1988_06

CORRELATION BETWEEN MEAN RADON LEVELS AND LUNG CANCER RATES IN U.S. COUNTIES
A TEST OF THE LINEAR - NO THRESHOLD THEORY

BY: Bernard L. Cohen
University of Pittsburgh
Pittsburgh, PA 15260

ABSTRACT

Mean radon levels in 415 U.S. counties were obtained from purchased measurements utilizing a high degree of selectivity to reduce bias. A linear regression of lung cancer rates for males and females vs. mean radon levels gives negative correlations with slopes over seven standard deviations more negative than zero, whereas all linear-no threshold theories predict a substantial positive slope. When data are segmented by states or by regions of the nation, negative correlations are very predominant. In addition, five studies by individual states find the same phenomenon of predominantly negative correlations. It is concluded that this represents an important breakdown of the linear - no threshold theory of radiation carcinogenesis. The limitations of ecological studies are discussed, but it is concluded that they are not applicable here.

THE DATA

The principal data analyzed here are from the data base of measurements purchased from the University of Pittsburgh Radon Project and its successor, The Radon Project, Inc. The data are for measurements in the living areas (not basements) of houses in which there had not been a previous measurement and the client did not know of a house with a high radon level within five miles. The geometric mean of all measurements in a county is used. The correlation between these and those obtained from random selection - no charge studies of the same county are shown in Figure 1. We see there that the correlation is quite strong even where the mean is derived from less than ten measurements, and that purchased measurements give typically 1.5 times higher means than random selection.

Figure 2 shows the mean radon level in the 415 counties where there were at least ten measurements as the abscessa. The ordinate is the age-adjusted female lung cancer mortality rate for 1950 - 1969. The abscissa is divided into ranges as shown at the top of Figure 1 - ≤ 1.0 pCi/L, 1.0-1.5 pCi/L etc. - with the number of counties included in the range listed. For each abscessa range, the median ordinate value is shown by a cross, and the first and third quartiles are shown by open circles. The solid line is the least suares fit of all data points to a straight line. One standard deviation (SD) in its intercept is indicated at the left and one SD in its slope is indicated at the right end. Note that the slope of the least squares fit, which is more sensitive to values far from the centroid has roughly the same slope as a line through the medians which are most sensitive to values close to the centroid.

The dashed line shows the prediction of BEIR-IV on the assumption that the average person who dies in a county has spent half of her life in that county (or a neighboring county with similar radon levels). The "1/2" here is a guesstimate; if it were 1/4, the slope of the line would be cut in half. The predictions are based on a comparison with the average radon level in a county, whereas the data points are plotted as the geometric mean. This compensates for the fact derived from Figure 1 that purchased measurements, even with our bias removal selections, give about 1 1/2 times higher mean radon levels than random selection studies; average radon levels in a county are typically about 2.0 times the geometric means.

Figure 3 is similar to Figure 2 except that the lung cancer rates are for males rather than females.

The most striking feature of Figure 2 and 3 is that BEIR-IV (and all other linear-no threshold theories) predict a strong positive slope, whereas the data exhibit a strong negative slope, as though exposure to radon protects against lung cancer. Statistics are not an issue here; the slopes are less than zero by more than 10 SD for males and 7 SD for females. No linear-no threshold theory can give a negative slope.

The most obvious confounder that could explain the negative slope is a chance correlation between smoking and radon exposure. For example, it might be that for the state of New Jersey there is more cigarette smoking in areas near New York City and Philadelphia, and these areas, for some unrelated reason, have lower radon levels than the rest of the state. This would be a believable coincidence for one small section of the country, but not for the nation as a whole as too many coincidences would be required. This raises the question "How general a phenomenon is that negative slope?"

Table 1 is largely an attempt to answer this question. In it, data are given for individual states or contiguous groups of states having at least ten counties with enough measurements to be included in Figures 2 and 3. Groupings were selected before the data were input and analyzed. Column 2 of Table 1 gives the number of counties included in the grouping. Columns 3-6 give the coefficients of correlation for the male and female lung cancer rates with the geometric mean and arithmetic average radon levels in those counties. Columns 7 and 9 give the slope of the least square fit to a line through the data for geometric means or b in

ordinate = constant + b x abscissa.

Columns 8 and 10 give t, the number of standard deviations of b from b = 0.

We see from this part of Table 1 that the coefficients of correlation and the slopes are negative for 18 of the 23 groupings. The distribution of coefficients of correlation are shown in the top part of Figure 4. We see that they are heavily shifted toward negative values. The distributions of t-values are shown in the lower part of Figure 4, again heavily shifted toward negative values, which means that b tends to be negative by typically more than 1 SD for each state.

