Statistical Science
1999, Vol. 14, No. 3, 305–337

Analysis of Local Decisions Using Hierarchical Modeling, Applied to Home Radon Measurement and Remediation

Chia-yu Lin, Andrew Gelman, Phillip N. Price and David H. Krantz

Abstract. This paper examines the decision problems associated with measurement and remediation of environmental hazards, using the example of indoor radon (a carcinogen) as a case study. Innovative methods developed here include (1) the use of results from a previous hierarchical statistical analysis to obtain probability distributions with local variation in both predictions and uncertainties, (2) graphical methods to display the aggregate consequences of decisions by individuals and (3) alternative parameterizations for individual variation in the dollar value of a given reduction in risk. We perform cost-benefit analyses for a variety of decision strategies, as a function of home types and geography, so that measurement and remediation can be recommended where it is most effective. We also briefly discuss the sensitivity of policy recommendations and outcomes to uncertainty in inputs. For the home radon example, we estimate that if the recommended decision rule were applied to all houses in the United States, it would be possible to save the same number of lives as with the current official recommendations for about 40% less cost.

Key words and phrases: Bayesian decision analysis, hierarchical models, small area decision problems, value of information.

1. INTRODUCTION

1.1 Decision-making for Environmental Hazards

Associated with many environmental hazards is a decision problem: whether to (1) perform an expensive remediation to reduce the risk, (2) do nothing, or (3) take a relatively inexpensive measurement of the risk and use this information to decide whether to (a) remediate or (b) do nothing. This decision can often be made at the individual, household, or community level. Performing this decision analysis requires estimates for the risks. In particular, the more precise are the local risk estimates, the more feasible it is to construct localized decision recommendations that allow attention and effort to be focused on the individuals, households and communities at most risk.

In this paper, we present an analysis of the remediation–measurement decision problem in the context of a hierarchical model for estimating risk as a function of location and various covariates. We develop and illustrate our method for the problem of home radon, a recognized cancer risk, for which appropriate measurement and remediation strategies have been and continue to be the subject

Chia-yu Lin is a graduate student, Department of Biostatistics, Columbia University, New York, New York 10032. Andrew Gelman is Associate Professor, Department of Statistics, Columbia University, New York, New York 10027. Phillip N. Price is scientist, Lawrence Berkeley National Laboratory, Berkeley, California 94720. David H. Krantz is Professor, Department of Psychology, Columbia University, New York, New York 10027.
Address all questions to Andrew Gelman, gelman@stat.columbia.edu. See also http://www.stat.columbia.edu/radon for more information.

of debate. In addition to its own importance, the radon problem shares several features with other environmental hazards: (a) the risks are geographically dispersed but have strong spatial patterns; (b) information exists to identify risky areas, but one cannot easily identify individual households at high risk; (c) it is possible to perform local measurements to identify the risks of individual households, but it would be expensive to measure every household.

By "hierarchical model," we mean a statistical model in which a separate statistical parameter is assigned to each predicted unit (each county, in the present case). To clarify, consider a standard nonhierarchical approach to predicting mean radon concentrations by county: measured radon concentrations are regressed on, say, measured surficial uranium concentration, and two regression coefficients (a slope and an intercept) are determined. These coefficients are then used to predict mean radon concentrations over an area such as the United States, with uncertainties estimated by the standard error of the regression. Although such an approach yields a different predicted mean radon level for each county, this is not a hierarchical model because the underlying statistical parameters—the regression coefficients—do not vary by county.

In contrast, a hierarchical model includes the regression coefficients and also, for each county, an additional coefficient allowing the predicted mean for that county to differ from the regression prediction. To see why this makes sense, consider a county with many radon measurements, whose observed mean concentration differs substantially from the regression prediction. Using the regression prediction in such a county is not reasonable, since (given enough samples) the prediction is known to be erroneous. In the hierarchical Bayesian approach, each county's estimate is a compromise between the regression prediction and the measured value, with the relative weighting determined by the amount of data in the county and the overall accuracy of the regression predictions. Moreover, a hierarchical model allows a separate uncertainty estimate for each county.

