SMOKING AS A CONFOUNDER IN ECOLOGIC CORRELATIONS OF CANCER MORTALITY RATES WITH AVERAGE COUNTY RADON LEVELS

J. S. Puskin*

Abstract—Cohen has reported a negative correlation between lung cancer mortality and average radon levels by county. In this paper, the correlation of U.S. county mortality rates for various types of cancers during the period 1970–1994 with Cohen’s radon measurements is examined. In general, quantitatively similar, strongly negative correlations are found for cancers strongly linked to cigarette smoking, weaker negative correlations are found for cancers moderately increased by smoking, whereas no such correlation is found for cancers not linked to smoking. The results indicate that the negative trend previously reported for lung cancer can be largely accounted for by a negative correlation between smoking and radon levels across counties. Hence, the observed ecological correlation provides no substantial evidence for a protective effect of low level radon exposure.

Key words: radon; cancer; cigarettes; health effects

INTRODUCTION

Numerous epidemiological studies of underground miners, as well as laboratory studies of exposed rats, clearly demonstrate that inhaled radon progeny cause lung cancer (NRC 1999). The miner studies and most of the animal studies suggest that the excess cancer incidence per unit radon progeny exposure is maximal at low exposure rates. This finding is consistent with a wide range of experimental evidence, including studies of mutagenesis, cell transformation, and carcinogenesis, for an inverse dose rate effect in the case of (high-LET) alpha particle radiation (NCRP 1990; NRC 1988, 1999). Thus, other uncertainties aside, it would be expected that the extrapolation of risk estimates derived from the miner studies to the case of the reduced exposure rates found in homes should not overestimate the risks from residential radon.

Case-control studies, which compare estimated past residential radon exposures of lung cancer cases and controls, are generally consistent with the risk estimates extrapolated from the miner studies (Lubin 1999).

Another approach to the question of risks from residential radon has been advanced by Cohen (Cohen 1990, 1995; Cohen and Colditz 1994). In this “ecological” approach, lung cancer rates by county are plotted against the measured average radon level for that county. Since the risk of lung cancer for individuals is projected to increase linearly with radon exposure, other things being equal from county to county, one might expect a linearly increasing county lung cancer rate with increasing average county radon. Furthermore, exposure-response models derived from the miner cohort studies can be used to predict the slope of the relationship.

However, the radon measurements obtained for each county by Cohen were limited and do not represent a random sample; moreover, due to the mobility of the population and changes in housing characteristics over time, even a true average radon determination for a county in a particular year would not properly reflect the average past exposure of people in that county. Nevertheless, while these types of errors in the estimates of average radon level would arguably reduce the slope of the exposure-response relationship, in the absence of confounding [or complications arising from within county correlations between radon and smoking (Lubin 1998)], one would still expect to obtain a positive slope so long as the measured average radon levels are at least positively correlated with the true average exposures.

Instead, a strong negative correlation between county lung cancer mortality and measured average radon levels was found (Cohen 1990, 1995; Cohen and Colditz 1994). The most obvious explanations for this finding are (1) confounding by a negative association across counties between radon and other risk factors for lung cancer, particularly cigarette smoking, which is a causal factor in the great majority of lung cancer deaths, or (2) a protective effect of alpha particle radiation at low
dose rates. A third explanation has also been proposed, based on the observed synergism between smoking and radon in causing lung cancer in the underground miners coupled with a posited negative association between radon and smoking within counties (Greenland and Robins 1994; Lubin 1998).

In the absence of county-specific data on smoking, Cohen has tested the first hypothesis above by looking at the effect of various indicator variables from county census data on the lung cancer-radon regression. Some of these variables are certainly correlated with smoking, e.g., urban/rural index, county population size, educational level, and income. A quantitative estimate of smoking prevalence in each county was also constructed based on state cigarette sales data and the observed dependence of state lung cancer rates on the fraction of the population living in urban areas (Cohen and Colditz 1994; Cohen 1995). Possible limitations with respect to the radon measurements, the indicator variables, and the estimated smoking parameters have been discussed by Smith et al. (1998). The problem of confounding between smoking and indoor radon has also been addressed by Stidley and Samet (1994) and by Darby and Doll (2000).

