HIGH RADON HOUSES: IMPLICATIONS FOR EPIDEMIOLOGY AND RISK ASSESSMENT

Robert Goble* and Robert Socolow

January 1990

PU/CESS Report No. 245

A short version of this report will appear in the Proceedings of the International Symposium on Radon and Radon Reduction Technology, February 19-23, 1990, Atlanta, GA.

*CENTED, Clark University, Worcester, MA 01610

Center for Energy and Environmental Studies
School of Engineering and Applied Science
Princeton University
Princeton, NJ 08544

HIGH RADON HOUSES: QUESTIONS ABOUT LOG NORMAL DISTRIBUTIONS AND IMPLICATIONS FOR EPIDEMIOLOGY AND RISK ASSESSMENT

R. Goble
Center for Technology, Environment, and Development
Clark University, Worcester, MA 01610

R. Socolow
Center for Energy and Environmental Studies
Princeton University, Princeton, N.J. 08544

ABSTRACT

There is a very real possibility that, for the United States as a whole, the number of houses with indoor radon concentrations tens or hundreds of times greater than the national median is greatly underestimated by the traditional log-normal distributions used to characterize radon levels in houses. A fat tail could imply that a significant fraction of the total national health impact of indoor radon is being experienced by those living in very high radon houses, contrary to the conclusions drawn from log-normal distributions. There is also a very real possibility that geographically delimited distributions (across, say, on the order of a million houses in a region) have fat-tailed distributions, relative to log-normal. Fat-tailed distributions would permit case-control epidemiological studies to have much greater power, for a given level of effort, than is now presumed on the basis of log-normal distributions.

HIGH RADON HOUSES: QUESTIONS ABOUT LOG NORMAL DISTRIBUTIONS AND IMPLICATIONS FOR EPIDEMIOLOGY AND RISK ASSESSMENT

Robert Goble* and Robert Socolow

February 1990

PU/CEES Report No. 245

To appear in the Proceedings of the International Symposium on Radon and Radon Reduction Technology, February 19-23, 1990, Atlanta, Georgia.

A longer version of this paper is Cented Research Report No. 5, Hazard Assessment Group, Center for Technology, Environment, and Development (CENTED), Clark University, Worcester, MA 01610-1477.

*CENTED, Clark University, Worcester, MA 01610

Center for Energy and Environmental Studies
School of Engineering and Applied Science
Princeton University
Princeton, NJ 08544

INTRODUCTION

By now many measurements of radon concentrations in houses have been made. They have been taken by many different groups for different purposes, using different methods, different protocols, in different geographical areas, under different conditions. Many of them have been taken for "screening" purposes with procedures intended to maximize the observed radon levels. To what extent is it feasible to formulate a picture of how radon concentrations are actually distributed across a sample of houses?

At present the standard approach is based on the use of log-normal distribution functions. For many data sets they provide, as we shall see, a good fit covering the bulk of the data. They do not, however, provide very secure guidance for estimating numbers of houses with high concentrations of radon. In this paper we discuss evidence that indicates that the numbers of high radon houses may be underestimated by log-normal fits: we suggest an alternative approach to analyzing the data, and describe some of the reasons it would be desirable to make more realistic estimates for the number of houses in various samples with high concentrations. High radon houses are important in two contexts: in the area of public policy, they constitute an environment in which the public may be exposed to risks that are considered unacceptable, or strongly undesirable, and they may well represent the source of a significant fraction of the population exposure to inhaled radon decay products; in the broader public health area, they may provide the best opportunity to learn about radiation-induced carcinogenesis in the general public. A lengthier discussion by us of fitting high radon distributions and their implications is available (1).

