1996 – 04

MAPPING OF MEAN RADON CONCENTRATIONS, USING SURVEY DATA AND COVARIATES

Phillip N. Price and Anthony V. Nero
Lawrence Berkeley National Laboratory
Berkeley, CA

ABSTRACT

Although elevated indoor radon concentration can occur almost anywhere, some areas of the U.S. have a much higher fraction of homes with elevated levels than do other areas. Knowledge of what areas tend to have elevated indoor concentrations would allow education and regulatory measures (such as required housing construction techniques) to be focused on populations most at risk from radon. In this paper, we discuss a quantitative approach to predicting annual-average living area radon concentration distributions by county, using survey data (including both short-term basement measurements and long-term living-area measurements), geologic province boundaries from the U.S. Geological Survey, aerometric radium measurements, and climate information. The resulting predictions are unbiased, quantitative, and have known uncertainties that vary from county to county depending on the available information within each county. Computer programs to implement the statistical methods, as well as databases and documentation, are being made available through Lawrence Berkeley National Laboratory.

INTRODUCTION

Data from the National Residential Radon Survey (Marcinowski 1991) and elsewhere (Nero et al. 1986) show that the distribution of annual-average living-area radon concentrations in U.S. housing as a whole is very nearly lognormal, with a geometric mean (GM) of about 0.7 pCi/L, and a geometric standard deviation (GSD) of about 3.1. It has been known for some time that some areas of the country (such as the Reading Prong in Pennsylvania, the Red River Valley of North Dakota and Minnesota, the area around Spokane, Washington, and others) have mean indoor radon concentrations much higher than the country as a whole, and have a larger than usual fraction of homes with concentrations in excess of reference levels such as the EPA's recommended "action level" of 4 pCi/L. The fact that high-radon homes tend to cluster suggests that radon programs geared towards locating high-radon homes or promoting radon-resistant housing could be focused on certain areas, vastly improving efficiency of the programs, in terms of the number of protected homes per dollar spent. For this reason individual states (e.g. Perritt et al. 1990, Hughes 1994), the federal government (e.g., Gundersen et al. 1993), and others (e.g., Cohen 1994) have attempted to map either mean radon concentrations or radon "potential," either by using semi-quantitative techniques or by mapping observed averages from monitoring data. These past attempts to predict or map radon concentration distributions by area have had several problems including (for various mapping efforts):

1. reliance on "volunteer" survey data, which can lead to selection bias if people in known high-radon areas are more likely to participate than are others, or if measurement data include multiple measurements from the same homes (which occurs when someone obtains a high measurement, and makes a second measurement to confirm the first);
2. variation (error) in estimated mean concentrations, due to small sample sizes in many areas;
3. use of ad hoc qualitative methods for including geologic information and other data;
4. failure to make quantitative predictions of annual average concentrations in the living area of homes; the parameter of greatest interest for policy purposes;
5. and a lack of attention to quantifying uncertainties in parameter estimates, leading to an inability to tell where it might be worthwhile to collect additional data rather than drawing conclusions based on the available data.

Our methods address all of these shortcomings by combining random-sample radon survey data (both "screening" data and long-term living-area measurements) with geologic information, climate information, and data on house construction, in a formal statistical model. Use of quantitative modeling methods generates predicted mean concentrations by giving appropriate statistical influence to the various kinds of data, and produces quantitative uncertainty estimates that, for example, allow counties where the fraction of high-radon homes is known to be small to be distinguished from those where this fraction is believed to be small, but the fraction is highly uncertain and might be moderate or large. For a variety of reasons, such as determining whether certain counties should have more data collected or should be included in a particular program, meaningful quantitative uncertainties are highly desirable.

In this paper, we summarize our approach to (1) jointly analyzing short-term "screening" measurements and long-term living-area measurements in order to adjust for bias and over-variability of the short-term measurements, (2) use of categorical information such as geologic province indicator variables in conjunction with quantitative predictive variables such as surficial radium concentration, to predict county mean radon concentrations, and (3) summarizing predictions and uncertainties so as to simplify interpretation by policy makers.

JOINT USE OF SHORT- AND LONG-TERM DATA

The most common radon monitoring protocol in the U.S. uses a 3- to 7-day charcoal canister, placed either on the lowest level of the home or the lowest "livable" level of the home, frequently in winter. Such a measurement is usually called a "screening" measurement. Unfortunately, as is well known (Ronca-Battista and Chiles 1990, Klotz et al. 1993, White et al. 1994, Price 1995) screening measurements are extremely poorly correlated with annual average concentrations in the living areas of homes, suffering from both bias (due to seasonal effects and the fact that radon concentrations are highest on the lowest levels of the home) and variability (due primarily to temporal variation in indoor radon concentrations).

