posterior percentile ranking (B-PPR) method was first proposed in Lin et al. (2006). They assumed a Poisson-Gamma model with vague and independent priors on the unknown parameters. To implement their procedures, we use $\alpha \sim \text{LN}(3.69, 100)$, $r \sim \text{LN}(0.69, 100)$, and $\xi \stackrel{\text{iid}}{\sim} \text{LN}(-6, 100)$. Here $\text{LN}(\eta, \phi^2)$ is the lognormal density of a random variable whose logarithm has mean η and variance ϕ^2 . The top 100 γ % highest risk regions are identified by ranking the following posterior probabilities,

$$P(\text{rank}(\theta_i) \geq (1 - \gamma)I | O). \quad (6)$$

These posterior probabilities can be estimated by sampling from the joint posterior distribution of $\{\theta_i\}$ using a Gibbs or Metropolis-Hastings algorithm. In this study, the B-PPR analysis was conducted in WinBugs1.4.3 (Lunn et al.; 2000). Based on graphical assessment of convergence, the following procedure was used for each simulation. We discarded the initial 2500 samples and used every fifth sample thereafter to obtain 1000 samples of parameters.

In this paper, we propose an approximate posterior percentile ranking (PPR) method and compare it with B-PPR. For PPR, we approximate the joint posterior distribution of $\psi=\{\alpha, r$ and $\xi_j\}$ as a multi-normal distribution with means $\hat{\psi}$ and with covariance given by the inverse of the observed Fisher information matrix from the marginal likelihood based on (3), and we obtain an approximate posterior sample of ψ , namely ψ^* from this multi-normal distribution. Given ψ^* , the approximate posterior $\{\theta_i^*\}$ are conditionally independent and can be independently generated from (4). In order to approximate (6), this sample of $\{\theta_i^*\}$ is ranked and the regions with ranks in the top 100 γ % of ranks (and bottom 100 γ % of ranks) are recorded. We repeated this approximated posterior sampling procedure $N = 1000$ times and selected as “high” the γI regions that most often had among the top 100 γ % of ranks. Likewise we determined which regions were “low” as those that most often had among the lowest 100 γ % of ranks. To be precise, let $\{S_i\}$ be a sample from the posterior distribution of $\{\theta_i^*\}$ and for a particular region i consider $P\{\text{rank}(S_i) > (1-\gamma)I\}$. We estimate this probability

as above and rank the regions according to $PPHR_i = \text{rank}\{\hat{P}(\text{rank } S_i > (1-\gamma)I)\}$. The “high” regions are those with $PPHR_i > (1-\gamma)I$, and the “low” regions are those with $PPLR_i > (1-\gamma)I$, where now $PPLR_i = \text{rank}\{\hat{P}(\text{rank } S_i \leq \gamma I)\}$.

We compare our PPR approximation with B-PPR using the heart disease example described in the following section, and we find that PPR performs as well as or better than B-PPR in ranking regions. Moreover, PPR takes much less computing time than B-PPR.

3. Example of Mortality from Heart Disease

We studied mortality from heart disease (ICD codes I00-I09, I13, I20-I51 based on ICD-10, 1992) in white U.S. males for the period 1990–2005. The data were presented separately in 805 health service areas (HSAs), which are defined as a single county or cluster of contiguous counties that are relatively self-contained with respect to hospital care. Heart disease death counts and person-years for each HSA and 13 age intervals (25–29, 30–34, ..., 80–84, 85+ years) were obtained from the National Cancer Institute’s Surveillance Epidemiology and End Results (SEER) Program data at <http://seer.cancer.gov>.

We estimated the prevalence of current smoking $\{P_{ij}\}$ from aggregated 1995–2005 data from the Behavioral Risk Factor Surveillance System (BRFSS) (http://www.cdc.gov/brfss/technical_infodata/surveydata/1995.htm). Four HSAs in Alaska were eliminated because smoking data were too sparse. Twenty HSAs had total sample sizes less than 40; to estimate $\{P_{ij}\}$ for these HSAs, we aggregated their data with that of adjacent counties within the same state. For some HSAs, there were too few samples in particular age groups to estimate $\{P_{ij}\}$ reliably. Therefore, in all HSAs we aggregated data into three age groups (25–44, 45–64, and 65+ years) to estimate $\{P_{ij}\}$. Thus, several values of j share the same estimate P_{ij} . In the analysis we assume that the estimated $\{P_{ij}\}$ are known without error (see Section 5).

