

## Using a Mixed Effects Model to Estimate Geographic Variation in Cancer Rates

Gene A. Pennello,\* Susan S. Devesa, and Mitchell H. Gail

Biostatistics Branch, National Cancer Institute, 6130 Executive Boulevard,  
EPN 431, Rockville, Maryland 20892, U.S.A.

\*email: gxp@cdrh.fda.gov

**SUMMARY.** Commonly used methods for depicting geographic variation in cancer rates are based on rankings. They identify where the rates are high and low but do not indicate the magnitude of the rates nor their variability. Yet such measures of variability may be useful in suggesting which types of cancer warrant further analytic studies of localized risk factors. We consider a mixed effects model in which the logarithm of the mean Poisson rate is additive in fixed stratum effects (e.g., age effects) and in logarithms of random relative risk effects associated with geographic areas. These random effects are assumed to follow a gamma distribution with unit mean and variance  $1/\alpha$ , similar to Clayton and Kaldor (1987, *Biometrics* **43**, 671-681). We present maximum likelihood and method-of-moments estimates with standard errors for inference on  $\alpha^{-1/2}$ , the relative risk standard deviation (RRSD). The moment estimates rely on only the first two moments of the Poisson and gamma distributions but have larger standard errors than the maximum likelihood estimates. We compare these estimates with other measures of variability. Several examples suggest that the RRSD estimates have advantages compared to other measures of variability.

**KEY WORDS:** Delta method; Maximum likelihood; Method-of-moments; Overdispersion; Relative risk; Standardized rates.

### 1. Introduction

Epidemiologists have searched for geographic variation in cancer rates in efforts to identify areas with unusually high or low rates that might suggest leads to cancer etiology. For example, melanoma rates in the U.S. vary according to latitude, presumably because of variations in exposure to sunlight (Scotto, Fears, and Fraumeni, 1996), while lung cancer rates have been especially high in seaboard areas due to exposure to asbestos in shipyards (Tagnon et al., 1980).

A commonly used method of cancer mapping ranks rates standardized for confounding effects such as age, classifies the ranks into percentiles groups, and color codes geographic areas according to these groups. Although this technique is useful for identifying areas with high and low cancer rates, it provides information neither on the magnitude of the standardized rates nor on the absolute or relative variability of these rates. In this note, we compare several measures of geographic variability of standardized rates and propose a measure called the relative risk standard deviation (RRSD).

One could, of course, simply compute the sample variance of standardized cancer rates. However, this sample variance depends on the intrinsic variation of the underlying relative risks associated with the areas and on the random variation of observed counts. The RRSD method attempts to estimate the intrinsic variability, as this is the component of etiologic interest. We give parametric and method-of-moments estimators with standard errors for RRSD.

Other measures of geographic variability have been used. Fraumeni et al. (1993) used the ratio of the maximum to the minimum of standardized rates. Pickle et al. (1987) used the interquartile range of the cube roots of standardized rates (IQRCR). The cube root was chosen because the simple interquartile range increased substantially with the mean rate. However, as with the sample variance, both the ratio statistic and IQRCR reflect not only the intrinsic variation but the random variation. Likewise, a standard measure of overdispersion (McCullagh and Nelder, 1989) does not estimate intrinsic variability directly.

In this note, we define RRSD and other measures of geographic variability (Section 2). Using rates based on cancer deaths from the National Center for Health Statistics and population estimates from the U.S. Bureau of the Census, we give some examples that suggest that RRSD is advantageous (Section 3).

### 2. Statistical Notation and Methods

#### 2.1 Statistical Model

Index areas by  $i = 1, \dots, I$  and age strata by  $j = 1, \dots, J$ . In our examples, we stratify only on age and carry out the analyses within groups defined by race and gender, but one could also stratify on other quantities. Assume the age-specific mortality counts, denoted by  $\{O_{ij}\}$ , are independent with

$$O_{ij} \sim \text{Poi}(y_{ij}\gamma_i\xi_j), \quad (1)$$

where  $Poi(m)$  denotes the Poisson distribution with mean  $m$ ,  $y_{ij}$  is the number of person-years at risk,  $\xi_j$  is an age effect representing the underlying rate for age group  $j$ , and  $\gamma_i$  is an area effect representing the underlying relative risk for area  $i$ . Let  $E_{ij} = y_{ij}\xi_j$ , the expected value of  $O_{ij}$  when relative risk  $\gamma_i = 1$ . When  $E_{ij}$  is assumed known (i.e.,  $\xi_j$  known) for all  $j$ , the relative risk  $\gamma_i$  is estimated by the standardized mortality ratio (SMR)  $O_{i.}/E_{i.}$  of observed to expected counts summed over  $j = 1, 2, \dots, J$  (Breslow and Day, 1987, Chapter 2).

Under model (1), the variability in  $\{\gamma_i\}$  corresponds to intrinsic geographic variability. Large variability in  $\{\gamma_i\}$  suggests that further etiologic studies may be warranted.

2.2 The RRSD Measure and a Parametric Estimate of RRSD

Suppose each relative risk  $\gamma_i$  has been drawn independently from a superpopulation of values with unit mean and standard deviation RRSD. RRSD measures the intrinsic variability of the relative risks and represents the variance component of principal interest to epidemiologists.

To estimate RRSD, we consider a specific parametric model,

$$\gamma_i \sim \Gamma(\alpha, \alpha), \tag{2}$$

where  $\Gamma(\alpha, \alpha)$  denotes the gamma distribution with mean 1 and variance  $1/\alpha$ . The relative risks  $\{\gamma_i\}$  are thus random effects, and (1) and (2) define a mixed effects model. The gamma distribution is reasonable because it is right skewed, agreeing with the notion of a spread for relative risks for which the largest values are extreme due to exposure to localized risk factors. Furthermore, nonparametric estimates of the distribution of underlying relative risks have been right skewed (Clayton and Kaldor, 1987). A mean of one for the relative risks is reasonable when underlying age rates  $\{\xi_j\}$  are internally estimated and allows for unique estimation of parameters.