Although Cohen has found evidence of confounding by smoking, he concluded that the confounding could not be large enough to explain the negative slope (Cohen 1995, 1998). Therefore, a protective effect of increasing radon over the range of residential exposure levels would seem to be implied. If correct, this would mean that actions to reduce moderately elevated radon levels in homes are misguided. More generally, it would fuel speculation about possible beneficial effects of low level radiation, which, if validated, might have profound implications for the field of radiation protection.

One way to test whether or not there is strong confounding by smoking is to look at the relationship between radon levels and the rates of smoking-related cancers in tissues that receive no significant dose from inhaled radon progeny. Cohen has performed such an analysis, from which he concluded that the correlations of lung cancer rates with radon were much stronger than for other smoking-related cancers (Cohen 1993). Gilbert noted that the analysis showed that other smoking-related cancers also were negatively correlated with radon (Gilbert 1994), but Cohen insisted that the evidence for negative correlations between other smoking-related cancers and radon was weak and that the magnitudes of the correlations were not consistent with what would be expected were the negative trend for lung cancer due to confounding by smoking (Cohen 1994).

In this paper the relationship of county-specific cancer mortality rates with average radon levels, as measured by Cohen, is reexamined in light of more complete data and improved statistical methodology, which properly deals with the very sparse data for some types of cancers in certain counties. A consistent, strongly negative dependence of both lung cancer and other smoking-related cancer rates on radon concentration is found, indicating that the confounding induced by a negative correlation between smoking and radon levels between counties provides a reasonable explanation of Cohen’s inverse relationship between lung cancer rates and radon levels.

**METHODS**

Cohen’s average radon levels \((r_i)\) by county \((i)\) were obtained from the University of Pittsburgh web-site (Cohen 1996). Estimates of smoking prevalence by county were obtained from the same source (Cohen 1996: columns 58 and 59). Data on white male and female county level cancer rates and person-years \((PY_i)\) for the period 1970–1994 were obtained from the National Cancer Institute (Devesa et al. 1999a). Cancer mortality and person-year data were available for 1,585 of Cohen’s 1,601 counties. Depending on sex and type of cancer, data for some counties were sparse, with few, if any, cases recorded during the time period under consideration. In such cases, Poisson variation may account for most of the error in the estimated county-specific rate. To arrive at an improved estimate of the relationship between county level cancer mortality and radon, an iterative procedure was employed, which weights each county in the regression inversely by the estimated variance in the rate for the county due to the combination of the Poisson variance for that county plus a residual variance assumed to be constant across counties (Pocock et al. 1981). Unweighted regressions were also performed, as indicated.

In outline, the weighted regression procedure is as follows. The age-adjusted cancer rate per 100,000 PY (y) is modeled as a linear function of radon level, \(r\):

\[
y = \alpha + \beta r. \quad (1)
\]

An unweighted linear regression of \(y\) on \(r\) is performed on the data from the 1,585 counties to arrive at initial estimates of the intercept and slope \(\hat{\alpha}\) and \(\hat{\beta}\). Initial estimates of the Poisson variance in \(y\) for each county are calculated from eqn (2):

\[
r_i^2 = 10^5(\hat{\alpha} + \hat{\beta} r_i)/PY_i. \quad (2)
\]

The remaining variance not ascribed to Poisson variation is assumed to be the same for all counties. An initial estimate is found by averaging the difference
between the residual sum of squares and the sum of the estimated Poisson variance over all counties:

$$\sigma_0^2 = (1/1.585)[\sum (y_i - \hat{\alpha} - \hat{\beta}r_i)^2 - \sum \tau_i^2].$$  \hspace{1cm} (3)

(If the value on the right hand side is negative, \(\sigma_0^2\) is set equal to zero). An iterative procedure is then followed to arrive at an improved estimate of \(\sigma^2\), and a new regression is performed, where each county, \(i\), is weighted by \((\tau_i^2 + \sigma^2)^{-1}\). Based on the new estimates for the slope and intercept, the \(\tau_i^2\) are recalculated, and the process repeated.