In this example, we use the ranking procedures presented in the last section to identify regions in the top 10% and bottom 10% of relative risks. To highlight the “high” and “low” HSAs, we plotted those in the top decile in black and those in the bottom decile in grey in Figure 1a, 1b, 1c and 1d. Other areas are shown in white, except the four excluded HSAs in Alaska, which are light grey.

Figure 1a depicts SMRs that are not adjusted for smoking. Adjustment for smoking (Figure 1b) produces noticeable changes in the pattern of “high” risk HSAs especially in and above Texas. Likewise areas of “low” risk are changed in California and elsewhere. Of those 81 HSAs classified as “high” risk in Figure 1b, only 60 were classified as high without adjustment for smoking (Figure 1a). Of those 80 classified as “low” risk in Figure 1b, only 62 were so classified in Figure 1a. Thus, adjustment for smoking changes the identification of high and low risk HSAs appreciably.

The choice of ranking procedure had less impact than adjustment for smoking. Figure 1b, 1c and 1d depict HSA deciles of risk based on the SMR, EB and PPR methods of ranking respectively, all adjusted for smoking. Of the 81 “high” risk HSAs identified by the PPR method, 80 were also ranked “high” by the EB method, and 73 by the SMR method. Likewise, of the 80 HSAs ranked “low” by the PPR method, 79 were also ranked low by the EB method and 70 by the SMR method. Thus the choice of ranking procedure had relatively little impact.

To study factors that might influence the relative performance of the ranking procedures and compare the performance of B-PPR and PPR methods, we conducted simulations based on the heart disease data. Using the estimates from the original data, $\hat{\alpha}=42.43$, $\hat{\tau}=1.71$, and

$\xi=(0.281, 0.487, 0.987, 2.17, 4.65, 9.22, 15.94, 26.15, 43.01, 73.17, 114.48, 181.57, 303.08, 622.98)\times 10^{-4}$, and y_{ij} and P_{ij} , we simulated 1000 data sets. In these simulations we knew that “true” ranking of the $\{\theta_i\}$ and could therefore calculate the proportion of the truly high and truly low $\{\theta_i\}$ that were correctly classified by the ranking procedures, with $\gamma = 0.05, 0.10$, and 0.20 . Heart disease mortality rates are higher than rates for most diseases including cancers. We therefore divided y_{ij} by 10 and repeated the simulations.

Table 1 reports results for ranking procedures for smoking adjusted B-PPR, PPR, EB and SMR, and for SMR without smoking adjustment. Across different values of γ and y_{ij} , PPR performed slightly better than B-PPR. The B-PPR method took about 7 minutes per simulation, whereas PPR took less than one minute. We found that a slight change in the mean of the vague prior could change the ranking results for B-PPR slightly for a given data set. Because PPR was faster than B-PPR, performed at least as well as B-PPR and does not require special MCMC expertise, we focus on PPR hereafter.

In simulations reflecting the original heart disease mortality data (columns 2 and 3 in Table 1), the proportions of HSAs correctly classified as low or high risk were lower without smoking adjustment. The PPR and EB procedures performed comparably and very slightly better than SMR with adjustment for smoking. The advantage of the PPR and EB procedures was more visible, however, in the scenario in which person-years were reduced by a factor of ten. In this setting, PPR performed slightly better than EB, and both outperformed SMR noticeably. The performance of all the procedures was less satisfactory than with the original y_{ij} , because the variance of SMR_i was larger relative to the true dispersion of the $\{\theta_i\}$.

