Environmental Health Perspectives
Vol. 96, pp. 223–238, 1991

Assessment of Spatial Variation of Risks in Small Populations

by Wilson B. Riggan,* Kenneth G. Manton,† John P. Creason,* Max A. Woodbury,† and Eric Stallard†

Often environmental hazards are assessed by examining the spatial variation of disease-specific mortality or morbidity rates. These rates, when estimated for small local populations, can have a high degree of random variation or uncertainty associated with them. If those rate estimates are used to prioritize environmental clean-up actions or to allocate resources, then those decisions may be influenced by this high degree of uncertainty. Unfortunately, the effect of this uncertainty is not to add "random noise" into the decision-making process, but to systematically bias action toward the smallest populations where uncertainty is greatest and where extreme high and low rate deviations are most likely to be manifest by chance. We present a statistical procedure for adjusting rate estimates for differences in variability due to differentials in local area population sizes. Such adjustments produce rate estimates for areas that have better properties than the unadjusted rates for use in making statistically based decisions about the entire set of areas. Examples are provided for county variation in bladder, stomach, and lung cancer mortality rates for U.S. white males for the period 1970 to 1979.

Introduction

Evaluation of the geographic variation of disease-specific incidence and death rates across small areas is important in identifying potential environmental hazards and in determining priorities for responses to ameliorate such environmental hazards. The finer the geographic detail (i.e., the smaller the area), the greater is the capacity to identify potential environmental causes of disease risks. Unfortunately, there is a difficulty in using the observed rates for small areas to make such decisions, i.e., small areas also tend to have small populations. The precision of a rate estimate is inversely related to the size of the local population and number of index events in the area. If there is a wide range of population sizes over the set of small areas, we can expect the rates for the smallest populations to have the greatest variability and thus to be overrepresented in groups of areas with highest and lowest rates. Consequently, use of the observed rates for small areas may introduce systematic errors in decision making if decisions require the identification of areas with rates that are truly extreme.

For example, if one had resources to conduct detailed epidemiological studies in the 0.1 % of small areas (e.g., the 3061 U.S. counties) with the highest rates, the studies would be almost certainly targeted to the subset of areas with the smallest populations which would have the most extreme (both high and low) rates due to random variation. Likewise, if one wished to target clean-up activities in a given year to the 10% of small areas with the greatest environmental hazards, use of the observed rates would again lead to allocation of clean-up resources to areas with small populations and large random variation in rates.

The decisions based on the observed rates are even more problematic if considered in terms of targeting actions to benefit the largest number of persons. That is, the 10% most extreme rates would, by chance, tend to be concentrated in areas with very small populations which, being smaller than average, would contain far less than 10% of the total population.

In addition, in studying the patterns of environmental risks over small areas, the use of observed rates may lead to distortion of spatial patterns because small populations with extreme deviations will dominate those patterns. To identify the interrelation of risks over spatial domains, it is again necessary to adjust the rates for chance variations due to small population sizes.

To deal with these problems, specialized statistical procedures were developed to produce stabilized rate estimators for small populations that are more precise on average (i.e., across the total set of small area populations) than the usual independent maximum likelihood rate estimator for each area, i.e., the observed "rate" or the ratio of events to population exposure in the area. To improve the average level of precision, these procedures borrow information from the distribution of rates to adjust the rate estimator for each area, i.e., the average rate over all areas is combined with the observed rate in each local area to produce a stabilized rate for the area. The average rate is a biased estimator, and the observed rate is an unstable estimator (i.e., subject to large random variation) for each local area. Combining weights are calculated which simultaneously *a*) minimize the bias of the average rate and *b*) increase the stability of the observed rate. If

*Health Effects Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711.
†Center for Demographic Studies, Duke University, Durham, NC 27706.
Address reprint requests to W. B. Riggan, Health Effects Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, NC 27706.

the weights are appropriately selected, then stabilized rate estimates for the set of local area populations can be produced where the standard errors of the rates are independent (or nearly so) of the size of each area's population. Thus, decisions requiring the identification of truly extreme rates can be more confidently made on the basis of the stabilized rates. The statistical methods that can be used to produce composite rate estimates with the desired properties are based upon "empirical Bayes" principles [e.g., Morris (*1,2*)].

In the remainder of the paper we briefly describe two empirical Bayes procedures (*3,4*) and then apply them to data on U.S. cancer death rates at the county level (*5*). The effects of these adjustments are illustrated using maps of the between-county variation of three types of cancers for white males in the United States in 1970 to 1979.

Data

The data employed in the analyses are drawn from files of county-specific cancer death rates prepared by the U.S. Environmental Protection Agency (EPA) from detailed micro-data mortality files prepared by the National Center for Health Statistics and censal and intercensal population estimates provided by the U.S. Bureau of the Census (*5*).

Death rates were calculated for 18 age categories (0 to 4 years, 5 to 9 years, etc., up to 85+), race (whites versus nonwhites), sex, county, and 15 different types of cancer identified from the underlying cause of death coded on U.S. death certificates (*4,5*). Tabulations of death and population counts were prepared for 3061 counties (or county equivalents). These were adjusted to match the set of 3073 counties defined by the available mapping software (*5*). Rates were available for single calendar years between 1950 and 1979. Recently these files have been extended to 1987, the most recent date for which mortality data are available. We restricted the analysis presented below to three of the 15 cancers (bladder, stomach, and lung) for white males for the decade 1970 to 1979.

Preliminaries

There are several different definitions of the "observed death rate." Each implies a different treatment of age. One option, stratification by age, produces a vector of 18 rates for each county. A second option is to aggregate over age, obtaining the "crude death rate" (CDR) for the *ith* county as

$$\mathrm{CDR}_i = \sum_{j=1}^{J} y_{ij} \Big/ \sum_{j=1}^{J} n_{ij} \tag{1a}$$

$$= y_{i+} / n_{i+} \tag{1b}$$

where y_{ij} is the count of deaths in county i for the *jth* age group, n_{ij} is the population in that group, and $J = 18$. The + subscript indicates summation over age.

Though CDR_i is an observed death rate, it is unsatisfactory for comparing county rates since two counties with identical age-specific death rates can yield different crude death rates because of differences in the age-specific population counts, n_{ij} (*6*). Thus, it is conventional to perform "direct age standardization" of the vector of age-specific death rates (DASDR) using

$$\mathrm{DASDR}_i = \sum_{j=1}^{J} N_j\,(y_{ij}/n_{ij}) \Big/ \sum_{j=1}^{J} N_j \tag{2a}$$

$$= \sum_{j=1}^{J} \frac{N_j}{N_+} \cdot m_{ij}\,, \tag{2b}$$

where m_{ij} is the death rate for age group j in county i. N_j values are the age-specific standard populations used to weight the m_{ij} values [here the 1970 U.S. population was used as a standard (*5*)].

For comparison, the "marginal age standardized death rate" (MASDR) can be calculated which, when $N_j = n_{+j}$, is simply the crude death rate for the entire set of areas,

$$\mathrm{MASDR} = \sum_{j=1}^{J} \left(\frac{N_j}{N_+}\right)\left(\frac{y_{+j}}{n_{+j}}\right) \tag{3a}$$

$$= \sum_{j=1}^{J} \frac{N_j}{N_+}\, m_{*j}\,, \tag{3b}$$

where m_{*j} is the death rate for age group j at the national level.

Because comparisons of DASDR_i are not confounded by differences in the age structure of the population between counties, it is often used as the observed death rate. It is, however, only an estimate of the rate because its value depends on the choice of a standard population. More important, however, DASDR_i is actually more statistically unstable than CDR_i. Specifically, Eq. (2b) shows that when n_{i+} is small, n_{ij} is substantially smaller. Consequently, random variation in y_{ij} will be large relative to that of y_{i+}. Thus, there is additional instability in DASDR_i not present in CDR_i. There are three ways to deal with this additional instability when making comparisons.

First, North Carolina (*7*) presents maps of both the CDR_i vallues and DASDR_i values and cautions the reader to beware of discrepancies. Second, the National Cancer Institute (NCI) (*8,9*) performs statistical tests of the significance of elevated DASDR_i values and maps the results of the test (rather than the values of the rates) and indicates which counties had elevated rates and which of those with elevated rates had statistically significant elevated rates. In both cases, the user is required to simultaneously deal with two complex patterns of geographic information. The methods provide no procedures for ranking the county rates from lowest to highest.

A third strategy is to replace direct age standardization with indirect age standardization using

$$\mathrm{IASDR}_i = \frac{y_{i+}}{\sum_{j=1}^{J} n_{ij}\, m_{*j}}\,\mathrm{MASDR} \tag{4a}$$

$$= r_i\ \mathrm{MASDR}\,, \tag{4b}$$

where r_i is the standardized mortality ratio (SMR) in county i. This rate is generated by applying a "standard" mortality schedule to the population age structure in the county. The relative risk between the observed overall count of deaths (y_{i+})

and that expected from the application of the national mortality rates to the county population is represented by r_i.

Like $DASDR_i$ values, $ISADR_i$ values are comparable between counties, but under a more restrictive assumption [proportional hazards (*6*)]. On the other hand, they are more stable than $DASDR_i$ values because they depend only on the total death count, y_{i+}, not the 18 age-specific counts, y_{ij}. Thus, their stability is comparable to that of CDR_i values.

Because r_i is proportional to $IASDR_i$, the analysis of local area populations is often carried out using r_i (*10,11*). Instability in r_i is typically handled by aggregating adjacent areas to obtain an area with a larger population size. This, of course, loses some of the geographic detail that is necessary to relate the elevated mortality risks to possible environmental causes.

Methods

We briefly discuss two distinct forms of empirical Bayes (EB) analysis. The first, the quintile model, is based on an extension of the SMR model in Manton et al. (*3*: 810) to age-specific death rates. The second, the two-stage model, analyzes both total and age-specific death rates, but sequentially. The statistical details and theoretical justification of the second model are presented in Manton et al. (*4*). This section briefly examines EB concepts to introduce the nonstatistical reader to the basic principles of this method. In the results section we compare the empirical performance of the two methods.

