Technology, Vol. 7 pp. 657–672, 2000
Printed in the USA.

1072-9240/00 $20.00 + .00

www.cognizantcommunication.com

UPDATES AND EXTENSIONS TO TESTS OF THE LINEAR-NO THRESHOLD THEORY

Bernard L. Cohen

Department of Physics, University of Pittsburgh, Pittsburgh, PA 15260

Two weaknesses in a 1995 article on tests of the linear-no threshold theory of radiation carcinogenesis are pointed out. One is addressed by introducing more recent cancer mortality statistics, and the other is addressed by introducing 450 newly available potential confounding factors. The later cancer statistics give results very similar to the original ones for the nation as a whole, but do cause significant changes for some geographic areas. None of the new confounding variables helps to explain the large discrepancy with predictions of the linear-no threshold theory, nor does the use of more recent cancer statistics.

Key words: Linear-no threshold theory; Radiation carcinogenesis; Confounding factors

REVIEW OF ORIGINAL STUDY

A 1995 article (1), hereafter referred to as C-95, presented a test of the linear-no threshold theory (LNT) based on a compilation of radon measurements from available sources that gave the average radon level, *r*, in homes for 1729 US counties, well over half of all US counties and comprising about 90% of the total US population. Plots of age-adjusted lung cancer mortality rates, *m*, vs. these *r* are shown in Figure 1a and c. To avoid the confusion of showing over a thousand data points, data have been combined into groups consisting of all counties within set intervals of *r* values, and for the group in each interval, the mean value of *m* is plotted with error bars showing 1 SD of the mean. The first and third quartiles for the distribution are also plotted as an indication of the spreads of *m* values involved. We see in Figure 1a and c a clear tendency for *m* to decrease with increasing *r*, in sharp contrast to the increase expected from the fact that radon can cause lung cancer, shown by the line labeled "theory."

One obvious problem is migration: people do not spend their whole lives and receive all of their radon exposure in their county of residence at time of death where their cause of death is recorded. However, it is easy to correct the theoretical prediction for this, and the "theory" lines in Figure 1 have been so corrected. To minimize the migration problem, data for Florida, California, and Arizona were deleted because deaths there are frequently recorded for retirees who received their radon exposures elsewhere. This reduces the number of counties to 1601, but causes no appreciable change in the results.

A more serious problem is that this is an "ecological study," relating the average risk of groups (county populations) to their average exposure dose. In general, the average dose does not determine the average risk, and to assume otherwise is what epidemiologists call "the ecological fallacy." However, it is easily shown that the ecological fallacy (thus defined) does not apply in testing a linear-no threshold theory. This is familiar from the concept that, with LNT, "man-rem" determines the number of cancer deaths: dividing each by the population, man-rem gives the average dose, and number of deaths gives the mortality rate. Other problems with ecological studies that have been discussed in the epidemiology litera-

Address correspondence to Bernard L. Cohen. Tel: (412) 624-9245; Fax: (412) 624-9163; E-mail: blc+@pitt.edu

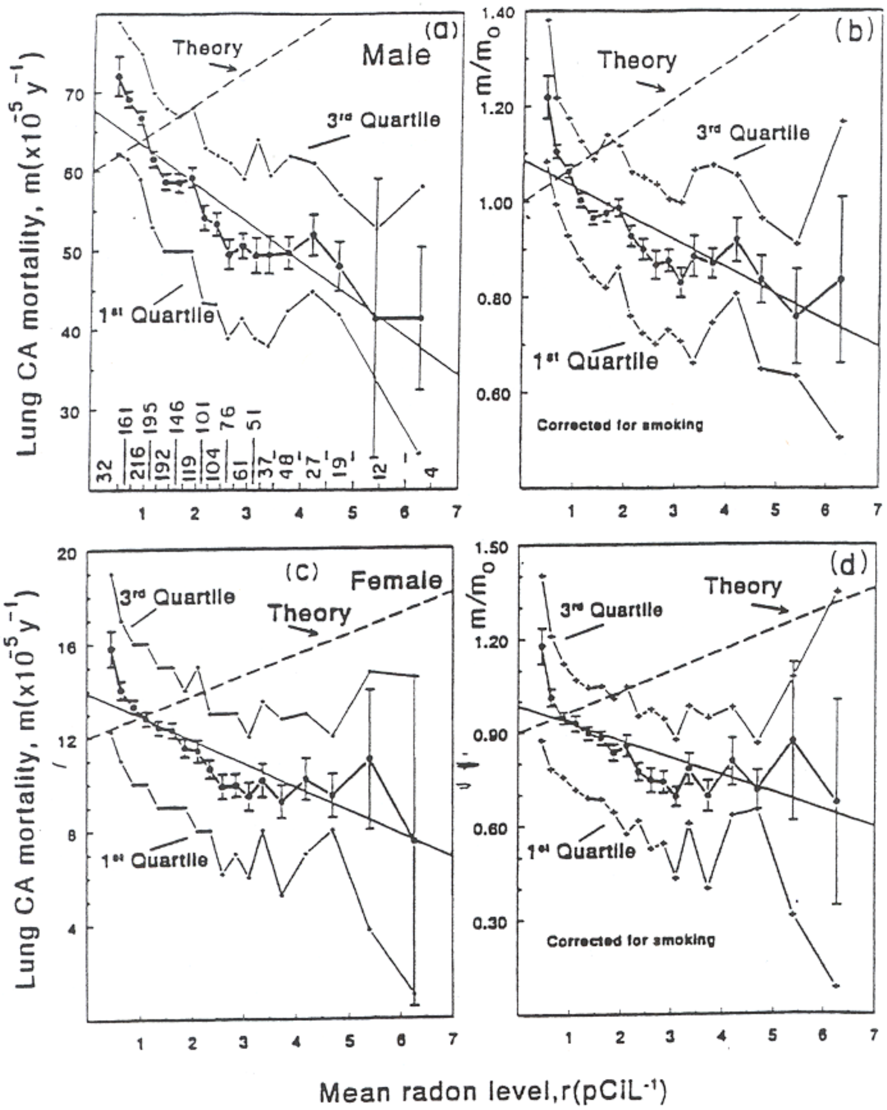

Figure 1. Lung cancer mortality rates vs. average radon levels in homes for US counties for males (a) and females (c). Data points shown are average ordinates for all counties within the range of *r* values shown on the basline of (a); the number of counties within that range is also shown there. Error bars are SD of the mean, and the first and third quartiles are also shown. Lung cancer rates for (b) males and (d) females corrected for smoking prevalence (m/m_o) are shown at right. Theory lines are arbitrarily normalized lines increasing at a rate of +7.3% per pCi/L.

ture were also investigated and found not to be applicable here.

