

U.S. Cancer Mortality 1950–1978: A Strategy for Analyzing Spatial and Temporal Patterns

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

There are a number of technical and statistical problems in monitoring the temporal and spatial variation of local area death rates in the United States for evidence of systematically elevated risks. An analytic strategy is proposed to reduce one of the major statistical concerns, i.e., that of identifying areas with truly elevated mortality risks from a large number of local area comparisons. This analytic strategy involves two stages. The first is a procedure for examining the entire distribution of local area death rates instead of simply selecting high risk "outliers." The second is the development of an analytic procedure to relate the temporal changes in the cross-sectional distribution of local area death rates to models of the disease process operating within the populations in those areas. The procedures are applied to data on cancer mortality for the 3050 counties (or county equivalents) of the United States over the period 1950 to 1978. A number of striking mortality patterns, both within the entire United States and within various regions and states, are identified. For example, perhaps the most persistent finding was that the risk increases in the death rates for respiratory cancer mortality were due to a "catching up" of nonmetropolitan county mortality rates with metropolitan area mortality rates.

Introduction

Frequently, geographic patterns of chronic disease death rates are monitored to identify areas with significantly elevated mortality risks (1). Once such areas are identified, more intensive investigations can be fielded to determine what life style, occupational, or environmental factors may have contributed to the local area elevation of rates (2).

Such a strategy is both feasible and highly cost-effective because the vital statistics system is an ongoing component of government, and all areas of the country are monitored for all causes of death. Though the basic strategy has proven highly effective, there are a number of analytic issues that can restrict the utility of the approach. For example, the quality of death certificate diagnoses varies both with the identity of the diagnosis and with the level of detail and specificity of the diagnosis. Thus, we must be aware of such variation in selecting diagnostic categories for analysis.

In this paper, a second set of analytic issues is

addressed. These issues arise because, in monitoring or screening the death rates for a number of different diseases in the 3050 counties (or county equivalents) (3) of the United States for up to 30 years, it is difficult to determine if a particular elevated (or depressed) pattern of cause-specific death rates has occurred by chance. Indeed, the large number of death rates makes it operationally difficult to systematically evaluate the set of rates for patterns. We propose a two-stage analytic strategy to perform this evaluation. The first stage is descriptive and involves examining the distribution of death rates across geographic areas and demographic variables to determine if nonrandom patterns exist and to determine if the patterns in the distribution of rates persist over different groupings of these local area rates. The purpose of this part of the overall geographic analysis could be described as "pattern recognition."

The second stage of our analysis involves assessment of the temporal variation of death rates within local areas (e.g., counties) to ascertain characteristics of the disease dynamics within the local area population and to relate local patterns to the disease dynamics in larger population groups (e.g., national or state populations). To accomplish this we employ detailed models of county level death rates developed from epidemiological and

*Center for Demographic Studies, Duke University, Durham, NC 27706.

†U.S. Environmental Protection Agency, Research Triangle Park, Durham, NC 27711.

biomedical theory and auxiliary data (4). These analytic models facilitate the tracking of temporal changes in mortality risks for individuals or cohorts. The primary benefit of analytic models describing cohort patterns of risk is that they increase our ability to screen out chance geographic variation in death rates by relating changes in the cross-sectional distribution of those rates to systematic age, period, and cohort changes in mortality risks. A second benefit is that such models provide a more realistic basis for projecting mortality trends over time and space (5).

The presentation of the analysis is organized around the three goals of this paper. First, we have examined changes in the distributions of cancer death rates for U.S. counties, both for the country as a whole, and for select states, to determine if there are persistent spatial and temporal patterns of variation. In this assessment we found a number of highly significant geographic patterns in the rate of change of county death rates. Of particular interest were metropolitan and nonmetropolitan differences in the rate of change of male respiratory cancer death rates. Second, we applied more detailed cohort models to analyses of county, state, and regional variation in site specific and total cancer death rates. In this type of analysis we were able to relate temporal changes in the cross-sectional distribution of cancer death rates to different cohort patterns of risk in these areas. Third, we evaluated the utility of the analytic models for identifying real deviations in cancer risks and for describing the mechanisms generating those risks.

Data

The data that we evaluated in this report are county level vital statistics data on cancer mortality and county level census data for the period 1950 to 1978. Using these data we can calculate age-, race-, and sex-specific death rates for each year in the period 1950 to 1978.

In particular, age-, race-, and sex-specific mortality counts were derived for 35 cancer sites for each of the 3050 counties (or county equivalents) for each year 1950 to 1978. The counts were prepared from computer tapes, prepared by EPA, which contained the individual death certificate diagnoses used by Mason et al. (1) for the period 1950 to 1969, and from NCHS public release tapes for the period 1970 to 1978. These were matched with (a) corresponding population counts from the 1950, 1960, and 1970 censuses and (b) corresponding population estimates obtained by linear interpolation between the censuses. For the period after 1970, population counts were taken from special post-censal estimates produced by the U.S. Bureau of the Census.

Rates of two types were generated from these data. First, direct age-standardized cancer death rates were produced for each of four race-sex groups for all U.S. counties. In all cases, the standard population used was the total U.S. population for 1970. These rates were calculated for three intervals: 1950 to 1959, 1960 to

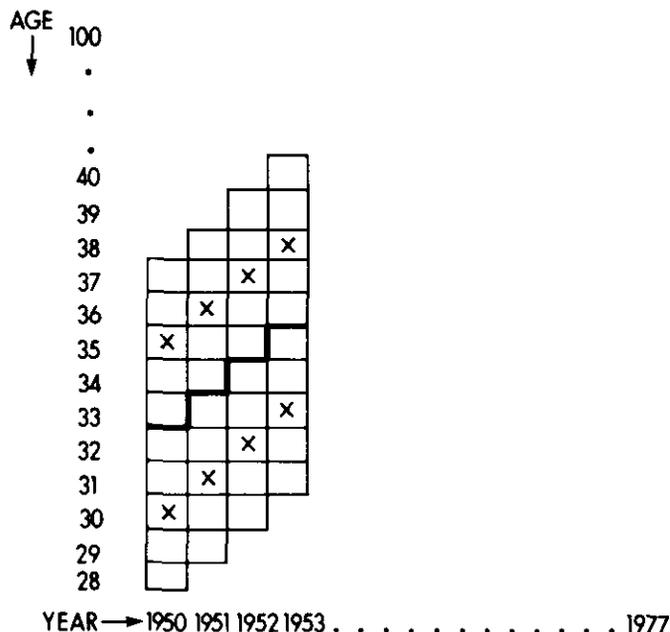


FIGURE 1. Data structure for defining cohort rates.

1969, and 1970 to 1978. For the computation of the frequency polygons exhibited below in Figures 2-7, these rates were weighted by the size of the population in each county for each time interval so that the weighted distribution of the rates reflects the distribution of the total population by risk levels.

The second rates calculated were sets of cohort specific death rates for each year in the period 1950 to 1977. Because data were available on a single year of date and single year of age basis, these cohort rates could be used to track specific cohorts over time. To understand how the cohort rates were calculated, consider Figure 1.

In Figure 1 we have presented a portion of an age-by-date matrix. Here we show two cohorts (aged 30 and 35 in 1950) followed for a four-year period (1950 to 1953). The cohort rate for each year is actually the average over a group of five single-year-of-age categories. The five-year groups are cross-temporally tracked on a single year basis. Thus, the death rate calculated for the first cohort actually represents the average death rate for persons 28 to 32 years of age in 1950. One year later, in 1951, this group was 29 to 33 years of age. We take the average death rate in each five-year group to represent the death rate at the middle age.