As an example of the relationship between short- and long-term measurements, Fig. 1 shows the results of screening and long-term measurements made in the same homes, in this case homes from the EPA/State Residential Radon Survey (SRRS) in Pennsylvania, which were monitored with both a winter screening protocol and with two alpha-track detectors placed for a year on the two lowest levels of the home. Note that in these data the short-term measurement almost always overestimates the long-term average measurement by a large fraction (the median multiplicative bias is a factor of 3.3), and that even homes with the same short-term measurement show large variation in long-term average concentration. Similar results have been found in other studies (cited above), but the details vary depending on house type, measurement protocol, climate, and other factors..

Due to the extremely poor correlation between screening measurements and long-term average measurements, a screening measurement obviously has very little utility in determining the annual average radon concentration in any individual home. However, screening measurements *can* be used to estimate average radon concentrations for a *group* of homes: if one adjusts for bias in the screening measurements, then the random sources of variation tend to cancel out over large numbers of homes. Thus a key step in using screening measurements to determine the radon distribution in an area is to determine the relationship between screening measurements and annual-average living area concentrations. Unfortunately, data sets that allow direct investigation of this relationship through analysis of screening data and long-term living-area data on the same home are rare. Data for the studies cited above, for example, either do not include long-term monitoring on every living level of the home, or do not distinguish which of the monitored levels are used as living space.

However, there is an alternative method for determining the relationship between screening measurements and long-term living-area average measurements, that does not require that both types of measurements be made in the same home. The method is described in detail in Price and Nero 1996, where we apply it to random-sample monitoring data from the SRRS and the National Residential Radon Survey (NRRS). We assume that a short-term screening measurement is equal to the long-term living-area concentration times a multiplicative bias and times a lognormal error term. We assume the bias depends on whether or not the home has a basement and whether the

basement is used as living space. Parameters (including the bias and variation for each type of home) are estimated from the data. We make separate estimates for each EPA region.

Although mathematically involved, the model is conceptually straightforward: if the geometric mean (GM) of the screening measurements within each county exceeds the GM of the annual-average living-area measurements by, say, a factor of 2.5, then we conclude that the screening measurements have a multiplicative bias equal to that factor. The relative variances of the screening measurements versus the long-term living-area measurements allows us to gauge the additional variability in the screening measurements. The mathematical complexity comes in when we make estimates for different types of homes (those with no basement, with a living-area basement, or a basement that is not used as living space) and adjust for small sample sizes in most counties to account for the regression effect (see Price 1995 for a discussion of the regression effect in another radon-related context).

The estimates of the bias for each type of home can be thought of as "conversion factors" for adjusting the GM of short-term measurements to predict the long-term GM. It's only useful to perform this conversion on aggregated data---short-term measurements are subject to so much variation that the long-term concentration in an individual home can only be estimated very poorly from a short-term measurement, as discussed above.

In Fig. 2 we show the relationship between short- and long-term GM measurements for counties in the northern U.S. that were sampled in both the NRRS (long-term monitoring) and the SRRS (short-term monitoring). The substantial bias and variation in short-term measurements is apparent in Fig. 2a, where nearly all counties fall to the right of the 45-degree line that would indicate agreement in the two measurements. Fig. 2b shows the result of applying the estimated "conversion factors" (Note that Fig. 2a and 2b have different scales). Both the bias and variation have been substantially reduced, although there is also evidence of modest problems with the statistical assumptions so that even after adjustment, the relationship between short-term monitoring GM and long-term living-area GM is not perfect. Unfortunately, in many areas short-term data are the only ones available, and the adjusted GMs are clearly far more useful than the raw GMs.

Better results might be obtained by collecting long-term data from a more spatially dense survey than the NRRS---for example, combining SRRS data from Minnesota with long-term data from a recent survey there (Turk et al. 1996, Price et al. 1996b).

We caution that to obtain meaningful results there must be little or no "selection bias" due to differences in radon concentrations between survey homes and non-survey homes. For example, obtaining data on screening measurements from companies that sell radon detectors can lead to serious problems if the data contain multiple measurements from some of the high-radon houses, as can occur if homeowners re-measure to confirm an initial high reading. Even more severe selection effects can occur if people are asked to report their radon measurement voluntarily to a third party---people with elevated radon measurements are far more likely to respond to such a request. For these reasons we restricted our work on this issue to data from the SRRS and the NRRS, both of which used stratified random sample protocols for determining participant eligibility, and are thus essentially free of selection bias. In principle the same procedures might be used on other data sets, such as Cohen's data set (Cohen 1994) which, although consisting of volunteer data, appears to have minimal selection bias due to attention to the issue in data collection.

