BIOMETRICS 41, 69–79
March 1985

Estimation of Cancer Mortality Rates: A Bayesian Analysis of Small Frequencies

Robert K. Tsutakawa

Department of Statistics, University of Missouri,
Columbia, Missouri 65211, U.S.A.

SUMMARY

A Bayesian method is presented for estimating mortality rates of specific diseases when the frequency of deaths over a specified time period is assumed to have a Poisson distribution with mean proportional to the population size. The estimators use information from related populations, each having its own rate which is assumed distributed according to a common prior distribution about which some information is available.

The study was motivated by an epidemiological study on the geographic variation of cancer mortality in the state of Missouri. Data from this study are used to illustrate the method and to compare it to a somewhat simpler empirical Bayes method.

1. Introduction

We consider making inferences about the occurrence of rare human events in a given population when we have information on similar occurrences in other populations. Mortality due to a specific disease such as lung cancer is one such example where the annual frequency in a small or average-sized city is quite low and the information from a single city is very limited. We present a method of using information from several cities with differing mortality rates that will yield better estimates of true mortality rates than the raw rates based on individual cities.

The current study was motivated by an epidemiological investigation (Marienfeld et al., 1980) of the effects of public drinking water on cancer mortality in the state of Missouri. Some of the data collected for this study are summarized in Table 1, which gives mortality frequencies due to lung cancer among males aged 45–54 from 1972 to 1981 in eighty-four of the largest cities in Missouri. The raw rates reported in the table are the annual rates per 10^6 population at risk. Note that these are quite variable, particularly among the smaller cities where a difference of even one or two deaths can have a large effect on the raw rates.

Regional variations in cancer mortality rates are well known through the tabulations of Mason and McKay (1974) of age-adjusted death rates in the United States for 1950–1969 by county, sex, race, and cancer type. These adjusted rates are useful for overall description, but can be unreliable in making regional inferences when the numbers of individuals in certain subgroups are small. Moreover, since most cancer types affect different age–sex groups quite differently, many epidemiologists find adjusted rates to be of limited value (Hill, 1977, p. 198).

Another approach to improving on the raw rates is to use a multiplicative model of the type proposed by Breslow and Day (1975). An example of this is to assume that the probability of death for a given individual can be factored into a component due to age and another due to geographic region. The components are then estimated by maximum likelihood. One consequence of this approach is that if there are no deaths in a region, an

Key words: Bayes; Cancer; Empirical Bayes; Mortality rates; Random effects.

Table 1
Lung cancer mortality in Missouri cities, males aged 45–54, 1972–1981

				Empirical Bayes				Bayes			
City	Size	Deaths	Raw rate*	$E(\theta) + 5$	$Sd(\theta)$	Expected deaths	Rate*	$E(\theta) + 5$	$Sd(\theta)$	Expected deaths	Rate*
1	1019	2	196	.009	.203	6.9	677	.024	.234	6.8	671
2	1512	8	529	.086	.187	11.2	742	.080	.202	11.2	739
3	1424	8	562	.109	.188	10.8	759	.104	.202	10.8	757
4	54155	402	742	.109	.042	404.0	746	.109	.043	404.2	746
5	447	1	224	.128	.218	3.5	778	.116	.239	3.4	772
6	1907	12	629	.120	.177	14.6	766	.117	.187	14.6	765
7	1755	11	627	.125	.180	13.5	770	.122	.191	13.5	769
8	5756	42	730	.134	.128	44.4	771	.135	.130	44.4	772
9	509	2	393	.152	.215	4.0	796	.142	.233	4.0	791
10	350	1	286	.165	.222	2.8	807	.155	.241	2.8	802
11	473	2	423	.165	.217	3.8	807	.156	.234	3.8	802
12	329	1	304	.173	.222	2.7	814	.163	.241	2.7	809
13	7137	55	771	.167	.117	56.8	796	.168	.118	56.8	796
14	430	2	465	.181	.218	3.5	820	.173	.234	3.5	816
15	304	1	329	.183	.223	2.5	822	.174	.241	2.5	818
16	163	0	0	.191	.230	1.4	830	.181	.250	1.3	826
17	163	0	0	.191	.230	1.4	830	.181	.250	1.3	826
18	159	0	0	.193	.230	1.3	832	.183	.250	1.3	827
19	281	1	356	.192	.224	2.3	830	.184	.242	2.3	826
20	154	0	0	.195	.230	1.3	834	.186	.250	1.3	829
21	889	6	675	.192	.201	7.3	826	.187	.212	7.3	824
22	260	1	384	.201	.225	2.2	837	.193	.243	2.2	834
23	371	2	540	.205	.220	3.1	840	.198	.236	3.1	837
24	232	1	432	.213	.226	2.0	848	.206	.243	2.0	845
25	228	1	439	.215	.227	1.9	849	.208	.244	1.9	846
26	343	2	584	.216	.221	2.9	849	.210	.236	2.9	847
27	454	3	660	.219	.216	3.9	851	.213	.230	3.9	848
28	323	2	619	.224	.222	2.8	856	.218	.237	2.8	854
29	311	2	643	.229	.222	2.7	861	.224	.237	2.7	859
30	784	6	765	.227	.203	6.7	856	.224	.213	6.7	854
31	426	3	705	.230	.217	3.7	860	.225	.231	3.7	859
32	184	1	545	.234	.229	1.6	866	.229	.245	1.6	865
33	181	1	553	.235	.229	1.6	867	.230	.245	1.6	866
34	177	1	564	.237	.229	1.5	869	.232	.246	1.5	867
35	177	1	566	.237	.229	1.5	869	.232	.246	1.5	867
36	291	2	688	.238	.223	2.5	868	.233	.238	2.5	867
37	170	1	587	.240	.229	1.5	871	.235	.246	1.5	870
38	158	1	632	.245	.230	1.4	876	.241	.246	1.4	875
39	274	2	730	.245	.224	2.4	875	.241	.239	2.4	874
40	150	1	667	.249	.230	1.3	880	.245	.247	1.3	879
41	265	2	755	.249	.224	2.3	878	.245	.239	2.3	877
42	257	2	779	.253	.225	2.3	881	.249	.239	2.3	881