If the observed rates are inadequate for mapping because of their large random fluctuations, one must find rate estimates that are more stable to replace them with. One might use rates for each area assuming that the rates are temporally stable. For example, in the analysis below we pooled each area's data by decade. However, pooling data over too long a period of time may cause temporal changes in risk to be missed. A decade was the longest period we felt could be substantively justified. However, even decade-specific rates were still often unstable. Thus, it was necessary to generalize the principle of averaging beyond the currently available data to a hypothetical case where the observed rate is one of an infinite number of outcomes that could have hap-

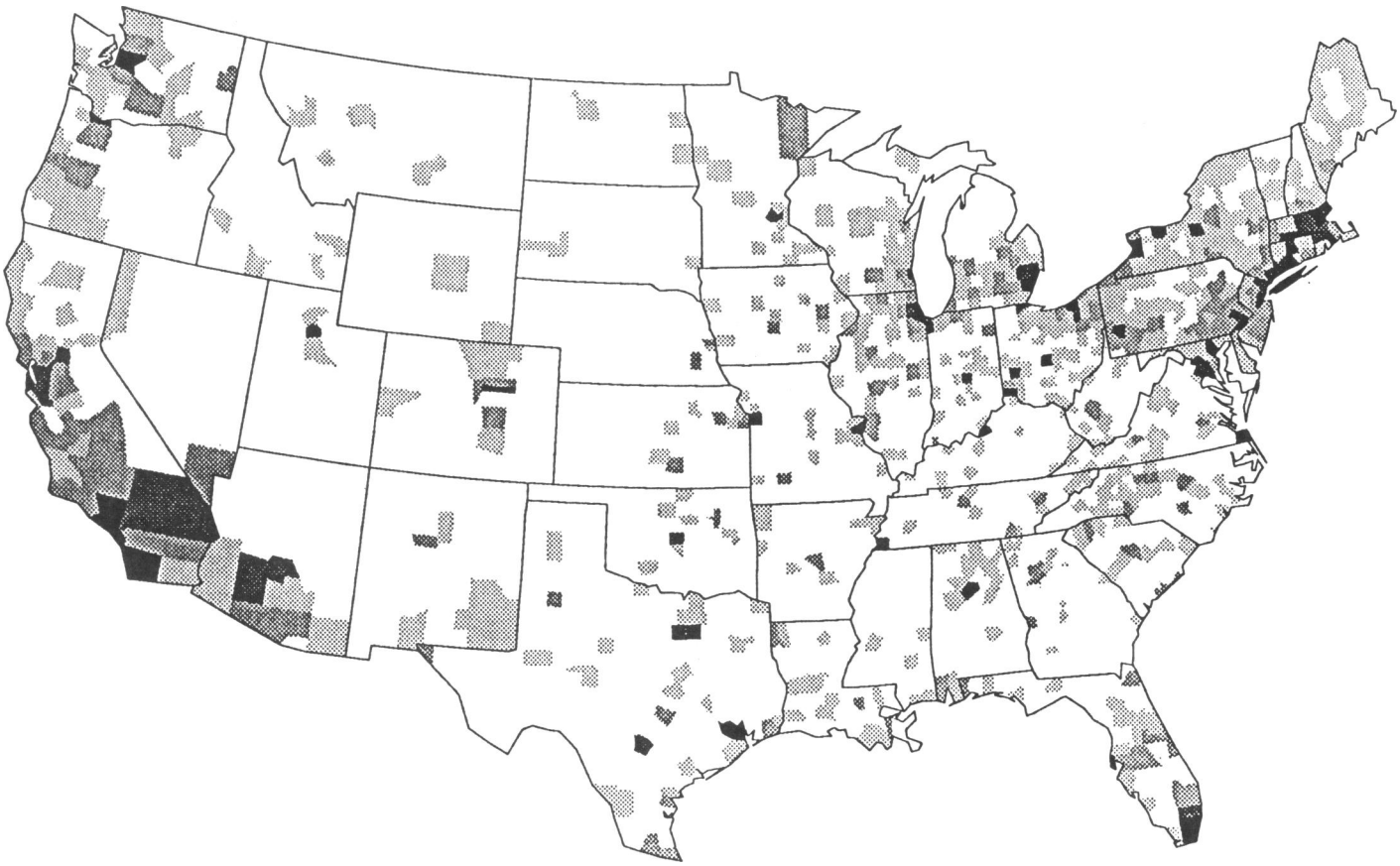

FIGURE 1. Classification of 3061 counties of the continental United States according to population size. Class 1 (white) contains 2340 counties with the smallest populations; class 2 (light gray) contains 497 counties with the next smallest populations; class 3 (middle gray) contains 148 counties with the next smallest populations; class 4 (dark gray) contains 57 counties with the next-to-largest populations; and class 5 (black) contains 19 counties with the largest populations. The population sizes are based on averages of the census counts for 1960 and 1970 for white males. The total populations in each class are equal, accounting for 20% of the overall total.

pened. The theoretically possible outcomes are described in different ways by different EB procedures and the actual data are used in different ways depending upon how the theoretical distribution of outcomes is assumed to be generated. This generalization is the underlying essence of the EB models.

The term "empirical Bayes" was introduced by Robbins (*12,13*) to refer to decision problems in which the identical conditions are faced repeatedly. For each decision, new data are provided and one wants to estimate the long run average over repeated, identical experiments. This problem is often referred to as "nonparametric empirical Bayes" (NPEB) (*1,14*) and is discussed by a number of authors (*15–17*).

NPEB is not direcdy applicable to the mapping problem because it requires multiple observed rates for each county for each decade produced under identical conditions. The NPEB procedure was developed for situations where the identical experiment could be repeated. We cannot repeat the identical conditions producing the set of cancer deaths in a county. Unlike the NPEB problem, however, we do have observations on multiple (n = 3061) counties and, although the conditions in these counties are not identical, the decisions to be made for each county are the same. If the age-specific population counts were constant over counties, the decision problems would be identical.

Morris (*1*) uses the term "parametric" empirical Bayes (PEB) to refer to EB problems where the conditions producing each event (e.g., a county cancer mortality rate) are similar in some respect, but not identical. By introducing a parametric distribution into the EB model the rate estimator for each county can be made dependent, through the parameters of the selected distribution, on the rate estimator for all other counties. In effect, this allows the information required to estimate one county's long run average to be obtained from the rates in all other counties (i.e., an average is obtained for the entire set of counties). This involves assuming that, say, a Poisson process governs the generation of cancer deaths in each count but the Poisson rates in each county may be different. Thus, the similarity is the nature of the generating process. Rates are made comparable by estimating the distribution of Poisson rates across counties.

PEB models are closely related to the Stein (*18*) effect and James–Stein estimator (*19*), as shown in Efron and Morris (*20*). The PEB approach readily generalizes for many contests. A James–Stein-type estimator was used by Fay and Herriot (*21*) to estimate small area income averages. PEB estimators have been used extensively for estimating vital rates in small areas (*3,4,22–29*). Thus, there is a developing consensus that PEB models can provide useful solutions to the problem of rate estimation for small local populations. Of course, the results of each application depend on the assumptions used in model specification. Knowledge of these dependencies is important to understand the properties of the different versions of this method.

Quintile Model

The quintile model involves sorting counties into five (some other grouping could have been used) size classes, with each class containing 20% of the total population (not 20% of the areas), i.e., class 1 contains 2340 counties; class 2 contains 497 counties; class 3 contains 148 counties; class 4 contains 57 counties; and class 5 contains 19 counties. On average, counties in class 5 are more than 120 times larger than counties in class 1. The counties and their class designations presented in Figure 1.

Because counties in each class are approximately equal in size, it is assumed that the PEB decision problem for the quintile is the same. This assumption could be improved by increasing the number of classes. Two additional assumptions are made: the variation of each area's rates is determined by the Poisson distribution and the total variation of all rates is proportional to the Poisson variation. The difference between the total variation and the "natural" variation is called the "excess" variation, (i.e., variation beyond that due to the Poisson with the rate parameter for an area) and can be tested using the methods of Collings and Margolin (*30*) and Dean and Lawless (*31*).

The assumption of Poisson-distributed counts of deaths in counties is consistent with the test procedures used by NCI (*8,9*). A more fundamental justification is provided by Brillinger (*32*). The assumption of proportional excess variation is consistent with the heterogeneity model proposed by Manton and Stallard (*33*) and Manton et al. (*34*). This model represents the rates for each local area as a composite of the risks for all individuals in that area. As small areas are combined to form larger areas, the rates for the different areas are weighted according to population six to form a composite rate, and the variances retain their proportionality to the Poisson variation.

Under these assumptions, the empirical Bayes age-standardized death rate for county i is

$$\mathrm{EBASDR}_i = W_{q_i} \cdot \mathrm{DASDR}_i + (1 - W_{q_i})\, \mathrm{IASDR}_{q_i}\,, \qquad (5)$$

where q_i denotes the quintile class for county i; W_{q_i} is the weight for this quintile class (*3*)

$$W_q = \alpha_q / (1 + \alpha_q)\,, \qquad (6)$$

which is the ratio of excess (α_q) to total ($1 + \alpha_q$) variation in the quintile class; and IASDR_{q_i} is the indirect age standardized death rate for all counties in the quintile class; i.e.,

$$\mathrm{IASDR}_q = r_q\, \mathrm{MASDR}\,, \qquad (7)$$

where r_q is the quintile SMR

$$r_q = \sum_{i \in \mathrm{Class}_q} y_{i+} \Big/ \sum_{j=1}^{J} \left(\sum_{i \in \mathrm{Class}_q} n_{ij} \right) \cdot m_{*j}\,. \qquad (8)$$

Estimation of r_q and α_q is conducted on age-specific data using maximum likelihood methods (*33*). The age strata are treated independently in estimation [unlike Manton et al. (*3*), where total deaths, y_{i+}, were analyzed], with the age standardization in Eq. (5) conducted as a final, separate step. The counties are also treated independently in estimation, so that no specific spatial correlation structure is assumed. If clusters of elevated rates are found in the maps, these are not the result of model assumptions, but reflect real variation (and covariation) between county rates.

One difficulty in interpreting patterns occurs when quintile SMRs are very different and quintile weights, W_q, are small. In this case, the spatial patterns can become dependent on the quintile classification system, i.e., the patterns are dependent on how

the counties are grouped into classes. The second method avoids this problem.

Two-Stage Model

The two-stage model derives from the quintile model by increasing the number of classes until each county is in its own class. In this case Eq. (5) becomes

$$EBASDR_i = W_i \cdot DASDR_i + (1 - W_i)\ IASDR_i\ , \qquad (9)$$

where $IASDR_i$ depends on r_i as in Eq. (4b). This is not satisfactory since r_i is unstable. Thus, the two-stage approach is *a*) generate EB estimates (ρ_i) of r_i to remove instability and *b*) use these estimates in Eq. (9) along with revised W_i values to make the final estimates:

$$EBASDR_i = W \cdot DASDR_i + (1 - W)\ \rho_i\ MASDR\ , \qquad (10)$$

where

$$\rho_i = B_i\ r_i + (1 - B_i)\ , \qquad (11)$$

$$B_i = \frac{\beta\ y_{i+}/r_i}{1 + \beta\ y_{i+}/r_i} \qquad (12)$$

and

$$W = \alpha/(1 + \alpha)\ . \qquad (13)$$

Thus, the EB SMR estimator ρ_i is a weighted average of the observed SMR (r_i) and the overall SMR (1.0). The weight B_i depends on the expected number of deaths (y_{i+}/r_i) in county i. The parameter β is the variance of ρ_i over all counties (*4*). The second-stage weight, W, depends on the parameter α, independently of quintile or county.