TABLE 1

		Correlation Coefficients				Slopes (b) and t-ratios			
STATES	Number	M/Mn	M/Av	F/Mn	F/Av	M-b	M-t	F-b	F-t
ME,NH,VT	17	+.27	+.07	+.14	+.17	+2.1	+1.1	+.49	+.57
MA	11	-.42	-.22	-.31	-.33	-7.7	-1.4	-1.6	-.97
CT, RI	13	+.05	+.21	-.05	+.22	+.70	+.17	-.17	-.16
NY	27	-.25	-.24	-.37	-.36	-2.9	-1.3	-.93	-2.0
NJ	17	-.47	-.46	-.36	-.31	-4.0	-2.0	-.46	-1.5
PA	46	-.18	-.12	+.09	+.03	-.84	-1.2	+.08	+.56
OH	17	-.48	-.46	-.22	-.27	-2.6	-2.1	-.10	-.87
MI	18	-.35	-.45	-.20	-.28	-3.5	-1.5	-.34	-.82
IL	11	-.01	-.25	-.54	-.71	-.11	-.03	-.96	-1.9
IN, WI	15	+.24	+.16	+.18	+.13	+3.5	+.88	+.29	+.65
IA	21	+.04	+.07	+.22	+.27	+.26	+.18	+.27	+.99
MO,MN,ND SD,NE,KS	16	-.69	-.66	-.68	-.58	-5.3	-3.6	-1.0	-3.4
MD, DE	16	-.48	-.42	-.34	-.32	-3.0	-2.1	-.43	-1.4
VA,WV,DC	15	-.49	-.41	-.24	-.25	-6.5	-2.1	-1.0	-.89
NC,SC,GA	18	-.46	-.45	-.58	-.55	-6.4	-2.1	-1.1	-2.8
FL	17	+.45	+.43	-.10	-.13	+3.0	+2.0	-.13	-.39
TN,KY, AL,MS	19	-.24	-.21	-.30	-.26	-2.2	-1.0	-.44	-1.3
TX,LA, AK,OK	12	-.55	-.50	-.44	-.36	-11.4	-2.1	-1.7	-1.6
ID	37	-.17	-.11	-.19	-.10	-1.3	-1.0	-.36	-1.1
CO	12	-.18	-.02	-.04	-.13	-.67	-.59	-.05	-.13
WY,MT,NM AZ,UT,NV	14	-.72	-.68	-.37	-.36	-4.7	-3.6	-.49	-1.4

CA	13	-.25	-.16	-.18	-.36	-3.6	-.90	-.63	-.60
WA,OR	13	-.38	-.45	-.17	-.20	-1.0	-1.3	-.09	-.57
New England	41	+.10	+.02	+.04	+.10	+1.0	+.61	+.15	+.27
Mid Atlantic	90	-.46	-.26	-.22	-.10	-3.1	-4.9	-.24	-2.1
Midwest	98	-.37	-.36	-.23	-.20	-2.4	-3.9	-.25	-2.3
Southeast	66	-.29	-.21	-.31	-.25	-2.9	-2.5	-.56	-2.6
South Central	31	-.47	-.44	-.46	-.42	-4.2	-2.9	-.67	-2.8
Mountain	63	-.23	-.17	-.18	-.14	-1.6	-1.9	-.30	-1.4
Pacific	26	-.46	-.45	-.36	-.32	-1.9	-2.5	-.40	-1.9
USA	415	-.46	-.24	-.35	-.16	-3.4	-10.4	-.49	-7.5
NJ	21	-.39	-.37	-.47	-.42	-1.4	-1.9	-.24	-2.3
NY (Basement)	35	-.43	-.48	-.46	-.54	-1.2	-2.7	-.30	-2.9
NY (Living)	44	---	-.23	---	-.44	-1.9	-1.5	-.85	-3.2
SC	28	---	-.16	---	-.24	-2.0	-.81	-.76	-1.3
FL	57	---	-.05	---	+.09	+.60	+.37	+.27	+.69

When the data are divided up into so many pieces the statistical significance of the results for each piece is reduced. For example, only a small fraction of the pieces have a slope deviating by more than 2 SD from zero.

As a compromise between this extensive fragmentation and the complete aglomeration of all data in Figures 2 and 3, an aglomeration into seven regions of the nation is shown in the next section of Table 1. Data for some of these regions are shown in Figure 5. We see that for six of the seven regions, b is negative by more than 1.4 SD for both males and females. In five of these six it is negative by more than 2.5 SD for males, and by more than 1.9 SD for females. The seventh region, New England, has a positive b, but only by 0.6 SD for males and 0.3 SD for females. Even for this case, the slope is less positive than the BEIR IV prediction by more than one SD.

The last entry in this part of Table 1, USA, is for all of the data combined as shown in Figures 2 and 3. We see that the slopes are negative by 10.4 SD for males and by 7.5 SD for females.

In addition to the data from the University of Pittsburgh studies, there are now data available from five separate state-sponsored studies. These are shown in Figures 6 and 7 along with a least squares fit to a straight line through the data, and the statistical analyses are included in the bottom part of Table 1. The BEIR-IV predictions are included in Figures 6 and 7 under the assumption that living areas have one-half the radon levels of basements, and arithmetic average radon levels are twice the geometric means or medians.