The standard deviation of  $\gamma_i$ , namely the RRSD, is  $\alpha^{-1/2}$  under (2). The maximum likelihood estimator (MLE) of RRSD is based on the marginal distribution of  $\{O_{ij}\}$ . This marginal distribution is obtained by integrating the joint distribution of  $\{O_{ij}\}$  and  $\{\gamma_i\}$  given by (1) and (2) over  $\{\gamma_i\}$ . We denote the MLE estimator of RRSD by  $\hat{\alpha}^{-1/2}$ . Iterative methods for obtaining  $\hat{\alpha}^{-1/2}$  and an estimated standard error of it are given in Appendix A.

2.3 Estimating RRSD Using Only First and Second Moment Assumptions

One can relax the parametric assumptions in (1) and (2) by relying only on their implied first and second moments for  $Z_{ij} = \log(O_{ij}/y_{ij})$  and  $\gamma_i^* = \log(\gamma_i)$ . We derived an estimate of RRSD that relies only on the approximations that  $E\gamma_i^* = 0$  and  $\text{var } \gamma_i^* = \alpha^{-1}$  and that, conditional on  $\gamma_i^*$ ,  $EZ_{ij} = \gamma_i^* + \xi_j^*$ , where  $\xi_j^* = \log(\xi_j)$  and  $\text{var } Z_{ij} = (E_{ij}\gamma_i)^{-1}$ . Thus, we have a classic linear mixed effects model except that the conditional variance of  $Z_{ij}$  is not constant. We extend method 3 of Henderson (1953) to compute an estimate  $\tilde{\alpha}^{-1}$  of  $\alpha^{-1}$ , the approximate variance of  $\gamma_i^*$ . An estimate of  $\alpha^{-1/2}$ , the RRSD, is  $(\tilde{\alpha}^{-1})^{1/2}$ , which we denote as  $\tilde{\alpha}^{-1/2}$ . We estimate  $\alpha^{-1}$  by equating the difference in weighted sums of squares for the full model,  $EZ_{ij} = \gamma_i^* + \xi_j^*$ , and the reduced model,  $EZ_{ij} =$

$\xi_j^*$ , to its marginal expectation (Appendix B). We use as weights the conditional variances of  $\{Z_{ij}\}$ , which we approximate by  $\{O_{ij}^{-1}\}$  (Breslow, 1984). When  $O_{ij} = 0$ , we substituted  $O_{ij} = 1/2$ . We also derived an estimated standard error of  $\tilde{\alpha}^{-1/2}$  that relies only on the moment assumptions (Appendix B).

2.4 The Interquartile Range of Cube Roots Measure

Pickle et al. (1987) used the interquartile range of the cube roots (IQRCR) of directly adjusted rates (DARs) to measure geographic variability. The DARs adjust for differences in age distributions among areas by using common reference weights  $\{w_j \geq 0, \sum_{j=1}^J w_j = 1\}$  to average the age-specific rates  $\{r_{ij} = O_{ij}/y_{ij}, j = 1, 2, \dots, J\}$  of each area  $i$ . These rates are typically expressed per 100,000 person-years. The DAR for area  $i$  is  $\sum_{j=1}^J w_j r_{ij}$  times 100,000. Pickle et al. applied the cube root transform to the DARs because, for her data, the interquartile range of untransformed DARs increased with national rate  $O_{..}/y_{..}$ , while IQRCR did not.

2.5 The Fractional Difference Measure

An alternative to IQRCR is to divide the interquartile range of the untransformed DARs by the national rate. This is called the fractional difference (FDIFF) measure, and it expresses the difference between the first and third quartiles as a fraction of the national rate.

2.6 A General Overdispersion Parameter  $\phi$

McCullagh and Nelder (1989) model overdispersion by multiplying the variance given by a generalized linear model by an overdispersion constant  $\phi$ . This measure has been suggested for detecting geographic variation (Olsen, Martuzzi, and Elliott, 1996). In our context, the parameter  $\phi$  can be estimated as the deviance from a Poisson model for  $O_{ij}$  with mean  $E_{ij}$  divided by the degrees of freedom  $IJ - J$ .

3. Cancer Mortality Examples to Compare RRSD with the Other Measures

3.1 Materials

We applied these measures to U.S. cancer mortality data for the time periods 1950-1969 and 1970-1994. County data for these time periods were grouped, respectively, into  $I = 506$  and 508 areas, called state economic areas (SEAs), as defined by the U.S. Bureau of the Census (1966). (Two SEAs representing Hawaii and Alaska were excluded for 1950-1969 because data were not available for the entire time period.) The SEAs consist of counties with similar economic, demographic, climatic, physiographic, and cultural factors and do not cross state boundaries.

Death counts and person-years were stratified into  $J = 18$  age intervals: seventeen 5-year age intervals beginning with 0-4 years of age and a final interval of 85+ years of age. The estimated person-year exposure was the cumulative total over annual population estimates. The reference age distribution  $\{w_j\}$  used to compute the DARs was taken to be that of the total 1970 U.S. population.

3.2 Test Case 1: Non-Hodgkin's Lymphoma

Non-Hodgkin's lymphoma (NHL) accounts for about 3% of initial AIDS diagnoses (Biggar and Rabkin, 1992) because

Table 1  
Comparisons of five measures of geographic variability on cancer mortality rates in white men

	DAR quartiles		$\hat{\alpha}^{-1/2}$	SE <sup>a</sup>	$\tilde{\alpha}^{-1/2}$	SE	IQRCR	FDIFF	$\hat{\phi}$
	Lower	Upper							
<b>Test Case 1: Non-Hodgkin's Lymphoma, 1970–1994<sup>b</sup></b>									
Age 20–54	2.49	3.25	0.147	0.009	0.164	0.030	0.125	0.265	1.522
Age 55+	25.91	30.55	0.085	0.005	0.073	0.006	0.167	0.161	1.284
<b>Test Case 2: Melanoma, All Ages</b>									
1950–1969	1.22	1.90	0.248	0.012	0.274	0.024	0.171	0.449	1.057
1970–1994	2.42	3.32	0.171	0.008	0.156	0.013	0.148	0.294	1.138
<b>Test Case 3: Colorectal Cancer, All Ages</b>									
1950–1969	16.42	26.18	0.288	0.010	0.271	0.038	0.428	0.425	4.467
1970–1994	20.06	25.99	0.178	0.006	0.170	0.017	0.245	0.242	3.099

<sup>a</sup> Estimated standard error.