A total of 45 birth cohorts born from 1878 to 1922 were tracked for 28 years. From these data, sets of age (a)- and cohort (c)-specific cancer death rates for nine five-year cohort groups aged 28-32, 33-38, . . . , 68-72 in 1950 were calculated from the corresponding death $D(a,c)$ and population $P(a,c)$ counts as indicated in Figure 1, i.e., by using the formula:

$$\lambda(a, c) = \sum_{i=-2}^{+2} D(a-i, c+i) / \sum_{i=-2}^{+2} P(a-i, c+i) \quad (1)$$

where $c = 1880, 1885, \dots, 1920$ and, in 1950, $a = 30, 35, \dots, 70$. Thus, these cohort death rates utilize all of the available data and provide rate estimates smoothed over the five-year-of-age window.

Descriptive Analyses

In the first stage of our analysis we examined temporal changes in the county-population weighted distribution of cancer death rates both in the U.S. as a whole and in select states. This can be illustrated in Figure 2, where the percent distributions of males in the U.S. white population at 21 levels of the age-adjusted death rate for all types of cancer are plotted for three time periods: 1950-59, 1960-69, and 1970-78.

In Figure 2 we see that the maximum cancer death rate for any U.S. county for any of the three time periods is nearly 270 per 100,000. This can be compared with an age-adjusted death rate for all cancer mortality of 180 per 100,000 for males in the U.S. white population in 1977. Thus the highest cancer death rate for a county is 50% higher than the national rate in 1977. In

terms of the modal rate we see that there has been an increase from about 160 per 100,000 to about 190 per 100,000 with the proportion experiencing this higher modal value increasing from 16 to 22%.

While there has been only a small increase in the proportion of the male population with cancer death rates between 230 and 270 per 100,000, we see that the maximum male death rate has not increased with time. The proportion of the male population with cancer death rates less than 120 per 100,000 has dropped greatly. The net effect of these changes is that the distribution of male cancer risks over U.S. counties has appeared to shift to the right and become more "peaked" as more of the male population reaches death rate levels between 170 and 230 per 100,000. Furthermore, the 1950-1959 plot shows evidence of an underlying second-distribution at about 120 per 100,000. This does not appear in the 1970-78 plot. This suggests that recent increases in U.S. white male cancer mortality have been a result of homogenization of cancer risks over local areas and not due to an overall upward displacement of the distribution. This suggests that cancer rates for certain local areas have increased rapidly over the 28-year period while the cancer death rates for other high risk areas have remained relatively static near the very high levels initially observed in 1950-59.

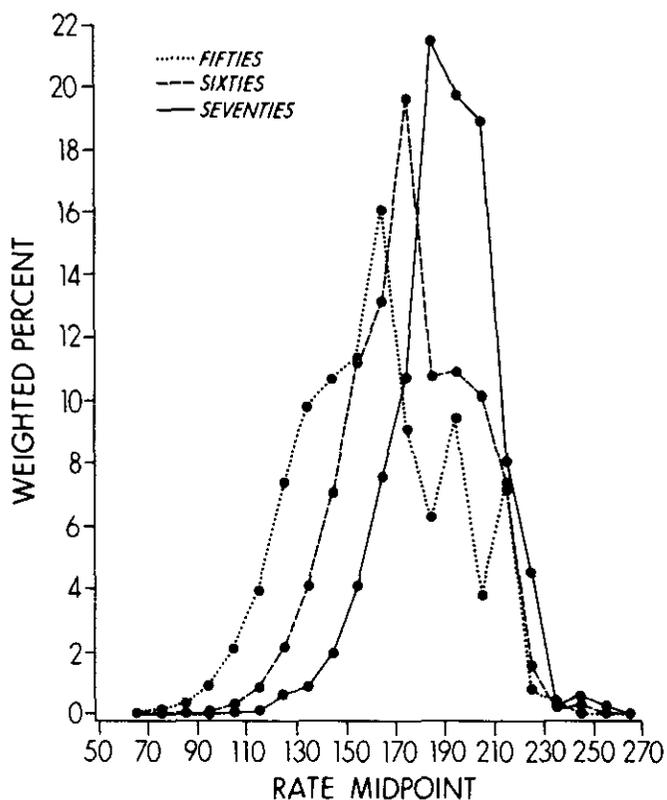


FIGURE 2. County-population weighted distributions of age-adjusted death rates for all cancer mortality among white males in the 3050 counties of the United States.

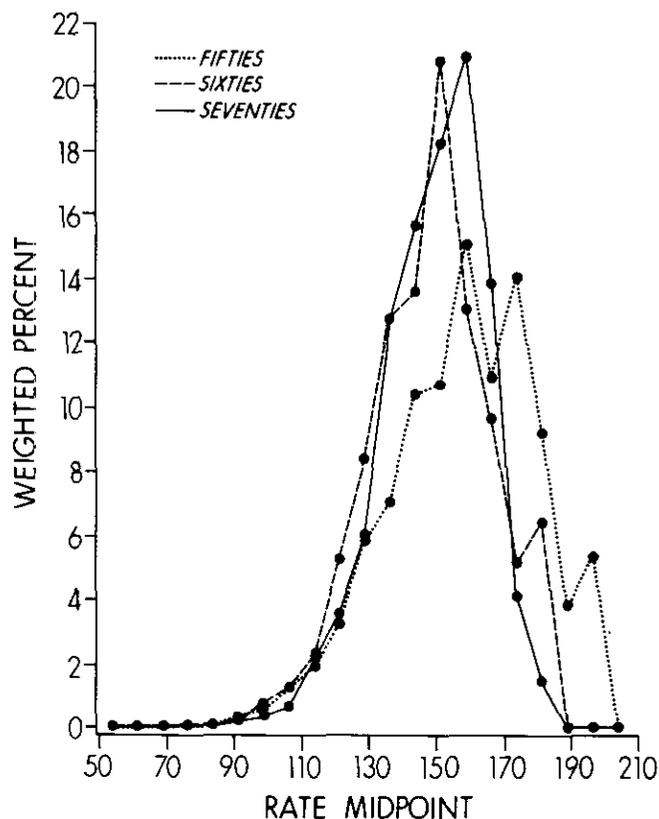


FIGURE 3. County-population weighted distributions of age-adjusted death rates for all cancer mortality among white females in the 3050 counties of the United States.

The corresponding county-population weighted distributions of age-adjusted death rates for all cancer mortality for females in the U.S. white population are presented in Figure 3.

There are a number of clear differences between the male (Fig. 2) and female (Fig. 3) distributions. First of all, the maximum female cancer death rates for counties are not as high as for males (e.g., 196 per 100,000 vs. 270 per 100,000). Second, the temporal increases in female cancer death rates are far less than for males. Indeed, there is little evidence in Figure 3 of a significant decrease in the proportion of the female population at the lowest rates (less than 130/100,000) over the three time periods, while the proportion of the female population at the highest levels of risk has actually decreased. The result of these changes is that the variance of the female distribution has decreased, and the distribution has become more peaked.

The differences between Figures 2 and 3 can be examined further by decomposing changes by cancer site. One type of cancer that has been argued to underlie sex differences in temporal trends at the national level is respiratory cancer. The sex specific population distributions of respiratory cancer among whites in the U.S. are shown in Figures 4 and 5.