1.2 Hierarchical Decision Analysis

Our approach to hierarchical decision analysis has four steps. First, a hierarchical model is fit to available data, resulting in a posterior distribution for exposure to the environmental hazard for any given household, as a function of locality and other household information. Second, the problem of "decision-making under certainty" is formulated: for example, for the radon remediation problem, the tradeoff between dollars and lives implies a willingness to remediate at an *action level* R_{action}, so that if the true exposure were known, one would remediate if and only if it exceeds R_{action}; depending on variations in risks and risk preferences, R_{action} can vary among households. Third, the "decision-making under uncertainty" problem is solved: for the radon problem, the measurement–remediation decision for any household is a function of its R_{action} and its posterior distribution of exposure level. If additional information is available at the household level—for example, a previous radon measurement—this can be incorporated into the posterior distribution. The fourth step of our analysis is to evaluate the effect of various decision recommendations, in terms of expected lives saved and expected costs, if applied within a larger geographic area (for example, the entire United States). Results can also be expressed in terms of expected marginal and aggregate cost per life saved. As always in decision analysis, sensitivity analysis is then done to see how the estimated costs and lives saved vary when assumptions are perturbed.

With the exception of the hierarchical modeling, the above steps follow the standard paradigm of expected-value or "Bayesian" decision analysis (see, e.g., Dakins, Toll, Small and Brand, 1996, and Englehardt and Peng, 1996, for a recent review and examples). The use of a hierarchical model for the spatially varying hazard allows us to incorporate modern Bayesian inference into a formal decision analysis. The standard regression-modeling approach does allow recommendations to vary as a function of the predictive variables and thus to vary with location, but hierarchical modeling has lower predictive error than standard regression approaches, and, more importantly, allows additional variation in recommendations, related to spatially varying uncertainties.

Our cost-benefit analysis of radon decisions makes geographically variable recommendations. The recommended localized actions are more cost-effective than the current single nationwide recommendation (see, in particular, Evans, Hawkins and Graham, 1988, who recognize that precise modeling of radon levels should allow targeted recommendations), but care is required in summarizing the decision analysis, which we do using maps and a series of graphs indicating costs for various decision options.

A characteristic of the hierarchical decision analysis is that aggregate outcomes of decision strategies can no longer be trivially derived from individual recommendations. In statistical terms, aggregating requires averaging over the predictive distribution of the thousands of model parameters that indicate

radon risk in the counties. We compute expected costs and lives saved by simulation from our posterior distribution.

1.3 Outline of This Paper

We develop our hierarchical approach to decision analysis in the context of measurement and remediation of home radon. Sections 2 and 3 of this paper provide background on the indoor radon problem and our previous work using hierarchical modeling of radon survey data to identify the houses that are likely to have high radon levels given information at the geographical and house level. Section 4 addresses the measurement and remediation decision for individual homeowners, and Section 5 presents the estimated aggregate consequences of following the recommended strategy and various alternatives if applied throughout the United States. For both the individual decisions and aggregate consequences, we develop a series of graphical displays that are potentially useful for hierarchical decision problems in general. In Section 6 we explore the sensitivity of our results to assumptions, and we conclude in Section 7 with a discussion of the specific relevance of our methods to the indoor radon problem and the general applicability to hierarchical decision problems.

2. THE RADON PROBLEM AND DECISION OPTIONS

2.1 Health Effects

Radon is a carcinogen, a naturally occurring radioactive gas whose decay products are also radioactive, known to cause lung cancer in high concentration and estimated to cause several thousand lung cancer deaths per year in the U.S. (see Nazaroff and Nero, 1988, for an overview of the radon problem; Cole, 1993, for a discussion of the governmental response to it; National Research Council, 1998, for an influential official report).

The well-documented dose-dependent excess of lung cancer among underground miners exposed to radon has convincingly demonstrated that exposure to very high concentrations of radon causes lung cancer. Levels in homes are usually lower than those in mines, miners are also exposed to other carcinogens, miners are overwhelmingly smokers and working miners generally breathe both harder and more deeply than people at home, so several assumptions and extrapolations are needed to estimate cancer risk at typical home levels (see National Research Council, 1991 and National Research Council (BEIR VI), 1998, Tables 3–6). These extrapolations (including an assumed linear dose-response function) suggests that about 15,400 additional lung cancer deaths occur annually in the United States due to radon, mostly among smokers, though this number is based on an unrealistic comparison to the number of deaths that would occur if nobody were exposed to any radon at all.

The miner studies demonstrate statistically significant elevated cancer risk at doses equivalent to lifetime residence in a home at about 20 picoCuries per liter (pCi/L). (An alternative notation is the international standard unit of Bequerels per cubic meter; 1 pCi/L = 37 Bq/m^3.) Estimates based on the miner studies, on experiments on animals and on biological and biophysical models suggest that, at least at high levels, lifetime exposure to each additional pCi/L of indoor radon adds a lifetime risk of about 0.0134, 0.0026, 0.0088 and 0.0018 of lung cancer for male ever-smokers, male never-smokers, female ever-smokers, and female never-smokers, respectively (National Research Council, 1998). See Table 1 for the parameters that we use in this paper for each sex–smoking category.