USE OF PREDICTIVE VARIABLES

In most buildings, most of the indoor radon originates in rocks and soil surrounding the building, and is carried into the building advectively in airflow driven by pressure differences between the soil gas and the indoor air (Nazaroff et al. 1988). These facts suggest that some types of information might help explain or predict indoor radon concentrations. For example, indoor radon concentrations might be related to geologic and soil information (radium content, soil permeability, etc.), information on house construction (presence or absence of a basement, materials used in constructing the building's foundation, etc.), and information related to the building's air exchange

rate (type of heating or cooling system, presence or absence of a direct connection between the basement and the first floor, etc.).

There have been several previous efforts to use explanatory information of the sort mentioned above to predict indoor radon concentrations or to classify areas by their "potential" to have homes with elevated concentrations. For example, the U.S. Geological Survey combined geologic, house construction, and radon monitoring data to construct "radon potential maps" of the United States, scoring each area on several factors and then summing the scores and characterizing each area as having high, medium, or low potential (Gundersen et al. 1993). This method does not make quantitative predictions of mean radon concentrations, and does not distinguish extremely high-radon areas from slightly elevated areas, but the work does serve as a first attempt to combine information from various sources in order to characterize radon concentrations in areas.

Use of explanatory variables has not allowed successful prediction of indoor concentrations in individual homes, but there has been some success at predicting average concentrations within areas. For example, our group at Lawrence Berkeley National Laboratory (LBNL) found that county mean surficial radium concentrations, as reported in the National Uranium Resource Evaluation (NURE), can be used to predict mean screening radon measurements in Minnesota counties with very low uncertainties (Nero et al. 1994, Price et al. 1996a). The approach used at LBNL is to determine the statistical influence to be given to different types of data---such as small numbers of monitoring data or county-average NURE value, for example---so as to best predict average radon concentrations.

Fitting formal statistical models has several advantages over qualitative or ad hoc scoring methods. First, models make quantitative predictions that can be compared to actual data. This allows incorrect predictions to be identified, so the reasons for the discrepancies can be examined. Second, formal models allow additional types of information to be readily included in the model as they become available. Third, determining statistical influence by fitting actual data prevents researchers from unconsciously biasing results to overfit models or to match prior prejudice. Finally, formal models allow quantitative assessment of uncertainties, which can identify areas that may need more attention in order to be completely characterized.

We use a linear regression approach known as Bayesian Hierarchical Modeling (see Gelman et al. 1995 for discussion) to create statistical models for mean indoor radon concentrations. A detailed explanation of this approach and its justification can be found in Price et al. 1996a. As with the joint analysis of short- and long-term data, the models are computationally involved but conceptually simple. The specific model that we fit depends on the available data; as an example, we consider predicting GM radon concentrations in the mid-Atlantic states (Pennsylvania, Maryland, Delaware, Virginia, and West Virginia---EPA's region 3). A paper on this subject has been recently submitted to Health Physics.

We define the "true SRRS GM" for each county to be the geometric mean radon concentration that would be found if every single-family detached home in the county were to be monitored with the SRRS screening protocol. Of course, the true GM is not known for any county, since only a small fraction of each county's homes were monitored in the SRRS---most counties in the region had fewer than 10 measurements in the SRRS. We assume the observed GM is equal to the true GM times an error term drawn from a known distribution that depends on the number of observations in each county.

We fit the model in log space, and write the logarithm of the true SRRS GM as a sum of terms:

1. a regression coefficient (to be estimated from the data) times the mean surficial radium concentration in the county,
2. a "geologic effect" (to be estimated from the data) associated with the geologic province containing the county, using provinces defined by the U.S. Geological Survey (Gundersen et al. 1993),
3. a regression coefficient (to be estimated from the data) times the fraction of homes that were monitored in the basement, and
4. a "county effect" that allows counties to have different true GMs even if the terms above are all equal.

Discussion of the techniques used to fit the model is outside the scope of this paper---see Price et al. 1996a and Gelman et al. 1995. The result of the model is a prediction, with uncertainty, for each county's true SRRS GM, as well as estimates for the parameters listed above, including the influence of each geologic province on indoor radon concentrations. Using conversion factors determined with the methods discussed above (after taking into account the fractions of non-basement homes in each county), we obtain a predicted long-term living-area GM for each county in the region. These estimates are mapped in Fig. 3, and additional information is presented for a random selection of counties in Table 1. The results can be manipulated in a variety of ways, for example to highlight counties with the greatest uncertainties in GMs, or those with the highest estimated fraction of homes over a reference level, and so on.

CURRENT EFFORTS

The work discussed above, which was performed in collaboration with many people in addition to the authors of this paper, represents a substantial improvement over prior art, with regard to characterizing and mapping living-area radon concentrations based on sparse data. Our current efforts are directed towards creating a package of software, documentation, and databases that will allow other groups to implement these methods. We are also investigating the extent to which these techniques are useful in predicting and mapping radon concentrations at smaller spatial scales such as census tracts or townships. Some discussion and examples of curent work and available data can be found on the world wide web at http:/eande.lbl.gov/IEP/high-radon/hr.html. The authors can be reached through email to pnprice@lbl.gov.