Because ρ_i is estimated in the first stage using y_{i+}, age strata are no longer treated independently. They are treated independently, conditional on ρ_i, in the second stage, but this is not the same as the fully independent treatment in the quintile model. Conditional independence is consistent with a model where county level effects are correlated over age via ρ_i. The quintile model could be modified to reflect county level effects by analyzing y_{i+}, rather than y_{ij}.

The counties are treated as independent in the two stage model, so that no spatial correlation structure is imposed. Furthermore, whatever spatial patterns do emerge cannot be confounded with the quintile classifications, since none is used. While small W_q value produce problems in the quintile model, small W values are desirable in the two-stage model. They imply that no excess variation (α) exists, beyond the natural variation, after calculating the first stage estimates. For W or $\alpha = 0$, Eq. (10) simplifies to

$$EBASDR_i = B_i\ IASDR_i + (1 - B_i)\ MASDR\ . \qquad (14)$$

Parameter Estimates

Quintile Model

The parameter estimates for the quntile model are presented in Table 1. The SMRs (r_q values) for all three cancer types increase monotonically over quintile class, except for class 5 for lung cancer. The lowest SMR is 80.7% for class 1 for bladder cancer. The highest SMR is 123.2% for class 5 for stomach cancer. The national death rates (MASDRs) are 7.3×10^{-5}, 8.9×10^{-5}, and 64.2×10^{-5} for bladder, stomach, and lung cancer, respectively. Over time the national death rates have *a*) been stable for bladder cancer (7.3×10^{-5} in 1950–59), *b*) declined about 4% per year for stomach cancer (from 20.1×10^{-5} in 1950–5), and *c*) increased about 4% per year for lung cancer (from 29.8×10^{-5} in 1950–1959). The different temporal patterns of change led us to expect different spatial patterns as well. The product of r_q and MASDR yields the indirect age standardized rate for quintile class q, as indicated in Eq. (7).

The second set of parameters, α_q, reflects the excess variance of the observed death rates relative to their natural variation (*32*). All parameters are statistically significant, indicating that excess variation is present. This excess variance increases monotonicaliy over quintile class, ranging from 2.4% for class 1 for stomach cancer to 477.5% for class 5 for lung cancer. The excess variance also increases monotonically over cancer type, except for class 1 for stomach cancer. Given the role of α_q in the weighting formula [Eq. (6)], these patterns of increase provide support for the assumption that the weight in formula [Eq. (12)] increases with the expected number of deaths (not necessarily population size) and that increases are cancer site dependent (through β). The weights implied by the α_q values in Table 1 range from a low of $W_q = 0.023$ to a high of $W_q = 0.827$, with a median value $W_q = 0.250$. Thus, the quintile specific rate dominates the observed rate in Eq. (5) in most cases considered.

Two-Stage Model

The parameter estimates for the two-stage model are presented in Table 2. The first stage involves estimation and testing of the variance of the set of 3061 county SMRs. The test is significant for the three cancer types. The estimated variance (β) ranges from 4.7 to 7.1%. Alternatively, the coefficient of variation ranges from 21.7 to 26.6%, which is consistent with the range of quintile specific SMRs (r_q values) in Table 1. Indeed, by equating the weights in Eqs. (6) and (12) and solving for α_q, we find that Table 2 predicts that the stomach cancer α_q values should be larger, on average, than the bladder cancer α_q values by the factor $1.8 = (7.1/4.7) \times (8.9/7.3)$ and the lung cancer values should be larger

Table 1. Parameter estimates, U.S. white males, 1970–1979, quintile model.

Quintile class (q)	$r_q \times 100$, Bladder	stomach	Lung	$\alpha_q \times 100$, Bladder	stomach	Lung
1	80.7	83.8*	93.6	3.8	2.4†	32.8
2	96.6	88.2*	98.4	11.9	12.8†	82.0
3	105.2	99.1	103.1	20.6	40.1†	140.9
4	110.9	110.3*	104.8	23.4	112.3†	376.8
5	112.4	123.2*	101.6	33.3	151.1†	477.5
MASDR $\times 10^{-5}$	7.3	8.9*	64.2			

MASDR, marginal age-standardized death rate.
*Parameters significantly different from 1.0 ($p < 0.05$).
†Parameters significantly greater than 0.0 ($p < 0.05$).

Table 2. Parameter estimates, U.S. white males, 1970–1979, two-stage model.

Site	$\beta \times 100$	$\alpha \times 100$	MASDR × 100
Bladder	4.7*	0.1	7.3
Stomach	7.1*	0.0	8.9
Lung	5.0*	8.4*	64.2

MASDR, marginal age-standardized death rate.
*Parameter significantly greater than 0.0 ($p < 0.05$).

by the factor 9.4 = (5.0/4.7) × (64.2/7.3), using MASDRs from Table 1. The median factors in Table 1 are 1.9 (= 40.1/20.6) and 8.6 (= 32.8/3.8), respectively. Thus, the two sets of parameter estimates are in good qualitative agreement.

The second stage tests the residual excess variance (α) from the first stage. For bladder and stomach cancer, this component of variance is negligible: The first stage-model is adequate to characterize those cancers. For lung cancer there is a small but statistically significant residual component of excess variance (8.4%). This means that there is variation in the age-specific lung cancer death rates that is not captured in the proportional hazards assumptions of the SMR model. Previous investigations strongly suggest that cohort effects differentially operating over geographic area produced the excess variation for lung cancer (*35*).

The EB estimator for bladder (since $\alpha \approx 0$) and stomach (since $\alpha=0$) cancer obtains from Eq. (14). The EB estimator for lung cancer (since $\alpha > 0$) obtains from Eq. (10). Thus, the excess variance in the second stage is accounted for in the more complex weighting formula. It is not ignored.

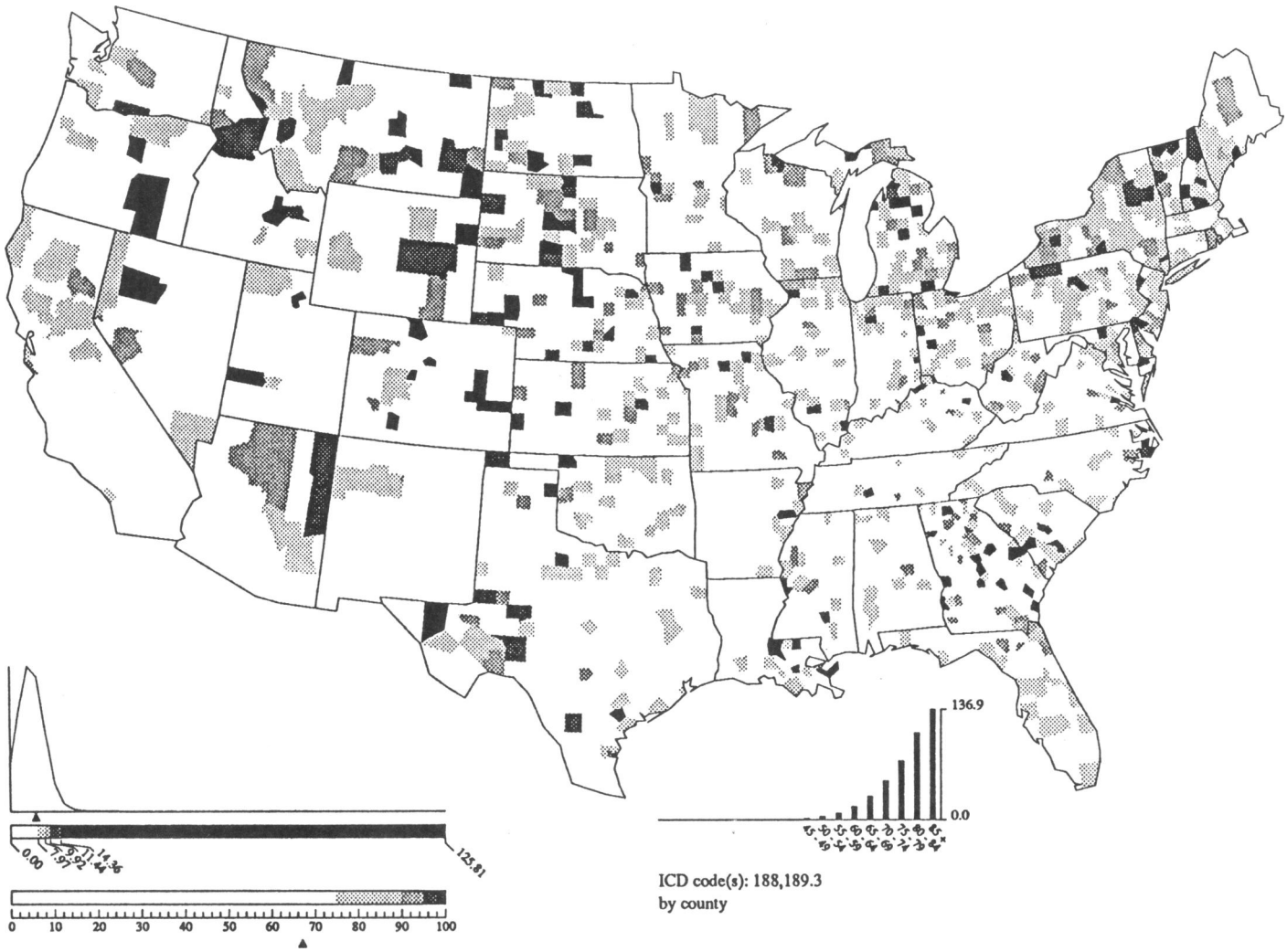

FIGURE 2. Guide to observed rates of cancer of the bladder and other urinary organs for U.S. white males, 1970–1979 map. The map key and frequency polygon for the 3061 county rates are in the lower left corner. Shading is as follows: black, counties ranked in the highest 2% of all counties (98th and 99th percentiles); dark gray, counties in the next highest 3% of all counties (95th to 97th percentiles); medium gray, counties in the next highest 5% of all counties (90th to 94th percentile); light gray, counties in the next highest 15% of all counties (75th to 89th percentile); white counties in the lowest 75% of all counties (0th to 74th percentiles). These cut points are graphically illustrated in the lower tone bar. The triangle below the lower tone bar indicates the relative ranking of the national rate. The upper tone bar relates these cut points to their locations on the frequency polygon constructed from the distribution of the 3061 county rates. The triangle between the upper tone bar and the frequency polygon indicates the location of the national rate. The histogram in the lower right of the map (in the Gulf of Mexico) shows the age-specific death rates for all age groups with significant numbers of deaths at the national level. The ICD codes below the histogram give the disease categories in the *9th Revision of the International Classification of Diseases*. Additional information is provided in Riggan et al. (*5*).