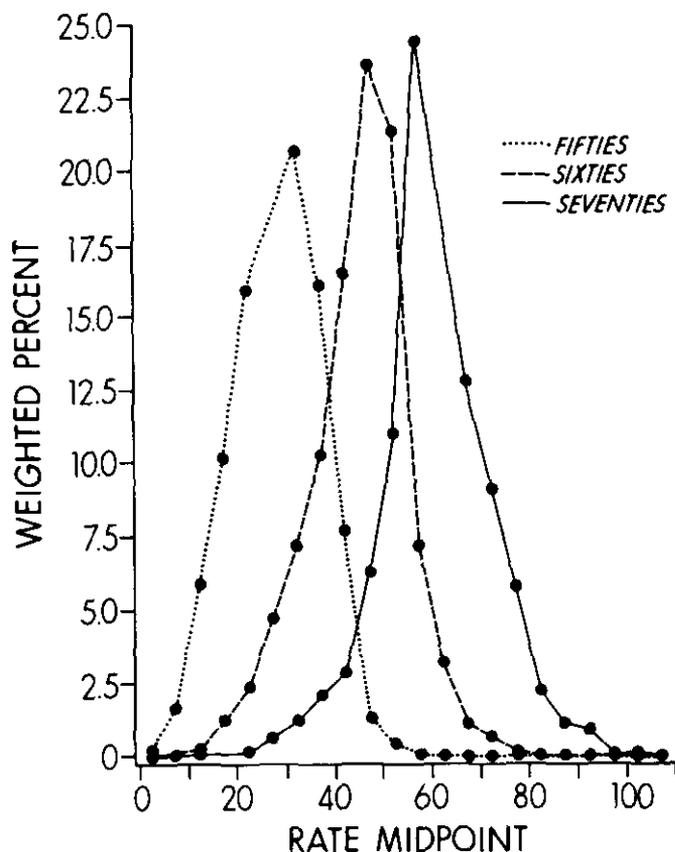


FIGURE 4. County-population weighted distributions of age-adjusted death rates for respiratory cancer mortality among white males in the 3050 counties of the United States.

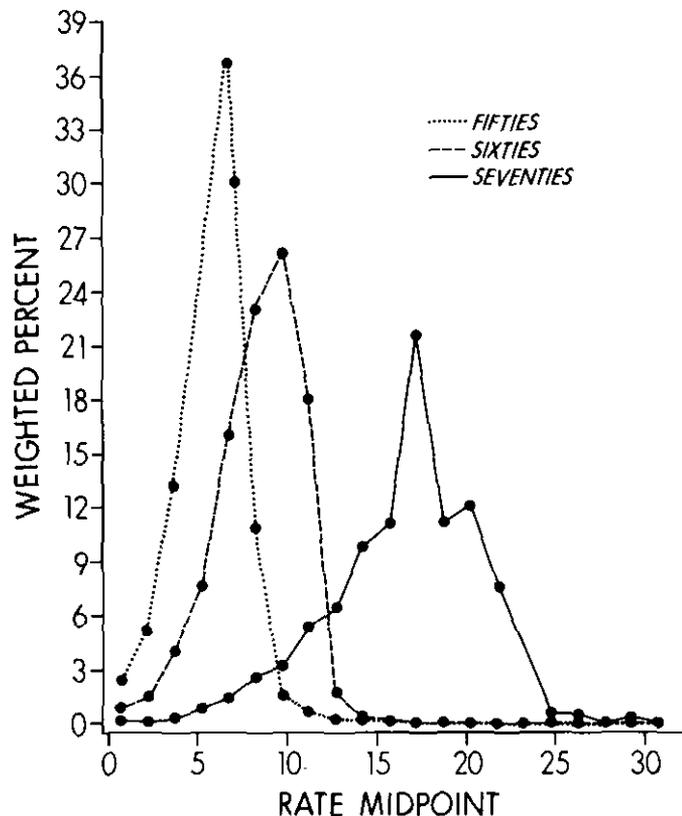


FIGURE 5. County-population weighted distributions of age-adjusted death rates for respiratory cancer mortality among white females in the 3050 counties of the United States.

The male and female respiratory cancer patterns are very different. For males there appears to be a reasonably constant upward shift in the distribution with time. There is also a tendency for the distribution to become more peaked with a greater proportion of males at the modal value. This parallels the pattern observed for all cancer mortality for males. The respiratory cancer mortality pattern for females is, however, quite different from the female pattern for all cancer mortality. For respiratory cancer, the variance of the female distribution increases rapidly over time, with the population becoming far more heterogeneous. This is apparently so because the respiratory cancer mortality rates for females were so low in 1950-59 that they were truncated on the lower side. With time, the distribution shifted to the right but with different areas changing at different rates so that the distribution flattened. This contrasts with the pattern for all cancer mortality for females, where there was no evidence of such a shift. The changes for all cancer mortality paralleled the changes for respiratory cancer mortality for males, but not for females, because a much larger proportion of total male cancer mortality is due to cancer of the respiratory system.

The distributions for the nation are difficult to interpret in detail because of the wide range of conditions affecting local area cancer death rates. By focus-

ing our attention on the distribution of county-population weighted death rates for a somewhat smaller aggregate (e.g., a state), we can better interpret the distributions and their change with time. In particular, there may be characteristics of particular local areas that will index probable exposures to various types of pollutants or to different life style factors (e.g., smoking, diet). The selection of such areas based on likely exposure differences provides us with a type of quasi-experimental logic to determine if geographic differentials in mortality risks are associated with exposure patterns (2). Given that the death rates can be calculated for different demographic categories, we can also control for various life style and occupational differences. For example, the effects of smoking might be assessed by comparing areas with high concentrations of Mormons (who are typically nonsmokers) to other areas. In a prior analysis this was done for Orleans Parish, LA, and Salt Lake County, UT (4).

A major advantage of the longitudinal, area-specific vital statistics data bases is that we can select areas according to a logic that allows us to approximate a retrospective or prospective type of analysis by using the county population as our unit of study. For example, by selecting sets of counties on the basis of exposure factors, e.g., selecting counties with high levels of ship building and counties without such industrial activity (2), one can evaluate prospectively differences in the pattern of change of rates in those areas. Alternatively, in a fashion analogous to the logic of retrospective studies, we could select areas on the basis of the level of their death rates and examine the distribution of exposure factors for areas with different types of mortality risks. In both cases, we could examine either cross-sectional or cohort death rates. Some caution is warranted, however, because the unit of analysis is an area population and not individuals. Thus, the use of such rates only approximates a retrospective or prospective logic because the population at risk can change due to migration. It has been shown, however, that the general effect of such movement is to diffuse risk differentials due to exposure (6) and, hence, the strategy ought to produce conservatively biased results.

To illustrate an analysis of death rates in a subgroup of counties, let us consider the county-population weighted distributions of age-adjusted death rates for all cancer mortality for white males in Illinois (Fig. 6).

Here we see that a secondary peak in 1950-59 moved to the right and merged with the primary peak in 1970-78. Illinois was selected for indepth analysis because it is a state with a major metropolitan area (Chicago) and a large down-state nonmetropolitan population. The primary peak in Illinois represents the elevated cancer risks in Chicago. It moves relatively little with time. Clearly the "down-state" counties have, in effect, caught up with the cancer risks in the Chicago area.

These rates can also be decomposed according to specific cancer sites. For example, in Figure 7, we present white male respiratory cancer death rates for Illinois counties.

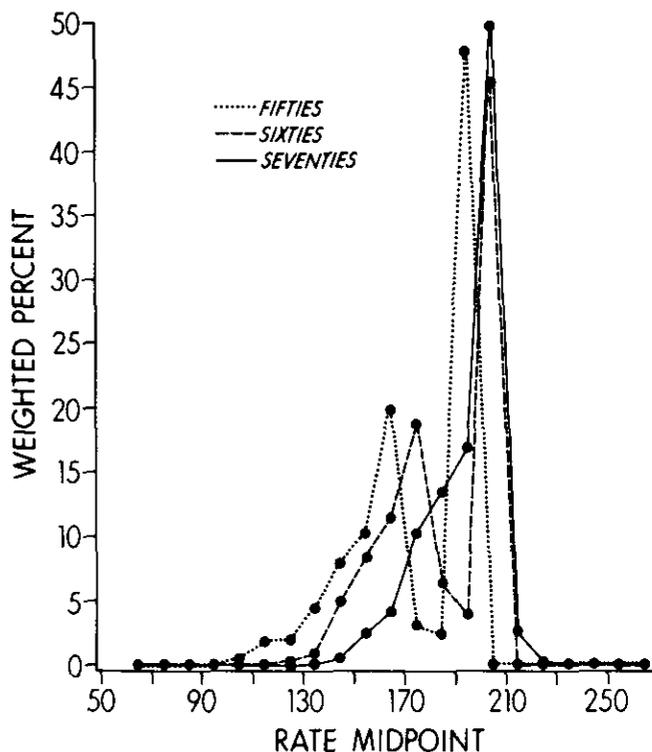


FIGURE 6. County-population weighted distributions of age-adjusted death rates for all cancer mortality among white males in the 102 counties of Illinois.

