INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF RADIOLOGICAL PROTECTION

J. Radiol. Prot. 22 (2002) 305–315 www.iop.org/Journals/jr

LETTERS TO THE EDITOR

Response to 'The potential for bias in Cohen's ecological analysis of lung cancer and residential radon'

Dear Sir

I am writing in response to the critique by Lubin [1] of my test of the linear no-threshold theory (hereafter LNT). That paper has been cited as providing an explanation for the discrepancy between my observational findings and the predictions of LNT, thus negating my test. Therefore, to stay within the length restrictions on letters to the editor, I limit my comments to addressing that issue.

In the simplified form adopted by Lubin, my procedure [2] starts with the lung cancer mortality risk for an individual male, m', expressed in the BEIR-IV version [3] of LNT as

$$m'_s = a_s(1 + br') \qquad \text{smokers}$$

$$m'_n = a_n(1 + br') \qquad \text{non smokers}$$

where r' is his radon exposure, and a and b are constants given in BEIR-IV. Mathematically summing the experiences of all males in a county, leads to

$$m = [S\,a_s + (1 - S)a_n](1 + br)$$

where m is the county lung cancer mortality rate, r is the average radon exposure in the county, and S is the fraction of adult males in the county that smoke cigarettes. Defining the important symbol R as

$$R = a_s/a_n$$

and applying minor corrections, this may be expressed as

$$M = m/a_n[RS + (1 - S)] = A + Br \tag{1}$$

where M is defined by the equation on the left, and A and B are constants, with A close to 1.0 and, according to BEIR-IV

$$B = +7 \qquad \text{LNT theory} \tag{2}$$

in units of % per 37 Bq m^{-3} (% per pCi/L).

Data on M and r are available for 1601 US counties and I use them to test the LNT theory by fitting them to equation (1) to determine an observational value of B. The result is

$$B = -7.3\ (\pm 0.56) \qquad \text{observation} \tag{3}$$

which is discrepant with the LNT-theory prediction, (2), by 26 standard deviations.

The 'scientific method' requires that theories be tested, where possible, with observational data, and that any discrepancies between observation and predictions of the theory must be given at least a possible explanation that is not implausible; failing that, the theory is invalid. In a series of papers, I have attempted to find such a possible plausible explanation for the discrepancy between (2) and (3), including evaluations of suggestions offered by others, but none of these explanations has been successful.

0952-4746/02/030305+11$30.00 © 2002 IOP Publishing Ltd Printed in the UK

Material may be protected by copyright law (Title 17, U.S. Code)

Lubin's recent suggestion [1] is that the discrepancy can be explained by systematic differences in average radon exposure to smokers, r_s, and non-smokers, r_n, within each county. I previously treated this general suggestion [4] by defining the effective average radon exposure, r_e, as

$$r_e = [RS\,r_s + (1 - S)r_n]/[RS + (1 - S)]$$

where the two terms in the numerator are the weighted radon exposures to smokers and non-smokers, and the denominator is the sum of the weightings. This differs from the measured county average radon exposure, r

$$r = S\,r_s + (1 - S)r_n.$$

If r_s and r_n are not equal, r_e instead of r should be used to fit the data with equation (1), which would change the observational value of B, and thus hopefully eliminate the discrepancy between (2) and (3). Up to this point, these are the same as the procedures used by Lubin [1].

I defined [4]

$$x = r_s/r_n$$

and treated as variable parameters the national average value of x, the width of the distribution of x-values among US counties, and the correlation between x and r. I explored the effects of varying these parameters within and well beyond their plausible ranges, but the maximal effect was to change the observational value of B from $B = -7.3$ to $B = -4.3$, still very discrepant with the LNT theory value, $B = +7$. I concluded that possible systematic differences between radon exposures to smokers and non-smokers could not explain the discrepancy between (2) and (3).

Lubin's paper [1] that I am responding to here demonstrates that the discrepancy between (2) and (3) can be explained if both R and x vary in a systematic way with r. In particular, this requires that R decreases in a systematic way with increasing r to 65% of its value at low r, and that x varies with r in a very complex way, some samples of which are in table 1, derived from figure 2 of [1].

Table 1. Variation of x with r.

r	30	50	75	90	110	150	200
x	7.0	2.7	1.0	2.2	1.0	0.36	0.27

The behaviour for $r < 30$, which includes about 20% of our data, is even more complex, but I could not read the plot with sufficient accuracy to list values here.

Lubin's paper [1] does not address the issue of plausibility. But it is extremely important to recognise that the 'scientific method' requires that an explanation be found for the discrepancy between predictions of the theory and observation that is *not implausible*. We therefore now consider the plausibility of Lubin's proposed explanation.

