Parallel Analyses of Individual and Ecologic Data on Residential Radon, Cofactors, and Lung Cancer in Sweden

Frédéric Lagarde1 and Göran Pershagen1,2

Divergent results from ecologic and analytic studies on residential radon and lung cancer have created uncertainty in terms of risk assessment. The authors performed concurrent analyses on individual and aggregated data from the nationwide case-control study of residential radon and lung cancer in Sweden. For data aggregated on the county level, the ecologic excess relative risk estimates per 100 Bq/m3 residential radon concentration ranged from -0.03 (95% confidence interval (CI) -0.21 to 0.15) to 0.00 (95% CI -0.21 to 0.21) with different adjustment for cofactors. For individual-level data, the average within-county excess relative risk estimates ranged from 0.07 (95% CI -0.01 to 0.15) to 0.11 (95% CI 0.01-0.27) with similar adjustment. Effect modification by differential county-level radon-smoking correlations appeared insufficient as an explanation for ecologic bias in the Swedish data. On the other hand, adjustment for latitude led to congruence between the two levels of analysis. The results confirm that ecologic studies may be misleading in studies of weak associations, even when major risk factors are accounted for. The large impact of latitude may be unique to Sweden and due to the correlation of latitude both with residential radon and other determinants of lung cancer risk. Am J Epidemiol 1999;149:268-74.

Combined analyses of case-control studies in the general population as well as extrapolation of results from cohort studies among miners support a positive association between radon exposure and lung cancer (1-3). Concurrently, data have accrued which indicate a negative slope from the regression of lung cancer rates on mean radon level in US counties, and these data now indicate a departure of about ten standard deviations from the null hypothesis of no radon effect (4-6). Issues of methodological limitations in the ecologic approach have been raised anew and have provided new insights into the difficulties of using aggregated data for risk assessment at the individual level (7-14). However, this negative ecologic association between radon and lung cancer, persistently put to the fore (6), may receive inordinate emphasis in setting policies for indoor radon and may even be misinterpreted as counterevidence (15, 16).

Indications regarding the magnitude of bias in ecologic studies have been obtained by comparing risk estimates from ecologic and analytic studies for several outcome events (17). Such evaluations were usually based on different data sets, but there are exceptions (18-23). Additional empirical comparisons may be useful (24, 25). In particular, no direct assessment has been made of the difference that may result from individual and aggregate level analyses of data on residential radon and lung cancer. The large case-control study of residential radon and lung cancer in Sweden, which reported a statistically significant individual-level association (26), is most appropriate for this purpose.

**MATERIALS AND METHODS**

The study was of case-control design and included 1,360 lung cancer cases aged 35-74 years, diagnosed from 1980 to 1984, as well as 2,847 population controls. For eligibility, subjects had to be resident at the time of selection in one of 109 municipalities that were considered to be areas likely to present either elevated or low radon levels in dwellings, and they had to have lived in Sweden on January 1, 1947. Incident lung cancer cases were age-matched, within 5-year age group intervals, to controls randomly sampled from the communities in the study base. Radon was monitored over...
a period of 3 months during the heating season in 8,992 dwellings occupied by study subjects for at least 2 years since 1947. Two rooms per dwelling were measured using solid-state alpha track detectors processed at the Swedish Radiation Protection Institute (26). Information on smoking habits and other risk factors for lung cancer were obtained from study subjects or next-of-kin through questionnaires supplemented with telephone interviews. More details are provided in prior publications (26, 27).

Individual time-weighted average radon concentrations had been previously computed as a weighted sum of measured radon levels, using as weights the relative length of residential times in relation to the total residential time with available measurements (26, 27). In the present analyses, individual time-weighted average radon concentrations among controls were averaged within counties to build an aggregate-level variable for radon. County affiliation was determined by the county where an individual had the longest total time of residence within the time frame of the study. In general, county-level aggregation of other covariates used for adjustment was not performed to avoid introducing further differences in the comparison between ecologic and individual-level risk estimates, besides the difference in the level of assessment for radon concentrations. Therefore, adjustment for urbanization, occupation, sex, and age categories (5-year intervals) was individually based according to previously defined classifications (26). However, individual- and aggregate-level adjustments for smoking were compared using either individual time-weighted average tobacco consumption or sex-specific county level averages as alternative model covariates. Latitude was included in some analyses, and refers to the latitude of the largest population center in each county.