The dose-response at low concentrations is difficult to estimate, because all case-control studies have been fairly small and because lifetime radon exposures are poorly estimated. Although a linear dose-response relation is plausible and is consistent with case-control data, current data are also consistent with a threshold model or even a small beneficial effect at low doses (Cohen, 1995; Bogen, 1997; Lagarde et al., 1997; Lubin and Boice, 1997).

Partly from necessity and partly for historical reasons, radon researchers use a fairly large and confusing assortment of units. For instance, the radiation absorbed by the body (the "dose") depends not just on the concentration of radon in the air, but also on the breathing rate and of course on the duration of exposure, so there is no simple conversion between radon concentration in indoor air and dose absorbed by a human body. Moreover, radon's decay products, rather than radon itself, deliver most of the radiation dose associated with radon, and the differential removal of decay products and radon itself can lead to variation in the relative concentrations of each. For clarity and convenience, we write "the radon dose" when we mean "the dose from radon and its decay products," where standard parameter estimates have been used to make all of the necessary adjustments. See Nazaroff and Nero (1988) for an overview of many of these issues.

We present cumulative exposures in terms of pCi/L-years, rather than the historical unit (also non-SI) which is "working level months (WLM)." The direct conversion, for breathing 1 pCi/L air for a year, is 1 pCi/L-year = 0.26 WLM, but it is stan-

TABLE 1
*Estimated absolute and relative risks of lung cancer death for lifetime indoor exposure to radon**

	Male		Female	
	Ever-smoker	**Never-smoker**	**Ever-smoker**	**Never-smoker**
Baseline absolute risk at 0 pCi/L	0.07409	0.00579	0.04349	0.00377
Excess relative risk for additional pCi/L	0.1149	0.2827	0.1280	0.2998
Excess absolute risk for additional pCi/L	0.0134	0.0026	0.0088	0.0018
Average numbers in U.S. households	0.30	1.07	0.27	1.16

*Derived from BEIR VI (National Research Council, 1998), which describes the absolute increment in lung cancer risk resulting from exposure to indoor radon beyond that from exposure to outdoor-background concentration of radon, and under the assumption that a person spends 70% of his or her time indoors. Model is absolute risk = baseline risk $\times$ (1 + excess relative risk), with excess relative risk proportional to radon exposure. Average household populations by sex and smoking status are derived from the *Statistical Abstract of the United States* and CDC (1994), combining populations of children and adults. An *ever-smoker* is defined as a person who has smoked at least 100 cigarettes or the equivalent in his or her lifetime.

dard to assume that an individual is only at home about 70% of the time, so that a home concentration of 1 pCi/L for a year leads to an exposure of 0.18 WLM.

2.2 U.S. Residential Radon Concentrations and Measurements

Residential radon measurements are commonly made following a variety of protocols. The most frequently used protocol in the United States has been the "screening" measurement: a short-term (2–7 day) charcoal-canister measurement made on the lowest level of the home (often an unoccupied basement), at a cost of about $15 to $20. (Under a more recent protocol, measurements are taken on the lowest living area level of the home.) Short-term measurements made at different times during the same season have an approximately lognormal distribution (i.e., the log measurements are normally distributed) with a geometric standard deviation (GSD, the exponential of the standard deviation of the log measurements) of roughly 1.6, primarily due to temporal variation in indoor radon concentrations.

In addition, because short-term measurements are usually made on the lowest level of the home and during the season of highest indoor radon exposure, they are upwardly biased measures of annual living area average radon level. The magnitude of this bias varies by season and by region of the country and depends on whether the basement (if any) is used as living space (White, Clayton, Alexander and Clifford, 1990; Klotz et al., 1993; Price and Nero, 1996); our estimated correction factors for winter-season, lowest-level measurements, known as "screening" measurements, appear in Table 2. (If the short-term measurement is not made in winter, then an additional seasonal correction factor is needed.) Due to the large temporal variability and other sources of variation, a short-term measurement can predict the long-term living area concentration only to within a factor of 1.8 or so, even after correcting for systematic biases.

A radon measure that is far less common than the screening measurement, but is believed to be much better for evaluating radon risk, is a twelve-month integrated measurement of the radon concentration. By monitoring on every living level of the home (that is, on every floor in which people spend more than a small amount of time each day), one can measure the "annual living area average radon concentration," or ALAA. For a typical home with two stories used as living space, such monitoring costs about $50. These long-term living area measurements are not subject to the biases and effects of day-to-day and seasonal variation that affect screening measurements. A national sample of ALAA measurements was collected in the National Residential Radon Survey (NRRS) (see Marcinowski, Lucas and Yeager, 1994).