Because epidemiologists normally study the mortality risk to individuals, this study started from that premise and derived, by rigorous mathematics, the correction for smoking in terms of the fraction of adults who are smokers, *S*. The data in Figure 1a and c, corrected for smoking, are shown in Figure 1b and d. It can be seen that there is a huge discrepancy between measurements and LNT theory, by about 20 SD. The theory predicts the slope of the line, *B* (in unit of percent per pCi/L), to be $B = +7.3$, whereas the data are fit by $B = -7.3 \pm 0.6$ and -8.3 ± 0.8 for males and females, respectively.

A wide variety of potential explanations for the discrepancy that we could develop or that were suggested by others were tested and found to be grossly inadequate. Three independent sources of radon data were used but all gave the same result. Three different sources of data on smoking prevalence similarly failed. In fact, it was found that if our best estimate of the width of the distribution of S values for US counties is correct, even a perfect negative correlation between radon and smoking prevalence eliminates only half of the discrepancy. If the S value distribution had the largest credible width, an essentially perfect negative correlation between radon levels and smoking prevalence in US counties would be required to explain the discrepancy. It was shown that anything approaching such a strong correlation is completely incredible.

It was shown that the strong correlation between radon exposure and lung cancer mortality, albeit negative rather than positive, is unique to lung cancer; no remotely comparable correlation was found for any of the other 32 cancer sites. We concluded that the observed behavior is not something that can easily occur.

To investigate effects of a potential confounding variable, data were stratified into quintiles on the values of that variable, and a regression analysis was done separately for each stratum. Because the potential confounder has essentially the same value for all counties in a given stratum, its confounding effect is greatly reduced in these analyses. An average of the slopes, B, of the regression lines for the five quintiles then gives a value for B that is largely free of the confounding under investigation.

This test was carried out for 54 socioeconomic variables and none was found to be a significant confounder. In all 540 regression analyses (54 variables $\times$ 5 quintiles $\times$ 2 sexes), the slopes, B, were negative and the average B value for the five quintiles was always close to the value for the entire data set. This means that the negative correlation between lung cancer rates and radon exposure is found if we consider only the very urban counties or if we consider only the very rural counties; if we consider only the richest counties or only the poorest; if we consider only the counties with the best medical care or only those with the poorest medical care; and so forth for all 54 socioeconomic variables. It is also found for all strata in between, as, for example, considering only counties of average urbanicity, only counties of average wealth, only counties of average medical care, etc.

The possibility of confounding by combinations of socioeconomic variables was studied and found not to be an important potential explanation for the discrepancy.

The stratification method was also used to investigate the possibility of confounding by geography, and of confounding by physical features such as altitude, temperature, precipitation, wind, and cloudiness, but these factors were of no help in explaining the discrepancy. The negative slope and gross discrepancy with LNT theory is found if we consider only the wettest areas, or if we consider only the driest; if we consider only the warmest areas, or only the coldest areas, etc.

The effects of the two principal recognized factors that correlate with both radon and smoking were calculated in detail: 1) urban people smoke 20% more but average 25% lower radon exposures than rural people; 2) houses of smokers have 10% lower average radon levels than houses of nonsmokers. These were found to explain only 3% of the discrepancy. Because they are typical of the largest confounding effects one can plausibly expect, it is extremely difficult to imagine a confounding effect that can explain the discrepancy. Requirements on such an unrecognized confounder were listed, and they make its existence seem extremely implausible. By far, the most plausible explanation that could be found for this discrepancy was that the linear-no threshold theory fails, grossly overestimating the cancer risk in the low dose, low dose rate region where there are no other data capable of testing the theory.

THE NEW STUDY

Two weaknesses in C-95 are addressed here by providing a great deal of further data and analysis. One weakness lies in the fact that lung cancer mortality statistics in C-95 were from 1970–1979, the latest age-adjusted rates available at that time. These deaths were presumably related to radon exposures in the 1940–1970 time period (or earlier) whereas the radon levels were measured in the 1986–1991 time period. Of course, this same problem applies to all other studies of the radon–lung cancer relationship, including the widely heralded case-control studies, and it is more serious for them because it is much more likely for the radon level in a single house to have changed over several decades (e.g., cracks develop, or become sealed) than for the average radon level for all houses in a county to have changed. Nevertheless, it is preferable if the disparity in time between relevant exposures and radon measurements is reduced by using more recent lung cancer mortality statistics; that is done here. This has the added advantage of testing the degree to which these statistics have been changing with time, and it gives insight into other issues.

The other weakness addressed is the fact that an ecological study like C-95 is susceptible to "cross level bias" as emphasized by Lubin (2). For example, a case-control study may match cases and controls by annual dollar income, but an ecological study like C-95 is limited to including average annual income for different counties, not for individuals who did or did not die from lung cancer. Perhaps very poor people are much more susceptible to radon-induced lung cancer than others, making the fraction of the population that is very poor an important confounding factor. This is not necessarily represented by average income, because the latter is influenced by the fraction of the population that is very rich. The solution to this dilemma in C-95 was to consider confounding by such other factors as percent of population below the poverty line and percent unemployment.

But suppose very poor people have compensating factors (e.g., better medical care in free clinics); maybe it is people with incomes in the $15,000–$25,000 range that are most susceptible. The obvious solution here is to consider confounding by the fraction of the population in that income range. One might concoct models in which the fraction of the population in any other income range (or group of ranges) is an important confounder. The solution is to treat that fraction as a potential confounding factor. To do this requires data on the fraction in each income range in each county.

Of course, annual income is used here only as an example. Similar problems may arise from percent of population in any particular age range, from educational attainment, from house characteristics, etc. These problems are addressed here by considering a very large number of potential confounding factors not available in C-95.

UPDATED LUNG CANCER RATES

The lung cancer mortality rates in C-95 were for whites only and from the time period 1970–1979; these were the latest age-adjusted rates available at that time. However, age-adjusted rates have now become available for all races from 1979 to 1994 from CDC (3), and we begin here by considering the effect of using them in the analyses. The basic equation derived and used for fitting data in C-95 was

$$m/m_o = A + B\,r \qquad (1)$$

where

$$m_o = [S\,a_s + (1 - S)a_n] \qquad (2)$$

and S is the fraction of the adult population that smoked cigarettes at the time their lung cancers were initiated and developed.