Conditional linear excess relative risk regression analyses were performed for different levels of stratification. Likelihood-based confidence intervals were used for statistically significant and positive linear excess relative risk estimates. In other situations, the iteration procedure may not converge, and the ordinary Wald-based confidence limits were reported.

Results from aggregate-level analyses may be biased because the usual models would only fit the situation where radon and smoking are independent within counties or their joint effect is additive, i.e., radon-smoking correlations within strata may bias the aggregate-level analysis (16). Therefore, to assess the extent to which the discrepancy between the two levels of analysis is related to the inadequacy of the aggregate-level analysis to account jointly for synergy and for within-group correlations between radon and smoking, we fitted an extension to multiple smoking categories of the model described by Lubin (16) for the situation where radon and smoking act multiplicatively and are correlated within counties.

In this model, the disease rate \( R_i \) in the \( k \)th county may be written:

\[
R_i = \alpha \Sigma \theta_a n_i^{-1}(1 + \beta \Sigma \theta_a w_a n_i^{-1}/\Sigma \theta_a n_i^{-1}) = \alpha \theta_a \{1 + \beta \text{avg}(\theta_a w_a)/\theta_a\},
\]

where \( \alpha \) denotes a common background rate across counties after adjustment for covariates, \( \theta_a \) is the smoking-related relative risk associated with the smoking category for the \( k \)th individual in the \( k \)th county, \( w_a \) is the individual's radon exposure, \( \theta_a = \text{avg}(\theta_a) \), \( n_i \) is the number of individuals in the \( k \)th county, and \( \beta \) is the radon-related excess relative risk.

For dichotomous smoking categories, this reduces to

\[
R_i = \alpha(p_{a\theta} + p_{a\theta})\{1 + \beta(p_{a\theta} w_{a\theta} + p_{a\theta} \theta w_{a\theta})/(p_{a\theta} + p_{a\theta})\},
\]

as in Lubin (16), where \( \theta \) denotes the relative risk in smokers compared with nonsmokers, \( p_{a\theta} \) and \( p_{a\theta} \) are the proportions of nonsmokers and smokers, respectively, and \( w_{a\theta} \) and \( w_{a\theta} \) are the respective average radon levels for nonsmokers and smokers in the county with index \( k \).

RESULTS

An example of the divergence between within-county and county-level associations is shown in tabular form in table 1, which is based on data from the subgroup of male smokers which includes the largest number of lung cancer cases. In addition to restriction concerning sex and smoking status to mitigate confounding, only those 14 out of a total of 24 counties were included for which both the numbers of male smoker cases and controls was larger than 10, because small numbers tend to induce more extreme values for statistics. The table shows that average radon levels within counties were generally larger for cases than for controls, which suggests a positive association between lung cancer and radon exposure. This contrasts with the apparent trend of increasing lung cancer odds ratios with decreasing county-mean radon levels. Similarly, the overall average radon exposure among counties was larger than among controls (111 Bq m\(^{-3}\) vs. 94 Bq m\(^{-3}\)), while the county-mean radon levels among counties with larger odds for lung cancer (equal to or above the overall odds) were smaller on average than among counties with lower lung cancer odds (82 Bq m\(^{-1}\) vs. 108 Bq m\(^{-1}\)). Thus, there appears to be a negative ecologic association between county-mean radon levels and county odds for lung cancer despite the positive association within counties.
TABLE 1. County-specific lung cancer odds ratios and average radon levels by disease status, among male smokers in the nationwide Swedish case-control study of residential radon and lung cancer, 1980–1984