<sup>b</sup> For FDIFF and IQRCR, the reference weights used to compute the DARs for the under and over age 55 categories were rescaled so that they summed to one within these categories.

the risk of NHL is greatly increased in persons infected with the human immunodeficiency virus (HIV). Because the HIV epidemic was initially concentrated among homosexual and bisexual men in urban settings, one could anticipate important geographic heterogeneity of NHL rates, especially among men under age 55. Gail et al. (1991) documented sharply rising NHL incidence rates among men 20–49 years of age between 1983 and 1987, with smaller increases among women in this age range. During the 1980s, older men and women in the U.S. have experienced much slower increasing rates of NHL incidence in line with long-term secular trends. Thus, one would expect that a good measure of intrinsic geographic variation would indicate greater variability in mortality rates for men under age 55 than for older men during 1970–1994.

The estimated standard deviation of  $\{\gamma_i\}$ , namely the RRSD estimate  $\hat{\alpha}^{-1/2}$ , is much greater for younger white men than for older white men, and this difference is statistically significant based on the estimated standard error of  $\hat{\alpha}^{-1/2}$  (Table 1). The method-of-moments estimate  $\tilde{\alpha}^{-1/2}$  gives similar results, but the estimates are much less precise than  $\hat{\alpha}^{-1/2}$ , especially in the younger age range 20–54. FDIFF and the overdispersion statistic  $\hat{\phi}$  also highlight the geographic variability in NHL mortality rates among younger men. The statistic IQRCR, however, fails to detect the increased geographic variability among younger men. This example suggests that  $\hat{\alpha}^{-1/2}$  and  $\tilde{\alpha}^{-1/2}$  may be useful measures of heterogeneity in area-specific intrinsic rates and may be preferable to IQRCR.

### 3.3 Test Case 2: Melanoma

Melanoma incidence rates for whites in the U.S. are higher in southern regions than northern regions, presumably as a result of differential sunlight exposure. However, the difference between the northern and southern rates has steadily decreased over time (Lee, 1997). Some possible explanations for this trend include increased outdoor recreational activity in the North, greater uniformity of medical diagnosis and care, increased mobility of the population, and depletion of the ozone layer. Thus, one might expect that the intrinsic

geographic variation in mortality rates was larger in earlier time periods.

The estimated RRSD  $\hat{\alpha}^{-1/2}$  for white men is much greater for 1950–1969 than for 1970–1994, and this difference is statistically significant based on estimated standard errors (Table 1). The estimate  $\tilde{\alpha}^{-1/2}$  gives comparable results, but again the precision is much less than for  $\hat{\alpha}^{-1/2}$ . FDIFF and IQRCR also indicate greater geographic variability for the 1950–1969 period. The overdispersion statistic  $\hat{\phi}$ , however, fails to detect a decrease in geographic variability over time. This example suggests that  $\hat{\alpha}^{-1/2}$  and  $\tilde{\alpha}^{-1/2}$  may be preferable to  $\hat{\phi}$ .

### 3.4 Test Case 3: Colorectal Cancer

The geographic variation of colorectal cancer has also diminished over time (Ziegler, Devesa, and Fraumeni, 1986). A possible explanation is increasing homogeneity in food consumption patterns (Potter, 1996). For white men, the RRSD estimates  $\hat{\alpha}^{-1/2}$  and  $\tilde{\alpha}^{-1/2}$  indicate a significant decline in geographic variability from 1950–1969 to 1970–1994 (Table 1). Again,  $\tilde{\alpha}^{-1/2}$  is much less precise than  $\hat{\alpha}^{-1/2}$ . IQRCR, FDIFF, and  $\hat{\phi}$  also show decreasing variability in this example.

### 3.5 Agreement Among Measures in 21 Cancers

One would like to use statistics like  $\hat{\alpha}^{-1/2}$  and  $\tilde{\alpha}^{-1/2}$  to rank cancers according to degree of intrinsic geographic variability. Presumably, those cancers exhibiting large variability in intrinsic area-specific rates warrant further specialized studies to identify area-specific factors of etiologic significance.

We have listed for white men for the years 1970–1994 cancers with national rates of at least one death per 100,000 person-years and pleural cancer, which is rare but known to vary greatly by geographic area (Table 2). Most of these cancers will be included in the forthcoming cancer mortality atlas of the U.S. from the National Cancer Institute. We plan to make the data on these cancers available at the Web site <http://rex-nci.nih.gov>. The values of  $\hat{\alpha}^{-1/2}$ ,  $\tilde{\alpha}^{-1/2}$ , and the other measures of variability are shown,

Table 2

Five measures of geographic variability applied to mortality data for 21 cancers in white men, 1970–1994

Cancer	DAR quartiles		$\hat{\alpha}^{-1/2}$			$\tilde{\alpha}^{-1/2}$			IQRCR		FDIFF		$\hat{\phi}$	
	Lower	Upper	Value	SE	Rank	Value	SE	Rank	Value	Rank	Value	Rank	Value	Rank
Pleura	0.13	0.33	0.532	0.024	1	0.457	0.034	1	0.179	5	0.716	1	0.722	21
Rectum	3.06	4.93	0.290	0.010	2	0.239	0.019	3	0.251	2	0.425	2	1.703	4
Oral	2.99	4.33	0.255	0.009	3	0.203	0.014	6	0.190	4	0.335	4	1.416	5
Other skin	0.93	1.40	0.235	0.011	4	0.263	0.020	2	0.144	12	0.406	3	1.087	15
Larynx	1.97	2.76	0.215	0.009	5	0.193	0.015	7	0.149	9	0.314	5	0.978	18
Liver	2.68	3.64	0.209	0.008	6	0.177	0.013	9	0.149	10	0.286	8	1.292	7
Stomach	5.63	7.33	0.197	0.007	7	0.204	0.024	5	0.164	8	0.232	13	1.810	3
Lung	62.25	80.73	0.196	0.006	8	0.158	0.011	10	0.359	1	0.264	12	6.353	1
Esophagus	3.91	5.27	0.185	0.007	9	0.154	0.011	13	0.165	7	0.282	9	1.175	11
Bladder	5.24	7.03	0.177	0.007	10	0.143	0.010	14	0.179	6	0.275	10	1.193	10
Melanoma	2.42	3.32	0.171	0.008	11	0.156	0.013	11	0.148	11	0.294	6	1.138	12
Colon	16.75	21.17	0.162	0.005	12	0.156	0.016	12	0.208	3	0.219	14	2.453	2
Hodgkin's disease	0.90	1.24	0.151	0.010	13	0.207	0.018	4	0.109	13	0.291	7	1.094	14
Connective tissue	0.92	1.22	0.106	0.010	14	0.187	0.018	8	0.097	16	0.271	11	1.097	13
Brain	4.81	5.70	0.102	0.005	15	0.108	0.010	16	0.099	15	0.169	16	1.238	8
Non-Hodgkin's lymphoma	6.29	7.34	0.085	0.004	16	0.076	0.008	18	0.097	17	0.148	18	1.342	6
Kidney	4.53	5.34	0.085	0.005	17	0.095	0.008	17	0.092	18	0.163	17	1.030	16
Multiple myeloma	2.89	3.45	0.080	0.006	18	0.121	0.012	15	0.087	20	0.181	15	0.796	20
Pancreas	9.43	10.68	0.076	0.004	19	0.074	0.007	19	0.089	19	0.121	19	0.954	19
Prostate	20.96	23.42	0.073	0.003	20	0.071	0.006	21	0.104	14	0.113	20	1.020	17
Leukemia	8.36	9.35	0.065	0.004	21	0.074	0.006	20	0.077	21	0.113	21	1.229	9
Rank correlation with $\hat{\alpha}^{-1/2}$					1			0.870		0.799		0.905		0.235