The determination of the constants a_n, a_s in Eq. (2) [cf. (4)] depends on the average values of m, r, and S at the relevant times, and hence must be changed to reflect changes in m when the time period is shifted. From 1970–1979 to 1979–1994, the US average m, in units of 10^{-5}/year, changed from 64.0 to 57.5 for males, and from 15.3 to 22.6 for females. This changes a_n from 8.7 to 7.8 for males, and from 3.47 to 5.13 for females, and it changes a_s from 104 to 93.5 for males and from 34.7 to 37.3 for females. With our correction for migration, the final expressions for m_o are $[8.1 + 89\,S]$ for males (vs. $[9 + 99\,S]$ in C-95) and $[5.3 + 48\,S]$ for females (vs. $[3.7 + 32\,S]$ in C-95).

In Eq. (1), m/m_o is the lung cancer rate corrected for smoking prevalence. In Figure 2, the left side shows the data for 1970–1979, reproduced from Figure 1b and d, and the right side shows the data for 1979–1994 presented in the same fashion. In accordance with Eq. (1), B is the slope of the best fit for a straight line through the data. The BEIR-IV model used in C-95 predicts B to be +7.3. In our analyses, the data from 1979–1994 mortality rates give $B = -7.7 \pm 0.51$ for males and $B = -8.2 \pm 0.67$ for females, close to the results from C-95, $B = -7.3$ and $B = -8.3$, respectively. The updated results are discrepant

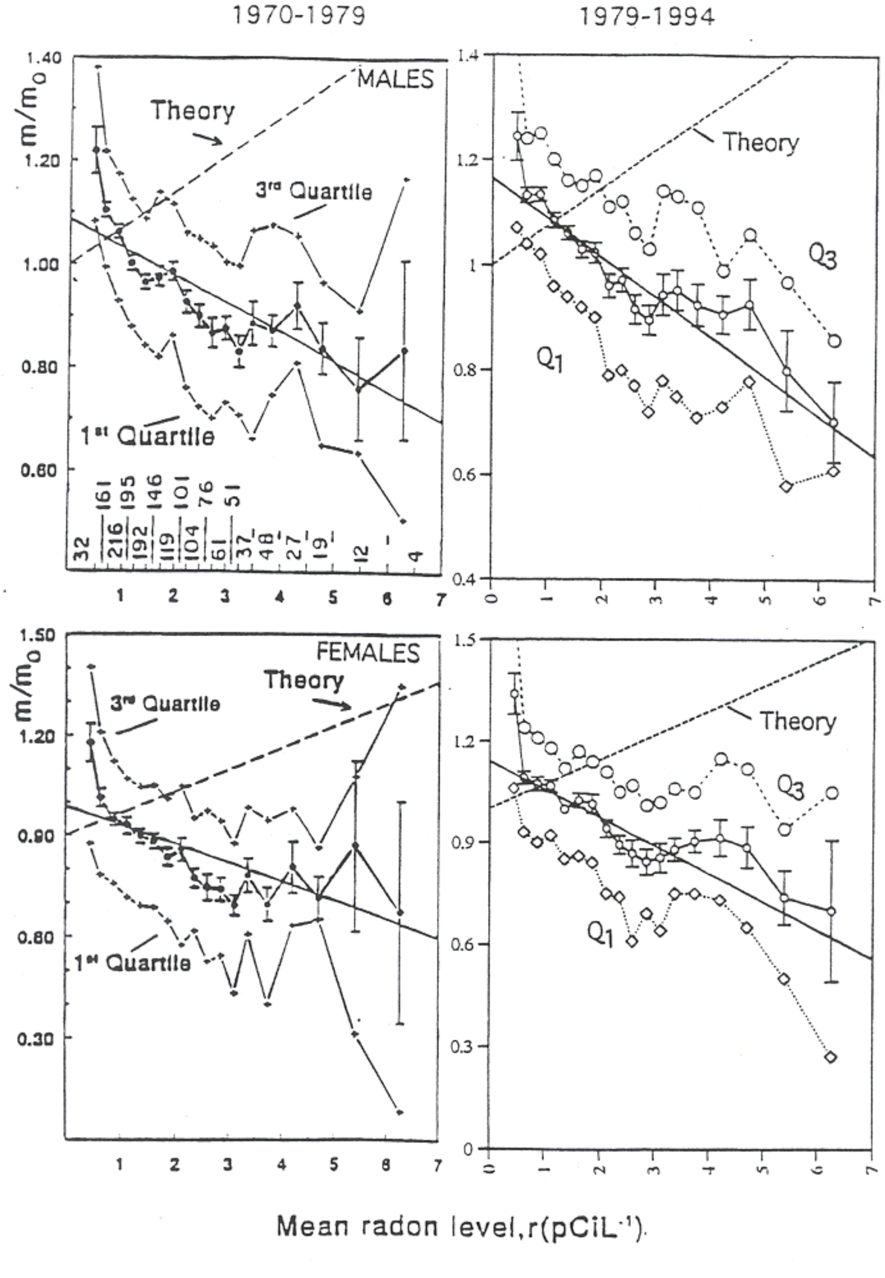

Figure 2. Comparison of results for mortality data from 1970–1979 and 1979–1994. The left side is from Figure 1b and d, lung cancer rates corrected for smoking prevalence using mortality data from 1970–1979, and the right side is a similarly treated analysis using mortality data from 1979–1994.

with the BEIR-IV prediction, $B = +7.3$, by 29 SD for males and by 23 SD for females. Further analyses will use the 1979–1994 lung cancer rates.

For some purposes, it is relevant to do similar analyses by states rather than by counties, as was done by Cohen and Colditz (4). Because of differences in the correction for migration (moving to a different state is less common than moving to a different county), the expressions for m_o for states are changed to [8.0 + 88 S] for males and [5.2 + 47 S] for females and the BEIR-IV prediction is $B = +8.6$. Fitting the data with Eq. (1) gives $B = -11.8 \pm 2.9$ for males, and $B = -13.7 \pm 2.5$ for females, discrepant with the BEIR-IV prediction by 7.0 and 8.9 SD, respectively.

The similarity in results for the two time periods may give the impression that lung cancer rates for the two time periods are nearly identical, but that is not the case. The coefficient of correlation between these rates is 0.73 for males and 0.53 for females. For the rankings of our 1601 counties by lung cancer rates it is 0.73 and 0.61 for males and females, respectively. The average age-adjusted lung cancer rates declined by 4.5% for males and increased by 69% for females between the earlier and the later period.

ADDED SOCIOECONOMIC VARIABLES

The analyses in C-95 involved 54 socioeconomic variables (SEV) from the 1980 census. Much more extensive data have recently become available in computer transferable form from the 1990 census. About 450 of these were judged to have some possible potential for confounding the lung cancer vs. radon relationship, and these plus 23 others to be used for normalizing were transferred into our data file. As examples of normalizing factors, the number of people age >24 years who are high school graduates is normalized to the population of age >24 years, the number of people of German ancestry is normalized to the number of people reporting an ancestry, the number of people working in each occupation is normalized to the number of employed people of age >16 years, etc. As in this last example, normalization factors are sometimes not exactly appropriate, but this was judged to be unimportant in comparing different counties.