THE LOG-NORMAL FITS TO RADON MEASUREMENTS IN HOUSES

Because of the differences in measurement methods and protocols, and, especially, because of differences in measurement goals, identification of a representative national sample for determining a national distribution of radon levels has not been easy. Results from two national surveys specifically addressing these difficulties have been published. They follow two different approaches: one, by Cohen (2), is an attempt to use a uniform protocol and consists of a measurement of about 450 homes of physics professors; the second, by Nero et al. (3) is a compilation of 800 measurements taken from nearly 40 data sets, where the compilers made a substantial analysis of the differences in measurement specifications. The two surveys are remarkably consistent: they find median concentrations of approximately 1 pCi/l, and the distributions fit surprisingly well a log-normal form with a geometric standard deviation of 2.4-2.8.

In Figure 1, we show Nero's fit to the data. The log-normal fit appears excellent, especially when one remembers that it is determined by only two parameters, the geometric mean and standard deviation. The excellence of the fit can, however, distract one from worrying about how well the tail of the log-normal distribution will match the numbers of high radon houses. Because the distribution is determined by only two parameters and because high radon houses constitute a small portion of the data, the log-normal fit essentially predicts the number of high radon houses based on the distribution of radon concentrations relatively near the median. There is no particular reason to expect such a prediction to hold. Neither the Cohen nor the Nero study serves to test the prediction because of the relatively small numbers of houses included.

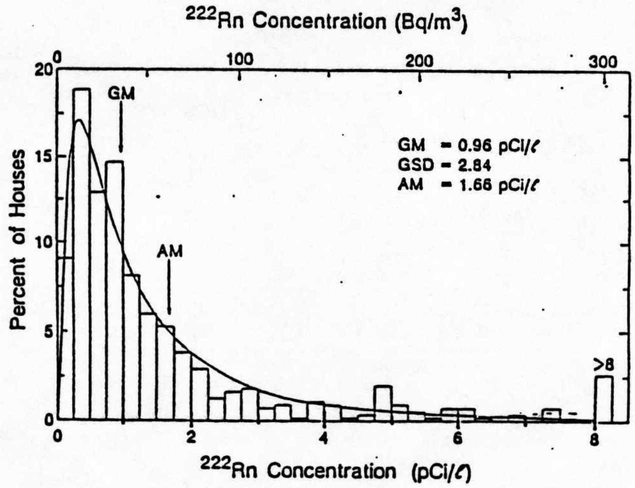

Figure 1. Log-normal fit to Nero's compilation of national radon data.

EVIDENCE FOR FATTER-TAILED DISTRIBUTIONS

Before discussing the evidence that log-normal distributions are likely to underestimate seriously the numbers of houses with high concentrations, it is worth stressing the problematic quality of the data. The radon concentrations of interest are not simple to define. The connection between house measurements and exposure to radon decay products is, unfortunately, neither direct nor certain. Most reported measurements are of radon concentrations taken in one or two locations of the house averaged over days, weeks, or months, since such long-term averages are obtained with relatively low cost methods. Most have been made for "screening" purposes with procedures intended to maximize the measured level (basement measurements, for instance, with windows kept closed for the measurement period). Research performed with more elaborate instrumentation has revealed that radon concentrations in a house can change substantially within a few hours (associated with rain or with diurnal changes in the difference between the indoor and outdoor temperature) and that there are seasonal variations caused by changes in soil conditions. Perhaps even more important are the large differences found in radon levels in different parts of a house at any particular time. Figure 2A shows some characteristic fluctuations in radon levels in a basement. Figure 2B illustrates the sort of variation which may be found in radon levels at different places in a house. Clearly where and when people spend their time at home will strongly influence the exposure they receive.