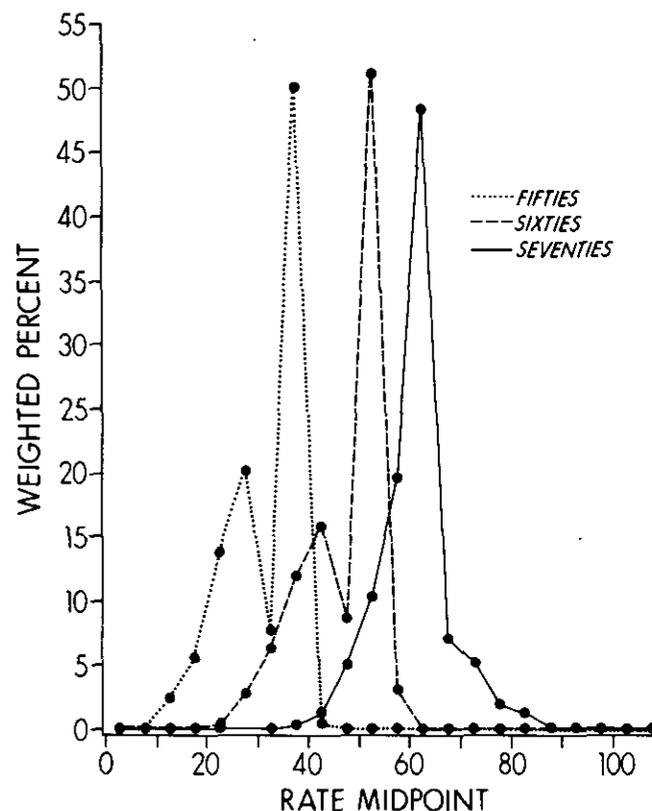


FIGURE 7. County-population weighted distributions of age-adjusted death rates for respiratory cancer mortality among white males in the 102 counties of Illinois.

The temporal changes in the distribution of male respiratory cancer death rates are of two types. First, the secondary peaks in 1950–59 and 1960–69 have merged with the primary peak in 1970–78. Thus the variance of the distribution has decreased, i.e., the level of respiratory cancer risks over counties has become more homogeneous. Second, the whole distribution shows a marked upward shift.

Comparisons of the respiratory cancer death rate distributions for males for the U.S. and Illinois show some similarities and some differences. The rapid increase of the respiratory cancer risks is evident in both distributions with a strong upward shift of the modal points of both distributions. In the Illinois distribution, however, there is clearly a secondary peak at lower death rate levels in 1950–59 whereas, in the U.S., though there is an elbow in the distribution, there is no separate peak. The bimodality of the distribution in Illinois shows the effect of a state population clearly divided between a large metropolitan area (Chicago) and the nonmetropolitan remainder. In the U.S. the distribution does not show such bimodality because there are many metropolitan and near-metropolitan areas to define a smoother gradient. A similar observation could be made for total cancer mortality.

Cohort-Specific Analyses

The descriptive analyses presented above provide an excellent sense of the nature of local area changes in cancer death rates. However, these changes in the cross-sectional distributions do not explicitly represent the dynamics of the disease process in cohorts. We analyzed these dynamics by using a model of cancer mortality risks that has been found to well describe the age dependence of cohort death rates for a number of different chronic diseases. This model, the Weibull hazard model, can also be related to the well known multistage/multi-hit model of carcinogenesis due to Armitage and Doll (7,8). This Weibull hazard function, modified to include cancer latency (the time between tumor onset and clinical expression), can be written in the form,

$$\lambda(a, c) = \alpha_c (a - l)^{m-1} \quad (2)$$

where $\lambda(a, c)$ is the cohort death rate defined in Eq. (1), α_c is a cohort-specific scale or proportionality factor, a represents age, l is an estimate of latency, and m is a shape parameter. (In the Armitage-Doll model, the m can be interpreted as the number of mutations needed to initiate a tumor; also a $\bar{\alpha}_c$ is an age invariant risk constant which is proportional to the product of the m annual mutation probabilities.)

In practice, Equation (2) often fails to predict cancer death rates at later ages accurately (9). Therefore we generalized Equation (2) to allow the mortality risks to differ over individuals within each cohort. To represent the effects of such risk heterogeneity on the age trajectory of the cohort death rates, we define $\bar{\alpha}_c$ as the average of individual risks, $\alpha_{i,c}$, within that cohort. In

order to develop this heterogeneous population generalization of the Weibull model fully, it is necessary to make an assumption about the form of the distribution of the $\alpha_{i,c}$. We assumed that the $\alpha_{i,c}$ were gamma-distributed, i.e., they have the density function

$$f(\alpha_{i,c}) = \left(\frac{\alpha_{i,c}}{\beta_c} \right)^{s_c} \frac{\exp\{-\alpha_{i,c}/\beta_c\}}{\alpha_{i,c} \Gamma(s_c)} \quad (3)$$

with mean $\bar{\alpha} = \beta_c s_c$ and variance $\beta_c^2 s_c$. This yields a modified form of the Weibull function (10):

$$\lambda(a, c) = \bar{\alpha}_c (a - l)^{m-1} / [1 + \bar{\alpha}_c (a - l)^m / (m s_c)] \quad (4)$$

One can see that the effects of risk heterogeneity are represented in Eq. (4) only through the gamma shape parameter s_c , which is seen from Eq. (3) to be the inverse of the squared coefficient of variation of the $\alpha_{i,c}$. The modified Weibull function in Eq. (4) is not unique, however, in the sense that alternative assumptions about the form of the distribution of the $\alpha_{i,c}$ lead to alternative modifications of the Weibull function. One alternative suggested by Hougaard (11) is the inverse Gaussian distribution which, like the gamma, is a member of the natural exponential families. Another alternative, suggested by Cook et al. (9), is the two-point susceptibility distribution in which only some given fraction of the population is susceptible to the initiation of a tumor. We have investigated the implications of these two alternative distributional assumptions and found that the fitted parameter values are reasonably robust with respect to those choices (12,13). Indeed, under comparable parameterizations it may be shown that the two-point distribution implies that s_c declines over time (12), that the inverse Gaussian distribution implies that s_c increases over time (13), whereas the gamma distribution implies that s_c is constant over time (14). Given that the constancy of s_c means that the coefficient of variation is also constant over time, one can see that the gamma distribution permits us to account for the effects of risk heterogeneity in a parsimonious manner which represents a consensus of the other plausible alternative distributional assumptions.

We employ two additional assumptions to adjust Eq. (4) for the effects of competing risks so that our model may be fit to the set of cohort specific death rates obtained by using Eq. (1). First, we assume that mortality from the specific cancer modeled in Eq. (4) is independent of mortality due to all other causes of death. This is a standard assumption that is frequently made to achieve identifiability in competing risk models (15–18). For the present model, this assumption means that the marginal hazard rate in Eq. (4) is equal to the conditional hazard rate, given that each exposed individual survives not only death due to the specified cancer but also all other causes of death up to at least age a . Second, because this conditional hazard rate obtains as the limit of the expected death rate as the exposure interval tends to zero, we assume that the conditional hazard rates at

the midpoint of each 1-yr age interval provide satisfactory approximations to the expected death rates for those intervals. This assumption is standard and derives from the work of Reed and Merrell (19).