From the bottom part of Table 1 (and Figures 6 and 7) we see that four of the five studies give a negative correlation, with the value of b negative by 0.8 to 2.7 SD for males and by 1.3 to 3.2 SD for females. The fifth case is Florida which has a slightly positive correlation, differing from zero by much less than 1 SD. Note that Florida data also gave a positive correlation in the purchased measurement study discussed above.

In summary, the data very clearly indicates that the correlation between lung cancer rates and radon levels in U.S. counties is, on the whole, negative. This conclusion would seem to be confirmed far beyond any reasonable doubt due to statistical uncertainties. It is clearly in strong and direct conflict with predictions of the BEIR-IV estimates and with the predictions of any other linear, no threshold theory of radiation carcinogenesis.

INTERPRETATION

In epidemiology, studies of relationships between exposures of population groups and their mortality rates are termed "ecological studies" and are given much less credibility than studies of individuals. The reasons for this fall into two categories:

(1). In ecological studies, the individuals who are exposed may be different from those who contribute to the mortality statistics. For example, the average exposure to a certain pollutant in a county might be far above average but there still may be none above the threshold for producing an effect, while another county may have a low average exposure but some above the threshold. However, this does not apply in a linear-no threshold situation where the mortality rate depends directly on the average exposure.

(2). Ecological studies are more susceptible to confounders. For example, smoking may cause windows to be opened more to release odors, while opening windows reduces radon levels. This particular problem was investigated and it was found that houses of smokers do have about 10% lower radon levels. But, a quantitative treatment showed that this was much too small an effect to change the correlations.

We have studied correlations between radon levels and a large number of factors, and these correlations are always too small to matter. Radon levels depend principally on geology and it is difficult to imagine how the amount of smoking can correlate with geology on a nationwide basis. Many possibilities have been considered but none can have nearly the required effect.

We therefore interpret the failure of linear-no threshold theory to explain the data as a breakdown of that theory. This breakdown should apply to all types of radiation.