4. Simulations and Analytical Studies to Compare Procedures for Selecting High Risk Groups

The example in Section 3 indicates that the proportion of truly high (or low) regions correctly detected decreases as the expected conditional variance of SMR_i which is $E(E_i|\theta_i)/E_i^2=1/E_i$, increases. Unreported simulations as in Section 3 indicate that the superiority of PPR over EB increases with increasing variability of $1/E_i$ across regions. These findings are analogous to results in Table 2 of Lin et al. (2006) for the Gaussian-Gaussian model. We now explore these relationships more systematically for the Gaussian-Gaussian and Poisson-Gamma models and provide excellent analytic approximations for the probability of correct classification of regions in the upper 10% of risk for the MLE (or SMR), EB, and PPHR ranking procedures. In what follows, we ignore the effects of smoking or other covariates.

4.1 Gaussian-Gaussian Model

The Gaussian mean is normally distributed $\mu_i \sim N(\mu_0, \tau^2)$. Given μ_i and the conditional variance σ_i^2 , we observe $X_i \sim N(\mu_i, \sigma_i^2)$. The conditional variance σ_i^2 might depend on the sample size in a survey in region i , for example. We assume that $\log(\sigma_i^2) \sim N(\eta, \varphi^2)$ is independent of μ_i . We assume μ_0, τ^2, η and φ^2 are known. The joint density is

$$f(X_i, \sigma_i^2, \mu_i) = \phi(\mu_i; \mu_0, \tau^2) \text{LN}(\sigma_i^2; \eta, \varphi^2) \phi(X_i; \mu_i, \sigma_i^2), \quad (7)$$

where $\phi(\mu_i; \mu_0, \tau^2)$ is a normal density with mean μ_0 and variance τ^2 and $\text{LN}(\sigma_i^2; \eta, \varphi^2)$ is the log-normal density of a random variable whose logarithm has mean η and variance φ^2 . In a

sample of I independent regions, we wish to identify those regions whose values μ_i have ranks $[0.9I], [0.9I]+1, \dots, I$, where $[\]$ is the integer part of a real number.

It is possible to estimate the proportion of regions correctly classified as having μ_i among the top 10% of ranks by simulation. The MLE procedure (analogous to SMR) ranks the X_i .

The empirical Bayes procedure ranks $\hat{\mu}_i^{EB} = E(\mu_i | X_i) = \rho_i \mu_0 + (1 - \rho_i) X_i$, where $\rho_i = \sigma_i^2 / (\sigma_i^2 + \tau^2)$. The PPHR procedure is based on 1000 simulations. In each simulation, a sample S_i is drawn from the posterior distribution of μ_i , namely $\phi(\mu_i; \rho_i \mu_0 + (1 - \rho_i) X_i, \kappa_i)$, where $\kappa_i = \tau^2 \rho_i$ is the posterior variance. The $\{S_i\}$ are ranked and the regions in the top 10% of ranks recorded. This process is repeated 1000 times, and the 0.1I regions most often in the top 10% of ranks are designated as “high” regions.

The left three graphs in Figure 2 show the simulation results of the percentages of correctly identified high regions for three procedures. Without loss of generality, we have set $\mu_0 = 0$ and $\tau^2 = 1$. We can see that the percentage correctly classified increases with $k = \tau^2 / GM(\sigma_i^2)$, which is a measure of signal to noise. $GM(\sigma_i^2)$ is the geometric mean of $\{\sigma_i^2\}$, which equals $\exp(\eta)$. For $\phi^2 = 0.1$, there is little variation among the $\{\sigma_i^2\}$, and the PPHR, EB and MLE rankings perform similarly. With $\phi^2 = 1$, PPHR and EB outperform MLE slightly. With $\phi^2 = 10$, PPHR and EB are markedly superior to MLE, and PPHR outperforms EM slightly.

Thus, when the variability in σ_i^2 is small, all ranking methods perform comparably, whereas when the variability in σ_i^2 is large, the PPHR method and EB are superior, and the more complex PPHR method is recommended when the variability of σ_i^2 is very large (cf. Lin et al., 2006).