* Annual rate per 10^6 population at risk.

event which commonly occurs in small populations, the probability is estimated as 0. Moreover, due to the absence of interaction effects, these models are inadequate for studying regional variations within age groups. Manton, Woodbury, and Stallard (1981) discuss a somewhat more complex model where the frequency of deaths is assumed to have a negative binomial distribution whose parameter depends on demographic as well as geographic variables, such as county and longitudinal gradient. As in the Breslow–Day study, the model is basically a fixed effects model, where the effects are estimated by maximum likelihood, and not well suited to dealing with small populations, many of which contain no deaths.

We propose to consider this problem from a Bayesian view of a random effects model, where true rates vary at random from city to city according to some distribution with

Table 1
(continued)

				Empirical Bayes				Bayes			
City	Size	Deaths	Raw rate*	E(θ) + 5	Sd(θ)	Expected deaths	Rate*	E(θ) + 5	Sd(θ)	Expected deaths	Rate*
43	254	2	786	.254	.225	2.2	882	.250	.240	2.2	882
44	28937	251	867	.260	.063	251.1	868	.260	.063	251.2	868
45	445	4	898	.269	.216	4.0	894	.266	.227	4.0	894
46	447	4	896	.268	.216	4.0	893	.266	.227	4.0	893
47	329	3	912	.270	.221	2.9	896	.268	.234	2.9	897
48	206	2	970	.275	.227	1.9	902	.273	.242	1.9	903
49	313	3	957	.277	.222	2.8	902	.275	.235	2.8	903
50	314	3	955	.277	.222	2.8	902	.275	.235	2.8	903
51	314	3	955	.277	.222	2.8	902	.275	.235	2.8	903
52	202	2	992	.277	.227	1.8	904	.276	.242	1.8	905
53	198	2	1009	.279	.227	1.8	905	.277	.242	1.8	907
54	183	2	1093	.286	.228	1.7	912	.285	.243	1.7	914
55	292	3	1029	.287	.222	2.7	911	.286	.236	2.7	913
56	178	2	1122	.288	.228	1.6	914	.288	.243	1.6	916
57	287	3	1046	.289	.223	2.6	913	.288	.236	2.6	915
58	282	3	1063	.291	.223	2.6	915	.290	.236	2.6	917
59	164	2	1219	.295	.229	1.5	920	.295	.244	1.5	923
60	164	2	1219	.295	.229	1.5	920	.295	.244	1.5	923
61	1923	18	936	.296	.170	17.5	910	.295	.173	17.5	910
62	3672	34	926	.300	.141	33.4	910	.300	.142	33.4	910
63	261	3	1150	.301	.224	2.4	924	.301	.238	2.4	927
64	581	6	1033	.303	.209	5.4	924	.303	.219	5.4	925
65	550	6	1091	.316	.210	5.1	935	.316	.220	5.2	938
66	431	5	1161	.321	.215	4.1	941	.322	.227	4.1	944
67	399	5	1252	.334	.217	3.8	954	.336	.228	3.8	958
68	286	4	1400	.338	.222	2.7	959	.341	.235	2.8	964
69	592	7	1181	.342	.208	5.7	959	.343	.217	5.7	963
70	246	4	1628	.357	.224	2.4	977	.362	.238	2.4	985
71	547	7	1281	.361	.209	5.3	978	.363	.220	5.4	982
72	438	6	1369	.363	.214	4.3	981	.367	.226	4.3	987
73	202	4	1978	.379	.226	2.0	999	.385	.242	2.0	1009
74	790	10	1266	.386	.199	7.9	1000	.388	.207	7.9	1004
75	648	9	1390	.403	.204	6.6	1018	.407	.214	6.6	1024
76	354	6	1694	.402	.217	3.6	1020	.408	.232	3.6	1030
77	730	10	1369	.409	.200	7.5	1024	.413	.210	7.5	1030
78	144	4	2769	.409	.229	1.5	1029	.419	.248	1.5	1045
79	1093	14	1281	.420	.187	11.3	1033	.422	.194	11.3	1036
80	384	7	1824	.434	.215	4.0	1052	.442	.231	4.1	1065
81	278	6	2156	.439	.221	2.9	1059	.449	.239	3.0	1074
82	596	10	1678	.466	.204	6.5	1084	.473	.219	6.5	1095
83	1889	28	1482	.572	.159	22.6	1195	.572	.169	22.6	1196
84	22514	334	1528	.804	.058	334.5	1486	.802	.059	333.7	1482