The new SEV are listed in the Appendix along with an item number, which is their column number in our data file. The 54 SEV listed in C-95 were, of course, still retained. To check on the possibility that our SEV are too finely categorized, additional SEV (items 560–577) were generated by combining several adjacent variables on annual income, value of house, ages of the population, etc. As a brief description, our personal characteristic SEV encompass age profile (31 age intervals from 0–1 years to >85 years); annual income (10 intervals from <\$5000/year to >\$150,000 /year); median and average income per capita, per family, and per household; percent below poverty level (persons, families, and households for age >65 years, with children <5 years, and with children 5–17 years); educational attainment of head of household (7 categories from <9th grade to graduate degrees). Our housing SEV include house values (20 intervals from <\$15,000 to >\$500,000), age of houses (8 intervals from <1 year to >50 years); years present occupants lived in house (6 intervals from <1 year to >30 years); housing units in structure [9 intervals from 1 to >50 (large apartment buildings)]; persons in household; age of head of household (7 intervals); number of bedrooms; 13 heating and plumbing characteristics; number of motor vehicles available; telephone availability; rent and rent as percent of income for rented houses; housing costs as percent of income for owner-occupied houses. Our ethnicity SEV include race (5 categories); ancestry (33 nationalities); languages spoken in home (17 categories); year of entry for foreign born (10 intervals); percent born in different state and different sections of US. Other categories of SEV include family type and presence and age of children, school enrollments, labor force characteristics, commuting methods and times, income type, marital status by sex and age, children ever born to women by age and by marital status, residence location 5 years previously, occupation (12 categories), industry of employment (17 categories), etc.

For a confounding factor to have an important effect on the lung cancer mortality rate (m) vs. radon (r) relationship, it must have a relatively strong correlation with m and/or r; we designate the coefficients of correlation with these Corr-m and Corr-r. All SEV for which either |Corr-r| or |Corr-m| for males is greater than 0.316 [because percent correlation = $(\text{Corr})^2$, this means correlation >10%] are listed in Table 1. Note that for every one of the 45 cases, Corr-m and Corr-r are of opposite sign, and only 4 of the 90 values are <0.20 (and these 4 are >0.17). For the 472 SEV under consideration, 81% of the |Corr-r| and 73% of the |Corr-m| are <0.20, so there is clearly a strong tendency for |Corr-m| to be large when |Corr-r| is large, and vice versa, and in all cases they are of opposite sign. This behavior is explained by the strong correlation between m and r seen in Figure 1. In numerical terms (Corr-m for r) = (Corr-r for m) = 0.40.

Table 1. Stratification on SEV With Large |Corr-*r*| or Large |Corr-*m*|

SEV No.	100 × Corr-*r*	100 × Corr-*m*	*B* Males Q_1	*B* Males Q_3	*B* Males Q_5	*B* Females Q_1	*B* Females Q_3	*B* Females Q_5
37	37	−37	−8.4	−8.4	−6.4	−7.0	−4.9	−7.1
82	43	−44	−8.8	−8.9	−5.3	−8.0	−6.4	−5.7
90	−34	39	−4.0	−4.4	−6.4	−7.0	−8.2	−4.1
93	20	−33	−7.9	−6.9	−7.1	−6.8	−5.0	−10.5
98	18	−34	−6.8	−5.7	−7.3	−5.7	−5.8	−7.2
111	32	−36	−7.0	−3.4	−3.5	−7.8	−7.3	−6.4
117	44	−36	−16.6	−7.2	−2.7	−13.0	−7.3	−2.6
120	−20	44	−9.6	−6.5	−1.4	−9.8	−4.9	−1.8
123	27	−40	−3.9	−5.6	−5.6	−5.1	−6.9	−9.3
132	26	−41	−5.0	−6.2	−4.2	−5.8	−9.3	−7.5
135	−24	32	−6.9	−4.8	−2.5	−8.6	−8.9	−1.8
154	37	−45	−5.5	−7.4	−2.7	−6.5	−3.7	−2.6
160	38	−46	−5.6	−5.4	−6.6	−5.4	−2.2	−8.4
178	−24	32	−8.3	−7.0	−2.6	−11.4	−6.6	−5.7
180	39	−46	−3.8	−5.4	−7.2	−3.2	−2.2	−8.7
222	33	−24	−12.6	−6.7	−7.3	−8.4	−6.3	−9.3
227	32	−31	−13.2	−2.7	−5.0	−6.8	−5.1	−10.1
228	28	−44	−8.3	−3.1	−6.5	−5.3	−5.4	−11.8
252	−17	36	−10.5	−8.9	−5.1	−12.0	−9.0	−5.7
282	44	−41	−9.0	−7.7	−5.2	−8.2	−5.0	−5.9
326	18	−34	−6.8	−5.7	−7.3	−5.7	−5.8	−7.2
342	26	−34	−3.0	−6.2	−8.2	−6.5	−7.3	−8.5
346	27	−35	−3.4	−5.2	−6.7	−4.9	−5.6	−8.6
377	32	−44	−2.9	−4.4	−7.4	+0.5	−5.5	−9.9
378	45	−48	−6.9	−7.5	−5.1	−8.1	−4.4	−4.8
388	43	−44	−7.6	−9.2	−5.3	−7.4	−6.5	−5.7
425	−32	37	−5.6	−3.2	−9.0	−5.6	−2.2	−7.5
431	−40	47	−4.7	−4.5	−8.1	−6.6	−2.8	−7.6
434	−22	41	−6.7	−6.1	7.5	−8.5	−6.4	−5.3
455	−23	36	−5.1	−7.6	−6.4	−7.5	−6.5	−8.1
459	41	−48	−8.7	−3.1	−5.7	−7.2	−5.8	−8.7
462	−39	48	−4.7	−3.5	−7.0	−7.9	−5.0	−7.4
463	−33	45	−6.5	−5.6	−10.1	−8.6	−7.5	−6.6
471	−33	42	−5.6	−5.9	−2.4	−9.5	−7.8	−3.7
490	35	−34	−13.8	−7.1	−5.3	−9.6	−8.2	−7.1
491	−36	44	−6.0	−4.4	−4.8	−6.8	−4.9	−8.2
535	26	−35	−2.5	−4.5	−7.8	−3.9	−6.8	−11.3
537	36	−46	−1.7	−6.6	−7.6	−3.3	−8.2	−9.0
538	32	−35	−5.5	−6.9	−7.2	−7.0	−6.9	−6.9
540	−28	36	−8.1	−10.4	−3.7	−8.5	−8.0	−3.3
541	−32	40	−7.3	−7.8	−2.7	−7.9	−7.7	−4.3
543	−32	33	−4.8	−6.1	−3.6	−5.8	−6.5	−5.6
549	43	−49	−3.8	−2.8	−6.2	−3.7	−2.7	−8.4
550	−32	36	−4.7	−7.7	−3.2	−5.9	−6.3	−4.7

B values for first, third, and fifth quintiles of counties from stratification on designated SEV.