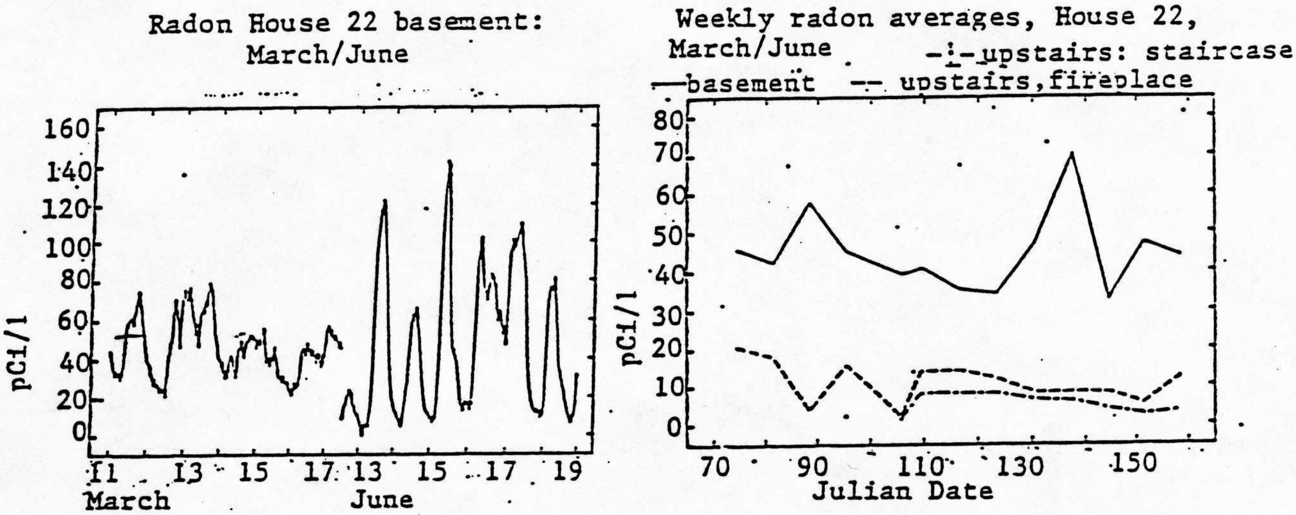

Figure 2. Examples of spatial and temporal variation in radon concentrations in a house.

Despite such problems with using general radon surveys for quantitative assessments, some recently available large data sets taken under non-representative circumstances provide a useful perspective on the distribution of radon exposures. One is a set of nearly 20,000 screening measurements conducted by the Pennsylvania Bureau of Radiation Protection (4) in five counties which include the Reading prong, the large geological feature on which many high radon houses have been found. A second is summary data of measurements made nationwide by the Terradex corporation (5). The particular subset of their data which we use is a compilation of about 10,000 measurements made throughout the U.S. from which 6 states (ME, NJ, NY, OR, PA, WA) were excluded. We chose this subset because it avoids areas where there have been large numbers of measurements within a small region and because the total Terradex compilation very substantially overlaps the Pennsylvania study.

We present in Figure 3 a reanalysis of the Nero et al. compilation, the Reading prong area data, and the Terradex subset.

On a log-log scale, the percent of houses having greater than the indicated concentration is plotted against that concentration, the latter made dimensionless by dividing by the median concentration. Since the median in the Nero data set is close to 1.0 pCi/liter, the x variable may be regarded as the concentration in pCi/liter. The median for the Reading Prong data set is roughly 5 pCi/l, while it is approximately 1.7 pCi/l for the Terradex data. Because we are interested in the high radon tail, we plot the data only for concentrations greater than four times the median concentration of that data set. This starting point is arbitrary, but we have found it useful for a large number of data sets. When the median value is 1 pCi/liter, the starting point is the EPA action level.

Together with each data set we show four theoretical distributions on each plot of Figure 3. Two are logarithmic normal distributions, named LnNrm 2.8 and LnNrm 3.5, which have geometric standard deviation 2.8 (Nero's fit (3)) and 3.5 (Alter's fit (5)) respectively. Two are power law distributions, named PWR 1.75 and PWR 1.25. They are specified by an exponent, 1.75 and 1.25, such that the percent of houses with concentration greater than x falls off as 1/x raised to that exponent.