<table>
<thead>
<tr>
<th>Odds ratio*</th>
<th>Average radon levels (Bq m⁻³)</th>
<th>No. of subjects‡</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Controls</td>
<td>Cases</td>
</tr>
<tr>
<td>0.64</td>
<td>146</td>
<td>183</td>
</tr>
<tr>
<td>0.73</td>
<td>133</td>
<td>184</td>
</tr>
<tr>
<td>0.78</td>
<td>66</td>
<td>114</td>
</tr>
<tr>
<td>0.93</td>
<td>127</td>
<td>144</td>
</tr>
<tr>
<td>0.97</td>
<td>107</td>
<td>234</td>
</tr>
<tr>
<td>0.97</td>
<td>96</td>
<td>104</td>
</tr>
<tr>
<td>1.01</td>
<td>65</td>
<td>58</td>
</tr>
<tr>
<td>1.14</td>
<td>82</td>
<td>115</td>
</tr>
<tr>
<td>1.20</td>
<td>118</td>
<td>102</td>
</tr>
<tr>
<td>1.37</td>
<td>94</td>
<td>75</td>
</tr>
<tr>
<td>1.41</td>
<td>91</td>
<td>187</td>
</tr>
<tr>
<td>1.54</td>
<td>87</td>
<td>85</td>
</tr>
<tr>
<td>1.70</td>
<td>59</td>
<td>59</td>
</tr>
<tr>
<td>1.71</td>
<td>70</td>
<td>88</td>
</tr>
</tbody>
</table>

* Relative to the overall odds and ranked in increasing order.
† Overall means in controls and cases were 94 Bq m⁻³ and 111 Bq m⁻³, respectively, among male smokers.
‡ Only counties with at least 10 subjects among cases and controls are included.

Table 2 compares individual- and aggregate-level risk estimates and shows that analyses at the county-level generally resulted in a negative trend despite increasing degrees of adjustment for cofactors, except when latitude was included in the model. The estimated decrease in risk of lung cancer was about 5 percent per degree of latitude \( e^{\theta} = 0.95, 95\% \text{ confidence interval (CI) 0.91–0.99} \). Using a linear gradient rather than a log-linear gradient was inconsequential. By contrast, there was a tendency for county-mean radon levels to increase with latitude, as shown in figure 1. These opposite trends explain why adjustment for latitude was instrumental in removing negative confounding from the ecologic risk estimate. Addition of longitude among model covariates did not affect the risk estimates.

We assessed the importance of the bias in the aggregate-level analysis that may be induced by radon-smoking correlations within strata, in the presence of synergy between radon and smoking. However, the aggregated-level excess relative risk estimate was left unchanged when regression analysis was based on adjusted county-mean radon levels accounting for the joint distribution of radon and smoking within counties and multiplicative interaction (e.g., when further adjusted for age, sex, urbanization, and occupation, the excess relative risk estimate was equal to -0.02 per 100 Bq m⁻³, 95\% CI -0.19 to 0.16). To shed further light on this issue, we plotted radon-smoking correlations against county-mean radon levels (figure 2).

Table 2 compares individual- and aggregate-level estimates of excess relative risks (RR) for lung cancer per 100 Bq m⁻³ residential radon and 95\% confidence intervals (CI), according to different adjustment for covariates, in the nationwide Swedish case-control study of residential radon and lung cancer, 1980–1984

<table>
<thead>
<tr>
<th>Covariates included for adjustment</th>
<th>Individual†</th>
<th>95% CI</th>
<th>Aggregated within county‡</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>In addition to stratification for age and sex*:</td>
<td>RR</td>
<td>95% CI</td>
<td>RR</td>
<td>95% CI</td>
</tr>
<tr>
<td>None</td>
<td>0.05</td>
<td>-0.02 to 0.11</td>
<td>-0.03</td>
<td>-0.21 to 0.15</td>
</tr>
<tr>
<td>Smoking (aggregated)</td>
<td>0.05</td>
<td>-0.02 to 0.11</td>
<td>-0.02</td>
<td>-0.21 to 0.17</td>
</tr>
<tr>
<td>Smoking (individual)</td>
<td>0.05</td>
<td>-0.02 to 0.13</td>
<td>-0.01</td>
<td>-0.21 to 0.20</td>
</tr>
<tr>
<td>Latitude</td>
<td>0.06</td>
<td>-0.01 to 0.13</td>
<td>0.04</td>
<td>-0.19 to 0.27</td>
</tr>
<tr>
<td>Latitude, smoking (aggregated)</td>
<td>0.06</td>
<td>-0.01 to 0.13</td>
<td>0.07</td>
<td>-0.18 to 0.32</td>
</tr>
<tr>
<td>Latitude, smoking (individual)</td>
<td>0.07</td>
<td>-0.01 to 0.15</td>
<td>0.12</td>
<td>-0.17 to 0.41</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>In addition to stratification for age, sex, urbanization, and occupation*:</th>
<th>RR</th>
<th>95% CI</th>
<th>RR</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>0.07</td>
<td>-0.01 to 0.15</td>
<td>-0.02</td>
<td>-0.21 to 0.17</td>
</tr>
<tr>
<td>Smoking (aggregated)</td>
<td>0.07</td>
<td>-0.01 to 0.15</td>
<td>-0.02</td>
<td>-0.21 to 0.17</td>
</tr>
<tr>
<td>Smoking (individual)</td>
<td>0.08</td>
<td>0.00 to 0.02</td>
<td>0.00</td>
<td>-0.21 to 0.21</td>
</tr>
<tr>
<td>Latitude</td>
<td>0.09</td>
<td>0.01 to 0.20</td>
<td>0.05</td>
<td>-0.19 to 0.29</td>
</tr>
<tr>
<td>Latitude, smoking (aggregated)</td>
<td>0.09</td>
<td>0.01 to 0.20</td>
<td>0.05</td>
<td>-0.20 to 0.29</td>
</tr>
<tr>
<td>Latitude, smoking (individual)</td>
<td>0.11</td>
<td>0.02 to 0.24</td>
<td>0.14</td>
<td>-0.17 to 0.45</td>
</tr>
</tbody>
</table>