* Annual rate per 10^6 population at risk.

parameters about which we have some prior information. For a given age–sex group within a city (which we consider to be a homogeneous target population), we assume that the frequency of deaths due to a given cause over a given time period has a Poisson distribution with mean λ. To account for groups of different sizes we let $\lambda = np$, where n is the size of the group. We assume that the p's vary randomly in such a manner that their logits, $\theta = \log[p/(1-p)]$, may be treated as a random sample from a normal distribution with unknown mean μ and standard deviation σ. We further assume that the hyperparameter (μ, σ) has a known prior distribution, enabling us to consider the posterior distribution of θ.

The Poisson model with the parametrization $\lambda = np$ has been used previously by Breslow and Day (1975). They consider p fixed, whereas we consider it random. Our study is related to works on contingency tables under empirical Bayes approaches by Good (1956) and Laird (1978), and under a Bayesian approach by Leonard (1972, 1975). It is also related to empirical Bayes analyses of regional mortality rates by Miao (in a Ph.D. thesis at Harvard University, 1977) and by Tsutakawa, Shoop, and Marienfeld (in a 1983 University of Missouri Department of Statistics technical report). One shortcoming of the empirical Bayes procedures is that they do not give a good measure of the reliability of the estimated mortality rates because they do not account for the uncertainty in the estimated prior distribution of θ (cf. Dempster, Rubin, and Tsutakawa, 1981). In this paper we extend the empirical Bayes solution of Tsutakawa et al. to a fully Bayesian one by using techniques presented by Deeley and Lindley (1981) and Lindley (1980). The extensive computational formulas needed to implement the Bayesian method are summarized in the Appendix.

The Bayesian approach is illustrated with the lung cancer data. The results show that the Bayesian estimates of mortality rates do not have the extreme fluctuation seen in the raw rates and the uncertainties in these rates may be studied through the posterior distributions of θ. The Bayesian method also provides a means of comparing rates among cities. A comparison with the empirical Bayes method, where (μ, σ) is estimated by maximum likelihood, shows that the posterior means of θ are generally quite close, but the posterior standard deviations are smaller for empirical Bayes, particularly for the smaller cities. The Bayesian approach presented here demonstrates that, by modeling true mortality rates as random variables having some underlying distribution, it is not necessary to obtain smoothed rates by the conventional methods of adjustment by age or fixed effects models, neither of which accounts for the differences in geographic variability from one age group to another. Moreover, in situations where some cities have no deaths, the Bayesian method gives positive rates which are low but consistent with the distribution of the rates of other cities.

2. Model and Different Approaches to the Problem

Let $\mathbf{Y} = (Y_1, \ldots, Y_k)$ be the numbers of deaths due to a specific cause during a given time period in k populations of sizes $n_1, \ldots, n_k$. Assume that for each i, Y_i has a Poisson distribution with mean $\lambda_i = n_i p_i$ and that $\theta_i = \log[p_i/(1 - p_i)]$ has a prior distribution which is normal with unknown mean and standard deviation (μ, σ). Given (μ, σ), assume that $\theta_1, \ldots, \theta_k$ are independent and given $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$, then $Y_1, \ldots, Y_k$ are also independent.

Given (μ, σ), the joint distribution of (θ_i, Y_i) is

$$p(\theta_i, Y_i \mid \mu, \sigma) = f(y_i \mid n_i, \theta_i) g(\theta_i \mid \mu, \sigma), \tag{2.1}$$

where

$$f(y_i \mid n_i, \theta_i) = [\exp(-n_i p_i)](n_i p_i)^{y_i}/y_i!,$$

$$p_i = [1 + \exp(-\theta_i)]^{-1},$$

and

$$g(\theta_i \mid \mu, \sigma) = (2\pi\sigma^2)^{-1/2} \exp[-(\theta_i - \mu)^2/(2\sigma^2)],$$

for $y_i = 0, 1, \ldots,$ and $-\infty < \theta_i < \infty$. The marginal probability function of Y_i is then

$$p(y_i \mid \mu, \sigma) = \int p(y_i, \theta_i \mid \mu, \sigma)\, d\theta_i \tag{2.2}$$

and, from the assumption of independence, the likelihood function of (μ, σ) is given by

$$\ell(\mu, \sigma) = \prod_{i=1}^{k} p(y_i \mid \mu, \sigma). \tag{2.3}$$

Our primary concern will be in drawing inferences regarding a particular θ_i, which in turn can be used to study p_i and λ_i. Without loss of generality, we assume that our interest is in θ_k. If (μ, σ) are known, the posterior probability density functions (pdfs) of $\boldsymbol{\theta}$ and θ_k are given by

$$g(\boldsymbol{\theta} \mid \mathbf{y}, \mu, \sigma) = \prod_{i=1}^{k} g(\theta_i \mid y_i, \mu, \sigma) \tag{2.4}$$

and

$$g(\theta_k \mid y_k, \mu, \sigma) = f(y_k \mid n_k, \theta_k) g(\theta_k \mid \mu, \sigma) / p(y_k \mid \mu, \sigma). \tag{2.5}$$