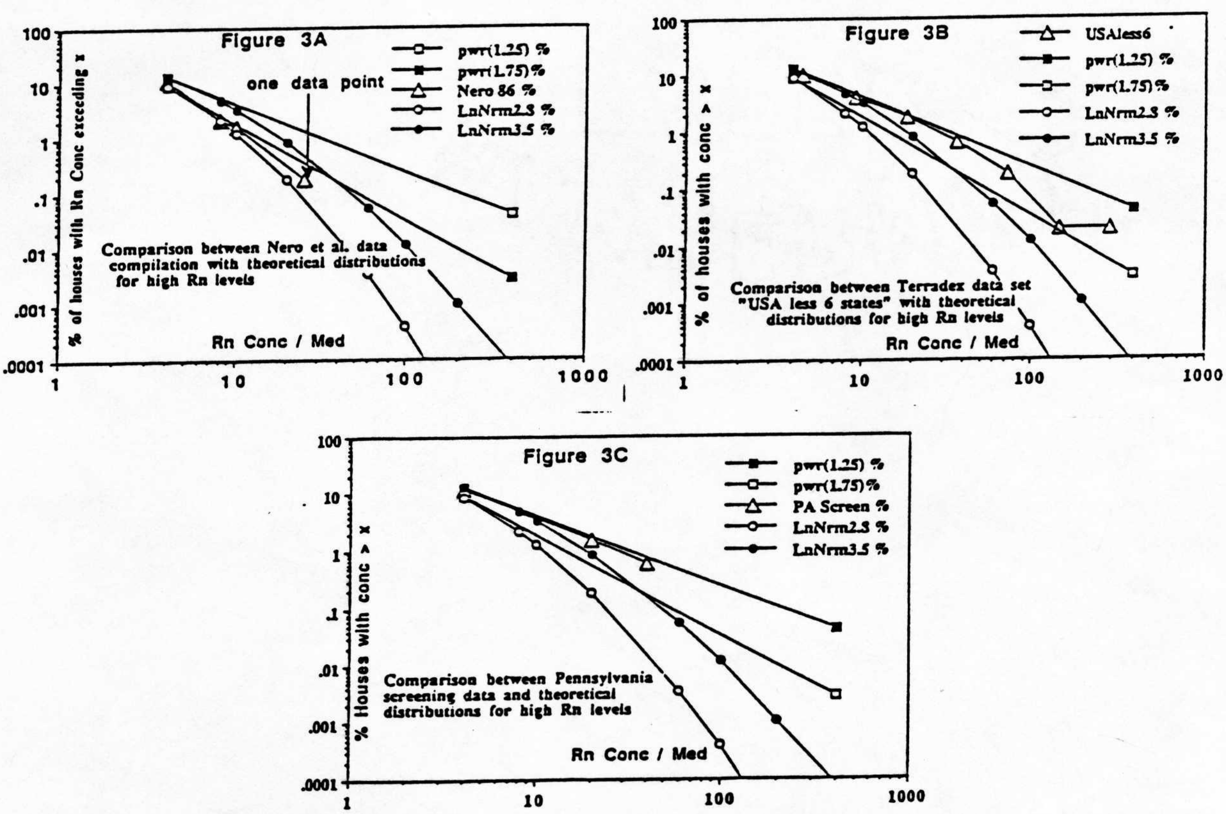

Figure 3. Comparison of four theoretical high radon distributions with three data sets.

The three high radon data sets in Figure 3 are reasonably well fit by a power law with an exponent bracketted by 1.75 and 1.25 and with the fractional number of houses with concentrations greater than four times the median roughly between 10 and 15%. The high radon tails of the Reading prong and Terradex data clearly appear fatter than log normal. In fact, log normal distributions fall off sufficiently at high levels that more houses with concentrations greater than 100 pCi/l have already been found in these studies than are predicted by the Nero et al. log normal fit for the entire nation, and the discrepancy at higher levels is even worse. Cohen's most recent survey results (6) support a power law distribution as well. He finds deviations from log normal; the deviations are greater the higher the observed levels, and his reported distributions for concentrations from 100 to above 1000 pCi/l are consistent with the range of power law exponents considered here. We also have found similar results in fitting a number of other non-representative survey data sets.

To summarize: in the absence of large surveys with protocol designed to assure the representativeness of the data set, these fits must be considered suggestive rather than conclusive evidence that houses with high radon concentrations will appear more frequently than expected from log-normal fits. For the remainder of the paper we argue that if the tails of the distribution are indeed fatter, there will be important applications. Thus careful study of high radon distributions is called for.