The following approximations are close to the results from simulations in Figure 2. For $X \sim N(\mu, \sigma^2)$, with $\mu \sim N(\mu_0, \tau^2)$ and $\log(\sigma^2) \sim N(\eta, \phi^2)$, and assuming μ_0, τ^2, η and ϕ^2 are known, the joint density $f(X, \sigma^2, \mu)$ is given by equation (6). Since the three ranking methods we consider are all monotone functions of X , we use $R(X)$ to denote a generic statistic with distribution

$$\begin{aligned} G_R(t) &= P\{R(X) \leq t\} = P\{X \leq R^{-1}(t)\} = E_{\mu, \sigma^2}[P\{X \leq R^{-1}(t) | \mu, \sigma^2\}] \\ &= \int_{\sigma^2=0}^{\infty} \int_{\mu=-\infty}^{\infty} \int_{-\infty}^{R^{-1}(t)} f(x, \sigma^2, \mu) dx d\mu d\sigma^2, \end{aligned} \quad (8)$$

and with p^{th} percentile $\zeta_p^R = G_R^{-1}(p)$.

Suppose μ^* is drawn from a $N(\mu_0, \tau^2)$ distribution but truncated to be at or above the p^{th} percentile $\delta = \mu_0 + z_p \tau$, where $z_p = \Phi^{-1}(p) = 1.645$, and Φ is the standard normal distribution. We approximate the probability that a region in the top 10% of risk will be correctly classified as high risk as the probability that an $R(X^*)$ corresponding to μ^* exceeds ζ_p^R , where $p = 0.9$. This probability is

$$\begin{aligned} P\{R(X^*) > \zeta_p^R\} &= 1 - P\{R(X^*) \leq \zeta_p^R\} = 1 - P\{X^* \leq R^{-1}(\zeta_p^R)\} = 1 - E_{\mu^*, \sigma^2}[P\{X^* \leq R^{-1}(\zeta_p^R) | \mu^*, \sigma^2\}] \\ &= 1 - \frac{1}{1-p} \int_{\sigma^2=0}^{\infty} \int_{\mu^*=\delta}^{\infty} \int_{-\infty}^{R^{-1}(\zeta_p^R)} f(x^*, \sigma^2, \mu^*) dx^* d\mu^* d\sigma^2. \end{aligned} \quad (9)$$

$f(X^*, \sigma^2, \mu^*)$ is given by equation (6) except that μ^* is from the truncated normal distribution. The specific functions $R(X)$ are defined as follows. The MLE procedure ranks $R_{MLE}(X) = X$, the EB procedure ranks $R_{EB}(X) = \rho\mu_0 + (1-\rho)X$, where $\rho = \sigma^2/(\sigma^2 + \tau^2)$, and the PPHR procedure ranks $R_{PPHR}(X) = 1 - \Phi[\{\delta - \mu_0\rho - (1-\rho)X\}/\sqrt{\kappa}]$, with $\kappa = \rho\tau^2$. As seen in the right three graphs in Figure 2, the approximation (8) gives results very close to the simulated values, as all pairs fall on the equiangular line by visual inspection.

4.2 Poisson–Gamma Model

We simplify the model in equations (1) and (2) to eliminate the covariate effect by setting r

$= 1$ and by assuming ξ_j and hence E_{ij} and $E_i = \sum_{j=1}^J E_{ij}$ are known. The underlying parameter of interest $\theta_i \sim \Gamma(\alpha, 1/\alpha)$ as before, and conditional on θ_i , O_i is Poisson with mean $\theta_i E_i$. We further assume $\log(E_i)$ are normally distributed $N(\eta, \phi^2)$, so that E_i are lognormally distributed and independent of θ_i . The joint density is thus

$$g(O_i, E_i, \theta_i) = \Gamma(\theta_i; \alpha, 1/\alpha) \text{LN}(E_i; \eta, \phi^2) \text{Poi}(O_i; E_i \theta_i). \quad (10)$$

We assume α , η , ϕ^2 , and E_i are known. Selection of high risk regions is based on the $SMR_i = O_i/E_i$, on the empirical Bayes estimate $E(\theta_i|O_i) = (\alpha + O_i)/(\alpha + E_i)$, and on the PPHR procedure. The posterior distribution of θ_i given O_i is $\Gamma(\alpha + O_i, 1/(\alpha + E_i))$, and O_i are conditionally independent given $\{E_i\}$ and α . Hence we can draw samples of S_i from these posterior distributions and rank them. From 1000 sets of samples of this type, we determine which regions are most often in the top 10% of ranks and call them “high risk” regions.