together with their ranks. FDIFF, IQRCR, and  $\tilde{\alpha}^{-1/2}$  have rank correlations of at least 0.799 with the maximum likelihood estimate of RRSD,  $\hat{\alpha}^{-1/2}$ , whereas  $\hat{\phi}$  has a rank correlation of only 0.235. The DAR quartiles shown indicate that  $\hat{\phi}$  is strongly correlated with the mean rate, whereas the other measures have been selected, in part, because they are not strongly correlated with the mean rate. This disagreement between  $\hat{\phi}$  and the other measures suggests that  $\hat{\phi}$  is not as useful for ranking the geographic variability of various cancers.

To evaluate the agreement of these measures of variability with true underlying RRSDs, we took the 21 estimates  $\hat{\alpha}^{-1/2}$  corresponding to Table 2 as the true RRSDs and simulated values  $\{\gamma_i\}$  and counts  $\{O_{ij}\}$  from (1) and (2). Equation (1) also required estimates  $\{\hat{\xi}_j\}$  and person-years  $\{y_{ij}\}$  from the original data. The mean rank correlations with the 21 "true" RRSDs from 15 such simulations were 0.99 for  $\hat{\alpha}^{-1/2}$ , 0.92 for FDIFF, 0.89 for  $\tilde{\alpha}^{-1/2}$ , 0.81 for IQRCR, and 0.26 for  $\hat{\phi}$ . Paired *t*-tests revealed that all differences in rank correlations were statistically significant. These simulations provide additional evidence that  $\hat{\phi}$  is not useful for our purposes.

FDIFF, IQRCR,  $\hat{\alpha}^{-1/2}$ , and  $\tilde{\alpha}^{-1/2}$  all rank pleural, rectal, and oral cancers among the top six in variability, while FDIFF,  $\hat{\alpha}^{-1/2}$ , and  $\tilde{\alpha}^{-1/2}$  all rank prostate cancer and leukemia as the bottom two (Table 2). These results are in accord with U.S. epidemiologic data for these tumors. Pleural cancer is primarily due to exposure to asbestos, which was once especially prevalent in the ship-building industry located in

coastal areas (Tagnon et al., 1980). Oral cancer, particularly among men, has been strongly linked with tobacco use and alcohol drinking, exposures that have varied substantially by region (Blot et al., 1996). The reason for the variation in rectal cancer is not entirely clear, but increasingly, it has been specified as intestinal cancer on death certificates, which is then categorized with colon cancer (Chow and Devesa, 1992). Whether this misclassification has varied geographically is unknown, but it may have contributed to the high variability observed for rectal cancer. In contrast, prostate cancer risk appears to be strongly influenced by host susceptibility hormonal and genetic factors, although diet appears to play a role (Giovannucci, 1995). Leukemia rates tend to be higher in areas associated with agricultural activity (Pickle et al., 1987), but there are only 3 million farmers, and the relative risk associated with farming is only slightly elevated from 1.1 to 1.4 (Blair and Zahm, 1991).

#### 4. Discussion

RRSD is a measure of variability in intrinsic area-specific relative risks under the mixed effects model given by (1) and (2). RRSD is therefore an appealing parameter for identifying cancers that may be influenced by the prevalence of localized risk factors. Estimates of RRSD like  $\hat{\alpha}^{-1/2}$  and  $\tilde{\alpha}^{-1/2}$  complement graphical depictions of geographic areas with color or shading that indicate percentile groups of risk. Such visual displays point to areas of high or low risk, while estimates of RRSD quantify the geographic variability.

Our methods of inference for RRSd are closely related to other works. Our model given by (1) and the one-parameter gamma distribution in (2) is identical to the mixed effects model in Clayton and Kaldor (1987) except that they used a two-parameter gamma distribution and a constraint different than a mean of one to uniquely estimate parameters. Clayton and Kaldor did not concentrate on estimating intrinsic geographic variability. Martuzzi and Hills (1995) estimated intrinsic variability using the Clayton and Kaldor model except that they assumed that stratum (e.g., age) effects were known, eliminating the need for an identifiability constraint. Martuzzi and Hills did not provide standard errors of the estimates but instead computed confidence intervals by profile likelihood. Muirhead and Butland (1996) estimated intrinsic variability and gave standard errors using a different model for the unconditional variance of the counts and also assumed that stratum effects were known.