The exact relationship between the ALAA concentrations and the occupant exposures is not known; people spend different amounts of time in different areas of the home, long-term measurements are still subject to some error, even on the same floor different rooms can have slightly different radon levels and so on. For the purposes of this paper, we assume that an ALAA measurement (i.e., the arithmetic mean of long-term measurements made on each occupied level of the home) estimates each resident's exposure to within a multiplicative error with a geometric mean (GM) of unity and a GSD of 1.2.

The distribution of annual-average living area home radon concentrations in U.S. houses, as measured in the NRRS, is approximately lognormal

TABLE 2
*Correction factors by which one must divide a short-term winter radon measurement to estimate annual-average living-area level**

Region	No basement	Basement is a living area	Basement is not a living area
New England	2.2 (1.3)	1.7 (1.1)	3.4 (1.3)
New York/New Jersey	1.6 (1.3)	1.6 (1.3)	3.0 (1.3)
Mid-Atlantic	1.6 (1.1)	1.6 (1.1)	2.8 (1.1)
Southeast	1.3 (1.1)	1.9 (1.1)	2.3 (1.1)
Midwest	1.2 (1.1)	1.6 (1.1)	2.2 (1.1)
South	1.3 (1.1)	1.8 (1.2)	1.7 (1.1)
Central Plains	1.5 (1.1)	1.7 (1.1)	3.1 (1.1)
Big Sky and Plains	1.2 (1.1)	2.1 (1.1)	3.1 (1.1)
Southwest	1.3 (1.1)	1.8 (1.2)	2.6 (1.1)
Northwest	1.2 (1.1)	1.9 (1.1)	4.0 (1.1)

*Geometric standard errors of estimation for the correction factors are in parentheses; even if the correction factors were known perfectly, the annual average living area concentration would still be subject to large uncertainty due to temporal variability in the short-term measurements. From Price and Nero (1996).

with geometric mean (GM) 0.67 pCi/L and geometric standard deviation (GSD) 3.1 (Marcinowski, Lucas and Yaeger, 1994). (Throughout, we use the term "house" to refer to owner-occupied ground-contact homes.) These data suggest that between 50,000 and 100,000 homes have radon concentrations in primary living space in excess of 20 pCi/L. This level causes an annual radiation exposure roughly equal to the occupational exposure limit for uranium miners. Thirty years' occupancy of such a house would yield an added estimated risk of lung cancer of about 2.4% among never-smokers and 12.1% among ever-smokers. The lung cancer risks from radon are very high compared with the risks estimated for other kinds of environmental exposures regulated by the EPA (for comparison, see U.S. Congress, Office of Technology, 1993).

2.3 Radon Remediation

Several radon control techniques have been developed, tested and implemented (Henschel and Scott, 1987; Prill, Fisk and Turk, 1990), and long-term performances of these systems were reported (Turk, Harrison and Sextro, 1991). The currently preferred remediation method for most homes, "sub-slab depressurization," costs about $1000–$1500 to install and requires constant use of a small electric fan; the net present value of such a system is about $2000, including the heating and cooling costs associated with increased ventilation. Although long-term experience with these systems is lacking, for purposes of our analysis we will assume that such a system remains effective for 30 years. We are not aware of any large-scale randomized studies on the effect of remediation on radon levels, but many small nonrandomized studies have been conducted and are summarized in an EPA report (Henschel, 1993). These studies suggest that almost all homes can be remediated to below 4 pCi/L, while reductions under 1 pCi/L are rarely attained with conventional methods, for homes with a very wide range of preremediation levels. For simplicity, we make the assumption that remediation will reduce radon concentration to 2 pCi/L. For obvious reasons, little is known about effects of remediation on houses that already have low radon levels; we will assume that if the initial annual living area average level is less than 2 pCi/L, then remediation has no effect.

Recommendations for radon remediation vary by country, with Sweden setting a recommended action level for the annual living area average (ALAA) indoor radon concentration of 10 pCi/L and Canada recommending action at 20 pCi/L, compared to the U.S. level of 4 pCi/L. The current U.S. recommendations, if fully implemented, would cost on the order of $10–$20 billion in measurement and remediation costs (see Nero, Gadgil, Nazaroff and Revzan, 1990). In Section 5, we discuss the efficiency of various policies in terms of estimated dollars per life saved.