To test whether an SEV is an important confounder, we stratify the data into five "quintiles" of 320 counties each, based on that SEV. In each quintile, all values of the SEV are about the same, so its confounding effect is greatly reduced. Analyzing the data in each quintile independently to determine the slope, *B*, then gives values of *B* largely free of confounding by that SEV.

Results of these analyses for the first, third, and fifth quintiles, Q_1, Q_3, Q_5 from stratification on each SEV are listed in Table 1. For Q_3, all values of the SEV are about average and the spread is minimal. For Q_1, the SEV has its smallest values and for Q_5 it has its largest values. Table 1 lists results for (45 SEV × 3 quintiles × 2 sexes = 270 analyses, and in 269 of these cases, B is negative and hence grossly discrepant with the prediction of LNT theory, $B = +7.3$). The one exception, $B = +0.5$, is still below the LNT prediction by 4 SD.

Cases where B is more positive than −3.0 for any Q_1, Q_3, or Q_5 in Table 1 are investigated further by a finer stratification into 10 deciles ($D_1, D_2, \ldots, D_{10}$) of 160 counties each. For these cases, B values for D_1 and D_2, and for D_9 and D_{10} are listed in Table 2. For SEV No. 135, 154, 180, 471, 535, 537, and 541, the finer stratification into deciles does not confirm a strong monotonic relationship between B and the SEV. For SEV No. 178, there does seem to be a trend for males, but not for females. For the three remaining cases (117—percent of German ancestry, 120—percent of Irish ancestry, 377—non-farm self-employed) there does seem to be a definite trend. For all of these, the strong correlations with radon are explainable by the urban–rural effect; people of German ancestry tend to live in rural areas, those of Irish ancestry tend to live in urban areas, and urban people are less likely to be self-employed. But this does not explain the correlation between B values and the SEV; this correlation is not found in other SEV that reflect strong urban–rural differences.

The SEV listed in Table 2 were also investigated by multiple regression (MR). The B value (coefficient of r) in the double regression of m/m_o on r and the SEV are listed in the columns labeled MR in Table 2. In all but one of the 22 cases, the discrepancy with the B value from LNT, $B = +7.3$, is less than the B value from single regression on r but in no case is the lessening of the discrepancy large enough to be important. In fact, even a 12 variable multiple regression of m/m_o on r plus all 11 of the SEV listed in Table 2 gives $B = -4.1 \pm 0.50$ for males and $B = -4.5 \pm 0.73$ for females, still discrepant with the LNT prediction by 23 and 16 SD, respectively. The problem of determining B by multiple regression including numerous variables was investigated in C-95, where it was found that most of the reduction of the discrepancy between these B values and those predicted by LNT is due to methodological problems with multiple regression, and is therefore spurious.

The most important problem with multiple regression is that its relevance depends on a linear relationship. But, for the three suspicious cases, SEV No. 117, 120, and 377, a linear relationship seems quite plausible. This, plus the fact that B values listed in Table 2 for even the most extreme deciles are still much less than the LNT prediction (+7.3), leads one to conclude that these SEV cannot explain an appreciable part of our discrepancy.

Table 2. Finer Stratification and Multiple Regression Results for Selected SEV

	B Males					*B* Females				
SEV No.	D_1	D_2	D_9	D_{10}	MR	D_1	D_2	D_9	D_{10}	MR
117	−16.1	−17.5	−2.6	−2.3	−6.6	−11.0	−15.3	−1.0	−2.7	−6.4
120	−9.2	−9.1	−2.1	+1.3	−5.8	−10.0	−8.3	−2.0	−2.2	−6.4
135	−4.3	−8.6	−10.9	+3.6	−5.9	−6.5	−9.7	−6.7	+2.9	−7.9
154	−7.2	−3.8	−8.9	−2.6	−6.3	−5.3	−7.7	−8.4	−5.6	−6.0
178	−6.7	−9.3	−3.2	−1.6	−6.8	−9.5	−12.2	−8.1	−3.1	−7.7
180	−5.3	−2.2	−10.8	−3.1	−6.1	−1.6	−3.3	−11.1	−3.7	−6.0
377	−5.4	−0.3	−5.6	−8.6	−5.4	0.0	+0.9	−12.8	−7.7	−7.1
471	−1.8	−8.7	−4.6	+0.1	−5.5	−4.4	−13.8	−3.6	−3.8	−6.5
535	−3.8	−1.7	−6.7	−8.4	−6.3	−4.8	−3.3	−10.3	+1.8	−7.4
537	−8.4	+1.4	−10.8	−3.6	−5.9	−8.8	−0.4	−10.2	−6.1	−6.4
541	−6.0	−7.9	−1.7	−4.0	−6.6	−7.1	−8.2	−3.9	−4.7	−7.1

B values from analyses of first, second, ninth, and tenth deciles of counties stratified on designated SEV, and from double regression.

CONFOUNDING BY GEOGRAPHY

The one variable that is known to correlate strongly with radon levels is geography. It can be seen from Figure 1 that county average radon levels vary by an order of magnitude, and in C-95 it was found that *B* values obtained from analysis of separate geographic areas can vary substantially. When 1979–1994 lung cancer statistics are used, these variations were found to be somewhat enhanced.

Table 3 shows *B* values from the various regions and divisions of the US as defined by the Bureau of Census. We see that average *B* values are somewhat less negative than the –7.7 for males and –8.2 for females obtained for the nation as a whole, and the difference is increased if the geographic stratification is finer, into eight "divisions" rather than into only four "regions" (the Pacific division of the West region is not listed as it includes only WA and OR and is dominated by the former, for which results are given in Table 4).

The MINITAB statistical package we use identifies outlying data points as having large standardized residuals and extreme radon levels, which give them a strong influence on *B* values. The *B* values obtained when these are deleted are listed in the last two columns of Table 3. The one case that stands out as important is a single county in New Hampshire (Carroll county) whose deletion changes *B* for females in New England from +9.4 to –0.9. In all other cases, deleting outliers has minimal effect, although always increasing the discrepancy with LNT.