Taken together, these two assumptions provide a tractable method of adjusting for the effects of competing risks. Furthermore, because the midpoint approximation refers to the conditional hazard rate, its validity does not depend on the validity of the independence assumption. It is useful to consider how violations of the independence assumption would affect results derived from our model. This can be done because the heterogeneity model provides an appropriate conceptual framework for considering the effects of dependence among several causes of death (11,20,21). For example, in addition to respiratory cancer, Klebba (22) identifies 13 other diseases associated with smoking. If each individual is characterized by a vector of disease specific susceptibility constants, e.g., $\beta\alpha_{i,c}^T = (\alpha_{1,i,c}, \dots, \alpha_{14,i,c})$

then the assumptions concerning dependence or independence of the 14 diseases may be stated in terms of the distribution of $\alpha_{i,c}$ and, conditional on $\alpha_{i,c}$, the times to death from each of the diseases may be assumed independent. Practical considerations suggest that models of dependence of the elements of $\alpha_{i,c}$ be restricted to the case where these elements can be expressed as functions of additional covariates that are available on an area and cohort specific basis. Otherwise, if the elements of $\alpha_{i,c}$ are independently gamma-distributed, then both the marginal and conditional hazard rates for each disease will be of the form of Eq. (4), as we have assumed.

The parameters of the gamma/Weibull function in Eq. (4) can be estimated by using a Poisson likelihood function based on the counts of the number of deaths and the person years at risk (10,23). For the Poisson function, the maximum likelihood estimates of $\lambda(a,c)$ for the fully saturated model (i.e., same number of parameters as observations) are precisely the observed death rates, computed as in Eq. (1).

To generate more parsimonious descriptions of these data, we parameterized $\lambda(a,c)$ as shown in Eq. (4) to examine the variation of cancer death rates within local area populations over age and time. This model was also used to examine the geographic variability of the local area rates. This was done by estimating the parameters of the model for the nation, state, or region, and testing to see if those parameters could satisfactorily be applied to a local area population. If the use of parameter estimates from larger population groupings did not cause significant deterioration of the fit to the local area death rates, then we employed the parameter estimates from the larger populations. This, in effect, provided an empirical test of the expected bias of the parameter estimates for the larger population when applied to the local population. Besides being one way of clearly defining local areas that had mortality risks that were significantly different from the larger population, it also helped

us to understand the structure and source of those differences.

The fit of the model in Eq. (4) to U.S. cohort data for white males is illustrated in Figure 8.

We see that the fit to the national cohort patterns is quite good. The increase in cohort levels of risk for younger cohorts is evident. It is also clear that the largest increases in cohort risks occurred over the five oldest cohorts (i.e., the birth cohorts 1880 to 1900). Consequently, a search for risk factors to explain the recent increases in respiratory cancer death rates should focus upon factors which became elevated for these cohorts at "susceptible" early ages. The likelihood that there exists a set of most susceptible ages for exposure derives from two observations: (a) the observation of enhanced susceptibility at certain ages to specific carcinogens, e.g., the enhanced susceptibility to cancer

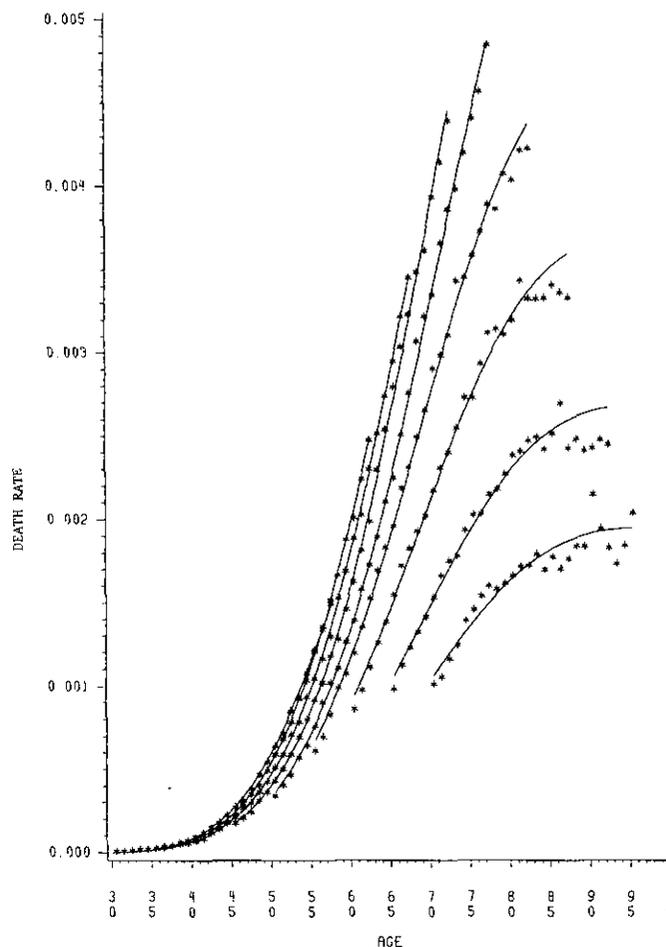


FIGURE 8. Observed and predicted age-specific death rates for respiratory cancer mortality in the United States from 1950 to 1977 for nine white male cohorts aged 30 to 70 in 1950. The observed death rates, indicated by the 9 sets of asterisks, were computed by using Eq. (1). The predicted death rate functions, indicated by the nine smooth curves, were computed by using Eq. (4) with the maximum likelihood estimates of the cohort specific parameters $\bar{\alpha}$ and \bar{s} , and the pooled estimates of the global parameters $m (=6)$ and $l (=20.1 \text{ years})$. See text for explanation.

of persons aged 15 to 25 due to cigarette consumption (24), and (b) the observation that carcinogenic exposures at later ages do not often result in clinically manifest diseases because of lengthy latency times, e.g., in Peto et al. (25), the eventual clinical manifestation of mesothelioma was five times as high if the initial exposure occurred at age 20 than at age 40.

In considering the role of such risk factors in explaining cohort differentials in respiratory cancer mortality, it needs to be emphasized that positive dependence in susceptibilities to respiratory cancer and other causes of death will yield progressively lower observed respiratory cancer mortality rates at older ages than would occur under independence (21). Because this form of bias would affect the parameter s (which controls the "downturn" in the curve at older ages: see Fig. 8) more so than the parameter $\bar{\alpha}$ (which controls the initial average level of the cohort hazard function), we expect cohort comparisons based on $\bar{\alpha}$ to be relatively more insensitive to violations of the independence assumptions. Hence, in the following we will restrict our within- and cross-cohort comparisons to contrasts of the parameter $\bar{\alpha}$.

The cohort trends and differentials at the national level are quite strong. Thus it would seem important that analyses at the local area level be adjusted to reflect the national cohort trends. The adjustment of local area and state analyses for national trends can take one of two forms. First, it is possible to impose constraints on the parameters of cohort specific models estimated at the local area level. Logically, since the national level data provide much more precise estimates of parameters, this would suggest that those parameter estimates ought to be employed, unless a significant deterioration of fit can be demonstrated in using those parameters for a given area. This would be equivalent to using the model to demonstrate a statistically significant bias in applying the parameters estimated in the national model to the local area data. The second approach is to compare the $\bar{\alpha}$ for specific cohorts estimated from the national and local area data, but with the parameters m , l , and the ratio $\beta_c = \bar{\alpha}_c/s_c$ [see Eq. (3)] constrained to their national values in the local area estimation phase. Applying the constraint to β_c rather than to s_c in Eq. (4) means that the ratio of the local area estimate of $\bar{\alpha}_c$ to the national estimate of $\bar{\alpha}_c$ provides a measure of the age-independent relative risk of that cohort in the local area vis-à-vis the national cohort.