2.4 Individual and Public Decision Options

One can imagine an ideal world in which homeowners make monitoring and remediation decisions based on full knowledge of the current understanding of radon risk and remediation costs and effectiveness and taking into account their own risk tolerance and financial state. In the real world, though, there is a substantial cost (in time and hassle) associated with reaching that level of expertise, and it is reasonable for people to follow more general recommendations on whether to take action and what sort of action to take.

An important policy decision, well outside the scope of this paper, is just how general such recommendations should be in practice: should different recommendations be made for smokers and nonsmokers, for old and young people, for large and small families, and so on. Given the estimated risks from radon, and the cost and effectiveness of remediation, an individual homeowner (or home seller or buyer) can make decisions about measurement and remediation. In addition, national, state and local governments can make recommendations for individual decisions or, for stronger action, can require measurement or remediation of new houses or existing houses, either in the entire country or in targeted areas.

From a long-term public health perspective, an approach that does not depend on household composition makes some sense; children are born, people take up smoking, houses are sold to other owners and so on, so that basing recommended actions on average households is not totally unreasonable. Following the procedures of the present paper, one can make household-specific recommendations based on such factors, but we have not done so, choosing instead to focus only on geographic variation and a few house construction parameters. Realistically, we think that if the EPA were to change its radon policy, the most likely change would be to make geographically specific recommendations rather than to make recommendations that vary at the level of individual households.

3. GEOGRAPHIC MODELING OF INDOOR RADON LEVELS

3.1 Data Sources and Hierarchical Regression Model

Although radon is thought to cause a large number of deaths compared to other environmental hazards, the vast majority of houses in the United States do not have elevated radon levels that would be substantially reduced by remediation. Based on the NRRS data, about 84% of homes have ALAA concentrations under 2 pCi/L, and about 90% are below 3 pCi/L. A goal of some researchers has been to identify locations and predictive variables associated with high-radon homes so that monitoring and remediation programs can be focused efficiently. One such effort at the Lawrence Berkeley National Laboratory used Bayesian hierarchical modeling to analyze indoor radon measurements. These models include monitoring data, county indicators, a measure of surficial radium concentration, a climatological variable, and house construction information and were fit separately in 10 regions of the United States (Price, Nero and Gelman, 1996; Price, 1997; Revzan et al., 1998). These models were used to fit data from short-term measurements, which were calibrated to long-term living area averages as described by Price and Nero (1996). Combining short- and long-term measurements allowed us to estimate the distribution of radon levels in nearly every county in the United States, albeit with widely varying uncertainties depending primarily on the amount of monitoring data within the county.

Unfortunately (from the standpoint of radon mitigation programs), indoor radon concentrations are highly variable even within small areas. Given the predictive variables mentioned in the previous paragraph, the radon level of an individual house in a specified county can be predicted only to within a factor of at best about 1.9, with a factor of 2.3 being more typical (Price, Nero and Gelman, 1996; Price 1996), a disappointingly large predictive uncertainty considering the factor of 3.1 that would hold given no information on the home other than that it is in the United States. On the other hand, this seemingly modest reduction in uncertainty is still enough to identify some areas where high-radon homes are very rare or very common. For instance, in the mid-Atlantic states, more than half the houses in some counties have long-term living area concentrations over the EPA's recommended action level of 4 pCi/L, whereas in other counties fewer than 0.5 percent exceed that level (Price, 1996).

Various monitoring efforts demonstrate that the distribution of indoor radon concentrations for an area or region of almost any scale is reasonably well represented by a lognormal distribution, or sometimes the sum of two such distributions (Nero, Gadjil, Nazaroff and Revzan, 1990). Further, a large area's distribution is effectively a mixture of the individual distributions of the composite subareas, all of which are reasonably well represented by individual lognormal distributions, with geometric means (GM's) that vary from one subarea to another (see Nero, Schwehr, Nazaroff and Revzan, 1986; Price, Nero and Gelman, 1996, for example).

In each region of the country, a hierarchical linear regression model at the level of individual counties was previously fit to the logarithms of home radon measurements (see Price, Nero and Gelman, 1996; Price, 1997). We shall apply these models to perform inferences and decision analyses for previously unmeasured houses i, using the following notation:

R_i = ALAA radon concentration in house i;
$\theta_i = \log(R_i)$;

X_i = Vector of explanatory variables (including county-level variables, house-level variables, and county indicators) for house i;
β = Vector of regression coefficients;
τ^2 = Variance component in the model corresponding to variability between houses conditional on the predictors;
σ^2 = Variance component in the model corresponding to measurement variability within a house.