In C-95, going to finer stratification on geography by considering individual states reversed the trend to less negative *B* with finer stratification, but that is not the case here with the 1979–1994 lung cancer rates. Data for individual states are listed in Table 4 for all states with data from more than 15 counties (for all cases listed there are 19 or more). Rather than ignore states with fewer counties, as was done in C-95, the current study combined contiguous states into groups and listed results for these groups at the end of Table 4. This allowed inclusion of all states except Delaware (3 counties) in rather natural groupings, and with no logical alternative groupings (except perhaps by rearranging the New England states). There were only 3 cases where outliers would be deleted by the prescription used in Table 3; the effects of these deletions are listed in Table 4 with asterisks.

It is interesting to ask whether the width of the distributions of *B* values in Table 4, as represented by their SDs (8.0 for males and 8.2 for females), can be explained simply as statistical fluctuations. To test this we take many random selections of *N* counties from our 1601 county data file, determine the *B* value

Table 3. *B* Values for Census Bureau Regions and Divisions

		Males		Females		Outliers Deleted	
	Number Counties	*B*	SD	*B*	SD	*B* Males	*B* Females
Regions							
Northeast	215	–2.8	1.0	–7.5	1.7	–3.4	–7.5
North Central	612	–5.0	0.8	–5.4	1.0	–5.0	–5.7
South	566	–3.9	1.2	–6.4	1.5	–5.1	–7.4
West	204	–8.9	1.4	–10.4	2.3	–9.6	–10.4
Average		–5.2		–7.4		–5.8	–7.8
Divisions							
New England	65	–0.5	4.2	+9.4	5.9	–0.5	–0.9
Mid-Atlantic	150	–3.3	1.1	–9.6	1.9	–4.0	–9.6
East North Central	308	+1.8	1.2	+1.3	1.6	+1.8	+1.3
West North Central	304	–5.7	1.1	–6.8	1.4	–5.7	–6.8
South Atlantic	273	–6.9	1.9	–7.1	2.4	–7.7	–7.1
East South Central	135	+3.2	1.5	–0.9	2.9	+2.3	–0.9
West South Central	158	–15.5	2.7	–17.4	2.8	–12.9	–17.4
Mountain	167	–5.9	1.6	–5.5	2.6	–5.9	–5.5
Average		–4.1		–4.4		–4.1	–5.9

Table 4. *B* Values for Individual States

State	Number of Counties	Males *B*	Males SD	Females *B*	Females SD
AL	38	−0.8	3.5	+1.0	5.3
AR	36	−10.9	4.2	−8.5	5.9
CO	41	−5.8	3.1	−1.7	3.6
GA	52	−1.8	5.0	−22.2	6.0
ID	40	−5.3	3.2	−8.7	5.7
IL	54	−4.1	3.1	−3.1	2.8
IN	58	−2.8	2.7	−1.4	4.3
IA	98	−1.9	1.7	−0.7	2.7
KS	21	−6.9	5.2	−7.0	5.0
KY	32	+0.8	2.6	−7.3	7.9
LA	21	−35.2	20.8	−30.4	32.3
MD	24	−13.6	5.9	−19.4	6.6
MI	53	+2.2	2.3	+2.4	3.0
MN	64	−4.6	1.4	−10.2	2.7
MS	19	+1.2	10.8	−1.6	16.2
MO	31	+1.6	8.5	+5.6	9.7
NE	43	+5.5	3.3	+5.7	4.9
		+8.5*			
NJ	21	−2.3	3.9	+3.3	5.8
NM	30	−6.3	3.8	−2.8	5.3
NY	62	+1.7	3.0	+3.5	4.6
NC	54	−5.7	3.2	−4.3	3.1
ND	38	+0.7	3.5	−0.3	5.0
OH	88	−0.2	2.0	−1.0	3.6
OK	43	−9.1	8.4	−3.8	6.8
PA	67	−1.4	1.3	−4.7	2.5
SC	36	−18.8	6.5	−1.6	12.2
TN	46	+1.7	3.0	−0.2	3.8
		−2.2*			
TX	58	−12.7	3.3	−16.9	4.0
VA	66	−6.2	4.0	−7.1	5.1
WA	29	+5.0	3.6	+6.0	8.6
WV	37	+5.4	4.6	+1.4	6.1
WI	55	−7.3	4.4	−4.5	6.2
WY	21	−7.1	6.2	+0.2	7.5
CT+MA+RI	25	−8.3	10.2	+1.6	12.8
ME+NH+VT	40	−1.0	4.9	+9.4	7.4
				−3.3*	7.7*
NV+OR+UT	32	−20.0	4.7	−10.4	10.0
MT+SD	20	−12.6	4.1	−8.5	8.5
Average		−5.0	1.3	−4.0	1.3
		−5.0*		−4.3*	
SD of distribution		8.0		8.2	

for each selection, and determine the SD of the distribution of *B* values thus determined. For example, for $N = 30$, we analyzed 60 different random selections of 30 counties; the distribution of the 60 derived *B* values had SD of 4.3 for males and 5.6 for females. Because the number of counties in the 37 states listed in Table 4 averaged 43 and three fourths of these numbers were more than 30, we would expect the SD of the *B* values in Table 4 to be smaller than 4.3 for males and 5.6 for females if statistics was the only consideration. The fact that this conflicts with observation means that other factors are involved. By doing calculations of the type described above for different N values, we found that the SD of about 8 found for the B values in Table 4 was what would be expected if the data for each state included only about 12 counties randomly selected from our file.

This can be understood from the fact that radon levels in the counties of a given state are not independent of one another. Because radon levels are determined by geology, and neighboring counties frequently have similar geology, the effective number of independent areas in a given state is much smaller than the number of counties. The above results suggest that a state with 43 counties typically contains about 12 truly independent areas that serve as independent data points.

This implies that the SD for all B values derived in this article should be nearly twice as large as the SD derived from our statistical treatment. For example, the B values for our entire data set, given as -7.7 ± 0.51 for males and -8.7 ± 0.67 for females, should be viewed more properly as -7.7 ± 1.0 and -8.7 ± 1.3, respectively. This, of course, does little to resolve our discrepancy with the LNT prediction: +7.3.

Another aspect of Table 4 that begs for an explanation is that the average values of B (-5.0 ± 1.3 for males and -4.0 ± 1.3 for females) are quite discrepant with the values for our entire data file (−7.7 and −8.2, respectively). Of course, all SD given for B values in this article ignore uncertainties in the data points for individual counties, so these SD are somewhat understated. But none of our sets of random selections of counties had average B values differing nearly as much from the B values for our entire 1601 county data set.

It thus seems that, for some reason we cannot explain, B values for more restricted areas tend to be less negative than for the nation as a whole. However, even these less negative B values are still grossly discrepant with the LNT prediction: $B = +7.3$.

