

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.

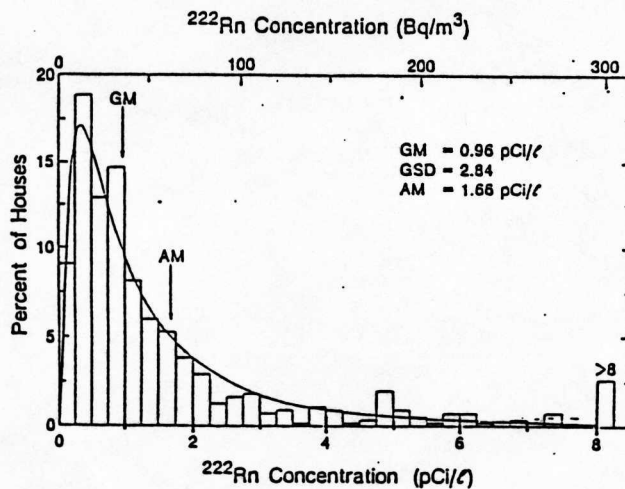


Figure 1. Log-normal fit to Nero's compilation of national radon data.

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.

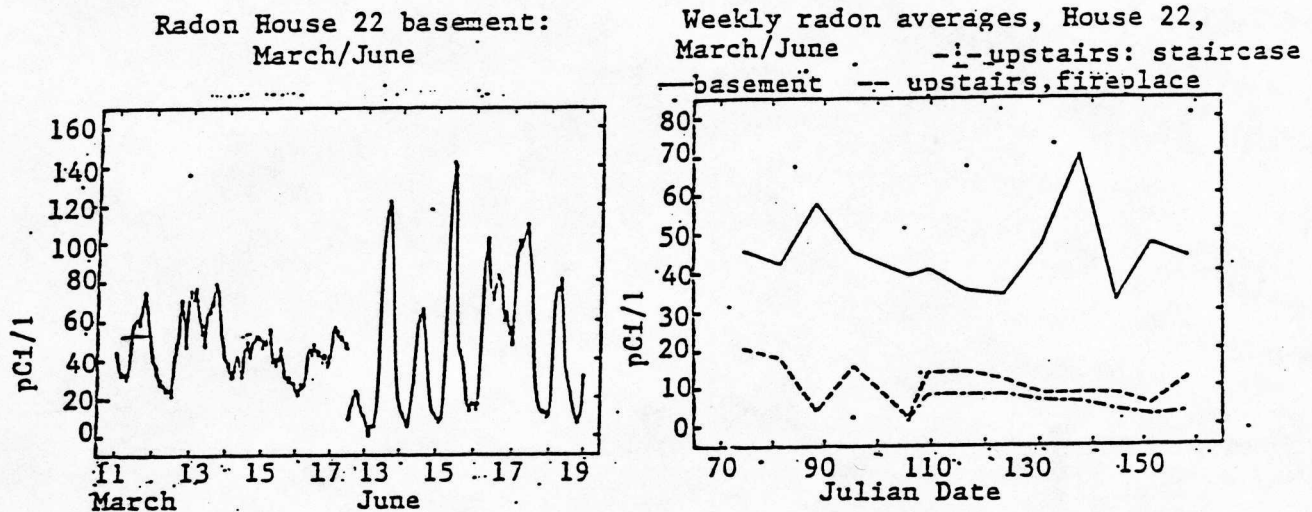


Figure 2. Examples of spatial and temporal variation in radon concentrations in a house.

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.

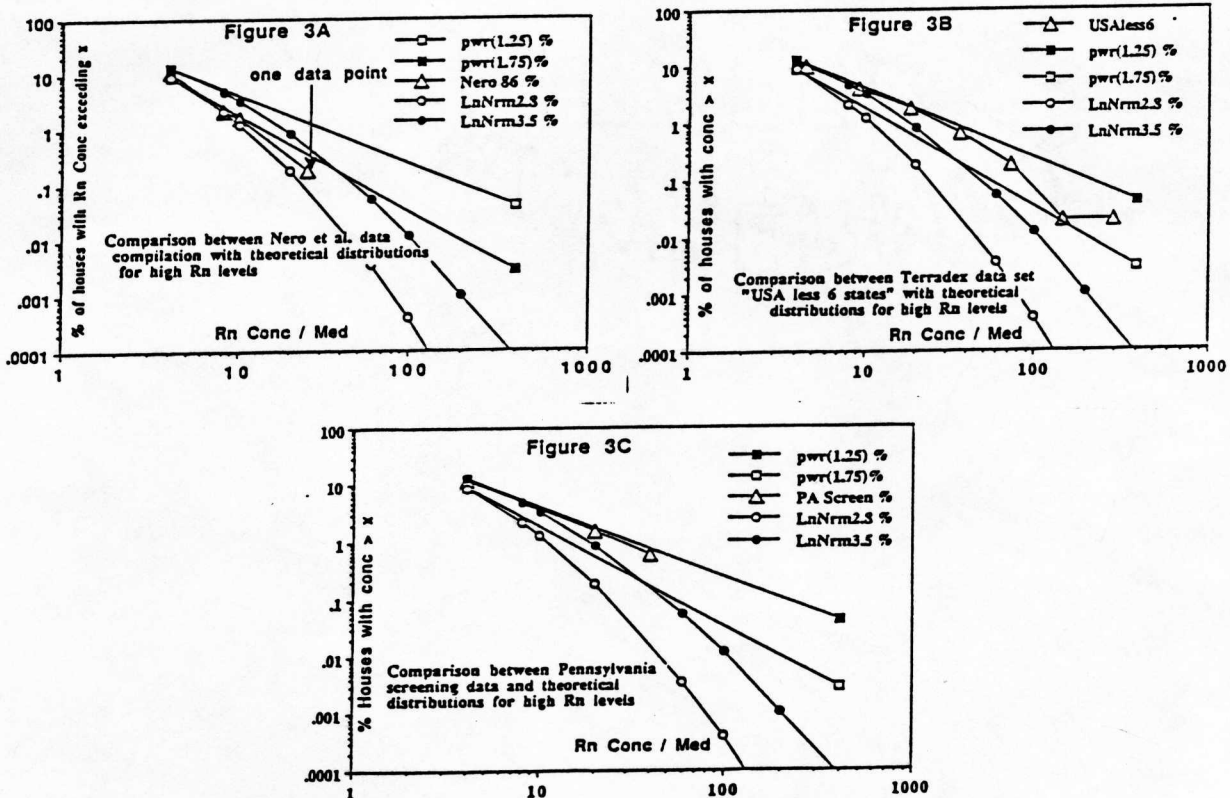


Figure 3. Comparison of four theoretical high radon distributions with three data sets.

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.