In the absence of known (μ, σ), a number of approaches for estimating θ_k may be considered. An empirical Bayes approach suggested in a paper by Good (1956) and refined by Tsutakawa et al. is to first find $(\hat{\mu}, \hat{\sigma})$, the maximum likelihood estimate of (μ, σ) with respect to the likelihood function (2.3), and then to estimate θ_k by its posterior expectation

$$\mathrm{E}(\theta_k \mid y_k, \hat{\mu}, \hat{\sigma}) = \int \theta_k g(\theta_k \mid y_k, \hat{\mu}, \hat{\sigma})\, d\theta_k,$$

where (μ, σ) is now replaced by its estimate $(\hat{\mu}, \hat{\sigma})$. Another empirical Bayes approach suggested by Laird (1978) for two-way contingency tables is to first compute the marginal maximum likelihood estimate $\tilde{\sigma}^2$ of σ^2, assuming μ has a flat prior, and then to estimate θ_k by its posterior mode when $\sigma = \tilde{\sigma}$. A Bayesian approach suggested by Leonard (1972, 1975), again for contingency tables, is to consider the marginal posterior distribution of $\boldsymbol{\theta}$ when (μ, σ) has a prior distribution such that μ is uniform over the real line and σ^2 is independent of μ with $\nu\tau\sigma^{-2}$ distributed as chi-square with ν degrees of freedom, where $\tau > 0$ and $\nu \geqslant 0$ are known parameters. We interpret τ to be a prior estimate of σ^2 and ν as a measure of the strength of this belief. We will first discuss the joint posterior distribution of $\boldsymbol{\theta}$ and then use Lindley's (1980) approximation to numerically study the marginal posterior distribution of θ_k.

3. Posterior Distribution of $\boldsymbol{\theta}$

Suppose now that the joint prior of (μ, σ) is given by the pdf

$$\pi(\mu, \sigma) \propto \sigma^{-\nu-1} \exp(-\sigma^2 \nu\tau/2), \tag{3.1}$$

where (τ, ν) is a known parameter, $\tau > 0$, and $\nu \geqslant 0$. Then the posterior pdf of $(\boldsymbol{\theta}, \mu, \sigma)$ is

$$h(\boldsymbol{\theta}, \mu, \sigma \mid \mathbf{y}) \propto \pi(\mu, \sigma) \prod_{i=1}^{k} p(\theta_i, y_i \mid \mu, \sigma). \tag{3.2}$$

By integrating h first with respect to μ and then with respect to σ, it follows [as in Leonard (1972), where binomial distributions are considered] that the marginal posterior pdf of $\boldsymbol{\theta}$ is

$$g(\boldsymbol{\theta} \mid \mathbf{y}) \propto \prod_{i=1}^{k} f(y_i \mid n_i, \theta_i) [\textstyle\sum (\theta_i - \bar{\theta})^2 + \nu\tau]^{-(\nu+k-1)/2}, \tag{3.3}$$

where $\bar{\theta} = \sum \theta_\ell / k$. Although the marginal of θ_k is not easy to derive, the joint mode, $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_k)$, when it exists can be shown to satisfy the equations

$$(y_i/n_i - p_i) q_i = (\theta_i - \bar{\theta})/(n_i s^2) \qquad i = 1, \ldots, k, \tag{3.4}$$

where $q_i = 1 - p_i$ and

$$s^2 = [\textstyle\sum (\theta_i - \bar{\theta})^2 + \nu\tau]/(\nu + k - 1). \tag{3.5}$$

When evaluated at $\hat{\boldsymbol{\theta}}$, s^2 may be conveniently interpreted as the posterior estimate of σ^2.

For our application, q_i will be very close to 1 and, when evaluated at the mode, p_i is approximately

$$\hat{p}_i = y_i/n_i - (\hat{\theta}_i - \hat{\bar{\theta}})/(\hat{s}^2 n_i), \tag{3.6}$$

the observed proportion, y_i/n_i, minus an adjustment inversely proportional to $\hat{s}^2 n_i$, where $\hat{\bar{\theta}}$ and $\hat{s}^2$ are $\bar{\theta}$ and s^2 evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. Since p_i must be nonnegative, $\hat{\theta}_i$ cannot exceed $\hat{\bar{\theta}}$ in the special case when $y_i = 0$.