CONCLUSION

In this article we have introduced a great deal of new data, lung cancer statistics for all races for 1979–1994 replacing those for whites only between 1970 and 1979, and about 450 new potential confounding factors in addition to the 60+ used in C-95. The results are essentially the same as in C-95, indicating failure of the linear-no threshold theory in grossly exaggerating the risk of low-level exposure to radon.

APPENDIX: NEW SOCIOECONOMIC VARIABLES (SEV) FROM 1990 CENSUS

Item	Description
79	Total population
80	Percent urban
81	Percent rural
82	Percent living on farms
	School enrollment
83	Persons in schools
84	Percent of these in preprimary
85	Percent of these in elementary or high school
86	Percent of these in private schools
87	Percent of these in college
	Educational attainment (percent of age >25)
89	<9th grade
90	9th–12th grade, no diploma
91	High school graduates
92	Some college
93	Associates degree
94	Bachelors degree
95	Graduate or professional degree
96	High school graduates or higher
97	Bachelors degree or higher
98	Children ever born/1000 woman age 35–44
99	Percent of population born in US
100	Percent born in state of residence
101	Percent born outside US
103	Percent who speak a language other than English
104	Percent who do not speak English well
	Ancestry (1st or 2nd) (percent of those reporting)
106	Arab
107	Austrian
108	Belgian
109	Canadian
110	Czech
111	Danish
112	Dutch
113	English
114	Finnish
115	French
116	French Canadian
117	German
118	Greek

119 Hungarian
120 Irish
121 Italian
122 Lithuanian
123 Norwegian
124 Polish
125 Portuguese
126 Rumanian
127 Russian
128 Scotch-Irish
129 Scottish
130 Slovak
131 Sub-Sahara African
132 Swedish
133 Swiss
134 Ukrainian
135 American
136 Welsh
137 West Indian
138 Yugoslavian
139 Other Ancestries

<u>Labor force</u>
140 Percent in labor force
141 Percent of males in labor force
142 Percent of males unemployed
143 Percent of females in labor force
144 Percent of females unemployed

<u>Occupation</u>
146 Executive, administrative, managerial
147 Professional specialty
148 Technicians
149 Sales
150 Administrative support
151 Private household occupations
152 Protective services
153 Service occupations (not 151,152)
154 Farming, forestry, fishing
155 Precision production, craft, repair
156 Machine operators. assemblers, inspectors
157 Transportation
158 Handlers, helpers, cleaners, laborers

<u>Industry (percent employed in)</u>
160 Agriculture
161 Mining
162 Construction
163 Manufacturing, nondurable
164 Manufacturing, durable
165 Transportation
166 Communication, other utilities
167 Wholesale trade
168 Retail trade
169 Finance, insurance, real estate
170 Business services, repair services
171 Personal services
172 Entertainment, recreation
173 Health services
174 Educational services
175 Other professional services
176 Public administration

<u>Class of worker (percent of those employed)</u>
178 Private wage and salary
179 Government
180 Self-employed
181 Unpaid family workers

<u>Household income (percent in each range)</u>
183 <$5000
184 $5000–$9999
185 $10,000–$14,999
186 $15,000–$24,999
187 $25,000–$34,999
188 $35,000–$49,999
189 $50,000–$74,999
190 $75,000–$99,999
191 $100,000–$149,999
192 >$150,000

<u>Annual income in dollars</u>
193 Median household
194 Median family (Fam)
195 Median family (Fm)
196 Per capita average
197 Mean wage and salary
198 Mean non-farm self-employment
199 Mean farm self-employment
200 Mean social security
201 Mean public assistance
202 Mean retirement

<u>Percent below poverty level</u>
203 All persons
204 Persons age >17
205 Persons age >64
206 With related children age <5
207 With related children age 5–17
208 All families
209 With related children age <18
210 With related children age <5
211 Female householder families
212 With related children age <18
213 With related children age <5

<u>Percent of housing units built in time periods</u>
215 1989–1990
216 1985–1988
217 1980–1984
218 1970–1979
219 1960–1969
220 1950–1959
221 1940–1949
222 Before 1939

<u>Percent with various numbers of bedrooms</u>
223 No bedroom
224 One

225 Two
226 Three
227 Four
228 Five or more

House characteristics (percent in category)
229 Lacking complete plumbing facilities
230 Lacking complete kitchen facilities
231 Part of condominium
232 Water from public system
233 Water from individual drilled well
234 Water from individual dug well
235 Public sewer
236 Septic tank
237 Heated by utility gas
238 Heated by bottled gas, tank, or LP gas
239 Heated by electricity
240 Heated by oil
241 Heated by coal or coke
242 No heating fuel

Year householder moved into house
243 1989–90
244 1985–88
245 1980–84
246 1970–79
247 1960–69
248 Before 1960

249 No telephone in unit
251 No vehicle available
252 One vehicle available
253 Two vehicles available
254 Three or more vehicles available
255 Median mortgage (dollars)
256 Net median mortgage (dollars)

Owner costs as percent of household income
258 <20%
259 20%–24%
260 25%–29%
261 30%–34%
262 35% or more

Rent (percent of rented units in range)
264 <$200
265 $200–$299
266 $300–$499
267 $500–$749
268 $750–$999
269 $1000 or more
270 Median rent

Rent as percent of household income
272 <20%
273 20%–24%
274 25%–29%
275 30%–34%
276 35% or more

278 Percent of housing units occupied
279 Percent of housing units vacant
280 Percent urban inside urban area
281 Percent urban outside urban area
282 Percent rural—farm
283 Percent rural—non-farm

Units in structure (percent in category)
284 1, detached
285 1, attached
286 2
287 3 or 4
288 5–9
289 10–19
290 20–49
291 50 or more
292 Mobile home or trailer

Value of owner occupied units (percent in range)
293 <$15,000
294 $15,000–$19,999
295 $20,000–$24,999
296 $25,000–$29,999
297 $30,000–$34,999
298 $35,000–$39,999
299 $40,000–$44,999
300 $45,000–$49,999
301 $50,000–$59.999
302 $60,000–$74,999
303 $75,000–$99,999
304 $100,000–$124,999
305 $125,000–$149,999
306 $150,000–$174,999
307 $175,000–$199,999
308 $200,000–$249,999
309 $250,000–$299,999
310 $300,000–$399,999
311 $400,000–$499,999
312 $500,000 or more
313 Median value
314 Median value with mortgage
315 Median value not mortgaged

Residence five years earlier (percent in category)
317 Same house
318 Different house in US
319 Different house in same state
320 Different house in same county
321 House in different county, same state
322 House in different state
323 Lived abroad

Children ever born per 1000 women
324 Women age 15–24 years
325 Women age 25–34 years
326 Women age 35–44 years