In the simulation, we fixed $\alpha = 42.43$, which is the MLE estimate from the heart disease mortality data, and we used $\phi_0^2 = 1.55$, which was obtained by matching the first two moments of E_i from the heart disease mortality data to those of a lognormal distribution. The unconditional variance of $SMR_i = O_i/E_i$ is $E(\theta_i E_i)/E_i^2 = 1/E_i$ which has geometric mean $\exp(-\eta)$. The variance of θ_i is $1/\alpha$, yielding a signal to noise ratio $k = (1/\alpha)/\exp(-\eta) = \exp(\eta)/\alpha$.

In Figure 3, the left three graphs illustrate the simulation results corresponding to three degrees of variability in $1/E_i$, namely $\phi^2 = 1.55/5 = 0.31$, 1.55 , and 1.55×7.75 . The percentage of high risk regions correctly classified increases with k for all procedures across different values of ϕ^2 . When there is little variability in $1/E_i$ (i.e. $\phi^2 = 1.55/5$), PPHR, EB and SMR perform comparably, but for $\phi^2 = 1.55$, PPHR and EB outperform SMR slightly, especially for small k . For $\phi^2 = 7.75$, the PPHR and EB are much better than SMR, even for large k . In Figure 3, we also indicated three points that correspond to 1/100, 1/10, and 1 times the death counts expected for heart disease. These points correspond respectively to an uncommon disease (e.g. esophagus cancer), a moderately common disease (e.g. breast cancer) and a common disease (e.g. heart disease). These results indicate that when there is variability in $1/E_i$ and the Poisson variation in the SMR estimate is large, PPHR or EB rankings outperform the SMR ranking, and, if the variability of $1/E_i$ is large, PPHR outperforms EB.

For $O \sim \text{Poi}(E\theta)$, with $\theta \sim \Gamma(\alpha, 1/\alpha)$ and $\log(E) \sim N(\eta, \phi^2)$, and assuming α , η and ϕ^2 are known, we approximate the performance of the three ranking methods, which are all monotone functions of O , denoted $R(O)$ with

$$\begin{aligned}
 G_R(t) &= P(R(O) \leq t) = P(O \leq R^{-1}(t)) = E_{\theta, E} \{P(O \leq R^{-1}(t) | \theta, E)\} \\
 &= \int_{E=0}^{\infty} \int_{\theta=0}^{\infty} \sum_{O=0}^{\{R^{-1}(t)\}} g(O, \theta, E) d\theta dE,
 \end{aligned} \tag{11}$$

and with p^{th} percentile $\zeta_p^R = G_R^{-1}(p)$.

Suppose θ^* is drawn from a $\Gamma(\alpha, 1/\alpha)$ distribution but truncated to be at or above the p^{th} percentile $\delta = \Gamma^{-1}(p; \alpha, 1/\alpha)$. We approximate the probability that a region in the top 10% of risk will be correctly classified as high risk as the probability that an $R(O^*)$ corresponding to θ^* exceeds ζ_p^R , where $p = 0.9$. This probability is

$$\begin{aligned}
 P\{R(O^*) > \zeta_p^R\} &= 1 - P\{R(O^*) \leq \zeta_p^R\} = 1 - P\{O^* \leq R^{-1}(\zeta_p^R)\} \\
 &= 1 - E_{\theta^*, E} \{P\{O^* \leq R^{-1}(\zeta_p^R) | \theta^*, E\}\} = 1 - \frac{1}{1-p} \int_{E=0}^{\infty} \int_{\theta^*=\delta}^{\infty} \sum_{O^*=0}^{\{R^{-1}(\zeta_p^R)\}} g(O^*, \theta^*, E) dO^* d\theta^* dE.
 \end{aligned} \tag{12}$$

Specifically, the SMR procedure ranks $R_{SMR}(O) = O/E$, the EB procedure ranks $R_{EB}(O) = (O+\alpha)/(E+\alpha)$, and the PPHR procedure ranks $R_{PPHR}(O) = 1 - \int_0^\delta \Gamma(s; O+\alpha, E+\alpha) ds$. As seen in the right three graphs in Figure 3, the approximation (12) gives results visually indistinguishable from the simulated data.