It is also instructive to see what happens in the hypothetical cases when $\hat{s}^2 = 0$ and $\hat{s}^2 = \infty$. Multiplying (3.6) by n_i and summing over i, we have $\sum \hat{p}_i n_i = \Sigma y_i$. Thus, when $\hat{s}^2 = 0$, all $\hat{p}_i$ are equal and are given by $\sum y_i/\sum n_i$, which is the overall proportion. On the other hand, when $\hat{s}^2 = \infty$, $\hat{p}_i = y_i/n_i$, which is the maximum likelihood estimate of p_i, based solely on population i.

4. Marginal Posterior Distributions of $\boldsymbol{\theta_k}$ and $(\boldsymbol{\theta_j}, \boldsymbol{\theta_k})$

Given the prior pdf $\pi(\mu, \sigma)$, the marginal pdf of θ_k is given by

$$g(\theta_k \mid \mathbf{y}) = \int \int g(\theta_k \mid y_k, \mu, \sigma) h(\mu, \sigma \mid \mathbf{y})\, d\mu\, d\sigma, \tag{4.1}$$

where h is the posterior pdf of (μ, σ), given by

$$h(\mu, \sigma \mid \mathbf{y}) \propto \ell(\mu, \sigma)\pi(\mu, \sigma). \tag{4.2}$$

A straightforward proof is presented in more general terms by Deeley and Lindley (1981). The joint marginal pdf and moments of (θ_j, θ_k) are similarly given by

$$g(\theta_j, \theta_k \mid \mathbf{y}) = \int \int g(\theta_j \mid y_j, \mu, \sigma) g(\theta_k \mid y_k, \mu, \sigma) h(\mu, \sigma \mid \mathbf{y})\, d\mu\, d\sigma \tag{4.3}$$

and

$$\mathrm{E}(\theta_j^r \theta_k^t \mid \mathbf{y}) = \int \int \mathrm{E}(\theta_j^r \mid y_j, \mu, \sigma) \mathrm{E}(\theta_k^t \mid y_k, \mu, \sigma) h(\mu, \sigma \mid \mathbf{y})\, d\mu\, d\sigma, \tag{4.4}$$

for $r = 0, 1, \ldots; t = 0, 1, \ldots;$ and $j \neq k$. Since the numerical evaluation of these marginals and their moments requires the evaluation of complicated multiple integrals, we turn to an approximation due to Lindley (1980), which reduces the problem to evaluating single integrals and appears adequate for practical work. Following Lindley (1980), let L_{rt} denote the (r, t)th partial derivative

$$\frac{\partial^{r+t}}{\partial \mu^r\, \partial \sigma^t} \log \ell(\mu, \sigma),$$

evaluated at $(\hat{\mu}, \hat{\sigma})$, $r = 0, 1, \ldots; t = 0, 1, \ldots;$ let

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} = \begin{bmatrix} -L_{20} & -L_{11} \\ -L_{11} & -L_{02} \end{bmatrix}^{-1}; \tag{4.5}$$

let (ρ_1, ρ_2) be the first partials of $\log \pi(\mu, \sigma)$ evaluated at $(\hat{\mu}, \hat{\sigma})$; and finally, for an arbitrary real-valued function $u(\mu, \sigma)$ of (μ, σ), let u, u_r, and u_{rt} denote $u(\mu, \sigma)$ and its first two partials evaluated at $(\hat{\mu}, \hat{\sigma})$ [e.g., $u_2 = \partial u/\partial \sigma$ and $u_{11} = \partial^2 u/\partial \mu^2$, evaluated at $(\hat{\mu}, \hat{\sigma})$]. Then to terms of order $O(1/k)$, Lindley shows that

$$
\begin{aligned}
\mathrm{E}[u(\mu, \sigma) \mid \mathbf{y}] \approx u &+ \tfrac{1}{2} \Sigma \, (u_{rt} + 2u_r\rho_t)\sigma_{rt} \\
&+ \tfrac{1}{2} \{L_{30}[u_1\sigma_{11}^2 + u_2\sigma_{11}\sigma_{12}^2] \\
&+ L_{21}[3u_1\sigma_{11}\sigma_{12} + u_2(\sigma_{11}\sigma_{22} + 2\sigma_{12}^2)] \\
&+ L_{12}[u_1(\sigma_{11}\sigma_{22} + 2\sigma_{12}^2) + 3u_2\sigma_{12}\sigma_{22}] \\
&+ L_{03}[u_1\sigma_{12}\sigma_{22} + u_2\sigma_{22}^2]\}, \qquad (4.6)
\end{aligned}
$$

where the first term omitted is $O(k^{-2})$. We can apply this approximation to (4.1), (4.3), and (4.4) by taking $u(\mu, \sigma)$ to be $g(\theta_k \mid y_k, \mu, \sigma)$, $g(\theta_j \mid y_j, \mu, \sigma)g(\theta_k \mid y_k, \mu, \sigma)$, and $\mathrm{E}(\theta_j^r \mid y_j, \mu, \sigma)\mathrm{E}(\theta_k^t \mid y_k, \mu, \sigma)$, respectively. These approximations require the numerical evaluation of a large number of single integrals. Numerical integration by the Gauss–Hermite quadrature method has been found to be quite adequate for practical work.