Variation of R with r

The quantity R is the ratio of lung cancer risks for smokers and non-smokers, irrespective of their exposure to radon. It is applicable to individuals. According to Lubin's proposal, the ratio of these risks is lower if they live in a county with high average radon levels than if they live in a county with low average radon levels, even if they have the same personal exposure to radon, or even if they have no personal exposure to radon. Surely an individual's risk depends on his personal exposure to radon, not on the average exposure to people in his county of residence. The variation of R with r proposed in [1] is therefore completely implausible.

Variation of x with r

Average radon exposure in a county, r, depends on geological factors, house characteristics, and ventilation practices (e.g. window opening). The only way I can imagine these being systematically different for smokers and non-smokers, thus affecting the value of x, is in ventilation practices. But it seems inconceivable that these can vary systematically and very sharply over a very wide range with average radon exposure in the county, especially since these average radon levels were not known to the people controlling the ventilation. The systematic variation of x with r illustrated in table 1 is surely completely implausible.

Importance of plausibility

Aside from its importance in applying the scientific method, plausibility is of the utmost importance in all epidemiological studies. For example, the choice of confounding factors to be considered in a study is based completely on plausibility considerations; a plausible confounding factor that is not considered can easily invalidate any such study. It is therefore difficult to understand how such wild violations of plausibility could have escaped consideration in [1].

Yours faithfully,

B L Cohen

References

[1] Lubin J H 2002 The potential for bias in Cohen's ecological analysis of lung cancer and residential radon *J. Radiol. Prot.* **22** 141–8

[2] Cohen B L 1995 Test of the linear-no threshold theory of radiation carcinogenesis for inhaled radon decay products *Health Phys.* **68** 157–74

[3] National Academy of Sciences Committee on Biological Effects of Ionizing Radiation 1988 *Health Risks of Radon and Other Internally Deposited Alpha Emitters (BEIR-IV)* (Washington, DC: National Academy Press)

[4] Cohen B L 1998 Response to Lubin's proposed explanations of our discrepancy *Health Phys.* **75** 18–22

Reply to Cohen's letter on 'The potential for bias in Cohen's ecological analysis of lung cancer and residential radon'

Dear Sir

In the above letter, Cohen [1] repeats arguments that he as made in the past; however, the thrust of those comments indicates a fundamental misunderstanding of the point of my article [2] and indeed of my prior articles on this topic. It was not and has never been my goal to identify a specific factor (or several factors) that induced the negative correlation between the estimates of county-level mean radon concentration and county lung cancer mortality rates. Other investigators have offered powerful examples, both theoretical and practical, that Cohen's results are indeed an artifact of ecological regression [3–9]. The primary purpose of

my contributions to this topic has been to demonstrate the fundamental deficiencies of the methodology itself, in particular the unboundedness of the bias [10]. The current article [2] demonstrates that the ecological fallacy always applies, that the addition of county-level adjustment variables does not reduce bias and increase validity, and that an observed ecological risk pattern can differ markedly from the true risk pattern. Because of these deficiencies, epidemiologists have never used ecological regression as a tool for confirmatory analysis. For radon and lung cancer, results of 25–30 analytic studies of individuals clearly prove the deficiency of the method.

There are numerous risks factors for lung cancer, including smoking, age, various occupational exposures, air pollution, previous lung diseases, and so on. In [2], I take an extremely simplified model for lung cancer in radon and smoking status, and demonstrate that the induced county-level model is non-linear. A more realistic, and complex, model for lung cancer would include all principal risk factors, and would also generate a non-linear model at the county level. Cohen's claims, which are based on his linear (or linear-quadratic) model for county lung cancer rates, are therefore built on a false premise, and his continual reference to the 'scientific method' is fatuous. One could equally fit and reject a sinusoidal relationship for county lung cancer rates and radon concentrations; the factual basis of an inadequate model is true, but of little inferential value for evaluating risk.

Given two counties with equal proportions of smokers and equal patterns of smoking, I showed that a positive association for radon and lung cancer at the individual level can be easily transformed into an observed negative relationship between lung cancer rates and mean radon levels at the county level through simple manipulations of the *within-county* correlation between smoking and radon [10]. Moreover, the within-county correlations can be extremely small, on the order of 0.05 to 0.10 (see table 1 in [10]). The current paper [2] extends that two-county analysis, and shows that the reversal in trend can be extended to all 1,599 counties in Cohen's regression, even when the ecological regression is assumed to perfectly fit the data with no residual variation. Those results demonstrate the potential for extreme distortion of any ecological regression.