REFERENCES

Cohen, B.L., Stone, C.A., Schilken, C.A. Indoor radon maps of the United States. Health Physics 66, 201-205; 1994.

Gelman, A., Carlin, J., Stern, H., Rubin, D. (1995) Bayesian Data Analysis. Chapman and Hall, New York, New York.

Gundersen, L.C., Otton, J.K., Szarzi, S.L. EPA Region 3 Geologic Radon Potential Summary. In Geologic Radon Potential of EPA Region 3, USGS Open-File Report 93-292-C, R.Schumann editor (1993).

Hughes, G. Radon in Washington. Washington State Department of Health Special Report, Olympia, Washington (1994).

Klotz, J.B., Schoenberg, J.B.; Wilcox, H.B. Relationship among short- and long-term radon measurements within dwellings: influence of radon concentrations. Health Physics 65:367-374; 1993.

Marcinowski, F. Nationwide survey of residential radon levels in the U.S. Radiation Protection Dosimetry 45, 419-424; 1991.

Nazaroff, W.W., Moed, B.A., Sextro, R.G. Soil as a source of indoor radon: generation, migration, and entry. In "Radon and its decay products in indoor air", (1988), W.W. Nazaroff and A.V. Nero, editors. Wiley and Sons, New York, New York.

Nero, A.V., Schwehr, M.B., Nazaroff, W.W., Revzan, K.L. Distribution of airborne radon-222 concentrations in U.S. homes. Science 234, 992-997; 1986.

Nero, A.V., Leiden, S.M., Nolan, D.A., Price, P.N., Rein, S., Revzan, K.L., Wollenberg, H.R., Gadgil, A.J. Statistically based methodologies for mapping of radon "actual" concentrations: the case of Minnesota. Radiation Protection Dosimetry 56, 215-219; 1994.

Perritt, R.L., Hartwell, T.D., Sheldon, L.S., Cox, B.G., Clayton, C.A., Jones, S.M., Smith, M.L. Radon-222 levels in New York State homes, Health Physics 58: 147-155; 1990.

Price, P.N. The regression effect as a cause of the nonlinear relationship between short- and long-term radon concentration measurements. Health Physics 69:111-114; 1995.

Price, P.N., Nero, A.V. Joint analysis of long- and short-term radon monitoring data from the Northern U.S. Environment International, in press (1996).

Price, P.N., Nero, A.V., Gelman, A. Bayesian prediction of mean indoor radon concentrations for Minnesota counties. Health Physics, in press (1996a). Also available from Lawrence Berkeley Laboratory National Laboratory, Berkeley, CA: LBL-35818 (1996).

Price, P.N., Fridlyand, J., Nero, A.V., Revzan, K.L., Turk, B., Oatman, L., Gelman, A. Preliminary results from a year-long radon survey in Minnesota homes. Lawrence Berkeley National Laboratory, Berkeley, CA: LBNL-38822 (1996b).

Ronca-Battista, M., Chiles, B. The relationship between winter screening and annual average radon concentrations in U.S. homes. Proceedings of the 1990 International Symposium on Radon and Radon Reduction Technology. Washington, D.C.: US EPA; Vol. I; 1990a: 2-3--2-14.

Turk, B., Nero, A.V., Price, P.N., Oatman, L., Brunelle, S. Operational aspects of the Minnesota high radon project. Lawrence Berkeley National Laboratory, Berkeley, CA: LBNL-38647.

White, S.B., Alexander, B.V., Rodman, N.F. Predicting the annual concentration of indoor Rn222 from one or more short-term measurements. Health Physics 66, 55-62; 1994.