5. Discussion

We investigated procedures for identifying the regions with the upper 100 γ % and lower 100 γ % of standardized mortality rates. In an example of heart disease mortality (Section 3), we found that adjusting for smoking prevalence produced notable changes in the regions identified as high and low risk. Once smoking prevalence was accounted for, however, ranking based on the SMR, on empirical Bayes estimates (EB), or on posterior percentile ranking (PPR) yielded very similar classifications (Figure 1b, 1c and 1d). The EB and PPR methods outperformed SMR, however, when the expected counts were reduced by a factor of 10, (Table 1). In scenarios representing less common diseases, PPR and EB outperformed SMR when expected counts varied over regions (Section 4), and when there was wide variation in expected counts, the PPR method was better than EB. We developed very accurate approximations for the probability of correct classification in the upper 100 γ % of risk for Gaussian-Gaussian and Poisson-Gamma models (Section 4).

The Poisson-Gamma empirical Bayesian framework was used by Clayton and Kaldor (1987). The choice of Gamma distribution for the random effect θ_i is mathematically convenient, because it is the conjugate prior for the Poisson distribution, and the posterior of θ_i is also gamma distributed. Clayton and Kaldor (1987) and Christiansen and Morris (1997) modeled the log-mean of θ_i as a linear combination of covariates. We suggested incorporating the effects of a covariate like smoking prevalence as a mixture that represented smokers and non-smokers. The Poisson-Gamma model helps to smooth θ_i by borrowing strength from the other regions. One could achieve further smoothing by allowing for spatial correlations. However, the Gamma prior cannot be easily generalized to for this purpose. In our analysis of mortality from heart disease, we aggregated data from nearby counties, reducing the number of units from 3,141 counties to 805 health service areas, and achieving some local smoothing.

Lin et al. (2006) presented data from a Gaussian-Gaussian model (their Table 2) that indicated better performance for picking high and low regions with B-PPR than by classification based on their posterior expected ranks. Conlon and Louis (1999) and Shen and Louis (1998, 2000) discussed procedures that simultaneously gave good estimates of region-specific rates, good estimates of the ranks of the rates, and good estimates of the cumulative distribution of the rates (“triple goal estimates”). Stern and Cressie (1999) described a loss function to identify regions with extreme risks.

Although the original B-PPR method was based on a fully Bayesian analysis, simulated comparisons using heart disease data in Section 3 suggests that our approximate PPR performs as well as or better than B-PPR with reduced computation time and easier implementation. Perhaps the fully Bayesian approach would perform better with different choices of parameters for the vague prior distributions and Gibbs sampling regime.

We assumed that the prevalence of smoking was known without error. Further work is needed to evaluate the impact of error and the use of hierarchical models to improve prevalence estimates. More experience is needed to assess the usefulness of the covariate adjustment and ranking procedures in this paper. The present work illustrates, however, that it is computationally feasible to adjust for covariates and use a preferred ranking procedure, PPR, which may be particularly advantageous for less common diseases with wide variation in expected counts.

Acknowledgments

The authors thank Professor Thomas A. Louis for pointing out some key references and providing reprints and the reviewers for helpful suggestions. This work was supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics of the National Cancer Institute.