Maps

In this section we present maps of the observed death rates, the quintile model rates, and the two-stage model rates for each of the three cancer types for U.S. white males for 1970 to 1979.

Bladder Cancer

Bladder cancer is of interest in that it is temporally stable, exhibits modest variation over counties, has been linked in correlation studies with chemical exposures in certain industries, and has been previously identified by NCI as concentrated in the Northeast (particularly New Jersey), around the Great Lakes, and in southern Louisiana (*36*). Figure 2 displays the observed rates for bladder cancer.

The rates in the NCI areas are moderately elevated, but higher rates are manifest in the western Great Plains and Rocky Mountains. Comparison with Figure 1 shows that these are almost all sparsely populated (class 1) areas. The NCI maps indicate that most of those highly elevated rates are statistically nonsignificant (*9*).

Figures 3 and 4 display the two sets of EB rate maps. Both shift the locations of the elevated rates to the NCI areas (*36*), so that there is a general concordance between the two models. Examination of the frequency polygon for county rates, however, shows that the quintile model has five modes (at 5.9, 7.1, 7.7, 8.1, and 8.2×10^{-5}). This is because the results for the quintile model depend on how the counties are grouped into six classes, with the dependence inversely related to the size of the α_q parameters. From Table 1, it follows that the effect is greatest for bladder cancer. This is undesirable for producing maps for a large number of different types of cancer. The two-stage model was adopted for use in the U.S. EPA mapping volume (*5*) because no *a priori* size grouping of counties was necessary.

Stomach Cancer

Stomach cancer is of interest because its rates are declining, with the declines associated with cohort differentials [i.e., younger cohorts have lower risks; see Manton and Stallard (*37*)]. Figure 5 displays the observed rates ($DASDR_i$ values) for stomach cancer.

Clearly there is a concentration of elevated rates (18.3×10^{-5}) in the Dakotas, Minnesota, Wisconsin, upper Michigan, New Mexico, Colorado, and several other central states. Comparison with Figure 1 shows that these are almost all sparsely populated (class 1) areas.

Figures 6 and 7 display the two sets of EB rate maps. Both shift the elevated rates to more populated areas. Examination of the frequency polygon shows that both models have unimodal distributions of predicted rates. For the quintile model, however, modes for class 1 at 7.5×10^{-5}, class 2 at 7.9×10^{-5}, and class 3 at 8.8×10^{-5} are discernable. The upper tail of the distribution is spread more than in Figure 3 because of the larger α_q values (Table 1). Less than 5% of counties in Figure 6 are above the national death rate, whereas in Figure 7 more than 25% are above that rate. in this case, it follows that the two-stage model preserves more of the natural variation in the rates. The fact that the statistical test at the second stage showed that there was no excess variation suggests that the two-stage model does better in determining the "correct" level of variation.

In Figure 7 there are considerable changes in the spatial pattern from that in Figure 5. The highest rate is now 15.9×10^{-5}, the lowest rate 4.4×10^{-5} instead of 0.0. There continues to be high rankings of rates in Minnesota, New Mexico, North Dakota, and upper Michigan, but now rates in northeastern states, southern California, Illinois, lower Michigan, and Louisiana have much higher rankings—features hidden by the extreme variation of rates in the smallest counties in Figure 5. A number of these areas are identifiable as being statistically significant in the NCI maps (*36*). Thus, we have additional evidence that the rate estimates from the two-stage model are reasonable and would lead to reasonable rankings of the counties with elevated rates.

Lung Cancer

Lung cancer is interesting because it has increased in risk over time, it is the most frequent cancer cause of death, and it was the one cancer type wher the second stage adjustments proved statistically significant (*4*). Figure 8 displays the observed death rates ($DASDR_i$ values) for lung cancer. The greatest concentration of risk is along the Mississippi and Ohio Rivers in eastern and Gulf coastal areas and in counties of Georgia. The observed death rates range from 0 to 269.1×10^{-5}.

Figures 9 and 10 display the two sets of EB rate maps. In both sets of maps the shifts in ranking are smaller than for stomach and bladder cancer. There is much greater concordance between the two EB rate maps. In Figure 10 the rate varies from 23.6 to 98.8 $\times 10^{-5}$. The shrinkage of the extreme variation, though still considerable, is less than for stomach cancer because of the greater number of lung cancer deaths (e.g., the highest stabilized rates are one-third of the highest observed rates; for stomach cancer, the highest stabilized rates were one-fifth of the highest observed rates). It is interesting that most counties in Montana, North and South Dakota, Colorado, and Utah lose their rankings as high risk areas because of stabilization. What is most interesting is the reinforcement of the high risk patterns in coastal areas in the east and along the Mississippi and Ohio Rivers.

Discussion

Maps based on the stabilized rate estimates allow new spatial features of cancer mortality risks to be identified that reflect absolute levels of risk by using composite estimators that weight the statistical evidence in several ways. The rate stabilization procedure helps in identifying broad spatial patterns suggestive of hypotheses about the sources and nature of environmental risks. In addition, the rate stabilization procedure produces rate estimates for specific areas that allow improved selection of groups of areas for specific types of actions to be implemented. This was illustrated by comparisons with Figure 1, which showed that rate stabilization reduces the assigned rankings for areas with small populations, thereby reducing the odds of incorrectly identifying a small area as a high risk area. Separate comparisons of maps produced by the two-stage model (*5*) with corresponding NCI maps (*9*) showed that NCI's statistical screening procedure yielded groups of elevated counties that closely matched the highest ranked counties under the EB stabilization procedure. Though it is comforting that the NCI procedure and our EB procedure identify comparable sets of highly elevated country rates

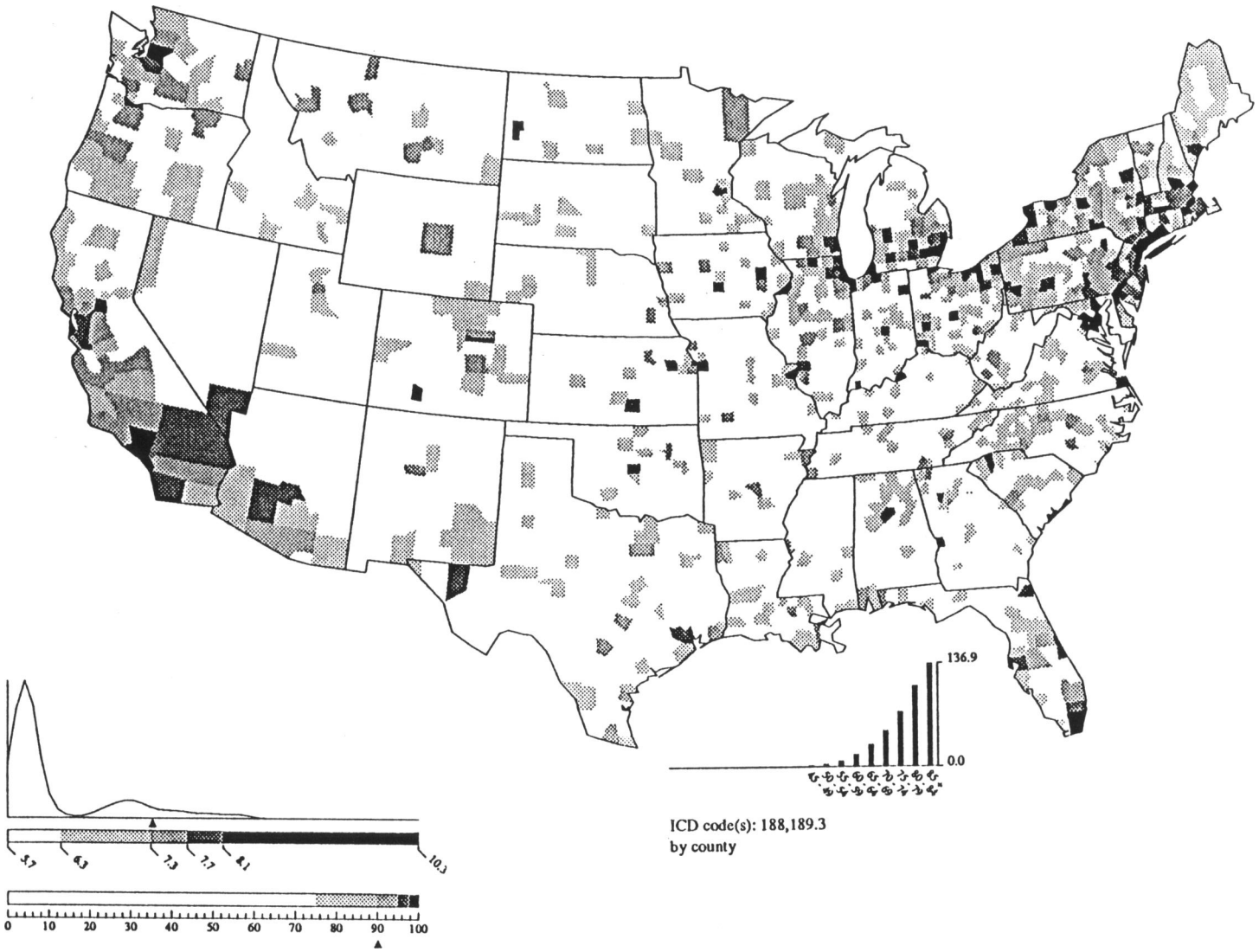

FIGURE 3. Guide to quintile model for cancer of the bladder and other urinary organs for U.S. white males, 1970–1979 map. The map key and frequency polygon for the 3061 county rates are in the lower left corner. Shading is as follows: black, counties ranked in the highest 2% of all counties (98th and 99th percentiles); dark gray, counties in the next highest 3% of all counties (95th to 97th percentiles); medium gray, counties in the next highest 5% of all counties (90th to 94th percentile); light gray, counties in the next highest 15% of all counties (75th to 89th percentile); white counties in the lowest 75% of all counties (0th to 74th percentiles). These cut points are graphically illustrated in the lower tone bar. The triangle below the lower tone bar indicates the relative ranking of the national rate. The upper tone bar relates these cut points to their locations on the frequency polygon constructed from the distribution of the 3061 county rates. The triangle between the upper tone bar and the frequency polygon indicates the location of the national rate. The histogram in the lower right of the map (in the Gulf of Mexico) shows the age-specific death rates for all age groups with significant numbers of deaths at the national level. The ICD codes below the histogram give the disease categories in the *9th Revision of the International Classification of Diseases*. Additional information is provided in Riggan et al. (5).