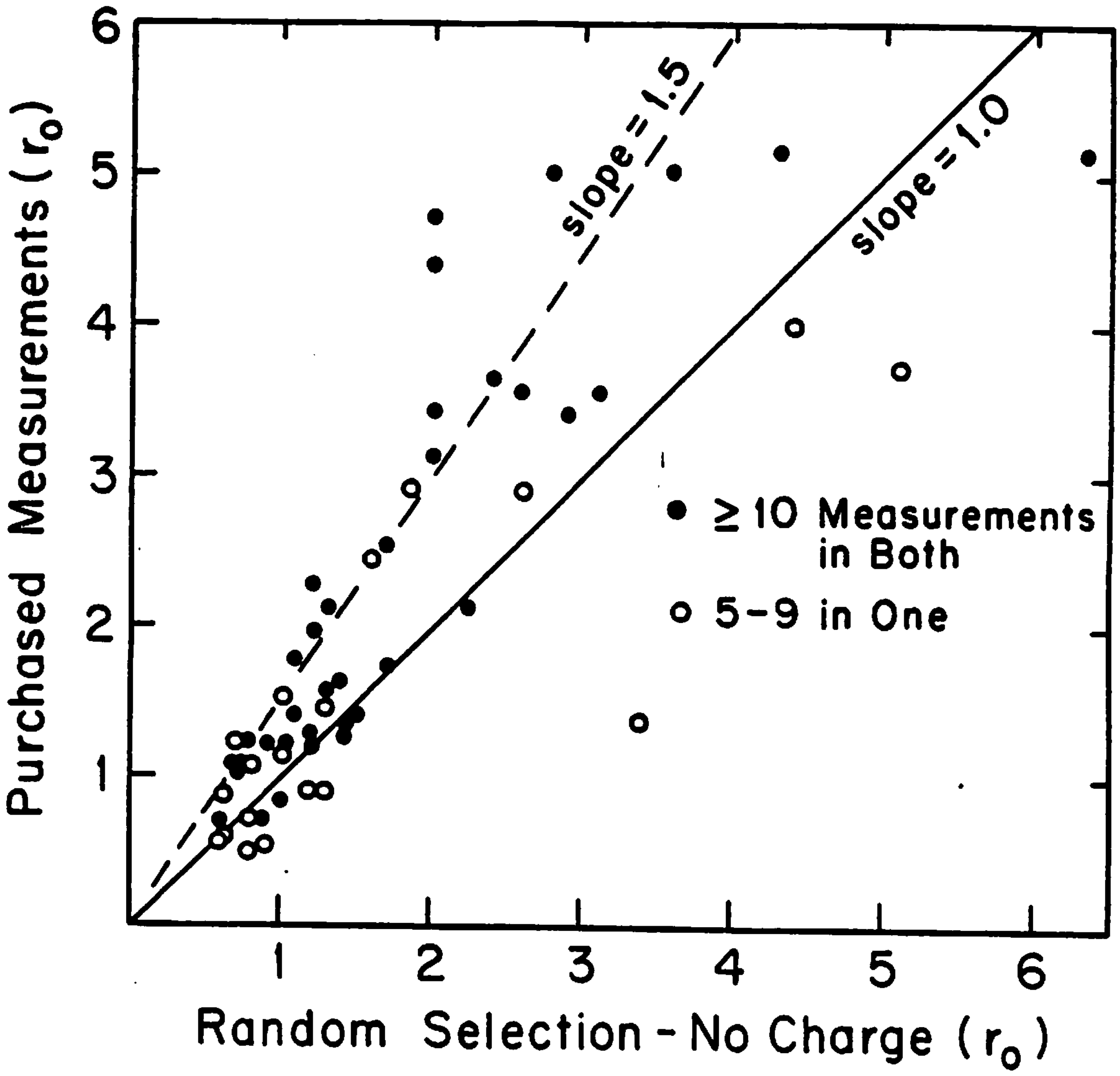

Purchased Measurements (r_0)
Random Selection - No Charge (r_0)
1
2
3
4
5
6
1
2
3
4
5
6
slope = 1.5
slope = 1.0
● ≥ 10 Measurements in Both
○ 5-9 in One

Figure 1.

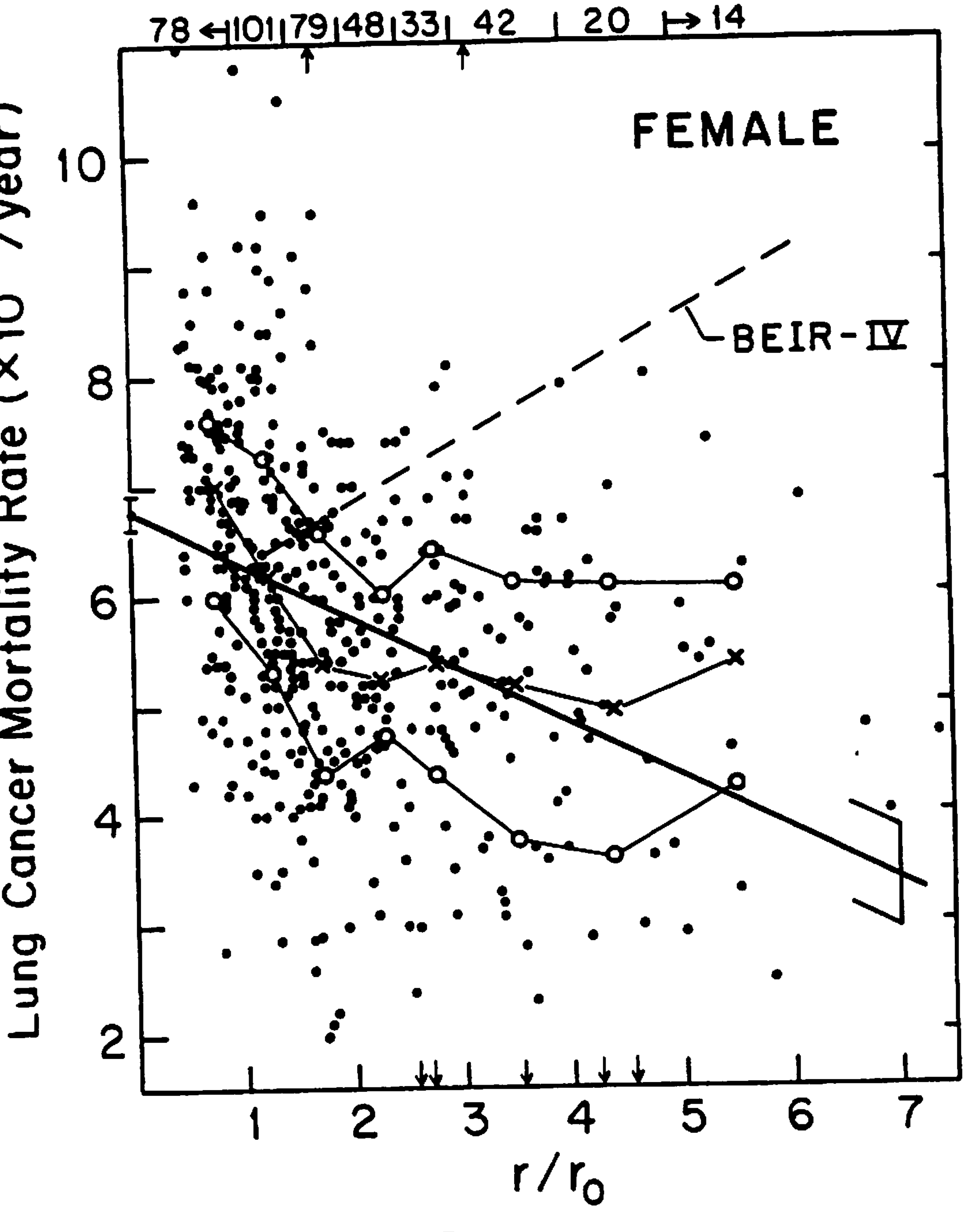

78
101
79
48
33
42
20
14
FEMALE
BEIR-IV
Lung Cancer Mortality Rate ($\times 10^{-5}$/year)
10
8
6
4
2
1
2
3
4
5
6
7
r/r_0

Figure 2.