If one is willing to assume only the first two moments of the Poisson and gamma distributions in (1) and (2), then the moments estimate  $\tilde{\alpha}^{-1/2}$  is an alternative to the maximum likelihood estimate  $\hat{\alpha}^{-1/2}$ . The estimate  $\tilde{\alpha}^{-1/2}$ , however, had a much larger estimated standard error than  $\hat{\alpha}^{-1/2}$  in our examples. We further compared  $\hat{\alpha}^{-1/2}$  and  $\tilde{\alpha}^{-1/2}$  on simulated counts from the mixed effects model (1) and (2) as described in Section 3.5. We generated 100 sets of simulated counts for 10 of the cancers in Table 2. The statistic  $\tilde{\alpha}^{-1/2}$  had a sample standard error 1–2 times that of  $\hat{\alpha}^{-1/2}$ , depending on the cancer. The mean of the estimated standard errors of  $\tilde{\alpha}^{-1/2}$  exceeded the sample standard error by 1.2–1.9 times. For Hodgkin's disease and multiple myeloma,  $\tilde{\alpha}^{-1/2}$  was very biased and yielded Wald confidence intervals for  $\alpha^{-1/2}$  that had coverage well below nominal levels. One possible explanation is that our moment approximations are sometimes very bad when rates are low. In contrast,  $\hat{\alpha}^{-1/2}$  showed little bias and its Wald confidence intervals for  $\alpha^{-1/2}$  had coverage close to nominal levels. We conclude that the potential advantage of the method of moments estimator implied by its fewer assumptions is often counterbalanced by its poorer properties, and we can recommend its use only for cancers with relatively high rates.

It is difficult to verify empirically that the estimates  $\hat{\alpha}^{-1/2}$  and  $\tilde{\alpha}^{-1/2}$  of RRSd are preferable to other measures of variability, such as IQRCR, FDIFF, or  $\hat{\phi}$ , because there are few cancers for which independent data, apart from the observed rates, are available to demonstrate geographic heterogeneity of intrinsic rates. We compared these measures based on the expectations that NHL mortality rates are more heterogeneous in younger men than in older men due to urban localization of AIDS-related diagnoses and that melanoma and colorectal cancer mortality rates have become less heterogeneous over time. In these cases, it appears that IQRCR performs poorly for NHL and  $\hat{\phi}$  performs poorly for melanoma, while  $\hat{\alpha}^{-1/2}$ ,  $\tilde{\alpha}^{-1/2}$ , and FDIFF are good indicators of intrinsic-rate heterogeneity.

Moreover, a survey of mortality rates for 21 cancers among white males (Table 2) indicates that  $\hat{\alpha}^{-1/2}$ ,  $\tilde{\alpha}^{-1/2}$ , IQRCR, and FDIFF are highly correlated. The quantity  $\hat{\phi}$ , however, is not highly correlated with these measures. Simulations in Section 3.5 show that  $\hat{\phi}$  also does not correlate highly with true underlying RRSds. A Taylor series expansion of the deviance

expression through quadratic terms shows that the expectation of  $\hat{\phi}$  under (1) is approximately

$$1 + \frac{2}{IJ - J} \sum_{i=1}^I \sum_{j=1}^J E_{ij} \gamma_i \log(\gamma_i / \bar{\gamma}). \quad (3)$$

(Details are available from the first author.) Thus, the magnitude of  $\hat{\phi}$  depends always, at least in part, on the Poisson means  $\{E_{ij} \gamma_j\}$  except in the homogeneous case when  $\gamma_i = 1$  for all  $i$ .

RRSD estimates, however, may not be in accord with some epidemiologic data. For example, estimates of relative risk based on incidence for areas in England and Wales have a range much narrower for Hodgkin's disease than for non-Hodgkin's lymphoma or leukemia (Cartwright et al., 1990), whereas for U.S. mortality data,  $\hat{\alpha}^{-1/2}$  is much higher for Hodgkin's disease among these cancers (Table 2). The discrepancy might be due to differences between incidence and mortality geographic variation in these cancers. This could be investigated by computing RRSd estimates for the European mortality data as summarized in Smans et al. (1992).

Nonetheless, there are theoretical reasons for believing RRSd may be preferable to other measures. In particular under (1) and (2), RRSd reflects the intrinsic variability in the relative risks  $\{\gamma_i\}$  as it is the standard deviation of the underlying distribution on  $\{\gamma_i\}$ . Quantities like  $\hat{\alpha}^{-1/2}$  and  $\tilde{\alpha}^{-1/2}$  estimate this standard deviation whether the absolute rates are large or small. The other measures of variability considered are less appropriate than  $\hat{\alpha}^{-1/2}$  or  $\tilde{\alpha}^{-1/2}$  because they reflect random variation in the counts  $\{O_{ij}\}$  as well as intrinsic variation (e.g., IQRCR, FDIFF), or worse, are highly correlated with the mean rate (e.g.,  $\hat{\phi}$ ).

It is important to note, however, that  $\hat{\alpha}^{-1/2}$  may converge to the boundary point zero, invalidating the standard error of this estimate, and  $\tilde{\alpha}^{-1/2}$  can be negative (Appendix B). These values of  $\hat{\alpha}^{-1/2}$  and  $\tilde{\alpha}^{-1/2}$  occur most often when rates are small and few deaths are observed. Of course, equations (1) and (2) may not hold precisely. The conditional distribution in (1) requires no interactions between age and area effects on a log scale, the random effects distribution in (2) may be incorrect, and the combined model only allows for overdispersion, not underdispersion. In (2), it is assumed that the relative risks of neighboring areas are uncorrelated. In the presence of autocorrelation, more elaborate models may be employed (e.g., Besag, York, and Mollié, 1991). When autocorrelation is important, however, there is no simple summary measure of geographic variation because the variance-covariance matrix of the relative risks is characterized by correlation parameters in addition to RRSd.

One can avoid parametric assumptions on the random effects distribution by computing the nonparametric MLE of the distribution on  $\{\gamma_i\}$ . Clayton and Kaldor (1987) used this approach to estimate  $\{\gamma_i\}$ , and Böhning and Schlattmann (1992) have developed software to compute this MLE. The standard deviation of the nonparametric MLE is an RRSd estimate. Further work is needed to compare the precision and bias of estimates from this approach with that of  $\hat{\alpha}^{-1/2}$  and  $\tilde{\alpha}^{-1/2}$ .

When (1) and (2) are not valid, RRSd is an imperfect summary of intrinsic variability. Even when (1) and (2) are

not precisely correct, however, estimates of RRSD may still be a useful descriptive summary for identifying cancers that warrant further study.

#### ACKNOWLEDGEMENTS

We wish to thank the referees for several useful comments, including the suggestion that we develop a method-of-moments estimate for RRSD. We also thank Professor Raymond Carroll for advice in estimating the variance of the method-of-moments estimators of RRSD.