REFERENCES

- Christiansen CL, Morris CN. Hierarchical Poisson Regression Modeling. *Journal of the American Statistical Association*. 1997; 92:618–632.
- Clayton DG, Kaldor J. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*. 1987; 43:671–682. [PubMed: 3663823]
- Conlon, EM.; Louis, TA. Addressing multiple goals in evaluating region-specific risk using Bayesian methods. New York: Wiley; 1999.
- Devesa, SS.; G, DG.; Blot, WJ.; Pennello, G.; Hoover, RN.; Fraumeni, JF, Jr. Atlas of Cancer Mortality in the United States, 1950–94. Washington, DC: 1999. US Govt Print Off [NIH Publ No. (NIH) 99-4564]
- Laird NM, Louis TA. Empirical Bayes ranking methods. *Journal of Educational Statistics*. 1989; 14:29–46.
- Lin R, Louis TA, Paddock SM, Ridgeway G. Loss function based ranking in two-stage, hierarchical models. *Bayesian Analysis*. 2006; 1:915–946. [PubMed: 20607112]
- Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS -- a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*. 2000; 10:325–337.
- Pennello GA, Devesa SS, Gail MH. Using a mixed effects model to estimate geographic variation in cancer rates. *Biometrics*. 1999; 55:774–781. [PubMed: 11315006]
- Shen W, Louis TA. Triple-goal estimates for disease mapping. *Statistics in Medicine*. 2000; 19:2295–2308. [PubMed: 10960854]
- Shen W, Louis TA. Triple-goal estimates in two-stage hierarchical models. *Journal of the Royal Statistical Society Series B*. 1998; 60:445–471.
- Stern, HS.; Cressie, N. Inference for Extremes in Disease Mapping and Risk Assessment for Public Health. Lawson, A.; Biggeri, A.; Bohning, D.; Lesaffre, E.; Viel, J-F.; Bertollini, R., editors. Chichester: John Wiley and Sons; 1999. p. 63-84.

Figure 1. Rankings of heart disease mortality relative risks in 805 Health Service Areas (HSAs) for white men, 1990–2005. Those HSAs in the *top* decile are shown in black; those in the *bottom* decile are shown in grey; others are in white. Four HSAs in Alaska with insufficient smoking data are shown in light grey.

Figure 2.

Gaussian-Gaussian model: left hand panels plot percent of regions correctly classified as having among the top 10% of means versus the signal to noise ratio, k ; right hand panels plot the percent correctly classified estimated from simulations with 1000 replications against the estimate from the analytic approximation in Section 4. The mean and variance of μ were $\mu_0 = 0$, and $\tau^2 = 1$.

Figure 3.

Poisson-Gamma model: left hand panels plot percent of regions correctly classified as having among the top 10% of relative risks versus the signal to noise ratio, k ; right hand panels plot the percent correctly classified estimated from simulations with 1000 replications against the estimate from the analytic approximation in Section 4. The variance of θ_i is $1/\alpha = 1/42.43$.

Table 1

Mean percentages (with standard error) of correctly classified HSAs in simulated cardiovascular mortality data^a.

100% Ranking procedure	Original person-years		Person-years/10	
	Low risk	High risk	Low risk	High risk
5% Smoking adjusted				
B-PPR	86.16 (0.127)	87.30 (0.118)	62.69 (0.182)	63.40 (0.191)
PPR	86.45 (0.126)	87.77 (0.117)	65.05 (0.178)	67.87 (0.183)
EB	86.40 (0.127)	87.70 (0.119)	64.26 (0.179)	67.29 (0.186)
SMR	85.28 (0.139)	86.46 (0.130)	56.91 (0.195)	61.69 (0.188)
No Smoking adjustment				
SMR	80.43 (0.151)	76.28 (0.160)	56.77 (0.199)	56.85 (0.198)
10% Smoking adjusted				
B-PPR	88.55 (0.080)	89.18 (0.080)	67.76 (0.123)	68.07 (0.122)
PPR	88.80 (0.079)	89.52 (0.075)	70.04 (0.120)	71.89 (0.117)
EB	88.77 (0.080)	89.56 (0.075)	69.56 (0.121)	71.58 (0.118)
SMR	87.98 (0.085)	88.63 (0.081)	65.77 (0.133)	68.48 (0.123)
No Smoking adjustment				
SMR	83.31 (0.094)	80.57 (0.102)	65.12 (0.129)	64.70 (0.124)
20% Smoking adjusted				
B-PPR	91.02 (0.051)	91.38 (0.049)	73.80 (0.082)	74.24 (0.081)
PPR	91.26 (0.049)	91.62 (0.049)	75.78 (0.079)	76.97 (0.080)
EB	91.25 (0.049)	91.61 (0.048)	75.59 (0.080)	76.85 (0.080)
SMR	90.73 (0.053)	91.00 (0.055)	74.32 (0.079)	75.78 (0.080)
No Smoking adjustment				
SMR	86.62 (0.062)	85.17 (0.064)	73.34 (0.080)	73.23 (0.079)

^aBased on 1000 simulations