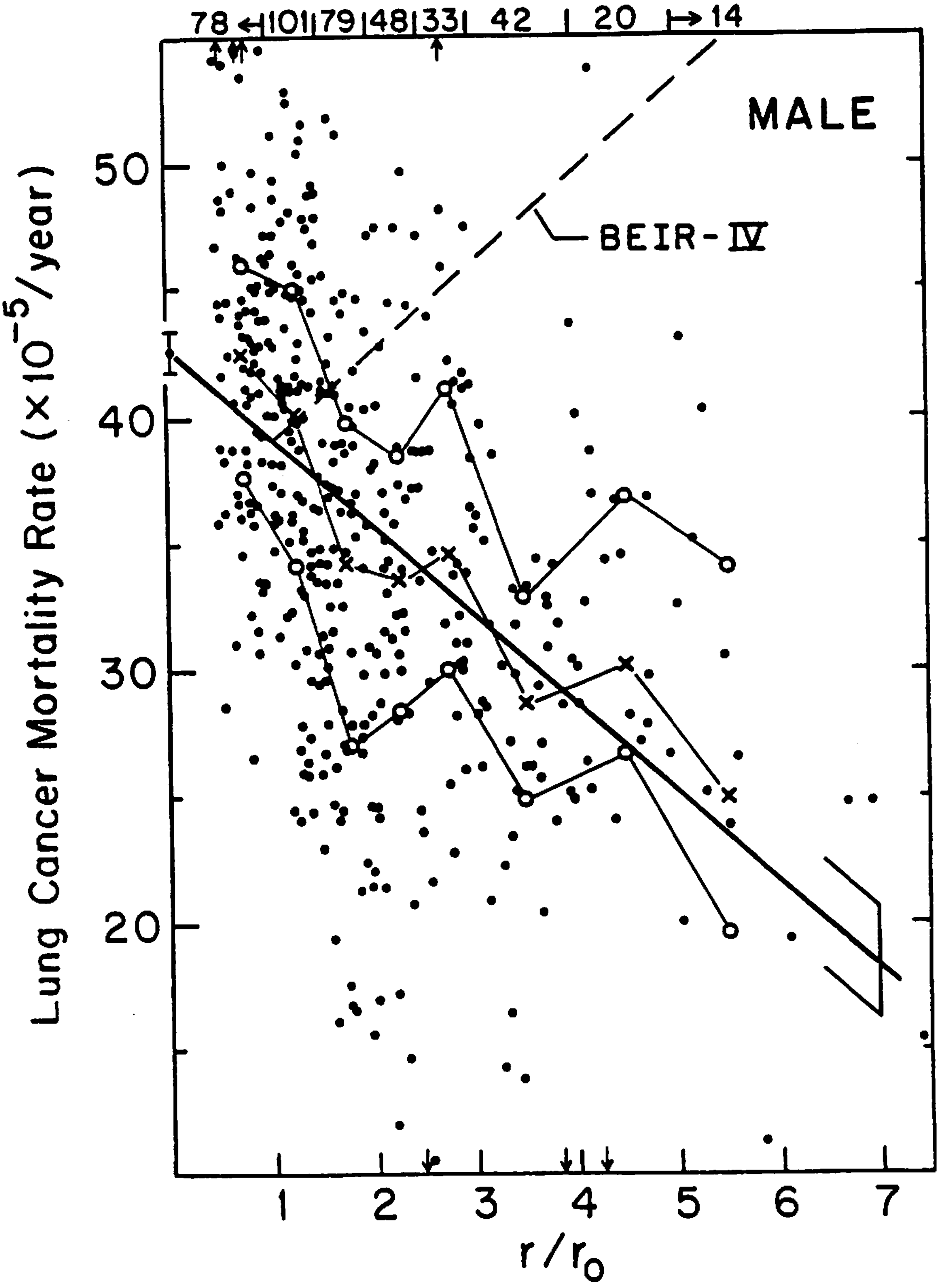

78
101
79
48
33
42
20
14
MALE
BEIR-IV
Lung Cancer Mortality Rate ($\times 10^{-5}$/year)
50
40
30
20
1
2
3
4
5
6
7
r/r_0

Figure 3.