Then the unknown θ_i has the predictive distribution,

$$\theta_i|X, \beta \sim N(X_i\beta, \tau^2). \quad (1)$$

There is some uncertainty in the coefficients β (particularly for the indicators corresponding to counties with few observations) and a small amount of posterior uncertainty in the variance components of the model. For the purposes of this paper, we need only know the predictive distribution for any given θ_i, averaging over all these uncertainties; it will be approximately normal (because the variance components are so well estimated), and we label it as

$$\theta_i \sim N(M_i, S_i^2). \quad (2)$$

We write $M_i = (X\hat{\beta})_i$, where $\hat{\beta}$ is the posterior mean from the analysis in the appropriate region of the country. The variance S_i^2 includes the posterior uncertainty in the coefficients β and also the within-county variance τ^2. The GSD of the unexplained within-county variation, e^τ, is estimated to be in the range 1.9–2.3 (depending on the region of the country) which puts a lower limit on e^S. To be precise, the GSD's e^S of the predictive distributions for home radon levels vary from 2.1 to 3.0, and they are in the range [2.1, 2.5] for most U.S. houses (the houses with $e^S > 2.5$ lie in small-population counties for which little information was available in the radon surveys, resulting in relatively high predictive uncertainty within these counties). The GM's of the house posterior predictive distributions, e^M, vary from 0.1 to 14.6 pCi/L, with 95% in the range [0.3, 3.7] and 50% in the range [0.6, 1.6]. The houses with the highest prior GM's are houses with basement living areas in high-radon counties; the houses with lowest prior GM's have no basements and lie in low-radon counties. See Price and Nero (1996) for more details on the characteristics of high- and low-radon houses.

3.2 Using the Model as Input to Decision Analysis

This paper focuses on decisions, not modeling. For the rest of the paper we work at the individual house level and use the posterior inference for house i from the model discussed above as our prior distribution for the subsequent analysis. Since we are considering decisions for houses individually, we suppress the subscript i for the rest of the paper.

The distribution (2) summarizes the state of knowledge about the radon level in a house given its county and basement information. Now suppose a measurement $y \sim N(\theta, \sigma^2)$ is taken in the house. (We are assuming an unbiased measurement. If a short-term measurement is being used, it will have to be corrected for the bias shown in Table 2, and for an addition seasonal correction factor, if the measurement was not made in winter (e.g., see Mose and Mushrush, 1997; Pinel, Fearn, Darby and Miles, 1995). In our notation, y and θ are the logarithms of the measurement and the true ALAA radon level, respectively. The posterior distribution for θ is

$$\theta|M, y \sim N(\Lambda, V), \quad (3)$$

where

$$\Lambda = \frac{M/S^2 + y/\sigma^2}{1/S^2 + 1/\sigma^2}, \qquad V = \frac{1}{1/S^2 + 1/\sigma^2} \quad (4)$$

(see, e.g., Gelman, Carlin, Stern and Rubin, 1995). We base our decision analysis of when to measure and when to remediate on the distributions (2) and (3).

Before moving to the decision analysis, we briefly discuss the relevance of the hierarchical aspect of our radon model. In a classical regression model, the estimated distributions of home radon levels vary across counties because of the geographic variation in the regression predictors (for our model, these are listed in the first paragraph of Section 3). In the hierarchical regression model, the county estimates are allowed to vary from the regression prediction, by an amount dependent on the observational data in the county. As a result, the recommended decisions within any county depend on the available data for that county as well as on the estimated regression coefficients.

4. INDIVIDUAL DECISIONS ON WHETHER TO MONITOR OR REMEDIATE

The suggestion that every home should monitor is highly conservative (we might also say highly "protective"), based on the knowledge that homes with elevated radon concentrations have been found in every state, so the only way to be sure that a home does not have an elevated concentration is to test. However, if the risk is low enough [i.e., if the predicted radon level $M_i = \exp(X_i\beta)$ is low for house i], then even the small cost of monitoring may not be worthwhile.

We now work out the optimal decisions of measurement and remediation conditional on the predicted radon level in a home, the additional risk of lung cancer death from radon, the effects of remediation and individual attitude toward risk. We follow a standard approach in decision analysis (see, e.g., Watson and Buede, 1987) by proceeding in two steps: first, decision-making under certainty—at what level would you remediate if you knew R, your home radon level?—and, second, averaging over the uncertainty in R.

4.1 Decision-making under Certainty

We shall express decisions under certainty in three ways, equivalent under a linear no-threshold dose-response relationship.