DISTRIBUTIONS OF RISK

Three separate questions about the distribution of radon concentrations in houses across the nation continually appear in policy analyses of the indoor radon problem. One question is how many high radon houses are there? What is the size of the population for which we have various levels of concern, or for which protective measures might be addressed? A second question is how are radiation exposures distributed across the population? What fraction of the overall dose to the public is accumulated in houses at various levels of exposure? The third question is how well known are the average levels of exposure?

In Table 1, we show how the answers to these questions differ for the four theoretical distributions shown in Figure 3. For each distribution we display representative values of the % of houses exceeding benchmark radon levels, the fraction of total dose contributed by houses in each category, and the ratio of the arithmetic mean to the median dose.

TABLE 1: FRACTION OF HOUSES WITH HIGH RADON CONCENTRATIONS AND THE FRACTION OF THE AGGREGATE RADIATION DOSE ASSOCIATED WITH THOSE HOUSES FOR FOUR THEORETICAL MODELS.

Concentrations of radon/median	% houses with conc. >	% of dose from these houses	% houses with conc. >	% of dose from these houses
	Fit: PWR(1.25)		Fit: LnNrm (gsd = 3.5)	
4	13.	72.	13.4	56.
10	4.1	57	3.3	28
20	1.7	48	.84	12.8
60	.44	37	.054	2.2
100	.23	32	.012	.77
200	.098	27	.0012	.15
arithmetic mean/median		3.6		2.2
	Fit: PWR(1.75)		Fit: LnNrm (gsd = 2.8)	
4	9.	44.	8.9	38.
10	2.7	22	1.3	11.
20	.54	13.	.18	3.
60	.079	5.8	.0035	.2
100	.032	3.9	.00029	.03
200	.0096	2.4	.00001	.002
arithmetic mean/median		1.9		1.7

Table 1 shows that if a distribution of radon concentrations across high radon houses is similar to one of the power law distributions:

1) the 10% to 12.5% (6-7 million) of houses with concentrations greater than 4 times the median may contribute half or more (50% to 70%) of the expected aggregate exposure,

2) the roughly 1% (600,000) of houses with concentrations greater than 20 times the median could be expected to contribute 15% to 45% of the aggregate exposure, and,

3) as many as 20,000-120,000 houses may have levels greater than 100 times the median and contribute 5% to 30% to the aggregate exposure.

These values for the power law fits differ strikingly from those for log normal distributions, where only a few tenths of a percent of houses have concentrations greater than 20 times the median and these contribute only 2% to 6% to the aggregate exposure. Only the values for log normal distributions are considered in EPA's present perspective (7). The fatter tail also increases the ratio of mean to median dose, but the difference is substantial only for the slowly falling (exponent = 1.25) power law.

IMPLICATIONS FOR EPIDEMIOLOGY

The presence of substantial numbers of houses with radon concentrations that present high risks suggests that there should be a significant number of lung cancers associated with those houses and that such an association should be observable. Thus far, attempts to observe such associations between environmental radon exposure and lung cancer have been ambiguous at best (8,9). There are, however, several reasons why it is difficult to observe the association.

- It is difficult to estimate exposures:
  - people move from house to house
  - the most recent 5-10 years of exposure may be irrelevant
  - patterns of occupancy in a house strongly affect exposures
  - radon concentrations vary widely in houses in a single neighborhood
  - the characteristics of houses may change as a result of changes in heating systems or other modifications
- Radon causes only a fraction of lung cancers
- There is great regional and local variability in other causes of lung cancer, particularly smoking.

Fortunately, in our view, these complications have not prevented some case-control studies from getting under way (10). The public health community is pessimistic that radon-induced lung cancer will be detectable but it has been willing to look. Successful studies could be very valuable. A determination of the rate at which environmental radon causes cancer should clarify significantly the debate on public policy. Environmental exposures will be the only source of reasonable data on effects of exposures on women, and on exposures of children and the elderly; they are the most likely source of information to clear up the controversies about the interactive effect of smoking.