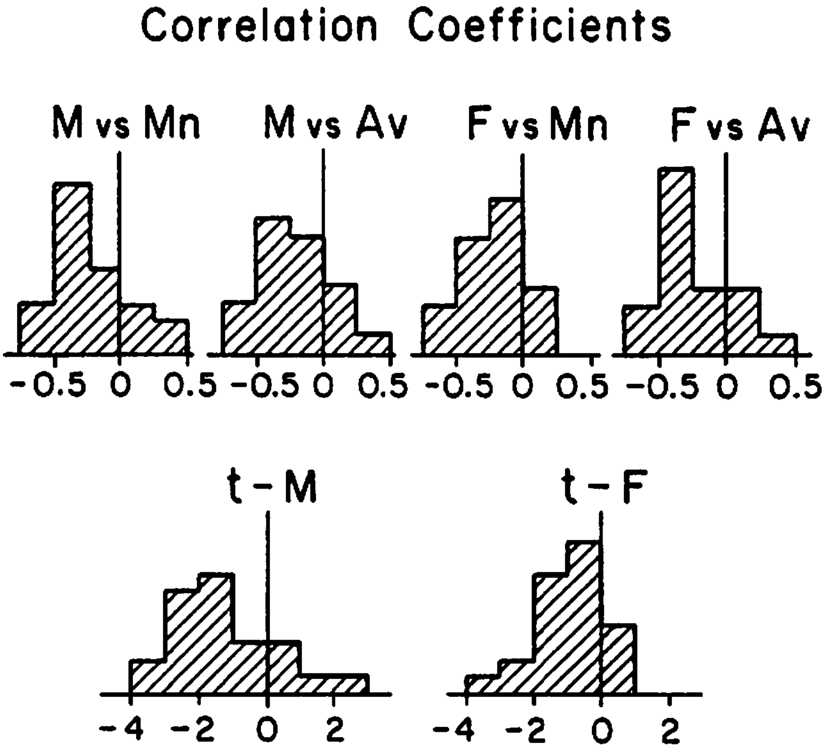
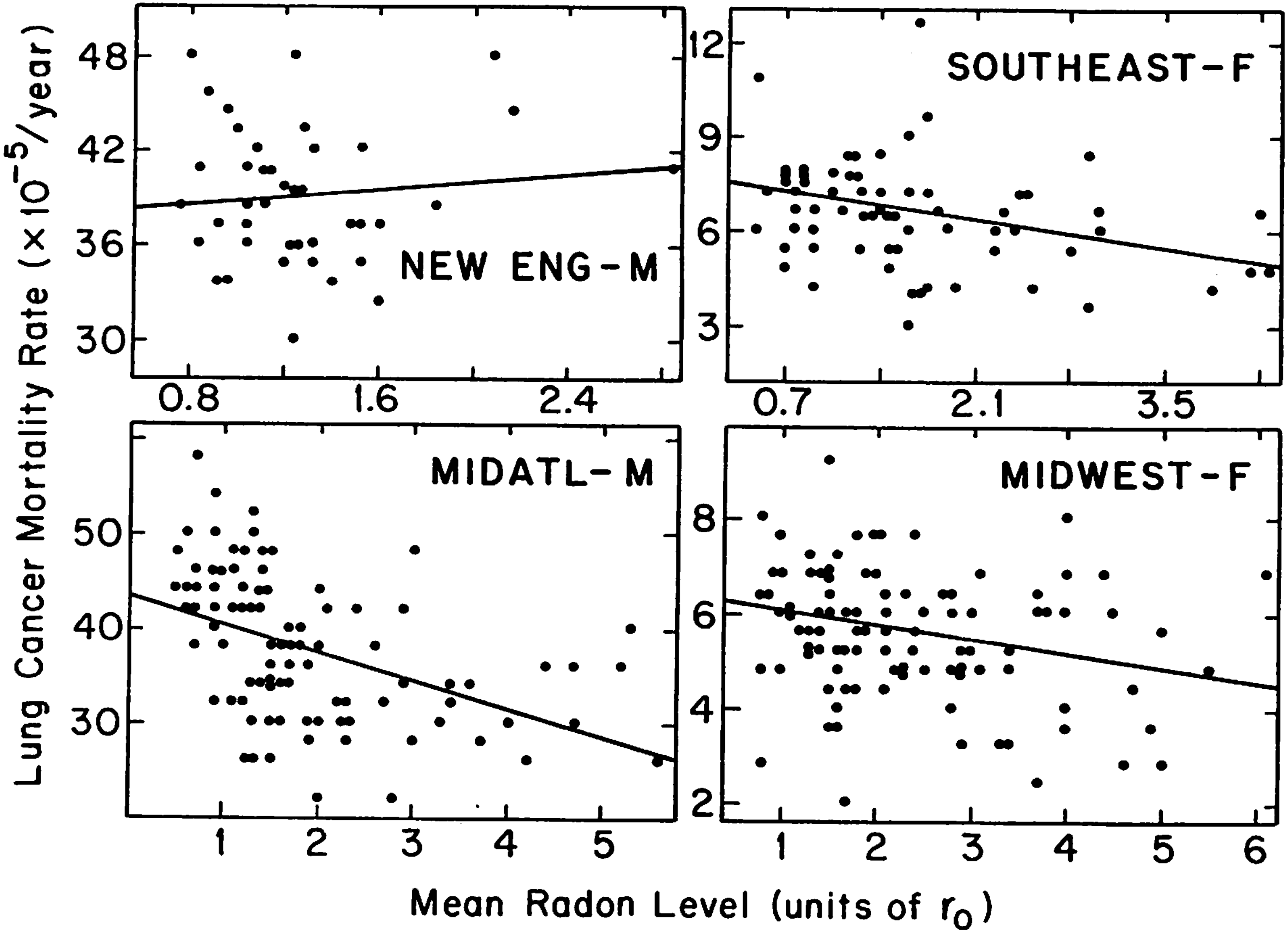