5. Lung Cancer Example

We now present an illustration using the lung cancer data, where the prior on (μ, σ) is approximated by the noninformative prior given when $\nu = 0$ in (3.1). Although this is an improper prior, it is an appropriate choice in (4.6) to represent small values of ν. Provided $\hat{\sigma} > 0$, we then have $\rho_1 = 0$ and $\rho_2 = 1/\hat{\sigma}$, the latter being the limiting value of ρ_2 as $\nu \rightarrow 0$.

Using the technique given in the technical report by Tsutakawa et al., we have for these data $(\hat{\mu}, \hat{\sigma}) = (-4.7327, 0.2384)$. The corresponding posterior mean or Bayes estimate obtained from (4.6) is $(-4.7352, 0.2459)$.

Table 1 summarizes the posterior means and standard deviations of θ_i, estimated annual rates, and expected deaths for both the empirical Bayes and Bayesian methods. The annual rate is estimated by the posterior mean of $10^6 p_i/10$ and the expected deaths (for the 10-year period) by the posterior mean of $n_i p_i$.

The estimated annual rates range from 677 to 1486 for empirical Bayes and from 671 to 1482 for Bayes, in contrast to the raw rates, which range from 0 to 2769. It is interesting to note that cities with no deaths have positive estimates and, in terms of posterior means, rank above several larger cities with a positive number of deaths. Some idea of the fit of the model can be obtained by observing the closeness of the expected number of deaths to the actual number of deaths.

To assess the uncertainties in these estimates, we consider the posterior standard deviations of θ_i. The difference between the empirical Bayes and Bayes methods now becomes more obvious, with the standard deviation generally being larger for the latter and by as much as 15% for city 1. This is due to the uncertainty in $(\hat{\mu}, \hat{\sigma})$, which is disregarded by the empirical Bayes method (Deeley and Lindley, 1981). This discrepancy becomes negligible, however, for the larger cities, where there is close agreement in terms of both the mean and standard deviation. The difference is illustrated more graphically in Fig. 1, which shows the posterior densities of θ_i, which are based on (2.5) for empirical Bayes and approximated by (4.6) for Bayes.

The posterior moments of $\boldsymbol{\theta}$ may be used to find an approximate interval estimate of an annual rate and to compare rates in different cities. Under the normal approximation, we can first derive an interval estimate for θ_k, then find the corresponding interval for the annual rate. For example, with city 62, a 95% interval for θ_{62} is $-4.700 \pm 1.96(.142)$ or $(-4.978, -4.422)$. Since $p_k = 1/[1 + \exp(-\theta_k)]$ the interval for the annual rate, $10^5 p_{62}$, is (684, 1187).