Suppose that we postulate a geographical region where high radon houses follow the sort of power-law distribution considered above, and further postulate that cancer incidence is proportional to radon dose (the "linear" hypothesis which underlies so much of radiation risk assessment). In such a region the capability of epidemiological studies to determine the effects of environmental radon exposures is greatly enhanced, relative to the capability of such studies in a region where high radon houses obey a log-normal distribution.

The argument is a natural extension of the analysis presented in Table 1: the larger the fraction of the total radon dose to the population absorbed by those living in the houses with the very highest radon levels (say, in the houses with radon levels 20 or more times the median, which may be roughly one percent of the houses), the more frequently research seeking to link lung cancer to radon in houses will find that the trail leads to high radon houses.

The case-control experiments we consider here would begin with a patient ("case") who has been diagnosed with lung cancer and a control who is matched with appropriate characteristics. The key step in the research is to establish the radon exposures in the houses in which the case and the control lived. Greatly complicating these studies are the difficulties involved in documenting these exposures, as discussed above.

We present an example to show numerically how much more productive such an investigation will be whenever the power-law rather than the log-normal distribution more closely represents reality. We find that confirmation of today's expectations about the dose-response relationship is obtainable only from studies involving tens of thousands of matched cases and controls if there is a log-normal distribution of high-radon houses, but will be achievable from studies involving roughly one thousand matched cases and controls if there is a power-law distribution of high-radon houses.

The example assumes that 10% of cancers in the sample are radon induced and that the dose response relation is linear and independent of age at exposure except for a latency period. These assumptions are consistent with EPA and other risk estimates. To obtain representative numbers, we invent a simple and arbitrary model of exposure histories.

- Every case and control moves precisely every ten years, and dies at age 70.

- A ten-year latency period is assumed,

- Accurate exposures could be established for only the most recent three of seven houses: consequently, the most recent house is not of interest, precisely two of the pertinent six houses are investigated, and thus one-third of the integrated effective radon exposure is documented.

- There is no correlation in house radon level across moves.

From this model, distributions of radon levels in housing "data" were derived for both cases and controls using each of the four theoretical distributions for radon levels in houses. The predicted fraction of cases and of controls whose measured exposures would exceed a value x is plotted against the exposure x, for two of the distributions, PWR (1.25) and LnNrm (3.5) in Figures 4A,4B. The numerical details of this exercise with several further examples are summarized in our larger paper (1).

Figure 4A

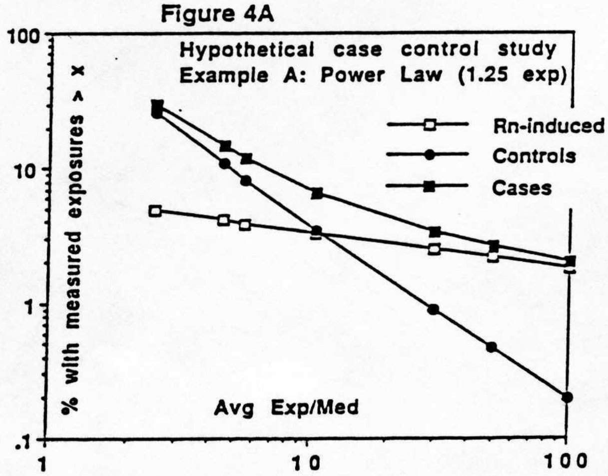

Figure 4B

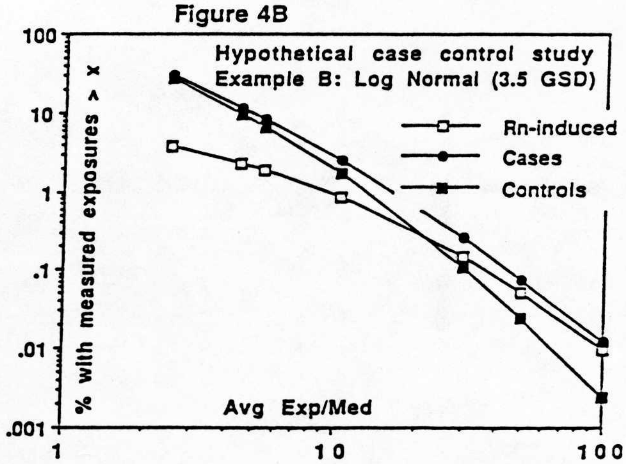

Figure 4. Calculated distributions of exposure for two hypothetical case control studies.