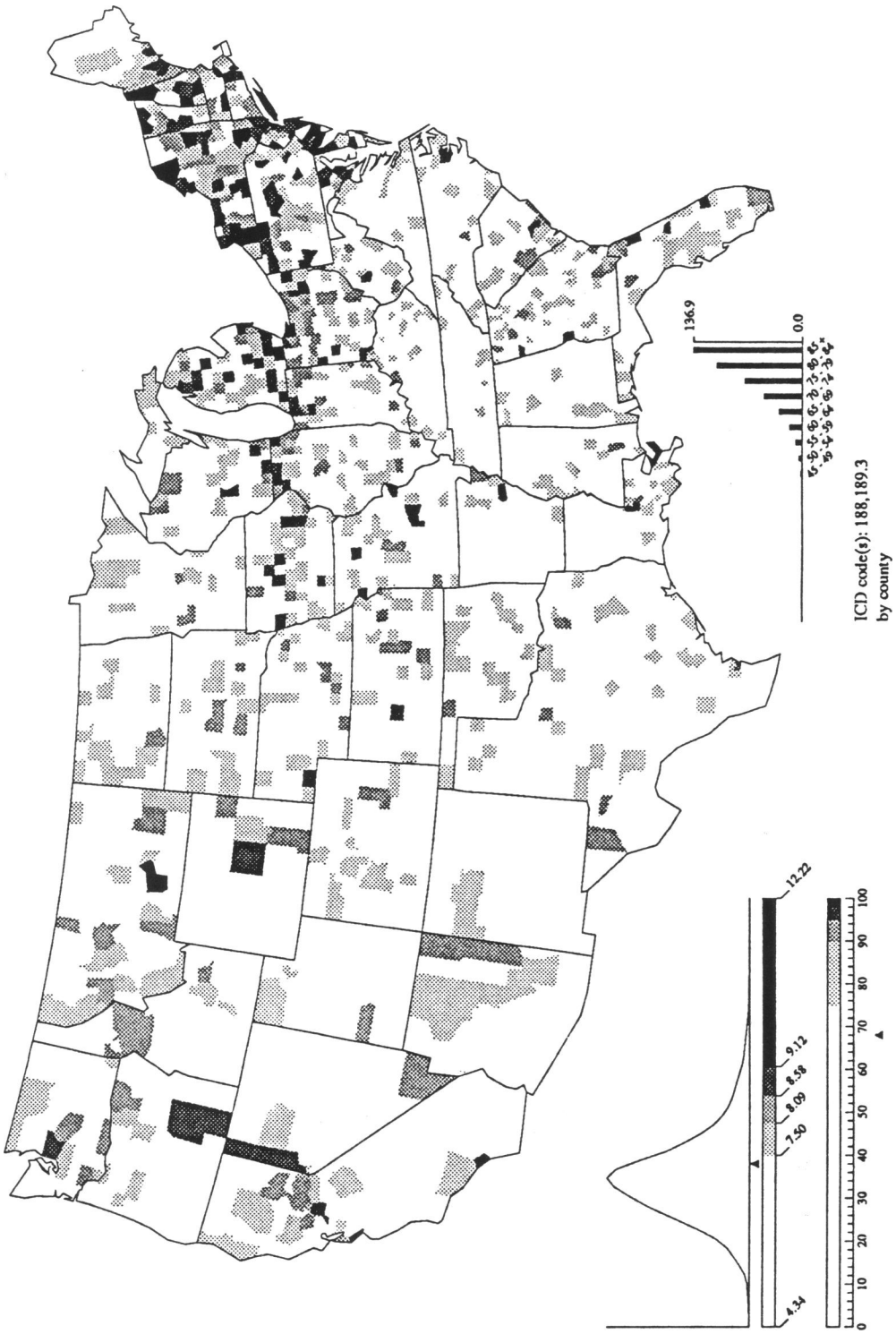

FIGURE 4. Guide to two-stage model for cancer of the bladder and other urinary organs for U.S. white males, 1970–1979 map. The map key and frequency polygon for the 3061 county rates are in the lower left corner. Shading is as follows: black, counties ranked in the highest 2% of all counties (98th and 99th percentiles); dark gray, counties in the next highest 3% of all counties (95th to 97th percentiles); medium gray, counties in the next highest 5% of all counties (90th to 94th percentile); light gray, counties in the next highest 15% of all counties (75th to 89th percentile); white counties in the lowest 75% of all counties (0th to 74th percentiles). These cut points are graphically illustrated in the lower tone bar. The triangle below the lower tone bar indicates the relative ranking of the national rate. The upper tone bar relates these cut points to their locations on the frequency polygon constructed from the distribution of the 3061 county rates. The triangle between the upper tone bar and the frequency polygon indicates the location of the national rate. The histogram in the lower right of the map (in the Gulf of Mexico) shows the age-specific death rates for all age groups with significant numbers of deaths at the national level. The ICD codes below the histogram give the disease categories in the *9th Revision of the International Classification of Diseases*. Additional information is provided in Riggan et al. (5).

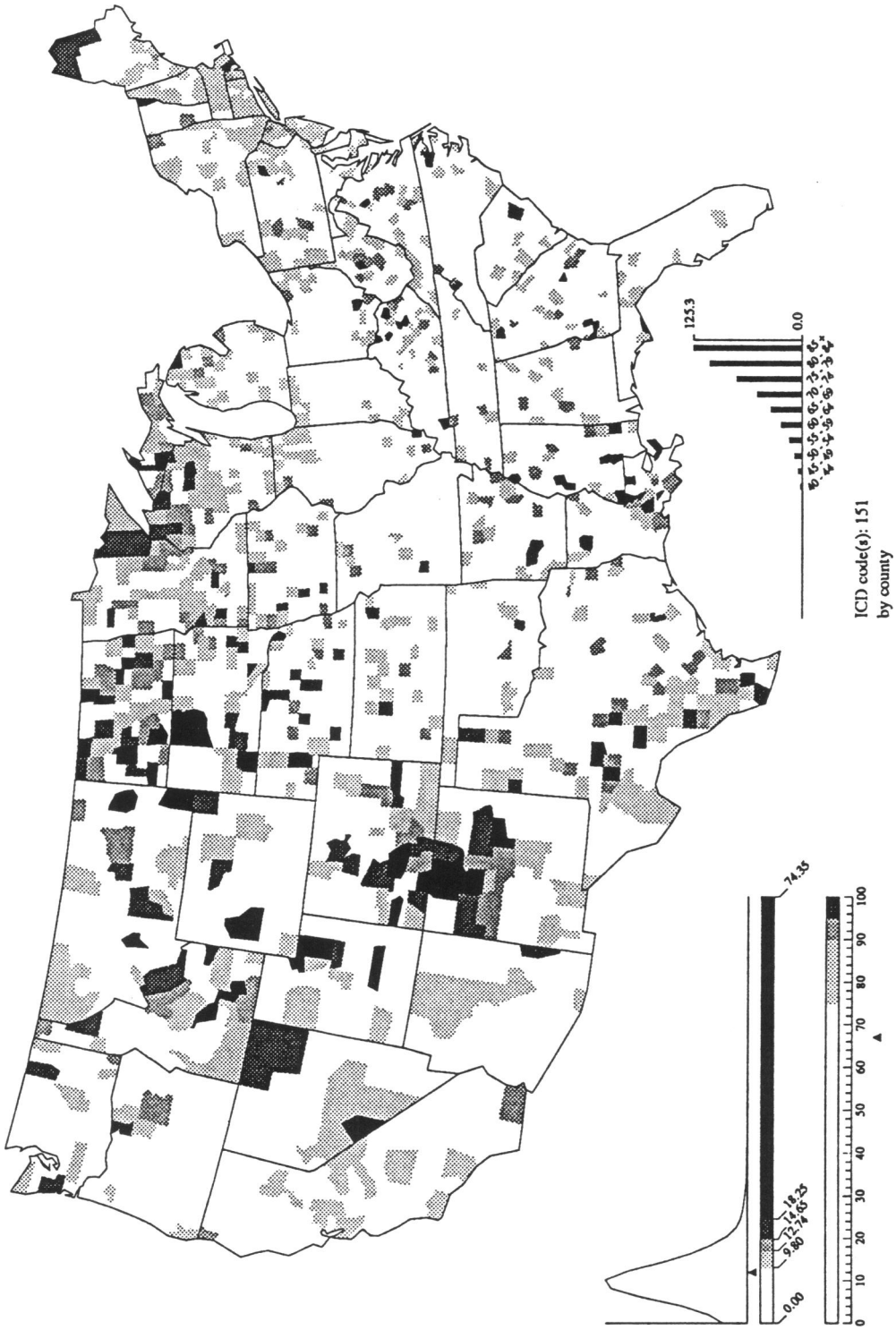

FIGURE 5. Guide to observed rates of cancer of the stomach for U.S. white males, 1970–1979 map. The map key and frequency polygon for the 3061 county rates are in the lower left corner. Shading is as follows: black, counties ranked in the highest 2% of all counties (98th and 99th percentiles); dark gray, counties in the next highest 3% of all counties (95th to 97th percentiles); medium gray, counties in the next highest 5% of all counties (90th to 94th percentile); light gray, counties in the next highest 15% of all counties (75th to 89th percentile); white counties in the lowest 75% of all counties (0th to 74th percentiles). These cut points are graphically illustrated in the lower tone bar. The triangle below the lower tone bar indicates the relative ranking of the national rate. The upper tone bar relates these cut points to their locations on the frequency polygon constructed from the distribution of the 3061 county rates. The triangle between the upper tone bar and the frequency polygon indicates the location of the national rate. The histogram in the lower right of the map (in the Gulf of Mexico) shows the age-specific death rates for all age groups with significant numbers of deaths at the national level. The ICD codes below the histogram give the disease categories in the *9th Revision of the International Classification of Diseases*. Additional information is provided in Riggan et al. (5).