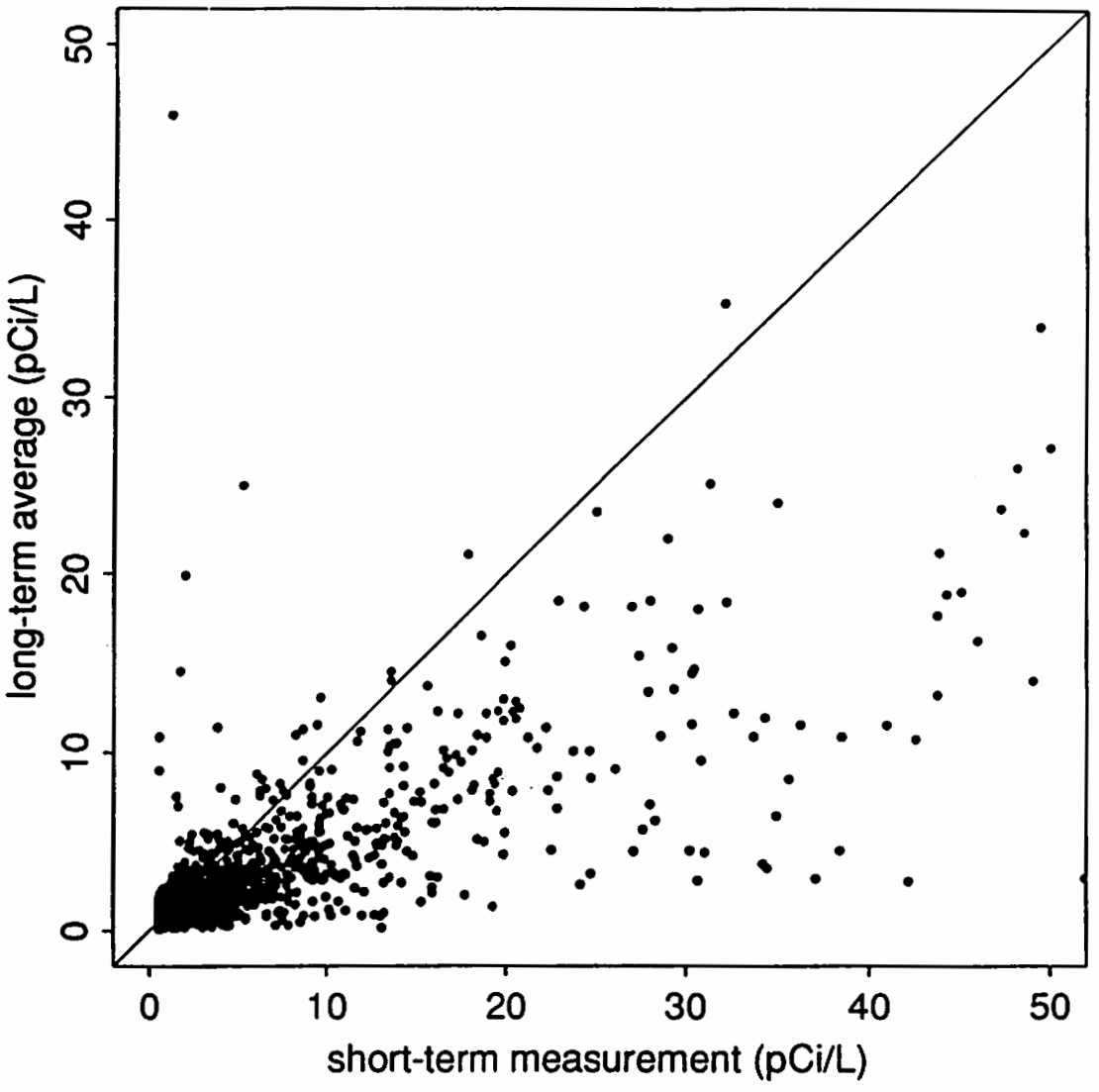

Fig. 1 Average of two year-long alpha-track measurements (on the two lowest levels of the home) versus a 3- to 6-day charcoal-canister "screening" measurement from the EPA/State Residential Radon Survey, for a random selection of homes in Pennsylvania. Most of the short-term measurements are higher than the long-term average, and so fall to the right of the 45-degree line (shown). A screening measurement provides little information on a home's annual-average living-area radon concentration, as discussed in several of the cited works.