* Using individual-level data.
† Ranging from 0 Bq m⁻³ to 3,613 Bq m⁻³, with a standard deviation of 151 Bq m⁻³.
‡ Ranging from 17 Bq m⁻³ to 178 Bq m⁻³, with a standard deviation of 35 Bq m⁻³.

Am J Epidemiol Vol. 149, No. 3, 1999
radon level was apparent. In addition, although within-county correlations tended to be stronger than the overall correlation (-0.03), the correlations were rather small and mainly contained between -0.1 to 0.1, with negative correlations prevailing.

The overall individual-level risk estimate, unstratified for counties, is an implicit combination of the average within-county and between-county estimates and is not rigorously valid as an estimate of the individual-level effect if these differ. In table 3, the overall estimate can be seen to underestimate the average within-county estimate due to the negative bias in the county-level estimate. The overall estimate is approximately equal to an inverse variance weighted sum of the within- and between-county risk estimates. Thus, the lack of precision of the between-county risk estimate contributes to limit the impact of negative ecologic bias on the overall risk estimate unadjusted for county.
**DISCUSSION**

With the use of data on residential radon and lung cancer in Sweden, we assessed whether aggregate-level and individual-level analyses were congruent. Simple tabular presentation of data as well as regression analyses show that the direction of the association according to the aggregated data may be opposite to the one given by the underlying individual-level data. The excess relative risk estimates appeared quite different, although the wide confidence intervals for the ecologic estimates include the individual-level estimates. This lack of precision is not unexpected. When studying environmental exposures associated with small excess relative risks, careful attention must be given to the variability of exposure. Ecologic studies may be efficient when the exposures of interest are clustered into areas and the variation between areas thereby much larger than within areas (28). However, it is not unusual for the between-area variation in exposure to be much smaller than the within-area variation as exhibited here by the restricted range of variation for county-mean radon levels.

Under certain assumptions, an aggregate-level analysis may provide valid estimates for the corresponding individual-level parameters of interest. For a unifactorial linear exposure-disease relation, aggregate- and individual-level estimations of the slope may be theoretically expected to provide similar results (7, 17, 29), but further assumptions are necessary in multifactorial situations, particularly in respect to effect modification by covariates (12, 13, 16, 17, 24, 30, 31). If individual-level correlations between radon and cofactors that interact with radon differ between counties, ecologic bias may be induced by differences in synergistic effects across counties. For example, if the effects of smoking and radon are synergistic and the effect of radon alone is relatively small in comparison with the effect of smoking, the effect of radon on the lung cancer rates within ecologic analysis units is determined to a much larger extent by the joint distribution of radon and smoking within these units than by the prevalence of radon itself. Group-level adjustment for smoking, or inclusion of interaction terms between area-specific averages for radon and smoking, would leave this issue unresolved, because the relevant information concerns the joint distribution of smoking and radon within counties (8, 9, 12, 16, 24, 30, 31). For the present data set, this source of bias appeared to be limited, in spite of a close to multiplicative interaction between residential radon and smoking in relation to lung cancer (26). This can be explained by the small range of radon-smoking correlations coupled to the lack of covariation between correlation sizes and county-mean radon levels.