#### RÉSUMÉ

Les méthodes couramment utilisées pour détecter les variations géographiques des taux de cancers sont basées sur des classements et l'identification des classes où les taux sont élevés et faibles, mais sans fournir l'ordre de grandeur des taux ni leur variabilité. Pourtant, des mesures de variabilité pourraient être utiles en suggérant quel type de cancer mériterait des analyses approfondies pour quels facteurs de risque géographiques. On considère un modèle à effets mixtes où le logarithme de la moyenne de Poisson des taux de mortalité sur des strates fixées (par exemple des groupes d'âge) suit un modèle à effet additifs sur les logarithmes des risques relatifs associés aux unités géographiques. On suppose que ces effets aléatoires suivent une distribution gamma de moyenne 1 et de variance  $1/a$  comme le propose Clayton et Kaldor (1987, *Biometrics* **43**, 671-681). On présente les estimations du maximum de vraisemblance et de la méthode des moments en prenant  $a^{-1/2}$  comme erreur standard pour estimer la "déviations standard du risque relatif" (DSRR). On peut avoir confiance dans l'estimation des deux premiers moments de la loi de Poisson et de la loi gamma, mais les erreurs standards sont plus élevées que pour celles des estimations par le maximum de vraisemblance. On compare ces estimations à celles fournies par d'autres mesures de la variabilité. Plusieurs exemples suggèrent que les estimations par la DSRR présentent des avantages comparativement aux autres mesures de la variabilité.

#### REFERENCES

- Besag, J., York, J., and Mollié, A. (1991). Bayesian image restoration with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics* **43**, 1-59.
- Biggar, R. J. and Rabkin, C. S. (1992). The epidemiology of acquired immunodeficiency syndrome-related lymphomas. *Current Opinion in Oncology* **4**, 883-893.
- Blair, A. and Zahm, S. H. (1991). Cancer among farmers. *Occupational Medicine* **6**, 335-354.
- Blot, W. J., McLaughlin, J. K., Devesa, S. S., and Fraumeni, Jr., J. F. (1996). Oral and pharyngeal cancer. In *Cancer Epidemiology and Prevention*, 2nd edition, D. Schottenfeld and J. F. Fraumeni, Jr. (eds), 660-680. Philadelphia: Saunders.
- Böhning, D. and Schlattmann, P. (1992). Computer-assisted analysis of mixtures (C.A.MAN): Statistical algorithms. *Biometrics* **48**, 283-303.
- Breslow, N. E. (1984). Extra-Poisson variation in log-linear models. *Applied Statistics* **33**, 38-44.
- Breslow, N. E. and Day, N. E. (1987). *Statistical Methods in Cancer Research, Volume 2, The Design and Analysis of Cohort Studies*. IARC Publication 83. Lyon: International Agency for Research on Cancer (distributed by Oxford University Press).
- Cartwright, R. A., Alexander, F. E., McKinney, P. A., and Ricketts, T. J. (1990). *Leukaemia and Lymphoma: An Atlas of Distribution Within Areas of England and Wales, 1984-1988*. London: Leukaemia Research Fund.
- Chow, W. and Devesa, S. (1992). Death certificate reporting of colon and rectal cancers. *Journal of the American Medical Association* **267**, 3028-3028.
- Clayton, D. and Kaldor, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* **43**, 671-681.
- Fraumeni, Jr., J. Devesa, S., Hoover, R., and Kinlen, L. (1993). Epidemiology of cancer. In *Cancer: Principles and Practice of Oncology*, 4th edition, V. T. DeVita, Jr., S. Hellman, and S. A. Rosenberg (eds), 150-181. Philadelphia: J. B. Lippincott.
- Gail, M., Pluda, J., Rabkin, C., Biggar, R., Goedert, J., Horm, J., Sondik, E., Yarchoan, R., and Broder, S. (1991). Projections of the incidence of non-Hodgkin's lymphoma related to acquired immunodeficiency syndrome. *Journal of the National Cancer Institute* **83**, 695-701.
- Giovannucci, E. (1995). Epidemiologic characteristics of prostate cancer. *Cancer* **75**, 1766-1777.
- Henderson, C. R. (1953). Estimation of variance and covariance components. *Biometrics* **9**, 226-252.
- Lee, J. A. H. (1997). Declining effect of latitude on melanoma mortality rates in the United States. *American Journal of Epidemiology* **146**, 413-417.
- Martuzzi, M. and Hills, M. (1995). Estimating the degree of heterogeneity between event rates using likelihood. *American Journal of Epidemiology* **141**, 369-373.
- McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd edition. London: Chapman and Hall.
- Muirhead, C. R. and Butland, B. K. (1996). Testing for overdispersion using an adapted form of the Potthoff-Whittinghill method. In *Methods for Investigating Localized Clustering of Disease*. IARC Publication 135. F. E. Alexander and P. Boyle (eds), 40-52. Lyon: International Agency on Research for Cancer.
- Olsen, S., Martuzzi, M., and Elliott, P. (1996). Cluster analysis and disease mapping—Why, when, and how?: A step by step guide. *British Medical Journal* **313**, 863-866.
- Pickle, L. W., Mason, T. J., Howard, N., Hoover, R., and Fraumeni, Jr., J. F. (1987). *Atlas of Cancer Mortality Among Whites: 1950-80*. Publication 87-2900. Washington, D.C.: U.S. Government Printing Office, Department of Health and Human Services (National Institutes of Health).
- Potter, J. D. (1996). Nutrition and colorectal cancer. *Cancer Causes and Control* **7**, 127-146.
- Scotto, J., Fears, T. R., and Fraumeni, Jr., J. F. (1996). Solar radiation. In *Cancer Epidemiology and Prevention*, 2nd edition, D. Schottenfeld and J. F. Fraumeni, Jr. (eds), 355-372. New York: Oxford University Press.

Smans, M., Muir, C. S., and Boyle, P. (1992). *Atlas of Cancer Mortality in the European Economic Community*. IARC Publication 107. Lyon: International Agency for Research on Cancer.

Tagnon, I., Blot, W. J., Stroube, R. B., Day, N. E., Morris, L. E., Peace, B. B., and Fraumeni, Jr., J. F. (1980). Mesothelioma associated with the shipbuilding industry in coastal Virginia. *Cancer Research* **40**, 3875-3879.

U.S. Bureau of the Census. (1966). *U.S. Census of Population: 1960, Number of Inhabitants, United States Summary Final Report PC (1)-1A*. Washington, D.C.: U.S. Government Printing Office.