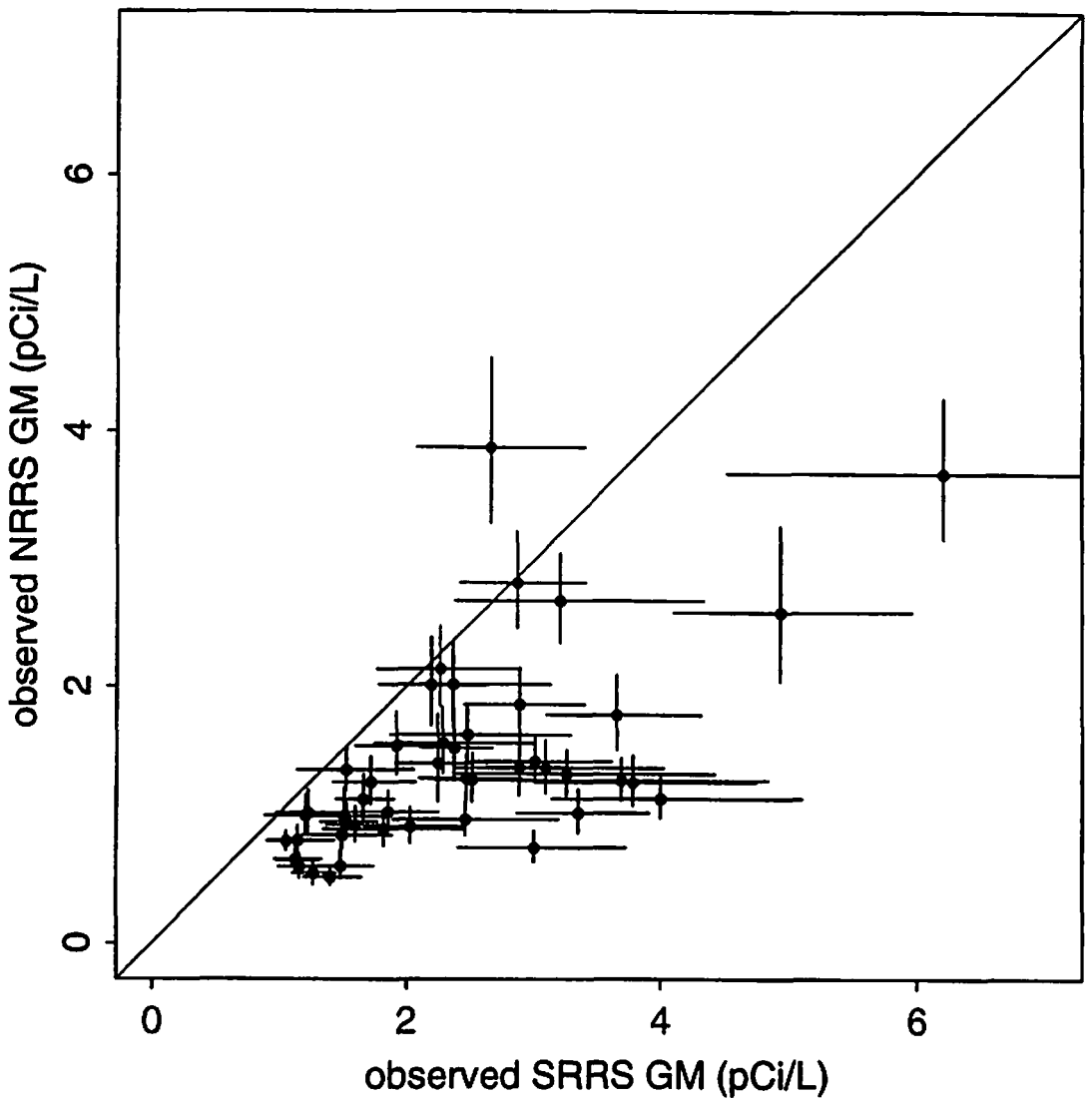

Fig. 2 (a) Geometric mean (GM) of annual-average living-area concentrations from the National Residential Radon Survey (NRRS), versus GM of screening measurements from the EPA/State Residential Radon Survey (SRRS), for NRRS counties in the Northern U.S. Note the very substantial bias, and the poor correlation between the two GM's, especially for counties with NRRS GMs below 2 pCi/L.