1. The dollar value D_d associated with a reduction of 10^{-6} in probability of death from lung cancer (the value of a microlife). If one applies a 5% per year discounting of the value of a life and an expected twenty-year lag to lung cancer death, then D_d corresponds to a net present value of a microlife of $1.05^{20} D_d = 2.7 D_d$.
2. The dollar value D_r associated with a reduction of 1 pCi/L in home radon level for a thirty-year period (the equivalent dollar cost per unit of radon exposure).
3. The home radon level R_{action} above which you should remediate if your radon level is known.

We need to work with all three of these concepts because, depending on the context, either D_d, D_r or R_{action} will be most relevant for individual decision-making. In any case, the essence of the radon decision is a tradeoff between dollars and lives.

Initially, we make the following assumptions.

- The increase of probability of lung cancer death is a linear function of radon exposure (consistent with current concepts of dose effects in high linear-energy-transfer radiation; see Upfal, Divine and Siemiatychi, 1995). The added risk differs for smokers and nonsmokers and for males and females; we use estimates $\gamma_{g,s}$ (g = male or female, and s = ever-smoking or never-smoking) for the additional lifetime risk per additional pCi/L exposure as derived from the Committee on Health Risks of Exposure to Radon (BEIR VI, National Research Council, 1998); see Table 1.
- Remediation takes a house's annual-average living-area radon level down to a level R_{remed} if it was above that, but leaves it unchanged if it was below that. We shall assume that R_{remed} has the value 2 pCi/L.
- Mitigation costs \$2000, including the net present value of future energy cost to run the mitigation system.
- Decisions will be made based on the consequences over the next 30 years.
- If a measurement is taken, it is a long-term measurement that is an unbiased measure of annual-average living-area exposure with a measurement GSD of 1.2, and it costs \$50.

We can now determine the equivalent cost D_r per pCi/L of home radon exposure and the action level R_{action} for remediation given the following individual information.

- The numbers of male and female ever-smokers and never-smokers in the house, $n_{g,s}$; see Table 1.
- The dollars D_d that would be paid to reduce the probability of lung cancer death by one-millionth. From the risk assessment literature, typical values for medical interventions are in the range of \$0.1 to \$0.5 (see, e.g., Eddy, 1989, 1990; Owens, Harris, Scott and Nease, 1996). Higher values are often found in other contexts, for example, jury awards for deaths due to negligence, values used in legislating industrial risks and risk tradeoffs between worker wages and fatality risks (see Viscusi, 1992, for an excellent survey of values of risks to life and health). However, we feel that the lower values are reasonable in this case since, like medical intervention, expenditure on radon remediation is voluntary and is aimed at reducing future risk rather than compensating for job fatality.

For any given household, the equivalent cost per pCi/L, D_r, can be computed as a function of the risk assumed above and the individual parameters and D_d:

$$D_r = \frac{30}{70}\left(\sum_{g,s} n_{g,s}\gamma_{g,s}\right)10^6 D_d, \tag{5}$$

where the fraction 30/70 is the ratio of the thirty-year decision period to a seventy-year life expectancy per occupant. For U.S. homes, the average value of $\sum_{g,s} n_{g,s}\gamma_{g,s}$ is 0.0113 (see Table 1). We can also compute the remediation concentration R_{action}, given the equivalent cost and the above assumptions of cost and effects of remediation:

$$R_{\text{action}} = \frac{\$2000}{D_r} + R_{\text{remed}}. \tag{6}$$

4.2 Individual Choice of a Recommended Remediation Level under Certainty

The U.S., English, Swedish and Canadian recommended remediation levels are $R_{\text{action}} = 4$, 5, 10 and

20 pCi/L, which, with $R_{\text{remed}} = 2$ pCi/L, correspond to equivalent costs per pCi/L of $D_r = \$1000$, \$670, \$250 and \$111, respectively. Setting the values of $n_{g,s}$ to the average numbers of male and female ever-smokers and never-smokers in a U.S. household implies dollar values per microlife of $D_d = \$0.21$, \$0.14, \$0.05 and \$0.02, respectively. This suggests that, to the extent that we believe the standard estimates of radon risk and remediation effects, the U.S. and English recommendations are on the low end for acceptable risk reduction expenditures, and the Canadian and Swedish recommendations are too cavalier about the radon risk. However, this calculation obscures the dramatic difference between smokers and nonsmokers, which arises entirely from the difference in risk per dose associated with the two groups. For example, a household of one male never-smoker and one female never-smoker that is willing to spend \$0.21 per person to reduce the probability of lung cancer by 10^{-6} should spend \$370 per pCi/L of radon reduction, implying an action level of $R_{\text{action}} = 7.4$ pCi/L. In contrast, if the male and female are both smokers, they should be willing to spend the much higher value of \$1900 per pCi/L, because of their higher risk per pCi/L and thus should have an action level of $R_{\text{action}} = 3.1$ pCi/L.