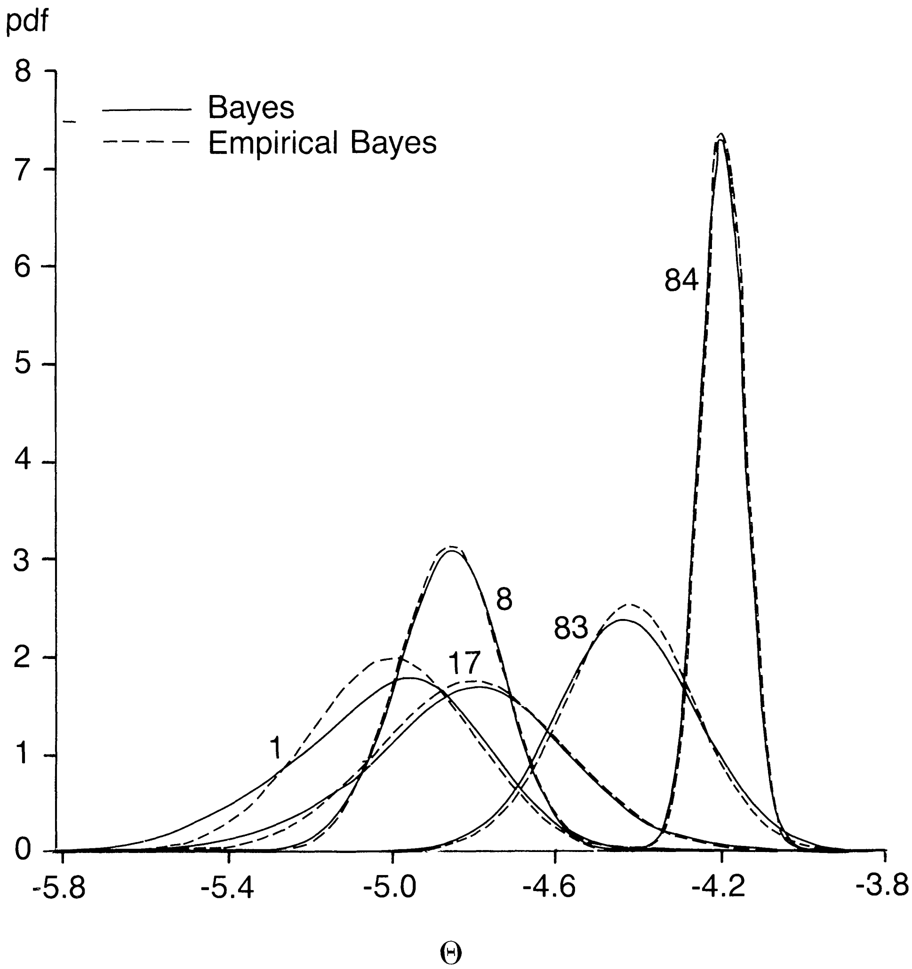

Figure 1. Posterior pdfs of θ under empirical Bayes and Bayesian methods for cities 1, 8, 17, 83, and 84.

Two cities, say j and k, may be compared in terms of the posterior distribution of $\theta_k - \theta_j$, whose mean and variance may again be obtained by Lindley's approximation. To compare cities $j = 8$ and $k = 83$, for example, consider the standardized difference

$$(\tilde{\theta}_j - \tilde{\theta}_k)/[v(\theta_j) - 2\ \mathrm{cov}(\theta_j, \theta_k) + v(\theta_k)]^{1/2}$$

$$= (4.428 - 4.865)/\{[1.7017 - 2(-.0726) + 2.8553]10^{-2}\}^{1/2} = -2.02,$$

where $\tilde{\theta}_j$ and $\tilde{\theta}_k$ are the posterior means from Table 1 and the variances and covariance are

Table 2
Posterior covariance matrix ($\times\ 10^2$) *of* $\boldsymbol{\theta}$ *for five cities*

	City number				
City number	1	8	17	83	84
1	5.4778	.2746	.6207	−.4279	−.1222
8		1.7017	.1756	−.0726	−.0226
17			6.2293	−.1012	−.0426
83				2.8553	.0800
84					.3440

taken from Table 2. Under the normal approximation, the posterior probability that city 83 has a higher rate than city 8 is 97.9%.

ACKNOWLEDGEMENTS

This work was supported by Grant No. 5 R01 CA 29765 from the National Cancer Institute. The author wishes to thank C. J. Marienfeld for suggesting the problem, D. V. Lindley for a brief but stimulating discussion, and G. L. Shoop and W. L. Lo for computation. He also wishes to thank the referees for many helpful suggestions.

RÉSUMÉ

Cet article présente une méthode bayésienne d'estimation de taux de mortalités spécifiques quand le nombre des décès sur un intervalle de temps donné a une distribution de Poisson dont la moyenne est proportionnelle à la taille de la population. Les estimateurs utilisent l'information fournie par des populations voisines, chacune ayant ses propres taux supposés distribués selon une même distribution a priori sur laquelle on dispose d'information.

La motivation de ce travail a été une étude épidémiologique de la variation géographique de la mortalité par cancer dans l'état du Missouri. Des données de cette étude permettent d'illustrer la méthode et de la comparer aux méthodes plus simples dites "bayésiennes empiriques."

REFERENCES

Breslow, N. E. and Day, N. E. (1975). Indirect standardization and multiplicative models of rates with reference to the age adjustment of cancer incidence and relative frequency data. *Journal of Chronic Diseases* **28**, 289–303.

Deeley, J. J. and Lindley, D. V. (1981). Bayes empirical Bayes. *Journal of the American Statistical Association* **76**, 833–841.

Dempster, A. P., Rubin, D. B., and Tsutakawa, R. K. (1981). Estimation in covariance component models. *Journal of the American Statistical Association* **76**, 341–353.

Good, I. J. (1956). On the estimation of small frequencies in contingency tables. *Journal of the Royal Statistical Society, Series B* **56**, 113–124.

Hill, A. B. (1977). *A Short Textbook of Medical Statistics*, 10th ed. Philadelphia: Lippincott.

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

Leonard, T. (1972). Bayesian methods for binomial data. *Biometrika* **59**, 581–589.

Leonard, T. (1975). Bayesian estimation methods for two-way contingency tables. *Journal of the Royal Statistical Society, Series B* **37**, 23–37.

Lindley, D. V. (1980). Approximate Bayesian methods. *Trabajos Estadistica* **31**, 223–237.

Manton, K. G., Woodbury, M. A., and Stallard E. (1981). A variance components approach to categorical data models with heterogeneous cell populations: Analysis of spacial gradients in lung cancer mortality rates in North Carolina counties. *Biometrics* **37**, 259–269.

Marienfeld, C. J., Collins, M., Wright, H., Reddy, R., Shoop, G., Roberts, K. K., and Rust, P. (1980). Cancer mortality and public drinking water in St. Louis city and county. *American Water Works Association Journal* **72**, 649–654.

Mason, T. J. and McKay, F. W. (1974). *U.S. Cancer Mortality by County, 1950–1969*. Washington, D.C.: DHEW. Publication Number (NIH)74-615.

Received January 1983; *revised December* 1983 *and June* 1984.