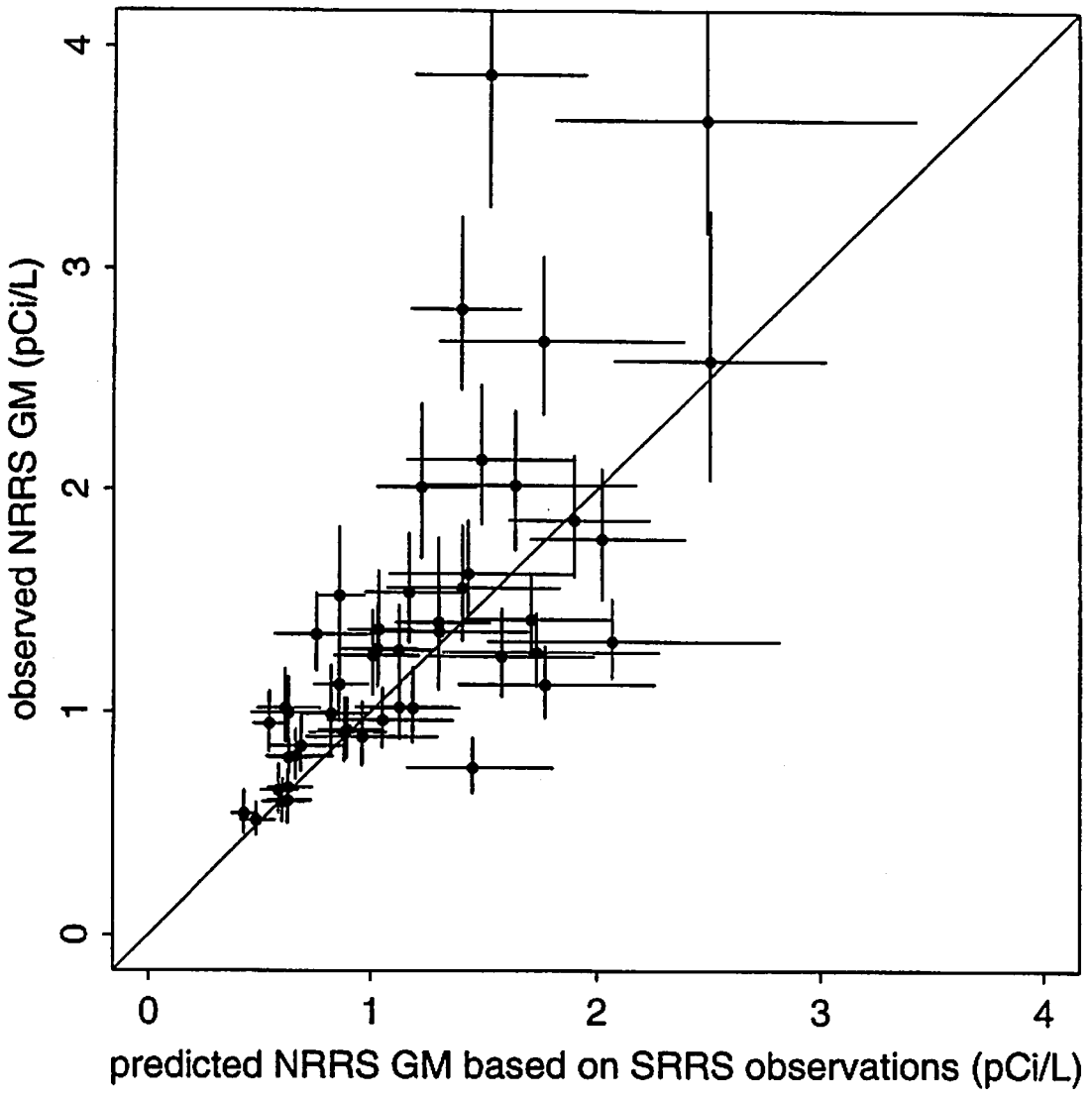

Fig. 2 (b) NRRS GM versus prediction from SRRS data after applying conversion factors for different types of homes, where conversion factors differ between EPA regions. Note that the scale on the plot is different from that for Fig. 2a. The bias has been removed, and there is a clear relationship between the two GMs. There are some minor systematic problems with the predictions, though. See Price and Nero (1996) for details and discussion.