The main feature of each Figure 4 is the crossover, the location at which the slope of the graph becomes less steep. This crossover appears in the lung cancer cases, as measured radon exposures become more important; it is absent in the controls. It is also much less apparent for log-normal distributions; although there is a shift in slope for the cases, it does not look like a break since the graph is already falling off rapidly.

One good identifier of the crossover is the value of x, call it x_0, above which half of the cases are associated with measured radon. For each theoretical distribution, Table 2 tabulates x_0 and gives the fraction of the houses above x_0; the larger this fraction, the smaller the sample size necessary to estimate the dose response coefficient to any given level of confidence.

To calculate actual, rather than relative, sample sizes, it is necessary to declare the size of some subsample desired. We choose the following power-of-experiment condition to establish sample sizes:

> At the radon concentration in houses above which half of the cases are radon-associated, there shall be 20 cases (10 radon induced).

There will then be roughly 10 controls. The corresponding sample sizes for each distribution are also tabulated in Table 2.

TABLE 2: THE DEPENDENCE OF THE NUMBERS OF CASES RECOMMENDED FOR A HYPOTHETICAL CASE CONTROL STUDY ON THE DISTRIBUTION OF HIGH RADON HOUSES FOR FOUR THEORETICAL DISTRIBUTIONS: RESULTS FOR BASE CASE ASSUMPTIONS USED IN FIGURES 5, ASSUMING 10% OF CANCERS ARE RADON-INDUCED AND ONE-THIRD OF LIFETIME EXPOSURE IS MEASURED IN THE STUDY.

THEORETICAL DISTRIBUTION	CROSS OVER POINT POINT X_0 (EXPOSURE/MEDIAN)	% OF POPULATION WITH EXPOSURE > X_0	NUMBER OF CASES
PWR(1.25)	10.75	3.4	300
PWR(1.75)	12.25	0.8	1,300
LnNrm(3.5)	21.0	0.3	3,500
LnNrm(2.8)	18.25	0.05	20,000

Table 2 shows an order of magnitude difference in sample sizes for experiments of equivalent statistical power when experiments are done in a world with an underlying power-law distribution, as compared with a world with an underlying log-normal distribution. For the power-of-experiment condition just stated, the sample sizes range from 3,500 to 20,000 for log-normal distributions, but from 300 to 1,250 for power-law distributions.

This order of magnitude increase in the power of the epidemiological experiments currently under way is the good news latent in the possibility, not otherwise a cheerful one, that there are many very high radon houses. It opens up the prospect of seeing a house-radon signal in lung cancer data many years earlier than would be expected on the basis of log-normal distributions.

Tables 3 and 4 further explore the sensitivity of the estimate of sample size: in Table 3, to differences in the % of cancer radon-induced (the objective of a study), and in Table 4, to different fractions of exposure history obtained. The dependence in each case is straightforward: the existence of fewer radon-induced cancers requires larger samples, so that a 5% induced cancer incidence means that even power law distributions could require up to several thousand cases, while 20% induced cancer incidence could be detectable even with lognormal radon distributions; better radon histories can compensate for all but the narrowest log normal distributions, while poor radon histories may make any study too difficult.