**RESULTS**

Fig. 1 shows a scatter plot of male lung cancer mortality for the time period 1970–1994 vs. average radon concentration. Following the notation of Cohen (1995), radon concentration \((r)\) is measured in units of \(r_0\), equal to 37 Bq m\(^{-3}\) (1 pCi L\(^{-1}\)). As reported by Cohen (Cohen 1990, 1995), the (unweighted) regression line shows a strong negative trend with nearly a 10% decrease in mortality per unit increase in \(r\) relative to the extrapolated mortality at \(r=0\).

Fig. 2 shows a similar plot of mortality from oral cancer (cancer of the oral cavity and pharynx), for which, like lung cancer, smoking is a strong risk factor (Rogot and Murray 1980; U.S. DHHS 1989). The similarity between Figs. 1 and 2 is striking. There is considerably more scatter in the data in Fig. 2, but this is not surprising in view of the much smaller number of oral cancer deaths. What is significant is that the relative falloff in mortality with increasing radon is about the same as for lung cancer. Since the dose from radon decay products to stem cells in the mouth is expected to be minimal, it is impossible to attribute the falloff in oral cancer to radon exposure. Rather, the comparable influence of smoking on lung cancer and oral cancer risks points to a (negative) correlation between smoking and radon as an explanation for the negative slopes seen in both Fig. 1 and Fig. 2. Support for this interpretation is shown in Fig. 3, which summarizes the results of weighted linear regressions of 1970–1994 mortality rates for different types of cancers against radon concentration (see Methods section). Qualitatively similar results were found in all cases based on unweighted regressions. For each sex and type of cancer, the ratio of 100 times the fitted slope \((\beta)\) to the fitted intercept \((\alpha)\) is plotted, which reflects the estimated percentage change in cancer mortality per unit change in radon level relative to the cancer mortality projected at zero radon concentration. A generally consistent picture emerges from this diagram. With the exception of male esophageal cancer, cancers for which the relative risks of smoking are known to be about 7 or higher (cancers of the lung, oral cavity and pharynx, larynx, and esophagus) (Rogot and Murray 1980) all show a large negative slope, with about an 8–10% decrease in rate per unit increase in radon. A similar dependence was also seen for nasopharyngeal cancer, but its link to smoking appears to be weaker than for the other cancers previously mentioned.
(Zhu et al. 1995; Vaughan et al. 1996). Bladder and pancreatic cancers, which have a relative risk for smoking of about 2 (Rogot and Murray 1980), show a smaller but statistically significant negative trend, as does male esophageal cancer. In contrast, cancers not linked to smoking (prostate, colon and breast) show no significant negative trend. Taken together, these data provide compelling evidence for a strong confounding effect of smoking on the regressions of county-level cancer rates on radon.

Cohen attempted to adjust for the effects of smoking confounding by introducing a separate independent variable, $s$, for smoking prevalence (Cohen 1993; Cohen and Colditz 1994). Inclusion of this variable in the analysis had little effect on the observed dependence of mortality on increasing radon for any of the cancers with high relative risks for smoking. This can be seen from the data in Table 1, which compares the results of a simple regression of the 1970–94 gender-specific mortality (deaths per 10^5 PY) for each cancer type on the estimated average radon level for each county.

$$y = \alpha + \beta r,$$

with the corresponding results of a multiple regression on average radon level and smoking prevalence,

$$y = \alpha' + \beta_1 r + \beta_2 s.$$

For this comparison, the regressions were unweighted.

As shown by the boldfaced columns of data in Table 1, the ratio of the radon regression coefficient ($\beta_1$) to the mortality projected for zero radon and average smoking ($\alpha' + \beta_2 s$), based on the multiple regression, differed only slightly from the ratio, $\beta : \alpha$, obtained from the simple regression. Thus, the smoking adjustment had little effect on the negative trends of the smoking-related cancer rates with increasing radon, indicating that the smoking variable $s$ did not substantially reduce the confounding by smoking (or by other risk factors negatively correlated with radon). This would suggest that $s$ is an inadequate measure of county-specific smoking rates—at least as they affect cancer mortality. That conclusion is supported by results showing that Cohen’s smoking variable predicts a relatively small fraction of the observed variation in lung cancer mortality across countries (Smith et al. 1998).