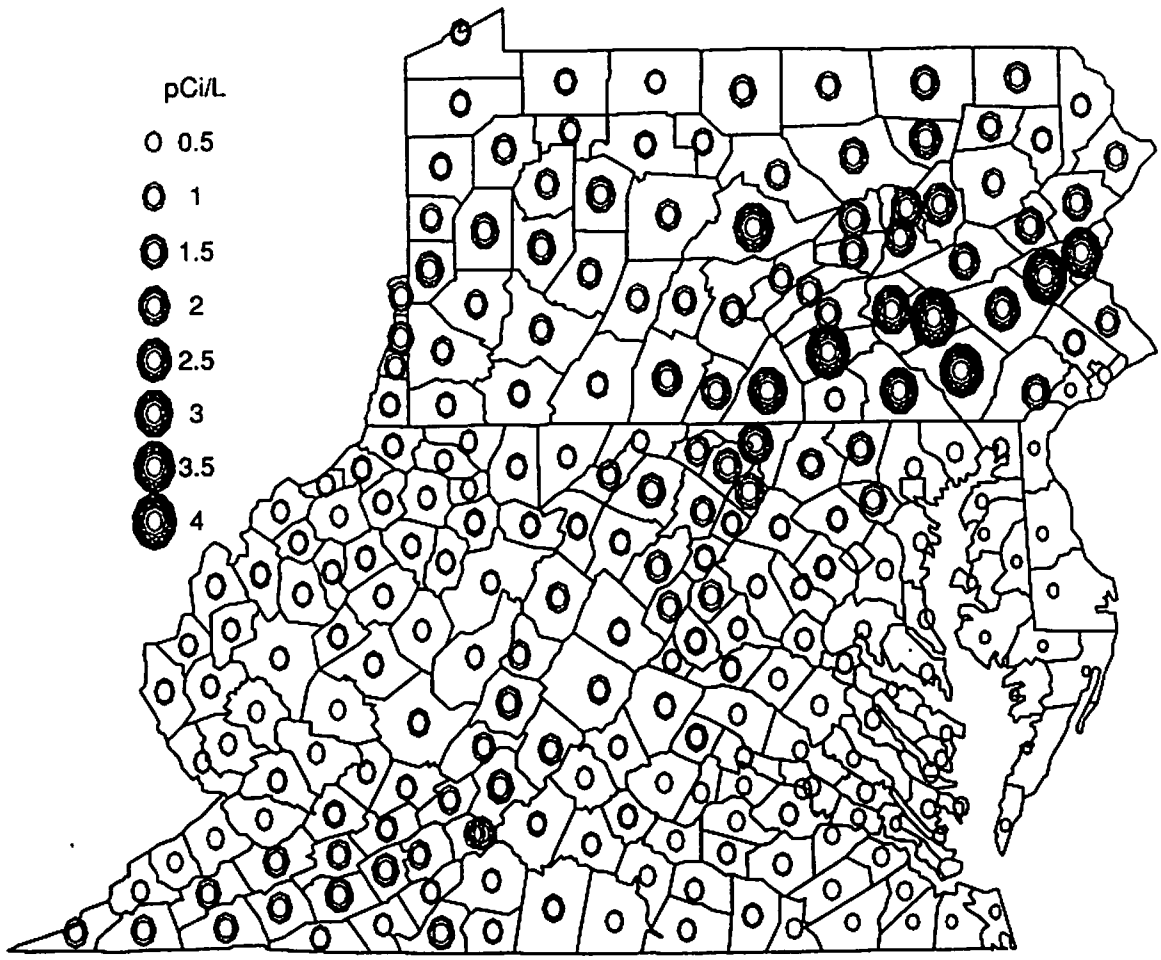

Fig. 3 Map of predicted geometric mean (GM) of annual-average living-area radon concentrations for ground-contact homes, by county, in EPA's region 3. Predictions are based on screening data, geologic province information, aerometric surficial radium measurements, and house construction information. Note the high values in most of southeastern Pennsylvania, the elevated values along the Appalachians in southern Virginia, and the generally low values on the coastal plain. Addition of climate information (heating infiltration degree-days) produces very similar results (not shown).

Table 1: County information for a random selection of counties in EPA's region 3. Counties are identified by their "fips" codes. The number of observations and the observed GM from the EPA/State Residential Radon Survey (SRRS), a number indicating the geologic province containing the county (province definitions not shown), estimated radium concentration (expressed in equivalent ppm of Uranium) from the National Uranium Resource Evaluation (NURE), the fraction of homes in which screening measurements were made upstairs, and the fraction of homes in which the screening measurement was made in a living area of the home, are shown in the leftmost columns of the table. These data are used in a statistical model to predict the GM that would be found if every home in the county were to be monitored using the SRRS protocol. Conversion factors are applied to adjust the predicted SRRS GM to a predicted long-term living-area GM ("NRRS GM") and multiplicative uncertainty, shown on the right.

county fips code	# of obs	obs. GM (pCi/L)	prov. #	NURE (ppm U)	upstairs frac	meas. living area	pred. SRRS GM (pCi/L)	pred. NRRS GM (pCi/L	mult. uncert.
24001	74	1.4	8	1.94	0.30	0.55	1.6	0.7	1.15
24011	23	0.4	23	1.21	0.83	0.95	0.4	0.2	1.22
24037	15	0.8	22	1.42	0.46	0.84	0.7	0.4	1.23
24045	50	0.3	23	0.82	0.78	0.88	0.3	0.2	1.15
42059	7	1.8	15	2.38	0.23	0.52	2.0	1.0	1.26
42077	23	9.2	13	3.31	0.21	0.52	8.7	4.2	1.22
42083	15	1.7	1	1.67	0.14	0.36	2.0	0.1	1.24
42095	26	8.1	13	3.28	0.02	0.36	8.7	3.8	1.20
42123	15	2.3	1	1.99	0.14	0.47	2.4	1.1	1.26
51015	19	2.0	10	1.80	0.61	0.69	2.0	1.1	1.23
51023	9	4.2	10	1.74	0.37	0.59	3.1	1.6	1.28
51036	1	1.1	20	1.69	0.57	0.63	1.0	0.5	1.34
51073	3	0.5	20	1.33	0.72	0.92	0.7	0.4	1.33
51139	5	2.0	10	1.84	0.44	0.77	2.4	1.3	1.28
51165	15	1.7	10	1.71	0.53	0.62	1.9	1.0	1.25
51735	1	0.5	20	1.99	0.82	0.88	0.9	0.5	1.35
54003	19	4.1	10	2.16	0.48	0.64	3.4	1.8	1.24
54013	3	0.9	15	2.08	0.71	0.87	1.2	0.7	1.27
54025	18	2.6	6	1.76	0.61	0.78	2.1	1.1	1.28
54099	16	1.8	15	1.89	0.65	0.75	1.5	0.8	1.23
54109	19	1.6	4	1.68	0.68	0.89	1.4	0.8	1.22