This second type of model can be used to make several types of comparisons. First, it can be used to make cohort risk comparisons within a state. For example, we saw in Figure 4 that U.S. white male respiratory cancer death rates increased in the period 1950–78. We also can see in Figure 8 that there is a tendency for these rates to increase for younger cohorts. In Figure 9 we present a plot of the ratio of cohort-specific $\bar{\alpha}$ for respiratory cancer for white males in Illinois, separately estimated for the entire state, for Cook County, and for all other counties in the state, to the corresponding $\bar{\alpha}$ -

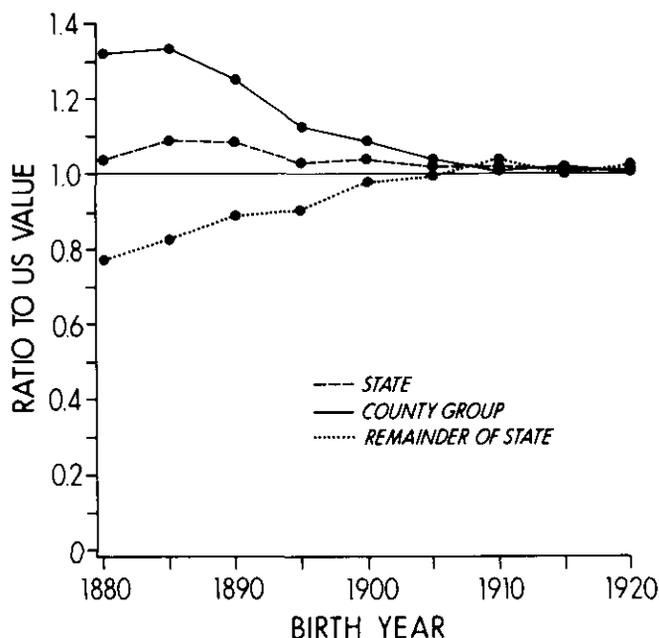


FIGURE 9. Cohort-specific relative risks of respiratory cancer mortality among white males in Illinois and in select counties in Illinois. Relative risks computed by using cohort specific risks of respiratory cancer mortality among all white males in the United States as the standard.

estimates obtained from the national data. The relative deviation of the subarea $\bar{\alpha}_c$ from the national value is plotted against the vertical axis of the figure while the year of birth of the cohort is plotted against the horizontal axis.

In Figure 9 we see that the state of Illinois is very near to the U.S. in terms of cohort specific levels of respiratory cancer mortality. However, we also see that this is a product of two very different trends. First, for Chicago (Cook County), we see that the oldest cohorts were much elevated with respect to the national values but that the differential disappears for younger cohorts. In contrast, the "down-state" counties initially had lower than average risks, in comparison with either the state or the nation. However, for younger cohorts the differential again disappears. This suggests a mechanism that may underlie the two aspects of change of the cross-sectional distribution of county specific respiratory cancer death rates. It suggests that, for birth cohorts up to about 1895, the major metropolitan area of Illinois had elevated risks. Thereafter, the metropolitan and other areas were quite similar. This explains why the secondary peak of the county specific respiratory cancer death rates observed in Illinois in 1950–59 and 1960–69 disappeared in 1970–78, and the variance of the distribution decreased. The upward shift of the rates with time is probably reflective of the general increase in the level of respiratory cancer death rates in younger cohorts. Such a progression is not manifest in Figure 9 because the relative risks have, in effect, had the na-

tional trends (i.e., increases in respiratory cancer mortality risks in younger cohorts) removed.

A second interesting comparison of cohort trends within a state can be made for the state of Alabama. This is illustrated in Figure 10.

Here the target counties are those around Birmingham. We see that the state and target county mortality risks for older cohorts are much below (20% below for Birmingham and 35% below for the state) the national values. We see that, by the birth cohort of 1890, the Birmingham cohort mortality risks exceed (by nearly 30%) those of the nation. We also see that the mortality risks for the remainder of the state begin to converge with those of Birmingham slightly later (i.e., beginning with the 1895 birth cohort). By the 1910 birth cohort, even these lowest mortality risks exceeded the national values.

It is interesting that similar types of patterns emerge for other southern states while the reverse exists for a number of major northeastern states (e.g., New Jersey, New York, Pennsylvania). In these northeastern states both the nonmetropolitan and metropolitan components are elevated for the older cohorts. Gradually, however, for younger cohorts the differentials both within state, and between the state and national rates, decrease.

We have shown how cohort trends can be compared between target counties and the remainder of the state and can be compared in a way that has the national trends removed. Though such a decomposition of the cohort specific mortality risks is of interest in explaining cross-temporal changes, it does not indicate the absolute

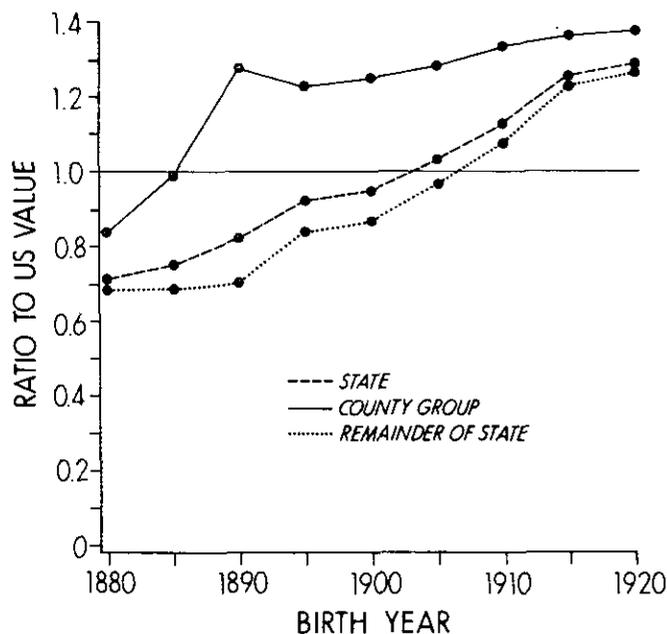


FIGURE 10. Cohort-specific relative risks of respiratory cancer mortality among white males in Alabama and in select counties in Alabama. Relative risks computed by using cohort specific risks of respiratory cancer mortality among all white males in the United States as the standard.

levels of risk expected. This can be studied by examining the size of the $\bar{\alpha}$ for different cohorts. These $\bar{\alpha}$ represent the age-independent increment in risk in our proportional hazards type of Weibull model. This establishes the absolute level of risk because, for a given tumor type, all cohort curves are assumed to have the same intrinsic curvature (i.e., all cohort m are assumed equal).

In Figure 11 we compare the cohort changes in the level of risk for the U.S. and four states (Alabama, New York, Louisiana, New Jersey). The dashed line for the U.S. indicates the trend in the increase in the absolute level of cohort risks which we removed in Figures 9 and 10. We see, for example, that there is about a fivefold increase in $\bar{\alpha}$ from the 1880 to 1920 cohorts. This means that the risk level at any given age will be five times higher in the 1920 cohort. Thus the trend removed in the earlier plots represented about a 12.5% increase in respiratory cancer risks for each year of cohort age. It is interesting to compare the trends in the four states with the national trend. We see that the two southern states have a much faster rate of increase in respiratory cancer mortality risks than for the U.S. as a whole. What is perhaps surprising is that the two northeastern states have a less rapid increase in risk with cohorts than the U.S. The regional differences may reflect the higher proportion of rural populations in the southern

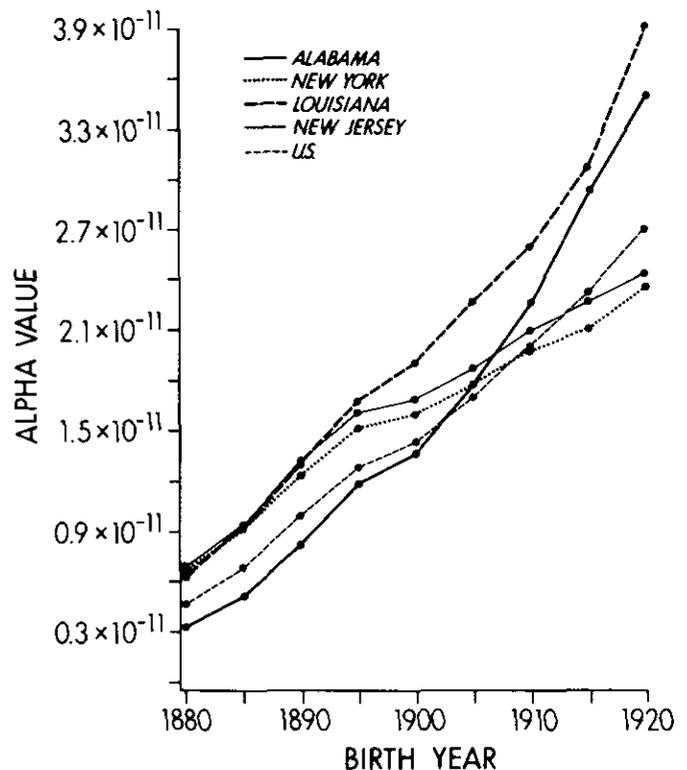


FIGURE 11. Cohort-specific risks of respiratory cancer mortality among white males in the United States and among white males in four select states. Absolute levels of risk measured in α units. See text for explanation.

states and the faster increase in risks for those areas.