**DISCUSSION**

The results presented here demonstrate a similar strong negative correlation between radon levels and mortality rates for cancers of the lung, oral cavity and esophagus (Smith et al. 1998). Diagnostically it had little effect on the observed dependence of mortality on increasing radon for any of the cancers with high relative risks for smoking. This can be seen from the data in Table 1, which compares the results of a simple regression of the 1970–94 gender-specific mortality (deaths per 10^5 PY) for each cancer type on the estimated average radon level for each county.

$$y = \alpha + \beta r,$$

with the corresponding results of a multiple regression on average radon level and smoking prevalence,

$$y = \alpha' + \beta_1 r + \beta_2 s,$$

For this comparison, the regressions were unweighted.

As shown by the boldfaced columns of data in Table 1, the ratio of the radon regression coefficient ($\beta_1$) to the mortality projected for zero radon and average smoking ($\alpha' + \beta_2 s$), based on the multiple regression, differed only slightly from the ratio, $\beta : \alpha$, obtained from the simple regression. Thus, the smoking adjustment had little effect on the negative trends of the smoking-related cancer rates with increasing radon, indicating that the smoking variable $s$ did not substantially reduce the confounding by smoking (or by other risk factors negatively correlated with radon). This would suggest that $s$ is an inadequate measure of county-specific smoking rates—at least as they affect cancer mortality. That conclusion is supported by results showing that Cohen’s smoking variable predicts a relatively small fraction of the observed variation in lung cancer mortality across countries (Smith et al. 1998).

**DISCUSSION**

The results presented here demonstrate a similar strong negative correlation between radon levels and mortality rates for cancers of the lung, oral cavity and esophagus (Smith et al. 1998). Diagnostically it had little effect on the observed dependence of mortality on increasing radon for any of the cancers with high relative risks for smoking. This can be seen from the data in Table 1, which compares the results of a simple regression of the 1970–94 gender-specific mortality (deaths per 10^5 PY) for each cancer type on the estimated average radon level for each county.

$$y = \alpha + \beta r,$$

with the corresponding results of a multiple regression on average radon level and smoking prevalence,

$$y = \alpha' + \beta_1 r + \beta_2 s,$$

For this comparison, the regressions were unweighted.

As shown by the boldfaced columns of data in Table 1, the ratio of the radon regression coefficient ($\beta_1$) to the mortality projected for zero radon and average smoking ($\alpha' + \beta_2 s$), based on the multiple regression, differed only slightly from the ratio, $\beta : \alpha$, obtained from the simple regression. Thus, the smoking adjustment had little effect on the negative trends of the smoking-related cancer rates with increasing radon, indicating that the smoking variable $s$ did not substantially reduce the confounding by smoking (or by other risk factors negatively correlated with radon). This would suggest that $s$ is an inadequate measure of county-specific smoking rates—at least as they affect cancer mortality. That conclusion is supported by results showing that Cohen’s smoking variable predicts a relatively small fraction of the observed variation in lung cancer mortality across countries (Smith et al. 1998).
pharynx, larynx, and the female esophagus, all of which have a high relative risk for cigarette smoking (Rogot and Murray 1980; US DHHS 1989). A similar negative dependence on radon was observed for nasopharyngeal cancer, which appears to be associated with smoking but less strongly than these other cancers (Nam et al. 1992; Zhu et al. 1995; Vaughan et al. 1996). In this case, additional confounding may arise from a negative association of other risk factors with county radon levels. A smaller, but statistically significant negative trend with increasing radon was also seen with male esophageal cancer, which is strongly associated with smoking. This exception might arise if another risk factor for the disease besides smoking is positively correlated with average radon. No negative trend was found for cancers of the colon, breast, or prostate, which are all cancers not linked to smoking (see Fig. 3).

These findings point either to a strong negative correlation between smoking and county level radon levels or to a protective effect of radon against all smoking-related cancers. Since the radiation doses from inhaled radon decay products are much higher to stem cells in the lung than to those in other organs, the latter interpretation is untenable.