Ziegler, R. G., Devesa, S. S., and Fraumeni, Jr., J. F. (1986). Epidemiologic patterns of colorectal cancer. In *Important Advances in Oncology 1986*, V. T. DeVita, Jr., S. Hellman, and S. A. Rosenberg (eds), 209-232. Philadelphia: J. B. Lippincott.

Received October 1997. Revised July 1998.  
Accepted August 1998.

APPENDIX A

Derivation of  $\hat{\alpha}^{-1/2}$  and Its Standard Error

A.1 The MLE of  $\alpha^{-1/2}$

From (1) and (2), it follows that the marginal density of  $\mathbf{O} = \{O_{i,j}\}$  given  $\alpha$  and  $\xi = \{\xi_j\}$  is

$$m(\mathbf{O}; \alpha, \xi) = \left( \prod_{i=1}^I \prod_{j=1}^J \frac{E_{ij}^{O_{ij}}}{O_{ij}!} \right) \prod_{i=1}^I \frac{O_i!}{E_i^{O_i}} f_i(O_i; \alpha, \xi),$$

where  $f_i(\cdot; \alpha, \xi)$  is the density corresponding to the negative binomial distribution with mean  $E_i$  and variance  $E_i(1 + E_i/\alpha)$ . With respect to  $\alpha$  and  $\xi$ ,

$$\begin{aligned} \log m(\mathbf{O}; \alpha, \xi) \propto & \sum_{i=1}^I \sum_{j=1}^J O_{ij} \log E_{ij} + \sum_{i=1}^I \sum_{k=0}^{O_i-1} \log(k + \alpha) \\ & + I \alpha \log \alpha - \sum_{i=1}^I (O_i + \alpha) \log(E_i + \alpha), \end{aligned}$$

where the summation over  $k$  is defined as zero when  $O_i = 0$ . The derivative of  $\log m(\mathbf{O}; \alpha, \xi)$  with respect to  $\alpha$  is

$$\begin{aligned} \frac{d}{d\alpha} \log m(\mathbf{O}; \alpha, \xi) = & \sum_{i=1}^I \sum_{k=0}^{O_i-1} (k + \alpha)^{-1} + I \log \alpha + I \\ & - \sum_{i=1}^I \left[ \frac{O_i + \alpha}{E_i + \alpha} + \log(E_i + \alpha) \right], \quad (A.1) \end{aligned}$$

and the derivative of  $\log m(\mathbf{O}; \alpha, \xi)$  with respect to  $\xi_j$  is

$$O_{.j}/\xi_j - \sum_{i=1}^I y_{ij}(O_i + \alpha)/(E_i + \alpha), \quad (A.2)$$

where  $O_{.j} = \sum_{i=1}^I O_{ij}$ . Setting (A.1) and (A.2) to zero and solving for  $\alpha$  and  $\xi$  yields the MLEs  $\hat{\alpha}$  and  $\hat{\xi}$ . One way to solve these equations is to set  $\hat{\xi}_j = O_{.j}/y_{.j}$  initially, compute  $\hat{\alpha}$  from (A.1) using  $E_i = \sum_{j=1}^J y_{ij} \hat{\xi}_j$ , update  $\hat{\xi}_j$  from (A.2)

using  $\alpha = \hat{\alpha}$ , etc., until convergence. By functional invariance, the MLE of  $\alpha^{-1/2}$  is  $\hat{\alpha}^{-1/2}$ .

A.2 The Standard Error of  $\hat{\alpha}^{-1/2}$

Let  $1 \times J$  vector  $l_{\alpha\xi} = (d/d\alpha)(d/d\xi) \log m(\mathbf{O} | \alpha, \xi)$  and similarly define  $l_{\alpha\alpha}$ ,  $l_{\xi\alpha}$ , and  $l_{\xi\xi}$  of  $1 \times 1$ ,  $J \times 1$ , and  $J \times J$  dimension. These are elements of the observed Fisher information for  $\alpha$  and  $\xi$ . From (A.1),

$$\begin{aligned} l_{\alpha\alpha} = & - \sum_{i=1}^I \sum_{k=0}^{O_i-1} (k + \alpha)^{-2} + I/\alpha \\ & - \sum_{i=1}^I [(O_i + \alpha)/(E_i + \alpha)^2 - 2/(E_i + \alpha)]. \end{aligned}$$

From (A.2),

$$\begin{aligned} l_{\alpha\xi_j} = & \sum_{i=1}^I y_{ij}(O_i + \alpha)/(E_i + \alpha)^2, \\ l_{\xi_j\xi_j} = & -O_{.j}/\xi_j^2 + \sum_{i=1}^I y_{ij}^2(O_i + \alpha)/(E_i + \alpha)^2, \end{aligned}$$

and, for  $j \neq k$ ,

$$l_{\xi_j\xi_k} = \sum_{i=1}^I y_{ij}y_{ik}(O_i - \alpha)/(E_i + \alpha)^2.$$

The asymptotic variance of  $\hat{\alpha}$  is  $-(l_{\alpha\alpha} - l_{\alpha\xi}l_{\xi\xi}^{-1}l_{\xi\alpha})^{-1}$ . From the delta method,  $-[(l_{\alpha\alpha} - l_{\alpha\xi}l_{\xi\xi}^{-1}l_{\xi\alpha})4\alpha^3]^{-1/2}$  evaluated at  $\hat{\alpha}$  and  $\hat{\xi}$  is an estimate of the standard error of  $\hat{\alpha}^{-1/2}$ .