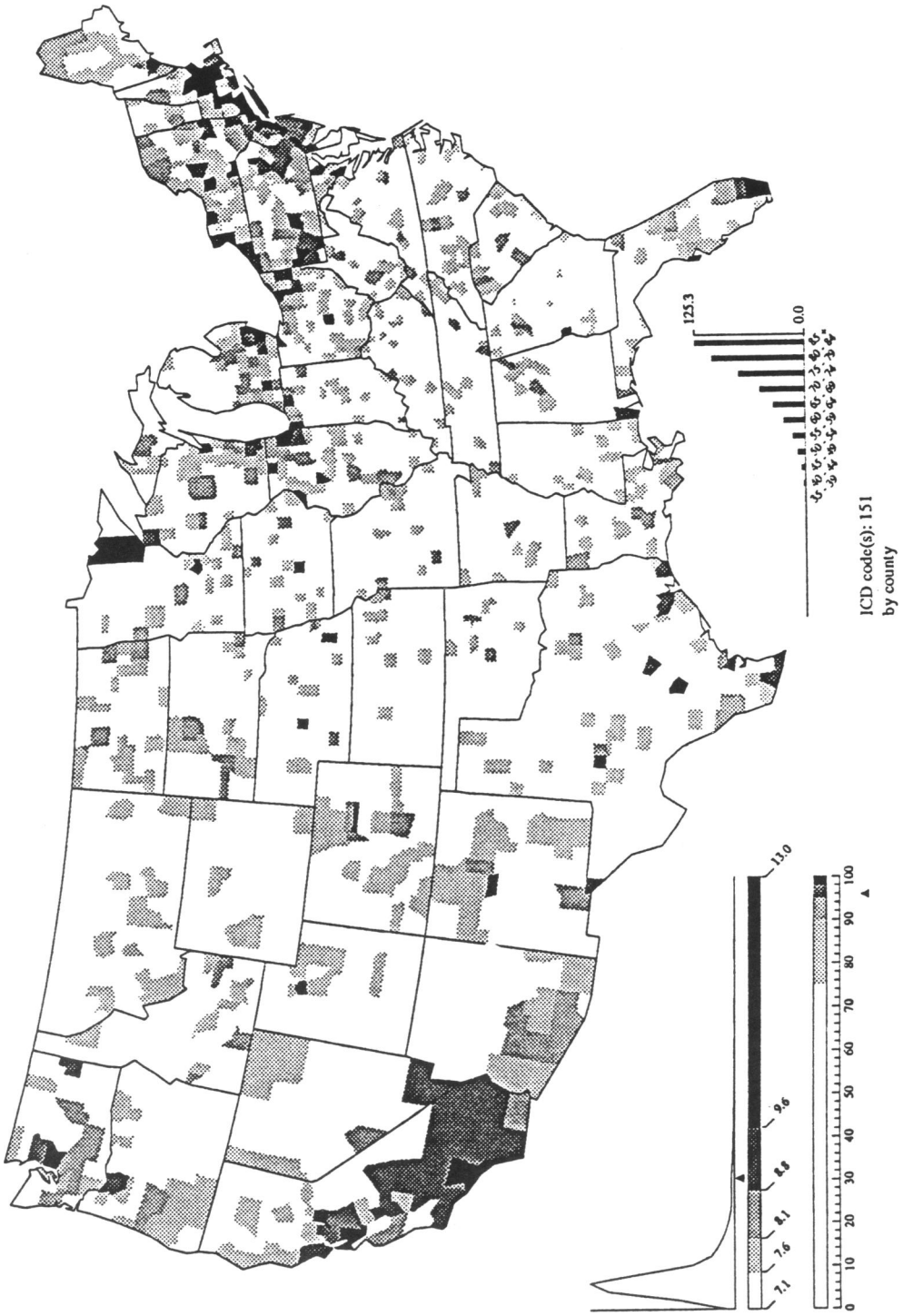

FIGURE 6. Guide to quintile model for cancer of the stomach for U.S. white males, 1970–1979 map. The map key and frequency polygon for the 3061 county rates are in the lower left corner. Shading is as follows: black, counties ranked in the highest 2% of all counties (98th and 99th percentiles); dark gray, counties in the next highest 3% of all counties (95th to 97th percentiles); medium gray, counties in the next highest 5% of all counties (90th to 94th percentile); light gray, counties in the next highest 15% of all counties (75th to 89th percentile); white counties in the lowest 75% of all counties (0th to 74th percentiles). These cut points are graphically illustrated in the lower tone bar. The triangle below the lower tone bar indicates the relative ranking of the national rate. The upper tone bar relates these cut points to their locations on the frequency polygon constructed from the distribution of the 3061 county rates. The triangle between the upper tone bar and the frequency polygon indicates the location of the national rate. The histogram in the lower right of the map (in the Gulf of Mexico) shows the age-specific death rates for all age groups with significant numbers of deaths at the national level. The ICD codes below the histogram give the disease categories in the *9th Revision of the International Classification of Diseases*. Additional information is provided in Riggan et al. (5).

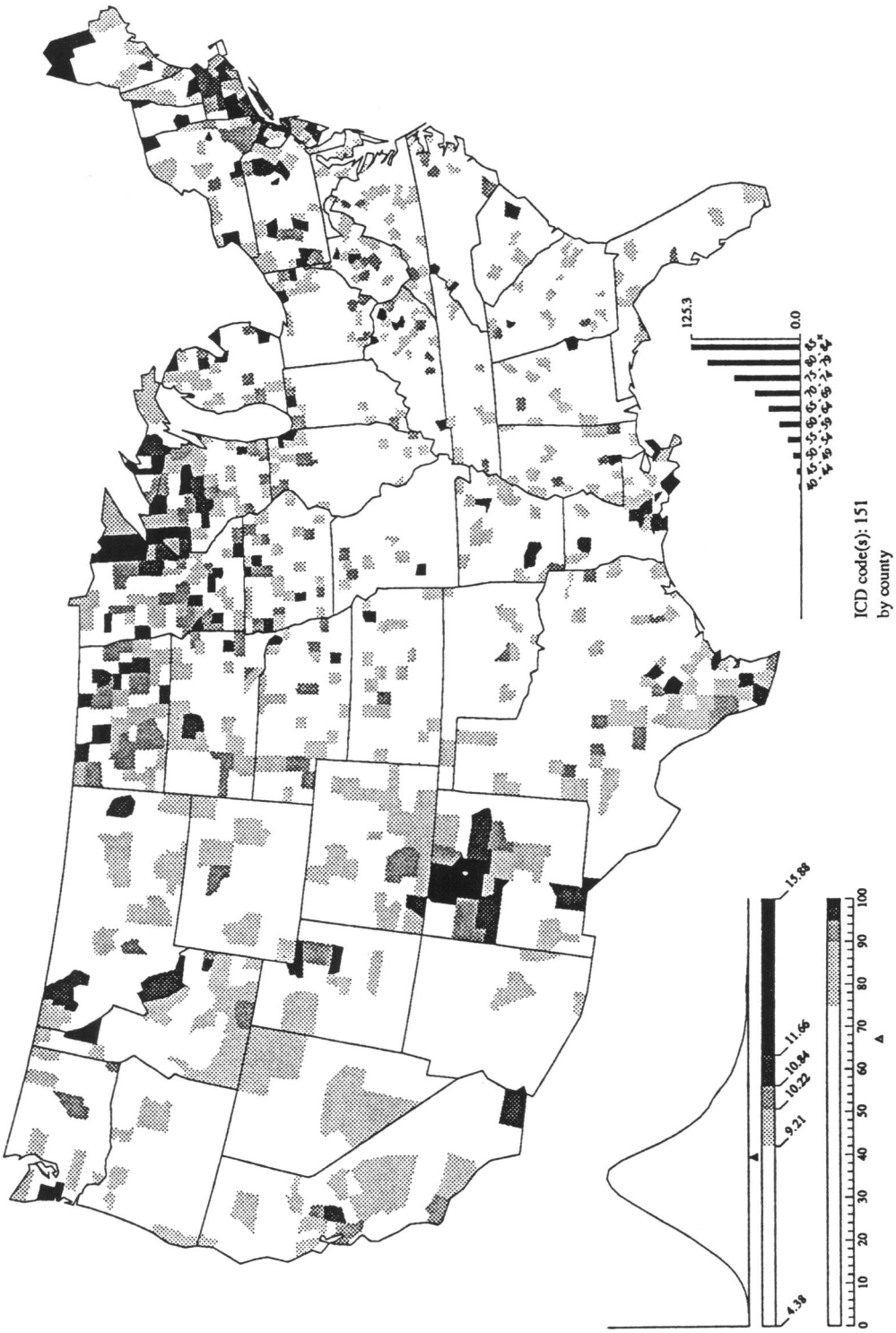

FIGURE 7. Guide to two-stage model for cancer of the stomach for U.S. white males, 1970–1979 map. The map key and frequency polygon for the 3061 county rates are in the lower left corner. Shading is as follows: black, counties ranked in the highest 2% of all counties (98th and 99th percentiles); dark gray, counties in the next highest 3% of all counties (95th to 97th percentiles); medium gray, counties in the next highest 5% of all counties (90th to 94th percentile); light gray, counties in the next highest 15% of all counties (75th to 89th percentile); white counties in the lowest 75% of all counties (0th to 74th percentiles). These cut points are graphically illustrated in the lower tone bar. The triangle below the lower tone bar indicates the relative ranking of the national rate. The upper tone bar relates these cut points to their locations on the frequency polygon constructed from the distribution of the 3061 county rates. The triangle between the upper tone bar and the frequency polygon indicates the location of the national rate. The histogram in the lower right of the map (in the Gulf of Mexico) shows the age-specific death rates for all age groups with significant numbers of deaths at the national level. The ICD codes below the histogram give the disease categories in the *9th Revision of the International Classification of Diseases*. Additional information is provided in Riggan et al. (5).