Correlation Coefficients
M vs Mn
M vs Av
F vs Mn
F vs Av
-0.5 0 0.5
-0.5 0 0.5
-0.5 0 0.5
-0.5 0 0.5
t - M
t - F
-4 -2 0 2
-4 -2 0 2

Figure 4.



NEW ENG-M
SOUTHEAST-F
MIDATL-M
MIDWEST-F
48
42
36
30
0.8
1.6
2.4
12
9
6
3
0.7
2.1
3.5
50
40
30
1
2
3
4
5
8
6
4
2
1
2
3
4
5
6
Lung Cancer Mortality Rate (× 10^{-5}/year)
Mean Radon Level (units of r_0)

Figure 5.

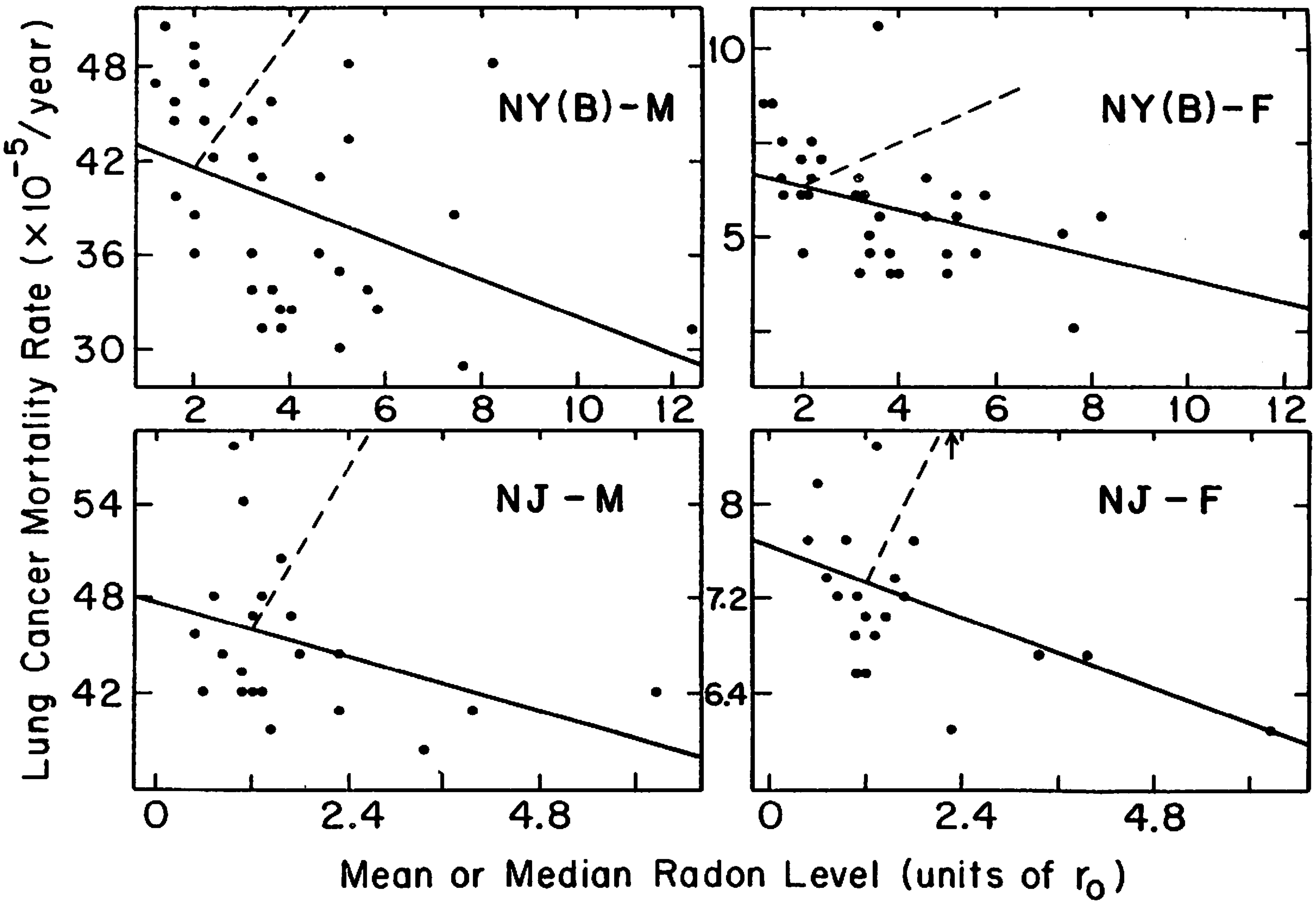
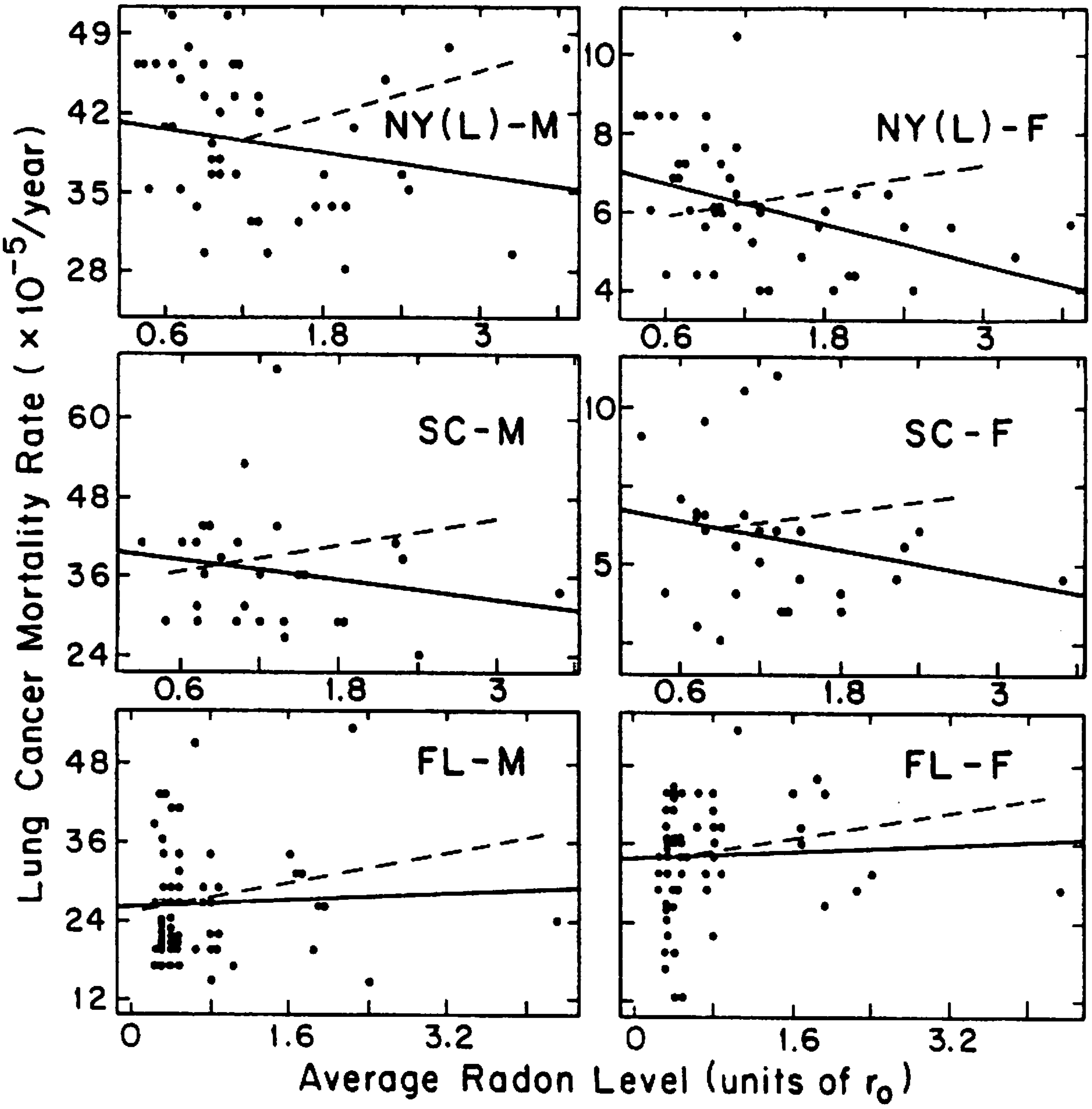

NY(B)-M
NY(B)-F
NJ - M
NJ - F
Lung Cancer Mortality Rate ($\times 10^{-5}$/year)
48
42
36
30
10
5
54
48
42
8
7.2
6.4
2
4
6
8
10
12
0
2.4
4.8
Mean or Median Radon Level (units of r_o)

Figure 6.



Lung Cancer Mortality Rate (× 10⁻⁵/year)
NY(L)-M
49
42
35
28
0.6
1.8
3
NY(L)-F
10
8
6
4
0.6
1.8
3
SC-M
60
48
36
24
0.6
1.8
3
SC-F
10
5
0.6
1.8
3
FL-M
48
36
24
12
0
1.6
3.2
FL-F
0
1.6
3.2
Average Radon Level (units of r₀)

Figure 7.