328 Percent of foreign born entered US 1980–1990

330 Percent of population speaking other than English
331 Percent not speaking English "very well"
332 Percent speaking Spanish
333 Percent speaking Spanish, not English "very well"
334 Percent speaking Asian language
335 Percent speaking Asian, not English "very well"
336 Civilian labor force
337 Percent of civilian labor force employed
338 Percent of civilian labor force unemployed
339 Percent of civilian labor force unemployed
340 Percent of labor force in armed forces

Commuting to work
342 Percent drive alone
343 Percent in carpools
344 Percent using public transportation
345 Percent using other means
346 Percent walk or work at home
347 Mean travel time (minutes)
349 Percent of employment by government
350 Percent employment by local government
351 Percent employment by state government
352 Percent employment by federal government

Percent of family households with income of
354 <$5000
355 $5000–$9999
356 $10,000–$14,999
357 $15,000–$24,999
358 $25,000–$34,999
359 $35,000–$49,999
360 $50,000–$74,999
361 $75,000–$99,999
362 $100,000–$149,999
363 >$150,000

Percent of nonfamily households with income of
365 <$5000
366 $5000–$9999
367 $10,000–$14,999
368 $15,000–$24,999
369 $25,000–$34,999
370 $35,000–$49,999
371 $50,000–$74,999
372 $75,000–$99,999
373 $100,000–$149,999
374 >$150,000

Percent of households with income from
376 Wage & salary
377 Non-farm self-employment
378 Farm self-employment
379 Social Security
380 Public assistance
381 Retirement plan

386 Percent of persons urban, inside urbanized area
387 Percent of persons urban, outside urbanized area
388 Percent of persons rural, farm
389 Percent of persons rural, non-farm
390 Percent of persons male
391 Percent of persons female

Percent of persons in age range
392 <1 year
393 1–2 years
394 3–4 years
395 5 years
396 6 years
397 7–9 years
398 10–11 years
399 12–13 years
400 14 years
401 15 years
402 16 years
403 17 years
404 18 years
405 19 years
406 20 years
407 21 years
408 22–24 years
409 25–29 years
410 30–34 years
411 35–39 years
412 40–44 years
413 45–49 years
414 50–54 years
415 55–59 years
416 60–61 years
417 62–64 years
418 65–69 years
419 70–74 years
420 75–79 years
421 80–84 years
422 85 years and over

Percent of persons age >14 in category
423 Male, never married
424 Male, married, spouse present
425 Male, married, spouse absent, separated
426 Male, married, spouse absent, other reasons
427 Male, widowed
428 Male, divorced
429 Female, never married
430 Female, married, spouse present
431 Female, spouse absent, separated
432 Female, spouse absent, other reasons
433 Female, widowed
434 Female, divorced
435 Female, never married, age 15–24
436 Femate, never married, age 25–34
437 Female, never married, age 35–44
438 Female, never married , age 45 and over
439 Female, ever married, age 15–24
440 Female, ever married, age 25–34
441 Female, ever married, age 35–44
442 Female, ever married, age 45 and over

Percent of persons in group quarters
443 Correctional institutions
444 Nursing homes
445 Mental hospitals
446 Juvenile institutions
447 Other institutions
448 College dormitories
449 Military quarters
450 Shelters for homeless
451 Visible in street locations

Percent of persons born in
452 State of residence
453 Other state in Northeast
454 Other state in Midwest
455 Other state in South
456 Other state in West

Percent of families with
458 Married couple, children >17 years
459 Married couple, no children >17 years
460 Male householder, no wife, children >17 years
461 Male householder, no wife, no children >17 years
462 Female householder, no husband, children >17 years
463 Female householder, no husband, no children >17 years
464 No workers
465 1 worker
466 2 workers
467 3 or more workers

Percent of households with
469 1 person
470 2 persons
471 3 persons
472 4 persons
473 5 persons
474 6 persons
475 7 persons or more

Percent of family households with householder age
476 15–24 years
477 25–34 years
478 35–44 years
479 45–54 years
480 55–64 years
481 65–74 years
482 75 years and over

Percent of nonfamily households with householder age
483 15–24 years
484 25–34 years
485 35–44 years
486 45–54 years
487 55–64 years
488 65–74 years
489 75 years and over

Percent of persons
490 White
491 Black
492 American Indian or Aleut
493 Asian or Pacific Islander
494 Other race
495 Hispanic origin, white
496 Hispanic origin, black
497 Hispanic origin, American Indian
498 Hispanic origin, Asian or Pacific Islander
499 Hispanic origin, other race

Percent of foreign-born persons entered US in
500 1987–1990
501 1985–1986
502 1982–1984
503 1980–1981
504 1975–1979
505 1970–1974
506 1965–1969
507 1960–1964
508 1950–1959
509 Before 1950

Language spoken in home
510 Only English
511 German
512 Yiddish
513 Other West Germanic
514 Scandinavian
515 Greek
516 Italian
517 French
518 Spanish
519 Polish
520 Russian
521 Chinese
522 Hungarian
523 Japanese
524 Korean
525 North American Indian
526 Vietnamese

Transportation to work
527 Drive alone
528 Carpool
529 Bus or trolley bus
530 Street car or trolley car
531 Subway or elevated
532 Railroad
533 Motorcycle
534 Bicycle
535 Walk
536 Other

Travel time to work (minutes)
537 Less than 5
538 5–9
539 10–14
540 15–19
541 20–24
542 25–29

543 30–34
544 35–39
545 40–44
546 45–59
547 60–89
548 90 or more
549 Work at home
550 Total

COMBINED CATEGORIES

560 Household income <$25,000
561 Household income $25,000–$75,000
562 Household income >$75,000
563 Housing built after 1979
564 Housing built 1950–1979
565 Housing built before 1950
566 10 or more units in structure

Owner-occupied housing unit value

567 <$40,000
568 $40,000–$100,000
569 $100,000–$250,000
570 >$250,000

Percent of population in age range

571 0–17 years
572 18–34 years
573 35–64 years
574 65 years or more

Percent of foreign born entered US

575 Since 1979
576 1965–1979
577 Before 1965

REFERENCES

1. Cohen, B. L. Test of the linear-no threshold theory of radiation carcinogenesis for inhaled radon decay products. Health Phys. 68:157–174; 1995.
2. Lubin, J. H. On the discrepancy between ecologic studies in individuals of lung cancer and residential radon and Cohen's ecologic regression. Health Phys. 75:4–10; 1998.
3. Centers for Disease Control and Prevention. Wonder searches and queries; 1997. http://wonder.cdc.gov/rchtml/Convert/data/AdHoc.html
4. Cohen, B. L.; Colditz, G. A. Test of the linear-no threshold theory of radiation carcinogenesis. Environ. Res. 64:65–89; 1994.