On the other hand, the disparity between the ecologic and individual risk estimates was largely removed when we accounted for latitude. The geographic gradient displayed by county-mean radon levels may be shared by extraneous risk factors, causing ecologic bias. The geographic distribution of environmental exposures is rarely random and a recurrent problem in ecologic studies is the likely covariation with the distribution of many other risk factors not included in the study (32). For example, in a comparison of ecologic and individual-level risk estimates for different pathologies in relation to smoking, the ecologic relative risk estimate for bladder cancer in France, which
exceeded the estimates obtained from case-control studies in several countries, was modified by adjustment for the latitude and longitude of the geographic units (17). Assessment of the robustness of the model for inclusion of latitude and longitude was also considered important in a study of cardiovascular disease mortality and water hardness for different areas in Great Britain, in view of the general northwest to southeast gradient exhibited by both water hardness and cardiovascular mortality (33). A further example is given by an ecologic study in the United States where cadmium levels in water and bowel cancer tended to be high in the northeast and lower in the mid-south. However, this spatial distribution was shared by industrial activity, increasing the likelihood for spurious ecologic associations between many environmental variables and bowel cancer (34). The impact of latitude in the present ecologic analyses may be due to the correlation of latitude with radon and with other determinants of lung cancer risk.

Appropriate ecologic-level adjustment for confounding may suffer from specification of models on the basis of a priori information from individual-level studies. Factors relevant for unbiased group-level analyses may be different from those that confound individual-level effects (30, 35). For example, in a study of the association between the proportion of woodworkers and lung cancer in Swiss cantons, smoking may induce confounding if it is heterogeneous across cantons and thus potentially correlated with the proportion of woodworkers, even if it is not associated with woodworking at the individual level within cantons (36). Associations between environmental variables tend to be stronger and more fortuitous at the group level than at the individual level, making ecologic risk estimates more vulnerable to confounding (21, 35, 37, 38).

The present study shows that, in practice, a negative ecologic association between radon and lung cancer may coexist with an underlying positive individual-level association. Results from aggregate-level analyses are thus potentially misleading in respect to individual-level associations. This strengthens the view that ecologic studies of residential radon and lung cancer should be given little weight in pondering the overall evidence (15, 16, 39–41). The particulars of each ecologic study grant a degree of specificity for sources of bias, and the identification of sources of bias in one study may be of little help in another situation. Although not apparent in our study, the potentially large impact of small differences in the joint distribution for radon and smoking within counties may be consequential for bias in previous ecologic radon studies. This was not addressed in any of these studies, nor similar considerations that may apply to the joint distribution for radon and other risk factors (16). Unfortunately, by design, individual-level data are not available in ecologic studies which may have to rely on Monte Carlo simulations to assess the sensitivity of their results to different sources of bias. The output of simulations depends on modeling assumptions, and sources of bias that are not known a priori are therefore not modeled. An improvement may be the integration of information from samples of individual-level data from each ecologic unit within a modified aggregate-level analysis (28, 38–40, 42). Another useful allocation of resources may be the examination of individual-level associations within one or more ecologic units, preferably those most influential in the ecologic regression (21).

It has been little recognized that some of the issues of confounding in ecologic studies may also be of concern for individual-level studies with geographically wide study bases. This follows from the observation that overall risk estimates unstratified for areas are implicit combinations of within- and between-area effect estimates (23). It may thus be useful to compare results with and without stratification for geographic divisions, which may also help to identify unrecognized confounders.

ACKNOWLEDGMENTS

This research was supported by grants from the Swedish Cancer Society and the Swedish Radiation Protection Institute. The nationwide Swedish study on residential radon and lung cancer constituted a collaboration between the Institute of Environmental Medicine at Karolinska Institute, the Department of Oncology at Umeå University Hospital, the Department of Occupational and Environmental Health at Linköping University Hospital, and the Swedish Radiation Protection Institute.

REFERENCES