The results here differ substantially from those reported by Cohen (1993). When he performed simple regressions of cancer mortality rates against average radon levels, he found an inconsistent pattern of slopes for the smoking-related cancers.

Table 2 summarizes the estimated slopes ($B$) and associated $t$-values ($t$) for those cancers, as well as the number of counties ($N$) for which data were available to be analyzed, as presented in Cohen (1993). Although $B$ is not defined in that reference, from work published elsewhere (Cohen 1995), it appears that $B$ represents the percentage change in mortality per unit change in radon as defined here.

The differences in $N$ by sex and cancer type are substantial. According to Cohen (1993), these differences reflect the lack of availability of mortality data for certain categories. This is puzzling because the National Cancer Institute data on cancer mortality for 1970–1994 used here (Devesa et al. 1999a) includes 99% of the 1,601 counties for which average radon determinations were provided. The values of $N$ in Table 2 are reduced for categories where the numbers of observed cases are relatively low. Although not stated explicitly in Cohen (1993), it would appear from a later reference (Cohen 1995) that Table 2 reflects 1970–1979 mortality data. From an examination of the National Center for Health Statistics cancer mortality data for that period (D. Grauman, private communication, National Cancer Institute, 6120 Executive Blvd., Executive Plaza South, Bethesda, MD 20892; July 2002), it was determined that $N$ includes only those counties reporting a nonzero number of cases.

Selecting out counties with no observed cases biases the estimated slope, particularly given the negative correlations of both radon level and cancer mortality with population size (Cohen 1990). For example, only 815 of the 1,585 counties reported any mortality from female nasopharynx cancer for the period 1970–1994. An unweighted regression of female nasopharynx cancer mortality against average radon level with all counties included yields a relative slope of $-9.1\%$ (S.E. = 2.5%), but when those counties with zero cases are removed the relative slope is $+9.8\%$ (S.E. = 3.7%). In this paper, the problem of sparse data is partly alleviated by analyzing data for a more extended time period. Moreover, with the weighting scheme outlined in the Methods section, an improved estimate of the slope is obtained, based on the data from all counties, including those where there were no recorded mortality from the cancer in question.

In conclusion, the previous failure to find a consistent negative trend between smoking-related cancer mortality and radon probably results from bias due to data selection.

Cohen argues that the correlation between lung cancer and radon is much stronger than for other types of

<table>
<thead>
<tr>
<th>Cancer type</th>
<th>Males</th>
<th>Females</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N$</td>
<td>$B$</td>
</tr>
<tr>
<td>Lung</td>
<td>1,592</td>
<td>-7.3</td>
</tr>
<tr>
<td>Oral cavity, tongue</td>
<td>1,478</td>
<td>-7.3</td>
</tr>
<tr>
<td>Nasopharynx</td>
<td>746</td>
<td>7.8</td>
</tr>
<tr>
<td>Esophagus</td>
<td>1,479</td>
<td>3.2</td>
</tr>
<tr>
<td>Larynx</td>
<td>1,400</td>
<td>4.6</td>
</tr>
<tr>
<td>Pancreas</td>
<td>1,577</td>
<td>2.8</td>
</tr>
<tr>
<td>Bladder, urinary organs</td>
<td>1,559</td>
<td>1.4</td>
</tr>
</tbody>
</table>
cancer including, e.g., oral cancer, which had a negative slope similar to that for lung cancer. His argument is based on a comparison of the proportion of total variance about the mean explained by the regression ($R^2$) and the statistical significance ($t$-values) of the negative trends for the two types of cancers. Such a comparison is largely irrelevant in the context of interest here. Of all the smoking-related cancers, lung cancer is the most common; consequently, the sampling errors are smaller than for the other cancers. For example, a comparison of Figs. 1 and 2 clearly shows a larger degree of scatter for oral cancer than for lung cancer. Thus, although oral cancer mortality exhibits as great a relative fall-off with increasing radon as does lung cancer mortality, a smaller fraction of the variance in oral cancer is accounted for by the linear trend, and a larger fraction by random error. The larger scatter will reduce $R^2$ and $t$, but this says nothing about the strength of the negative trend as it relates to the inverse correlation between smoking and radon.