The methods of cohort-specific analysis of site specific cancer mortality risks can also yield insight into the temporal trends of other cancer types. For example, so far we have concentrated on studying factors responsible for the change in the distribution of respiratory cancer death rates. Another important type of cancer that could be examined is stomach cancer. Stomach cancer might be examined because the etiological factors involved in its initiation are potentially quite different from those involved in respiratory cancers. The cohort plots for stomach cancer, in the U.S. and five states (Illinois, New York, Michigan, Pennsylvania, New Jersey), are provided in Figure 12.

Note that the size of the α_c for stomach cancer are much larger than those for respiratory cancer. This is because the Weibull shape parameter (m) is greater for respiratory cancer ($m = 6$) than for stomach cancer ($m = 4$), i.e., the stomach and respiratory cancer curves are not proportional. The larger shape parameter for respiratory cancer suggests that it will increase much more rapidly with age than stomach cancer. Consequently, the size of the age invariant component of risk is smaller.

It is evident that the pattern of stomach cancer risks over cohorts is quite different. It appears that the mean risk within cohorts declines nearly at an exponential

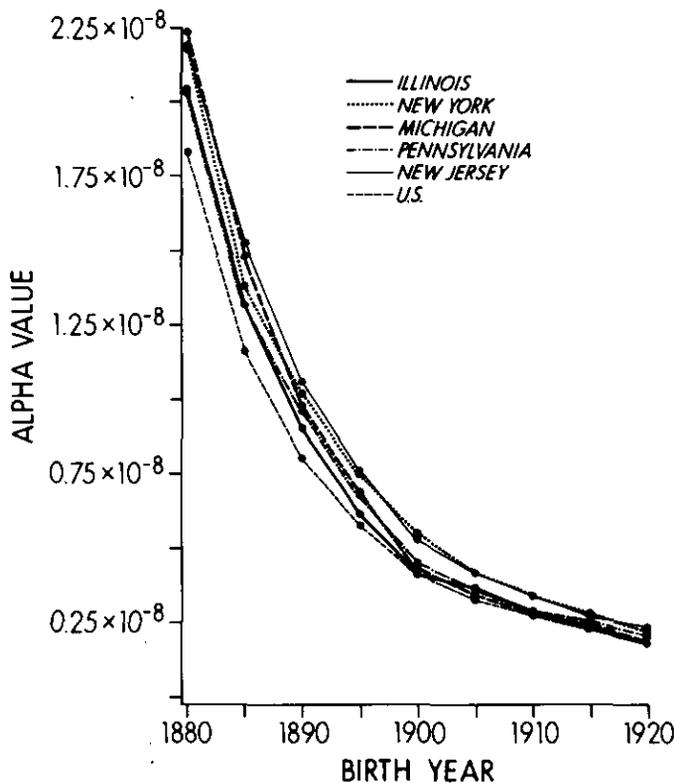


FIGURE 12. Cohort-specific risks of stomach cancer mortality among white males in the United States and among white males in five select states. Absolute levels of risk measured in α units. See text for explanation.

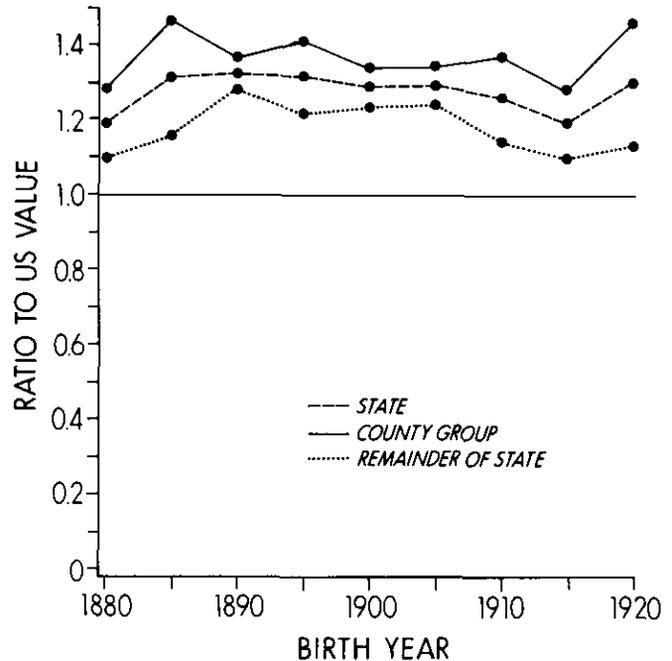


FIGURE 13. Cohort-specific relative risks of stomach cancer mortality among white males in New Jersey and in select counties in New Jersey. Relative risks computed by using cohort specific risks of stomach cancer mortality among all white males in the United States as the standard.

rate over cohort age. Thus, the cohort trends are nearly directly opposite between stomach and respiratory cancer. As with respiratory cancer, we can plot within state variation in cancer risks by casting them as a ratio to the national values. The detrended rates for stomach cancer in New Jersey are presented in Figure 13.

Here we see that the stomach cancer mortality risks, after removing the national trends (i.e., large declines for younger cohorts) are quite stable. That is, the elevation of stomach cancer mortality risks in the target counties (i.e., the metropolitan counties in the northeastern part of the state), vis-à-vis the remainder of the state, do not converge. Nor do they converge with the national values. This implies that the risk factors that are associated with the cohort declines in stomach cancer may operate nationally while those factors generating risk differentials within a state may affect all cohorts in a given area uniformly.

Discussion

In the preceding analyses we have attempted to accomplish two tasks. The first is the presentation of a methodology to describe and evaluate geographic, temporal, and cohort patterns of chronic disease mortality risks using available vital statistics data. The development of such a methodology was undertaken because (a) there exist a number of analyses of geographic differentials in chronic disease risks, (b) these prior analyses have often paid little attention to temporal and cohort patterns, and (c) the time series of chronic dis-

ease mortality risks specific to the county level has become lengthy enough to permit meaningful partial-cohort analyses (i.e., there are now 20 years of data available for all chronic diseases, 1962 to 1981, and 32 years available for malignant neoplasms, 1950 to 1981).

The methodology we propose involves a two-stage analysis of local area differences and trends in chronic disease death rates. The first involves assessing temporal changes in the distribution of chronic disease death rates in the population (by plotting the county-population weighted, age-adjusted death rates for individual counties). This is useful because it forces us to consider the deviation of local areas in terms of the overall distribution of local area death rates. Furthermore, it is useful operationally because a large number of rates can be screened rapidly.

The second stage of the analysis involves the development of models of the age progression of chronic disease mortality and morbidity risks observed in cohorts. These models should be developed on a disease-specific basis and should help us to unconfound age, period, and cohort changes in death rates at the local level. We illustrated this methodology using data on U.S. cancer mortality. We selected cancer because, for U.S. cancer mortality, there exists a lengthy time series of data, there is evidence of significant geographic differentials in risk, and there were available models of the age increase of cancer morbidity and mortality in cohorts.

In addition to illustrating the methodology, our empirical study of the geographic variation of cancer mortality identifies areal and temporal patterns that have not been previously described. These patterns suggest a number of interesting hypotheses about the exposure factors that generated them.