APPENDIX B

Derivation of  $\tilde{\alpha}^{-1/2}$  and Its Standard Error

B.1 The Method-of-Moments Estimator of  $\alpha^{-1/2}$

Let  $\mathbf{Z} = \{Z_{ij}\} = (Z_{11}, Z_{12}, \dots, Z_{1J}, Z_{21}, Z_{22}, \dots, Z_{IJ})^T$ ,  $I(\cdot)$  denote the indicator function, and  $D(\mathbf{h})$  denote a diagonal matrix with the elements of  $\mathbf{h}$  on the diagonal. Conditional on  $\gamma^*$ ,  $\mathbf{Z}$  has approximately the variance-covariance matrix  $\mathbf{W}^{-1}$ , where  $\mathbf{W} = D(\mathbf{O})$ , and mean  $\mathbf{A}\gamma^* + \mathbf{B}\xi^*$ , where  $\mathbf{A} = (\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_J)$ ,  $\mathbf{B} = (\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_J)$ ,  $\mathbf{A}_r = \{t_{ij}\}$  with  $t_{ij} = I(i = r)$ , and  $\mathbf{B}_s = \{u_{ij}\}$  with  $u_{ij} = I(j = s)$ . Let  $\mathbf{X}$  equal  $\mathbf{A}$  concatenated with  $\mathbf{B}$ . The difference between the weighted sums of squares for the full model  $E\mathbf{Z} = \mathbf{A}\gamma^* + \mathbf{B}\xi^*$  and the reduced model  $E\mathbf{Z} = \mathbf{B}\xi_0^*$  is  $\tilde{\theta}^T \mathbf{X}^T \mathbf{W} \mathbf{X} \tilde{\theta} - \tilde{\xi}_0^{*T} \mathbf{B}^T \mathbf{W} \mathbf{B} \tilde{\xi}_0^*$ , where  $\tilde{\theta} = (\tilde{\gamma}^*, \tilde{\xi}^*)$  and  $\tilde{\xi}_0^*$  are weighted least squares estimates of  $\theta = (\gamma^*, \xi^*)$  and  $\xi_0^*$  from the full and reduced models respectively. By matching this difference with its expectation, we obtain for  $\text{var } \tilde{\gamma}_i^* \simeq \alpha^{-1}$  the estimate  $\tilde{\alpha}^{-1} = [\tilde{\theta}^T \mathbf{X}^T \mathbf{W} \mathbf{X} \tilde{\theta} - \tilde{\xi}_0^{*T} \mathbf{B}^T \mathbf{W} \mathbf{B} \tilde{\xi}_0^* - (I - 1)]/t$ , where  $t$  is the trace of  $\mathbf{A}^T(\mathbf{I} - P_B)^T \mathbf{W}(\mathbf{I} - P_B)\mathbf{A}$  and  $P_B = \mathbf{B}(\mathbf{B}^T \mathbf{W} \mathbf{B})^{-1} \mathbf{B}^T \mathbf{W}$ . An estimate of  $\alpha^{-1/2}$  is  $(\tilde{\alpha}^{-1})^{1/2}$ , which we denote  $\tilde{\alpha}^{-1/2}$ .

B.2 The Standard Error of  $\tilde{\alpha}^{-1/2}$

Let  $\omega = (\theta, \xi_0^*, \alpha^{-1})$ . The least squares estimating equations for  $\theta$  and  $\xi_0^*$  and the moment equation for  $\tilde{\alpha}^{-1/2}$  are  $\mathbf{U}_1(\omega) =$

$\mathbf{X}^T \mathbf{W} \mathbf{X} \boldsymbol{\theta} - \mathbf{X}^T \mathbf{W} \mathbf{Z}$ ,  $\mathbf{U}_2(\boldsymbol{\omega}) = \mathbf{B}^T \mathbf{W} \mathbf{B} \boldsymbol{\xi}_0^* - \mathbf{B}^T \mathbf{W} \mathbf{Z}$ , and  $\mathbf{U}_3(\boldsymbol{\omega}) = \boldsymbol{\theta}^T \mathbf{X}^T \mathbf{W} \mathbf{X} \boldsymbol{\theta} - \boldsymbol{\xi}_0^{*T} \mathbf{B}^T \mathbf{W} \mathbf{B} \boldsymbol{\xi}_0^* - \alpha^{-1} t$  and have root  $\tilde{\boldsymbol{\omega}} = (\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\xi}}_0^*, \tilde{\alpha}^{-1})$ . Because the elements of  $\mathbf{Z}$  are independent, a first-order Taylor expansion shows that

$$\tilde{\boldsymbol{\omega}} \sim N \left\{ \boldsymbol{\omega}, \lim_{n \rightarrow \infty} \left[ \frac{d}{d\boldsymbol{\omega}} E\mathbf{U}(\boldsymbol{\omega}) \right]^{-1} \times \text{cov } \mathbf{U}(\boldsymbol{\omega}) \left[ \frac{d}{d\boldsymbol{\omega}} E\mathbf{U}(\boldsymbol{\omega})^T \right]^{-1} \right\}$$

in law provided the variance-covariance matrix is non-singular. It is singular because the columns of  $\mathbf{X}$  are not linearly independent, but this can be avoided by dropping one

of the columns and its corresponding parameter from  $\boldsymbol{\theta}$ . The asymptotic variance of  $\tilde{\alpha}^{-1}$  is obtained as the appropriate element of the variance-covariance matrix, which computes to  $4(\boldsymbol{\theta}^T \mathbf{L} \boldsymbol{\theta} - 2\boldsymbol{\xi}_0^{*T} \mathbf{M} \boldsymbol{\theta} + \boldsymbol{\xi}_0^{*T} \mathbf{N} \boldsymbol{\xi}_0^*)/t^2$ , where

$$\begin{aligned} L &= \mathbf{X}^T + \mathbf{X}^T \mathbf{W} \mathbf{A} \mathbf{A}^T \mathbf{W} \mathbf{X} \alpha^{-1}, \\ M &= \mathbf{B}^T \mathbf{W} \mathbf{X} + \mathbf{B}^T \mathbf{W} \mathbf{A} \mathbf{A}^T \mathbf{W} \mathbf{X} \alpha^{-1}, \end{aligned}$$

and

$$\mathbf{N} = \mathbf{B}^T \mathbf{W} \mathbf{B} + \mathbf{B}^T \mathbf{W} \mathbf{A} \mathbf{A}^T \mathbf{W} \mathbf{B} \alpha^{-1}.$$

By the delta method, an estimate of the standard error of  $\tilde{\alpha}^{-1/2}$  is  $[\alpha(\boldsymbol{\theta}^T \mathbf{L} \boldsymbol{\theta} - 2\boldsymbol{\xi}_0^{*T} \mathbf{M} \boldsymbol{\theta} + \boldsymbol{\xi}_0^{*T} \mathbf{N} \boldsymbol{\xi}_0^*)]^{1/2}/t$  evaluated at  $\tilde{\boldsymbol{\theta}}$ ,  $\tilde{\boldsymbol{\xi}}_0^*$ , and  $\tilde{\alpha}^{-1}$ .