APPENDIX

Computational Formulas

The implementation of Lindley's approximation requires the evaluation of many derivatives whose expressions we now summarize for computational purposes. From (2.1), (2.2), and (2.3), the log likelihood function is given by

$$L = L(\mu, \sigma) = \Sigma \log \int_{-\infty}^{\infty} f_i g \, d\theta,$$

where we now use the abbreviations

$$f_i = f(y_i \mid r_i, \theta) \quad \text{and} \quad g = g(\theta \mid \mu, \sigma).$$

Denote the partial derivatives $\partial^{r+t}g/\partial\mu^r\partial\sigma^t$ by g_{rt}, and let $z = (\theta - \mu)/\sigma$. Then the derivatives of g that are needed may be expressed as follows:

$$\begin{aligned}
g_{00} &= g \\
g_{10} &= \sigma^{-1}zg, \\
g_{01} &= \sigma^{-1}(z^2 - 1)g \\
g_{20} &= \sigma^{-2}(z^2 - 1)g, \\
g_{11} &= \sigma^{-2}z(z^2 - 3)g, \\
g_{02} &= \sigma^{-2}(z^4 - 5z^2 + 2)g, \\
g_{30} &= \sigma^{-3}z(z^2 - 3)g, \\
g_{21} &= \sigma^{-3}(z^4 - 6z^2 + 3)g, \\
g_{12} &= \sigma^{-3}z(z^4 - 9z^2 + 12)g, \\
g_{03} &= \sigma^{-3}(z^6 - 12z^4 + 27z^2 - 6)g.
\end{aligned}$$

When evaluated at $(\mu, \sigma) = (\hat{\mu}, \hat{\sigma})$, let

$$I_{irt} = \int f_i g_{rt}\, d\theta$$

and

$$Q_{irt} = I_{irt}/I_{i00},$$

for $r = 0, 1, 2, 3$ and $r + t = 0, 1, 2, 3$. Then

$$\begin{aligned}
L_{00} &= \sum \log I_{i00}, \\
L_{10} &= \sum Q_{i10}, \\
L_{01} &= \sum Q_{i01}, \\
L_{20} &= \sum (Q_{i20} - Q_{i10}^2), \\
L_{11} &= \sum (Q_{i11} - Q_{i10}Q_{i01}), \\
L_{02} &= \sum (Q_{i012} - Q_{i01}^2), \\
L_{30} &= \sum (Q_{i30} - 3Q_{i20}Q_{i10} + 2Q_{i10}^3), \\
L_{21} &= \sum [Q_{i21} - Q_{i20}Q_{i01} - 2Q_{i10}(Q_{i11} - Q_{i10}Q_{i01})], \\
L_{12} &= \sum (Q_{i12} - 2Q_{i11}Q_{i01} - Q_{i10}Q_{i02} + 2Q_{i10}Q_{i01}^2), \\
L_{03} &= \sum (Q_{i03} - 3Q_{i02}Q_{i01} + 2Q_{i10}^3).
\end{aligned}$$

The function $u(\mu, \sigma)$ considered in this paper can be classified as one of the following types:

$$u(\mu, \sigma) = \frac{f_i g}{D_i}, \tag{A.1}$$

$$u(\mu, \sigma) = \frac{\int c(\theta) f_i g\, d\theta}{D_i}, \tag{A.2}$$

or

$$u(\mu, \sigma) = \frac{\int c(\theta) f_i g\, d\theta \int c(\theta) f_j g\, d\theta}{D_i D_j}, \tag{A.3}$$

$i \neq j$, where $c(\theta)$ is some simple function of θ and

$$D_i = \int f_i g \, d\theta.$$

The derivatives for type (A.3) may be readily obtained from the derivatives for type (A.2). For (A.1) and (A.2), let A_{irt} be $f_i g_{rt}$ and $\int c(\theta) f_i g_{rt} \, d\theta$, respectively. Then the derivatives for (A.1) and (A.2) evaluated at $(\hat{\mu}, \hat{\sigma})$ may be summarized as follows:

$$u = A_{i00}/I_{i00},$$

$$u_1 = (A_{i10} - A_{i00} Q_{i10})/I_{i00},$$

$$u_2 = (A_{i01} - A_{i00} Q_{i01})/I_{i00},$$

$$u_{11} = (A_{i20} - 2A_{i10} Q_{i10} - A_{i00} Q_{i20} + 2A_{i00} Q_{i10}^2)/I_{i00},$$

$$u_{22} = (A_{i02} - 2A_{i01} Q_{i01} - A_{i00} Q_{i02} + 2A_{i00} Q_{i01}^2)/I_{i00},$$

$$u_{12} = (A_{i11} - A_{i10} Q_{i01} - A_{i01} Q_{i10} - A_{i00} Q_{i11} + 2A_{i00} Q_{i10} Q_{i01})/I_{i00},$$

$$u_{21} = (A_{i11} - A_{i01} Q_{i10} + A_{i10} Q_{i01} - A_{i00} Q_{i11} + 2A_{i00} Q_{i01} Q_{i10})/I_{i00}.$$