A relatively simple within-county adjustment was needed to show the compatibility of Cohen's regression with the BEIR VI extrapolation of risk. Contrary to Cohen's view, the vast majority (84 per cent) of within-county correlation coefficients for radon and smoking were between −0.3 and 0.3. In addition, those correlations are artificially elevated due to measurement error. The implicit assumption is that all covariates are measured without error. However, it is clear cigarette smoking is very poorly measured. A total of 85–90 per cent of all lung cancers are attributed to cigarette consumption, while Cohen's smoking variable explains only about 25 per cent of the variation in lung cancer rates among counties. Equation (3) in [2] defines the risk-adjusted radon concentration for a county (denoted $\tilde{w}$) as represented in the true county-level regression. The risk-adjusted mean radon depends on the proportion of smokers and the relative risk of smoking for the county. Thus, even if errors in smoking status and residential radon concentration were independent at the individual level, equation (3) shows that the proportion of smokers and $\tilde{w}$ are correlated. Thus, county-level correlations between smoking and radon are further distorted by the joint misclassification of two factors.

Finally, as a practical matter, it is worth noting that Puskin has recently offered a plausible explanation for Cohen's negative correlation [11] that agrees with the possible role of correlated errors. Puskin conducted ecological regressions of radon and smoking for several strongly smoking-related cancers (cancers of the lung, oesophagus, larynx, nasopharynx and oral cavity), weakly smoking-related cancers (cancers of the bladder and pancreas) and cancers unrelated to smoking (cancers of the colon, breast and prostate). He found strong negative correlations between county radon concentrations and cancers strongly linked to cigarette

smoking, weaker correlations between radon and cancers weakly associated with smoking, and essentially no correlation between radon and cancers not linked to smoking. Puskin concludes that the negative trend reported by Cohen for lung cancer is very likely explained by a negative correlation between smoking and radon levels across counties.

Yours faithfully,

J H Lubin

References

[1] Cohen B L 2002 Response to 'The potential for bias in Cohen's ecological analysis of lung cancer and residential radon' *J. Radiol. Prot.* **22** 305–7

[2] Lubin J H 2002 The potential for bias in Cohen's ecological analysis of lung cancer and residential radon *J. Radiol. Prot.* **22** 141–8

[3] Greenland S 1992 Divergent biases in ecologic and individual-level studies *Stat. Med.* **11** 1209–23

[4] Greenland S and Robins J 1994 Invited commentary: ecologic studies—biases, misconceptions, and counterexamples *Am. J. Epidemiol.* **139** 747–60

[5] Morgenstern H 1995 Ecologic studies in epidemiology: concepts, principles, and methods *Ann. Rev. Public Health* **16** 61–81

[6] Muirhead C R, Butland B K, Green B M R and Draper G J 1991 Childhood leukaemia and natural radiation (letter) *Lancet* **337** 503–4

[7] Piantadosi S, Byar D P and Green S B 1988 The ecologic fallacy *Am. J. Epidemiol.* **127** 893–904

[8] Stidley C A and Samet J M 1994 Assessment of ecologic regression in the study of lung cancer and indoor radon *Am. J. Epidemiol.* **65** 234–51

[9] Smith B J, Field R W and Lynch C F 1998 Residential Rn-222 exposure and lung cancer: Testing the linear no-threshold theory with ecologic data *Health Phys.* **75** 11–7

[10] Lubin J H 1998 On the discrepancy between epidemiologic studies in individuals of lung cancer and residential radon and Cohen's ecologic regression *Health Phys.* **75** 4–10

[11] Puskin J S 2002 Smoking as a confounder in ecological correlations of cancer mortality rates with average county radon levels *Health Phys.* at press

Comments upon human exposure to depleted uranium during and after the Gulf and Balkans conflicts

Dear Sir

Spratt [1, 2] has commented upon, and acted as Chairman to, The Royal Society working group reports [3, 4] on depleted uranium (DU) which was used in the 1991 Gulf (340 tonnes) and late 1990s Balkans (11 tonnes) conflicts. The Royal Society reports add nothing new to that which is already known (see Priest [5] and Hamilton [6] for overviews). In relation to Spratt [1]—where are we now?—we are more or less where we were two years ago; we have spent large sums of money to no avail and established scientific principles have and continue to be discarded.

In the mid 1970s the UK Ministry of Health Medical Research Council Radiological Protection Service, Sutton, Surrey (later the National Radiological Protection Board), together with other international organisations, had already laid down the principles for considering