Specifically, in the first stage of our analysis we were able to identify certain overall patterns of change that provided a useful context in which to describe the change of specific local areas. For example, there is an approximate "upper bound" on the death rates for all cancer mortality for white males. That is, the highest rates observed for any local area did not markedly increase from 1950 to 1978. Furthermore, there was a tendency for the variances of these distributions to decrease. Indeed, in several states with large metropolitan areas and large nonmetropolitan populations the distribution of rates was bimodal in 1950-59 with nonmetropolitan counties defining a secondary peak in the density distribution at much lower rates. The discrete secondary peak tends to disappear with time. Such "homogenization" did not seem to occur for females.

The second stage of the analysis involved modeling the cohort variation of cancer mortality risks within county. By using cohort-specific models we could examine the demographic and temporal details of differences in cancer risks between areas. This helped us accomplish two goals. First, by examining the demographic and temporal factors behind local area differences in risk, we can detail information to relate cancer mortality risks to changing exposure factors. Second,

if we determine that estimates of specific parameters of a model of the disease process are significantly different from the national estimates, we can be more confident in identifying the area as one of high risk. For example, we showed that the $\bar{\alpha}$ for each cohort represent an absolute risk level in a type of proportional hazards model. Thus the ratio of two cohort $\bar{\alpha}$ represents an age invariant relative risk. These risk measures permit the risk differentials in cohorts to be identified with exposure differentials. Such cohort models can be adjusted for latency and, under the assumption of independence, for the effects of competing risks. Furthermore, as we showed in several plots, the change in risk over cohorts can be "detrended" by casting them as a ratio to the national values. In this way, local area differences can be identified net of very large and persistent national trends that might otherwise tend to overwhelm local area differences.

By proceeding in this way we have been able to identify major differences in respiratory cancer trends at the local area level. First of all, within certain states, we tended to find large cohort differences in risk between metropolitan and nonmetropolitan areas for the oldest cohorts. These differences tended to disappear with decreasing cohort age. Second, we found major regional differences between the patterns within states. For major northeastern states (e.g., New York, New Jersey) we tended to find that both nonmetropolitan and metropolitan areas were above the national values for older cohorts. These differences tended to disappear for younger cohorts. In southern states, the older cohort rates tended to be lower in both metropolitan and nonmetropolitan counties. These too tended to converge with the national values. It appears that these patterns could be explained by a saturation of respiratory cancer risks in the white male population. Different patterns emerged for white females who were not at such a saturation level of risk. Other patterns would emerge for different cancer types. It appears that such modeling and decomposition of local area cancer death rates could help greatly in explaining local area risk patterns, help in guarding against accepting chance variation in local area death rates as real, and help in generating reasonable hypotheses before intensive field studies are initiated. This appears to be a useful extension of the proposed monitoring phase discussed in Blot et al. (2), i.e., by using detailed models of cancer mortality hazards, one is better able to filter out chance geographic and temporal variation in cancer mortality risks.

Dr. Manton's and Mr. Stallard's research was supported by NIA Grant No. 2-R01-AGO1159 and EPA cooperative agreement No. CR807465. Drs. Creason's and Riggan's research was supported by internal funds of EPA.

REFERENCES

1. Mason, T. J., McKay, J., Hoover, R., Blot, W. J., and Fraumeni, J. F. Atlas of Cancer Mortality for U.S. Counties: 1950-1969.

- DHEW Pub. No. (NIH) 75-780, U.S. Government Printing Office, Washington, DC, 1975.
2. Blot, W. J., Fraumeni, J. F., Mason, T. J., and Hoover, R. N. Developing clues to environmental cancer: a stepwise approach with the use of cancer mortality data. *Environ. Health Perspect.* 32: 53-58 (1979).
 3. Riggan, W. B., Van Bruggen, J., Acquavella, J. F., Beaubier, J., and Mason, T. J. U.S. Cancer Mortality Rates and Trends, 1950-1979, Vol. I-III. EPA 600/1-83-015c, U.S. Government Printing Office, Washington, DC, Sept. 1983.
 4. Manton, K. G., Stallard, E., and Riggan, W. B. Strategies for analyzing ecological health data: models of the biological risk of individuals. *Statistics Med.* 1: 163-181 (1982).
 5. Manton, K. G., and Liu, D. Projecting chronic disease prevalence. *Med. Care* 22: 511-526 (1984).
 6. Polissar, L. The effect of migration on comparison of disease rates in geographic studies in the United States. *Am. J. Epidemiol.* 3: 175-183 (1980).
 7. Armitage, P., and Doll, R. The age distribution of cancer and a multistage theory of carcinogenesis. *Brit. J. Cancer* 8: 1-12 (1954).
 8. Armitage, P., and Doll, R. Stochastic models for carcinogenesis. In: *Proceedings, Fourth Berkeley Symposium on Mathematical Statistics and Probability: Biology and Problems of Health*, Vol. IV (J. Neyman, Ed.), Univ. of California Press, Berkeley, 1961.
 9. Cook, P. J., Doll, R., and Fellingham, S. A. A mathematical model for the age distribution of cancer in man. *J. Cancer* 4: 93-112 (1969).
 10. Manton, K. G., and Stallard, E. Maximum likelihood estimation of a stochastic compartment model of cancer latency: lung cancer mortality among white females in the U.S. *Comput. Biomed. Res.* 12: 313-325 (1979).
 11. Hougaard, P. Life table methods for heterogeneous populations: distributions describing the heterogeneity. *Biometrika* 71: 75-83 (1984).
 12. Manton, K. G., and Stallard, E. A cohort analysis of U.S. stomach cancer mortality 1950-1977. *Int. J. Epidemiol.* 11: 49-61 (1982).
 13. Manton, K. G., Stallard, E., and Vaupel, J. W. Alternative models for the heterogeneity of mortality risks among the aged. *J. Am. Statist. Assoc.*, in review.
 14. Vaupel, J. W., Manton, K. G., and Stallard, E. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* 16: 439-454 (1979).
 15. Chiang, C. L. *Introduction to Stochastic Processes in Biostatistics*. Wiley, New York, 1968.
 16. Moeschberger, M. L., and David, H. A. Life tests under competing causes of failure and the theory of competing risks. *Biometrics* 27: 909-933 (1971).
 17. Gail, M. A review and critique of some models used in competing risk analysis. *Biometrics* 31: 209-222 (1975).
 18. Birnbaum, Z. W. *On the Mathematics of Competing Risks*. DHEW Pub. No. (PHS) 79-1351, U.S. Government Printing Office, Washington, DC, 1979.
 19. Spiegelman, M. *Introduction to Demography*. Harvard Univ. Press, Cambridge, MA, 1969.
 20. Manton, K. G., and Stallard, E. A stochastic compartment model representation of chronic disease dependence: techniques for evaluating parameters of partially unobserved age inhomogeneous stochastic processes. *Theor. Pop. Biol.* 18: 57-75 (1980).
 21. Vaupel, J. W., and Yashin, A. I. *The Deviant Dynamics of Death in Heterogeneous Populations*. International Institute for Applied Systems Analysis, Laxenburg, Austria, RR-83-1 (1983).
 22. Klebba, A. J. *Mortality from Diseases Associated with Smoking: United States, 1960-77*. DHHS Pub. No. (PHS) 82-1854, National Center for Health Statistics, Hayttsville, MD, 1982.
 23. Manton, K. G., and Stallard, E. A population-based model of respiratory cancer incidence, progression, diagnosis, treatment, and mortality. *Comput. Biomed. Res.* 15: 342-360 (1982).
 24. Doll, R., and Peto, R. The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. *J. Natl. Cancer Inst.* 66: 1192-1307 (1981).
 25. Peto, J., Seidman, H., and Selikoff, I. J. Mesothelioma incidence among asbestos workers: implications for models of carcinogenesis and risk assessment calculations. *Brit. J. Cancer* 45: 124-235 (1982).