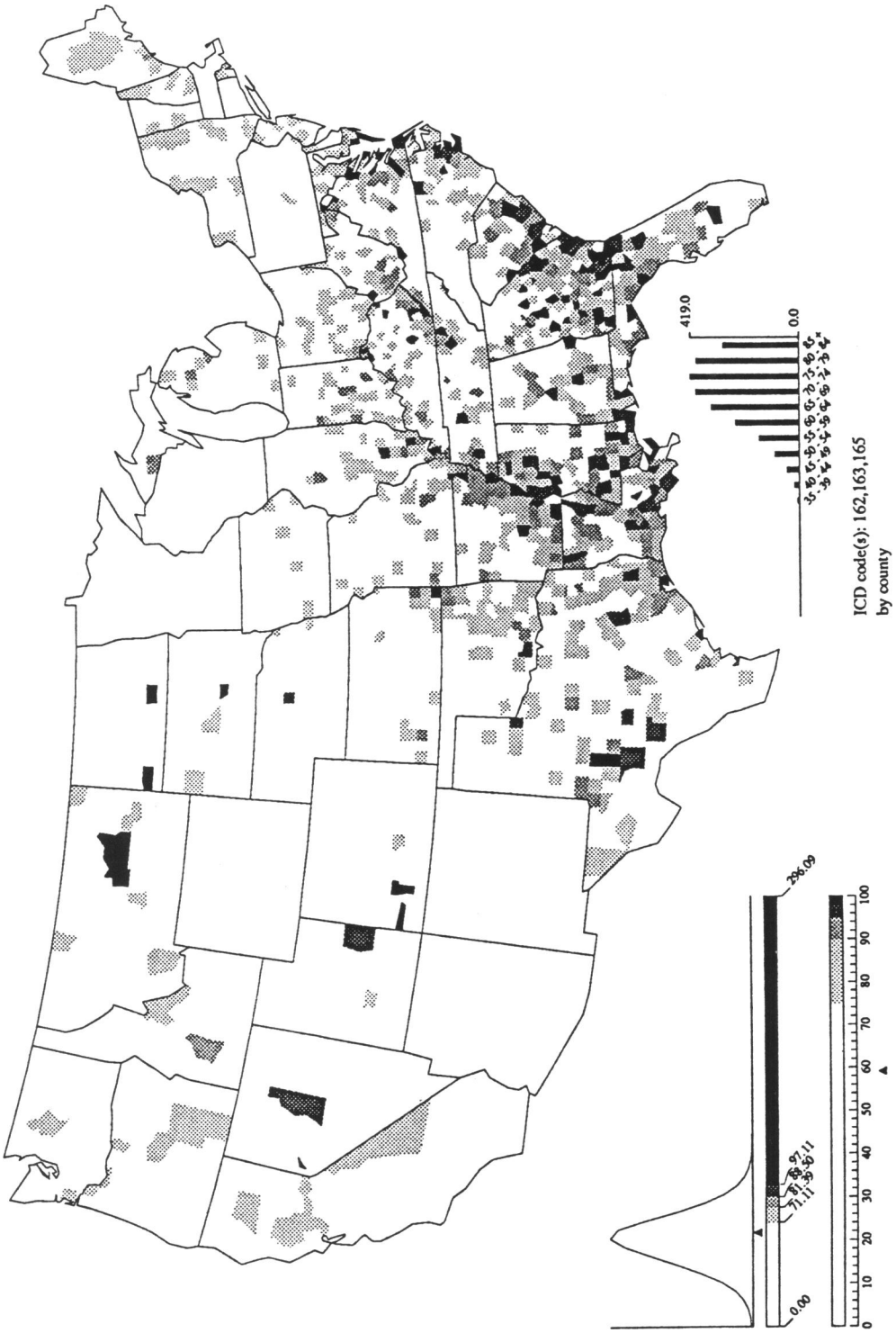

FIGURE 8. Guide to observed rates of cancer of the trachea, brochus, and lung, including pleura and other respiratory sites, for U.S. white males, 1970–1979 map. The map key and frequency polygon for the 3061 county rates are in the lower left corner. Shading is as follows: black, counties ranked in the highest 2% of all counties (98th and 99th percentiles); dark gray, counties in the next highest 3% of all counties (95th to 97th percentiles); medium gray, counties in the next highest 5% of all counties (90th to 94th percentile); light gray, counties in the next highest 15% of all counties (75th to 89th percentile); white counties in the lowest 75% of all counties (0th to 74th percentiles). These cut points are graphically illustrated in the lower tone bar. The triangle below the lower tone bar indicates the relative ranking of the national rate. The upper tone bar relates these cut points to their locations on the frequency polygon constructed from the distribution of the 3061 county rates. The triangle between the upper tone bar and the frequency polygon indicates the location of the national rate. The histogram in the lower right of the map (in the Gulf of Mexico) shows the age-specific death rates for all age groups with significant numbers of deaths at the national level. The ICD codes below the histogram give the disease categories in the *9th Revision of the International Classification of Diseases*. Additional information is provided in Riggan et al. (5).

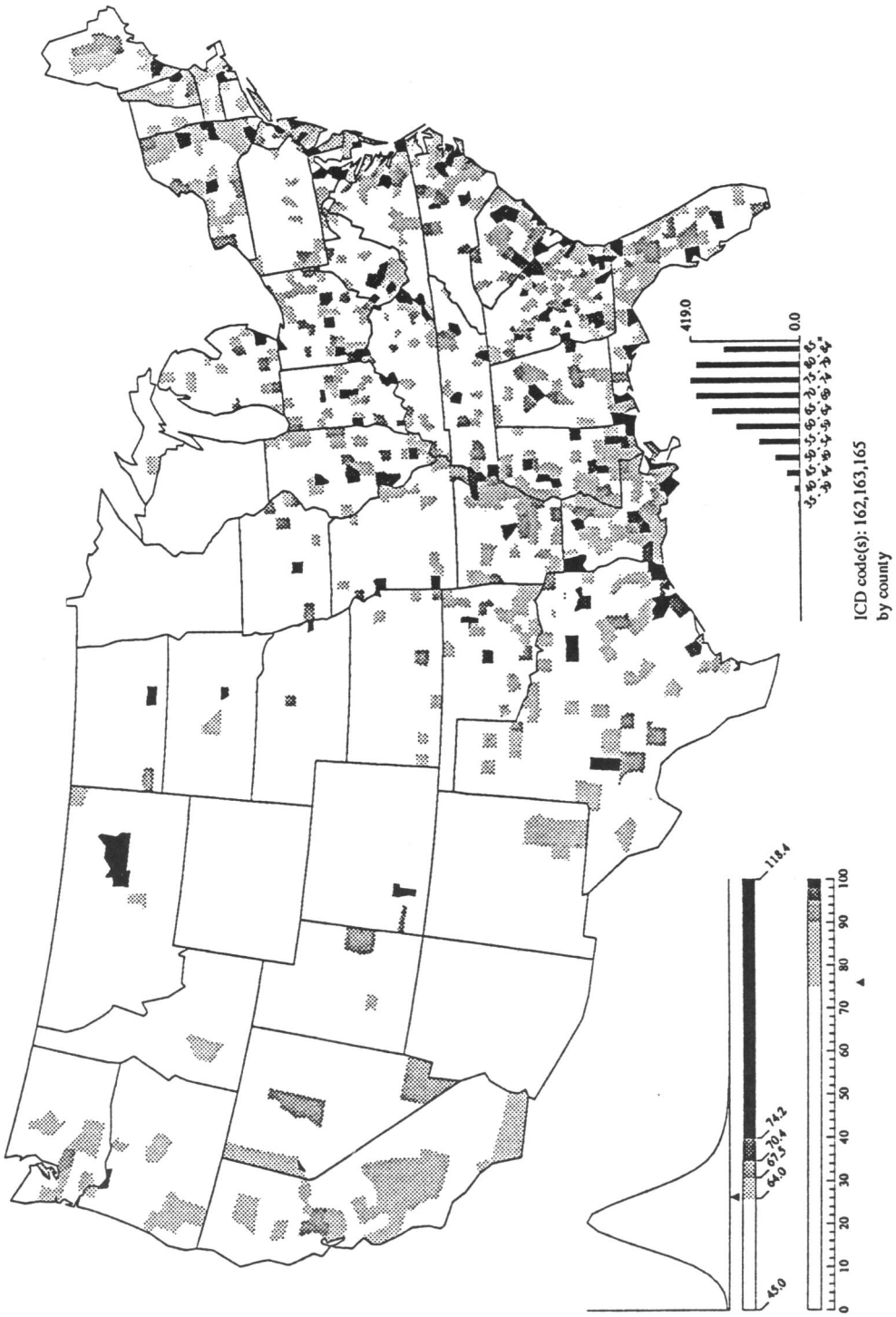

FIGURE 9. Guide to quintile model for cancer of the trachea, bronchus, and lung, including pleura and other respiratory sites, for U.S. white males, 1970–1979 map. The map key and frequency polygon for the 3061 county rates are in the lower left corner. Shading is as follows: black, counties ranked in the highest 2% of all counties (98th and 99th percentiles); dark gray, counties in the next highest 3% of all counties (95th to 97th percentiles); medium gray, counties in the next highest 5% of all counties (90th to 94th percentile); light gray, counties in the next highest 15% of all counties (75th to 89th percentile); white counties in the lowest 75% of all counties (0th to 74th percentiles). These cut points are graphically illustrated in the lower tone bar. The triangle below the lower tone bar indicates the relative ranking of the national rate. The upper tone bar relates these cut points to their locations on the frequency polygon constructed from the distribution of the 3061 county rates. The triangle between the upper tone bar and the frequency polygon indicates the location of the national rate. The histogram in the lower right of the map (in the Gulf of Mexico) shows the age-specific death rates for all age groups with significant numbers of deaths at the national level. The ICD codes below the histogram give the disease categories in the *9th Revision of the International Classification of Diseases*. Additional information is provided in Riggan et al. (5).