The evidence here indicates a strong negative correlation between smoking and measured radon level, which will distort the relationship between radon level and smoking-related disease mortality. Nevertheless, one might ask whether or not the negative slope for lung cancer can be fully accounted for by confounding while still maintaining current risk estimates for residential radon exposures. Depending on which of its preferred models is utilized, the BEIR VI Committee projects 10% or 14% of all lung cancer deaths in the U.S. attributable to residential radon (NRC 1999). This projection assumes a mean radon concentration in homes of 46.2 Bq m$^{-3}$ (1.25 pCi L$^{-1}$) (NRC 1999; Marcinowski et al. 1994). Thus, it might be estimated that, in the absence of confounding, radon-induced lung cancers should increase by about 7–11% per unit increase in radon concentration.

In view of the expected contribution from radon-induced lung cancers, why is the trend for lung cancer mortality vs. measured average county radon concentration about as strongly negative as for other smoking-related cancers, for which radon is not believed to be a causal factor? Several possibilities present themselves, some of which may be acting in combination.

First, the confounding by smoking may have a stronger influence on the lung cancer regressions than on those for other smoking-related cancers. The change in slope for each type of cancer will increase with the attributable fraction of the mortality that is smoking related. The attributable fraction due to smoking in a particular county and time period is a complex function of the characteristics of the population, especially the pattern of their past smoking histories. This complexity is reflected in the continually changing smoking-related relative risks observed for different types of cancer (U.S. DHHS 1989). Moreover, other risk factors (e.g., occupational and environmental exposures to chemical carcinogens, use of alcohol or tobacco products other than cigarettes) may themselves be statistically associated with radon level, which can change the observed slope for a particular cancer either up or down. As noted above, such confounding may explain why the negative dependence of mortality on increasing radon for male esophageal cancer was less steep than for other cancers known to have a high relative risk for cigarette smoking or why the negative trend for nasopharyngeal cancer was steeper than for other cancers moderately affected by smoking.

Second, the measured average radon levels by county may be an inadequate surrogate for past radon exposure, given population mobility, changes in the housing stock over time, and the possible non-representativeness of the residences for which radon determinations were made. These factors are likely to mask a positive effect of radon on lung cancer risk, whereas their influence on the degree of association between smoking and measured average radon levels is unpredictable.

Third, due to the positive synergism between radon and smoking in causing lung cancer, a negative association of radon level and smoking within counties may act to make the slope of the lung cancer vs. average radon curve more negative (Lubin 1998).

The results here provide compelling evidence of an inverse correlation between tobacco use and measured average radon levels by county. Some of this association appears to be related to population density and other components of urban/rural differences (Cohen 1990; Goldsmith 1999), with historically higher tobacco use and lower radon being found in urban areas. Other specific regional variations are also likely to be important. For example, smoking related disease is known to be low in Utah due to its high Mormon population, and an inspection of Cohen’s data file (Cohen 1996) shows that radon levels in the counties of Utah are substantially above average. Conversely, smoking related cancer rates are exceptionally high in Louisiana, but radon levels there are very low.

Profound demographic and regional shifts in smoking patterns have occurred over the past 50 y (Devesa et al. 1999b) and are likely to continue. Such shifts may significantly alter the correlation between radon and cancer mortality in the future.

**CONCLUSION**

The strong negative associations between mortality and average radon found for other smoking-related cancers indicate that the negative association observed for
l lung cancer can be explained in terms of confounding by smoking without invoking any kind of hormetic effect of low level radiation exposure. Given the uncertainty in the magnitude of the confounding and the other factors that can distort the association between lung cancer and measured average radon concentration, it is not possible to estimate residential radon risks from the ecological data. Current evidence from residential case-control studies, however, is supportive of the BEIR VI model estimates for residential exposures based on the epidemiological studies of underground miners (NRC 1999; Lubin 1999).

Acknowledgments—This work was performed while the author was on detail from the Environmental Protection Agency to the National Cancer Institute. The author gratefully acknowledges statistical advice and useful comments on the manuscript from Ethel Gilbert, Charles Land, and Jay Lubin of NCI, and David Pawel of EPA.

REFERENCES