TABLE 3: THE DEPENDENCE OF THE NUMBERS OF CASES RECOMMENDED FOR A HYPOTHETICAL CASE CONTROL STUDY ON THE DISTRIBUTION OF HIGH RADON HOUSES FOR FOUR THEORETICAL DISTRIBUTIONS: RESULTS SHOWING THE DEPENDENCE ON THE FRACTION OF CASES OF LUNG CANCER WHICH ARE RADON-INDUCED

THEORETICAL DISTRIBUTION	(% of cancers radon-induced / # of houses measured) 20%/ 2 houses	10%/ 2 houses	5%/ 2 houses
	NUMBER OF CASES		
PWR (1.25)	130	300	750
PWR (1.75)	350	1,300	4,400
LnNrm (3.5)	500	3,500	30,000
LnNrm (2.8)	1,600	20,000	450,000

TABLE 4: THE DEPENDENCE OF THE NUMBERS OF CASES RECOMMENDED FOR A HYPOTHETICAL CASE CONTROL STUDY ON THE DISTRIBUTION OF HIGH RADON HOUSES FOR FOUR THEORETICAL DISTRIBUTIONS: RESULTS SHOWING THE DEPENDENCE ON THE NUMBER OF HOUSES FOR WHICH THERE ARE MEASUREMENTS.

THEORETICAL DISTRIBUTION	(% of cancers radon-induced / # of houses measured) 10%/ 3 houses	10%/ 2 houses	10%/ 1 houses
	NUMBER OF CASES		
PWR (1.25)	120	300	1,500
PWR (1.75)	400	1,300	9,000
LnNrm (3.5)	700	3,500	62,000
LnNrm (2.8)	2,900	20,000	900,000

The details of any given research experiment will require modification of the analysis, particularly if additional information such as age, age at exposure, and smoking history is used to enhance the power of the study. The modifications should not change the overall message that a fatter tail at high dose makes the prospect of seeing a signal much more likely. There will always be a tradeoff between taking a larger sample and going further back in time in finding the houses occupied. There are also alternative ways of estimating actual past exposure, including using physiological radon indicators instead of house measurements.

CONCLUSIONS

Information about numbers of high radon houses is only a small fraction of the information that can be obtained about the distribution of radon concentrations. It is, however, information that is of direct significance in formulating public policy regarding radon exposures. Even more significantly, perhaps, it is information that is crucial to research planning that could lead to a better understanding of cancer induction by radiation. Data about relative numbers of high radon houses is best sought explicitly: the indications are that current practice underestimates the numbers, based on inappropriate extrapolations from the distributions of radon concentrations at lower levels.

The work described in this paper was not funded by the U.S. Environmental Protection Agency and therefore the contents do not necessarily reflect the views of the Agency and no official endorsement should be inferred.

REFERENCES

1. Goble, R. and Socolow, R. High radon houses: implications for epidemiology and risk assessment, CENTED Research Report, Clark University, Worcester, MA. 1990.

2. Cohen, B. L. A national survey of 222Rn in U.S. homes and correlating factors, Health Phys. 51: 175. 1986.

3. Nero A. V. Schwehr M.B., Nazaroff W. W. and Revzan K. L. Distribution of airborne radon-222 concentrations in U.S. homes, Science 234: 992, 1986.

4. Gerulsky T. M. Pennsylvania: protecting the homefront, Environment 29, No. 1,12.

5. Alter, H.W. and Oswald, R.A. Nationwide distribution of indoor radon measurements: a preliminary data base JAPCA, 37(3): 227, 1987.

6. Cohen, B.L., and Gromicko, N. Variation of radon levels in U.S. homes with various factors, JAPCA 38(2): 129, 1988.

7. Puskin, J.S. and Nelson, C. B. EPA's Perspective on Risks from Residential Radon Exposure, JAPCA, 39(7): 918, 1989.

8. Ballew, M.A. Health risks from radon progeny. In 18th annual national conference on radiation control. Frankfort, KY: Conference of Radiation Control Program Directors, Inc. Conf. Publication-87-2, pp.157-171. 1987.

9. Bodansky D. et al. Indoor Radon and Its Hazards, chapt.8, University of Washington Press, 1987.

10. Schoenberg, J., and Klotz, J. A case control study of radon and lung cancer among New Jersey women. Technical Report - Phase I, New Jersey State Department of Health, Trenton, NJ 1989.