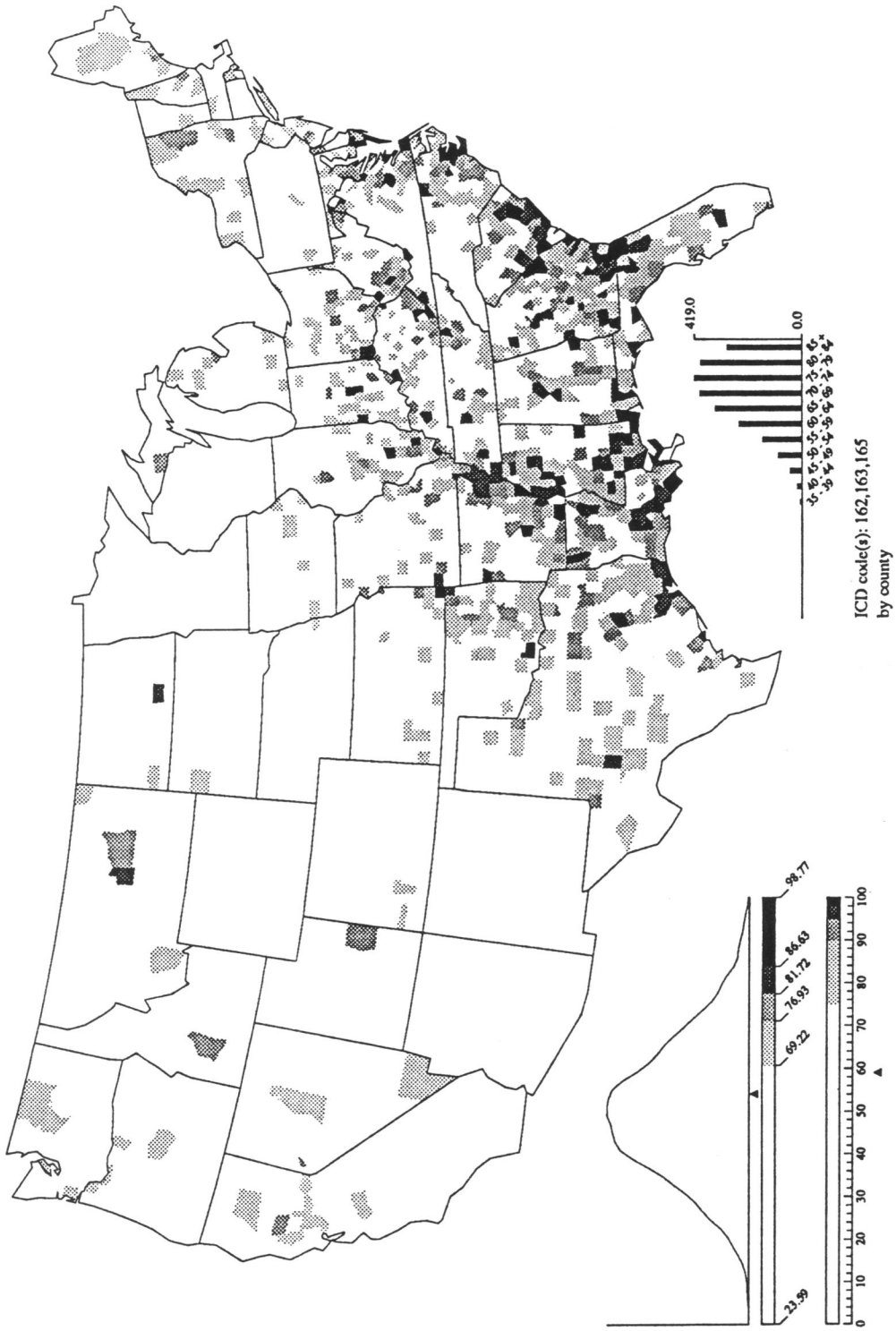

FIGURE 10. Guide to two-stage model for cancer of trachea, bronchus, and lung, including pleura and other respiratory sites, for U.S. white males, 1970–1979 map. The map key and frequency polygon for the 3061 county rates are in the lower left corner. Shading is as follows: black, counties ranked in the highest 2% of all counties (98th and 99th percentiles); dark gray, counties in the next highest 3% of all counties (95th to 97th percentiles); medium gray, counties in the next highest 5% of all counties (90th to 94th percentile); light gray, counties in the next highest 15% of all counties (75th to 89th percentile); white counties in the lowest 75% of all counties (0th to 74th percentiles). These cut points are graphically illustrated in the lower tone bar. The triangle below the lower tone bar indicates the relative ranking of the national rate. The upper tone bar relates these cut points to their locations on the frequency polygon constructed from the distribution of the 3061 county rates. The triangle between the upper tone bar and the frequency polygon indicates the location of the national rate. The histogram in the lower right of the map (in the Gulf of Mexico) shows the age-specific death rates for all age groups with significant numbers of deaths at the national level. The ICD codes below the histogram give the disease categories in the *9th Revision of the International Classification of Diseases*. Additional information is provided in Riggan et al. (5).

for the selected cut points, the EB procedure produces rankings of individual counties, whereas the NCI procedure does not. With rankings, we have the flexibility to select different cut points based on different decision problems. NCI's use of a selected statistical confidence level (e.g., 5%) does not mean that 5% of the counties will necessarily surpass that confidence level, under the null hypothesis of no significant differences in county rates. For example, depending upon the type of cancer and its frequency, one could have 4, 7, or more of the counties pass the 5% confidence limit. This will be problematic for certain decision problems such as when one has a fixed level of resources to conduct some actions in a fixed number of counties. The NCI procedure requires recomputation of all county classifications if the decision procedure, or a utility analysis, suggests that an alternative cut point may be more appropriate. Because it provides a fixed ranking of all counties, the EB procedures do not have his problem. Thus, the rate stabilization procedure has appropriate operating characteristics for accomplishing several types of scientific and environmental policy related tasks.

The applications of these procedures to death rates is not their only possible use. They could be used for analysis of any type of health, or health service, related event in small local populations, e.g., the risk of accidents or the use of renal dialysis. This can be accomplished without the necessity of making strong modeling or distributional assumptions about the spatial distribution of mortality or morbidity risks, using standard epidemiological measures of risk.

Research reported in this paper is supported by EPA Cooperative Agreement no. CR811090 and CR815811-01-1 and NIA grant AG01159.

REFERENCES

1. Morris, C. N. Parametric Empirical Bayes inference: theory and applications (with discussion). J. Am. Stat. Assoc. 78: 47–65 (1983).
2. Morris, C. N. Natural exponential families with quadratic variance functions: statistical theory. Ann. Stat. 11: 515–552 (1983).
3. Manton, K. G., Stallard, E., Woodbury, M. A., and Riggan, W. B. Statistically adjusted estimates of geographic mortality profiles. J. Natl. Cancer Inst. 78: 805–815 (1987).
4. Manton, K. G., Woodbury, M. A., Stallard, E., Riggan, W. B., Creason, J. P., and Pellom, A. Empirical Bayes procedure for stabilizing maps of US cancer mortality rates. J. Am. Stat. Assoc. 84: 637–650 (1989).
5. Riggan, W. B., Creason, J. P., Nelson, W. C., Manton, K. G., Stallard, E., Woodbury, M. A., Pellom, A. C., and Beaubier, J. U.S. Cancer Mortality Rates and Trends, 1950–1979, Vol. 4, Maps. U.S. EPA Health Effects Research Laboratory, EPA/600/1-83/015d. U.S. Government Printing Office, Washington, DC, 1987.
6. Fleiss, J. L. Statistical Methods for Rates and Proportions, 2nd ed. John Wiley and Sons, New York, 1981.
7. North Carolina Department of Human Resources. Leading Causes of Mortality: North Carolina Vital Statistics, 1978, Vol. 2. Division of Health Services, Administrative Services Section, Public Health Statistics Branch, Raleigh, NC, 1979.
8. Mason, T. J., McKay, F. W., Hoover, R., Blot, W. T., and Fraumeni, J. F. Atlas of Cancer Mortality for U.S. Counties, 1950–1969. DHEW Publication No. (NIH) 75-780. U.S. Government Printing Office, DC, 1975.
9. Pickle, L. W., Mason, T. J., Howard, N., Hoover, R., and Fraumeni, J. F. Atlas of U.S. Cancer Mortality Among Whites, 1950–1980. DHHS Publication No. (NIH) 87-2900. U.S. Government Printing Office, Washington, DC, 1987.
10. Breslow, N. E., and Day, N. E. Indirect standardization and multiplicative models for rates, with reference to the age adjustment of cancer incidence and relative frequency data. J. Chron. Dis. 28: 289–303 (1975).
11. Gail, M. The analysis of heterogeneity for indirect standardized mortality ratios. J. R. Stat. Soc. Ser. A 141: 224–234 (1978).
12. Robbins, H. The Empirical Bayes approach to statistics. In: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1. University of California Press, Berkeley, CA, 1955, pp. 157–164.
13. Robbins, H. The Empirical Bayes approach to statistical decision problems. Ann. Math. Stat. 35:49–68 (1964).
14. Casella, G. An introduction to Empirical Bayes data analysis. Am. Stat. 39: 83–87 (1985).
15. Robbins, H. Some thoughts on Empirical Bayes estimation. Ann. Stat. 11: 713–723 (1983).
16. Maritz, J. S. Empirical Bayes estimation for the Poisson distribution. Biometrika 56: 349–359 (1969).
17. Susarla, V. Empirical Bayes theory. In: Encyclopedia of Statistical Sciences, Vol. 2 (S. Katz, N. L. Johnson, and C. B. Read, Eds.), 1982, pp. 490–503.
18. Stein, C. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1. University of California Press, Berkeley, CA, 1955, pp. 197–206.
19. James, W., and Stein, C. Estimation with quadratic loss. In: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1. University of California Press, Berkeley, CA, 1955, pp. 361–379.
20. Efron, B., and Morris, C. Data analysis using Stein's estimator and its generalizations. J. Am. Stat. Assoc. 70: 311–319 (1975).
21. Fay, R. E., and Herriot, R. A. Estimates of income for small places: an application of James-Stein procedures to census data. J. Am. Stat. Assoc. 74: 269–277 (1979).
22. Louis, T. A., and DerSimonian, R. L. Health statistics based on discrete population groups. In: Regional Variations in Hospital Use (D. L. Rothberg, Ed.), Lexington Books, Lexington, MA, 1982, pp. 205–236.
23. Louis, T. A. Estimating a population of parameter values using Bayes and Empirical Bayes methods. J. Am. Stat. Assoc. 79: 393–398 (1984).
24. Tsutakawa, R. K. Estimation of cancer mortality rates: a Bayesian analysis of small frequencies. Biometrics 41: 69–79 (19855).
25. Tsutakawa, R. K. Mixed model for analyzing geographic variability in mortality rates. J. Am. Stat. Assoc. 83: 37–42 (1988).
26. Tsutakawa, R. K., Shoop, G. L., and Marienfeld, C. J. Empirical Bayes estimation of cancer mortality rates. Stat. Med. 4: 201–212 (1985).
27. Manton, K. G., Woodbury, M. A., Stallard, E., and Dowd, J. E. A composite estimation model for producing stabilized health rate estimates for small area using sample surveys: experience from health surveys in Ethiopia, India, and Indonesia. In: Small Area Statistics, Proceedings of an Internal Symposium, 1985 (R. Platek and M. P. Singh, Eds.), Laboratory for Research in Statistics and Probability, Carleton University, Ottawa, Canada, 1986, pp. 182–206.
28. Lui, K. J., Martinez, B., and Mercy, J. An application of empirical Bayes to direct-adjusted rates: a note on suicide mapping in California. Suicide Life Threat. Behav. 20: 240–253 (1990).
29. Kaldor, J., and Clayton, D. Role of advanced statistical techniques in cancer mapping. Recent Results Cancer Res. 114: 87–98 (1989).
30. Collings, B. J., and Margolin, B. H. Testing goodness of fit for the Poisson assumption when observations are not identically distributed. J. Am. Stat. Assoc. 80: 411–418 (1985).
31. Dean, C., and Lawless, J. F. Tests for detecting overdispersion in Poisson regression models. J. Am. Stat. Assoc. 84: 467–472 (1989).
32. Brillinger, D. R. The natural variability of vital rates and associated statistics (with discussion). Biometrics 42: 693–734 (1986).
33. Manton, K. G., and Stallard, E. Methods for analysis of mortality risks across heterogeneous small populations: examination of space-time gradients in cancer mortality in North Carolina counties 1970–75. Demography 18: 217–230 (1981).
34. Manton, K. G., Woodbury, M. A., and Stallard, E. A variance components approach to categorical data models with heterogeneous cell populations: analysis of spatial gradients in county lung cancer mortality rates in North Carolina counties. Biometrics 37: 259–269 (1981).
35. Manton, K. G., Stallard, E., Creason, J. P., and Riggan, W. B. U.S. cancer mortality 1950–1978: a strategy for analyzing spatial and temporal patterns. Environ. Health Perspect. 60: 369–380 (1985).
36. National Cancer Institute, Office of Cancer Communications. Research Contributions Made Possible by the NCI Cancer Atlases Published in the 1970s. Backgrounder Series, June 9, 1987, NCI, Bethesda, MD.
37. Manton, K. G., and Stallard, E. A cohort analysis of U.S. stomach cancer mortality, 1950–1977. Int. J. Epidemiol. 11: 49–61 (1982).