Other sources of variation in R_{action}, in addition to varying risk preferences, are (a) variation in the number of smokers and nonsmokers in households, (b) variation in individual beliefs about the risks of radon and the effects of remediation, and (c) variation in the perceived dollar value associated with a given risk reduction. From a public policy standpoint, one might wish to ignore the variation attributable to (a), since over the thirty-year period of assumed remediation effectiveness the household composition is likely to change, and indeed the house is likely to be sold to several sets of new owners with possibly different smoking habits. However, as a practical matter, the homeowners are likely to perform remediation only if they foresee major risk reductions for themselves, or if they are planning to sell their house and fear that an elevated radon concentration will reduce its value. As illustrated above, a male-female never-smoking couple might choose an action level of 7.4 pCi/L or higher, depending on their willingness and ability to pay for risk reduction, whereas most smokers may be more willing to risk lung cancer than are nonsmokers and thus might be unwilling to remediate at levels as low as 3.1 pCi/L.

Through the rest of the paper, we use 4 pCi/L as an exemplary value, but rational informed individuals might plausibly choose quite different values of R_{action}, depending on smoking habits, risk tolerance, financial resources and the number of people in the household.

4.3 Decision-making under Uncertainty

Given an action level under certainty, R_{action}, we now address the question of whether to pay for a home radon measurement and whether to remediate. The decision of whether to measure depends on the prior distribution, (2) of radon level for your house, given your predictors X. The decision of whether to remediate depends on the posterior distribution, (3) if a measurement has been taken or the prior distribution, (2) otherwise. In our computations, we shall make use of the following results from the normal distribution: if $z \sim N(\mu, s^2)$, then $\mathrm{E}(e^z) = \exp(\mu + (1/2)s^2)$ and $\mathrm{E}(e^z | z > a)\Pr(z > a) = \exp(\mu + (1/2)s^2)(1 - \Phi((\mu + s^2 - a)/s))$, where Φ is the standard normal cumulative distribution function.

The decision tree is set up as three branches. In each branch, we evaluate the expected loss in dollar terms, converting radon exposure to dollars using $D_r = \$2000/(R_{\text{action}} - R_{\text{remed}})$ as the equivalent cost per pCi/L for additional home radon exposure.

1. *Remediate without monitoring.* Expected loss is remediation cost + equivalent dollar cost of radon exposure after remediation,

$$\begin{aligned} L_1 &= \$2000 + D_r \mathrm{E}(\min(R, R_{\text{remed}})) \\ &= \$2000 + D_r\big[R_{\text{remed}}\Pr(R \geq R_{\text{remed}}) \\ &\qquad + \mathrm{E}(R | R < R_{\text{remed}})\Pr(R < R_{\text{remed}})\big] \\ &= \$2000 + D_r\left[R_{\text{remed}}\Phi\left(\frac{M - \log(R_{\text{remed}})}{S}\right)\right. \\ &\qquad + \exp\left(M + \frac{1}{2}S^2\right) \\ &\qquad \left.\cdot\left(1 - \Phi\left(\frac{M + S^2 - \log(R_{\text{remed}})}{S}\right)\right)\right]. \end{aligned} \tag{7}$$

2. *Do not monitor or remediate.* Expected loss is the equivalent dollar cost of radon exposure,

$$L_2 = D_r \mathrm{E}(\exp(\theta)) = D_r \exp\big(M + \tfrac{1}{2}S^2\big). \tag{8}$$

3. *Take a measurement y* (measured in log pCi/L). The immediate loss is measurement cost (assumed to be \$50) and, in addition, the radon exposure during the year that you are taking the measurement [which is 1/30 of the thirty-year exposure (8)]. The inner decision has two branches.
   (a) *Remediate.* Expected loss is computed as for decision 1, but using the posterior rather than the prior distribution,

$$
\begin{aligned}
L_{3a} = \$50 &+ D_r \frac{1}{30} \exp\left(M + \frac{1}{2}S^2\right) + \$2000 \\
&+ D_r \Bigg[R_{\text{remed}} \Phi\left(\frac{\Lambda - \log(R_{\text{remed}})}{\sqrt{V}}\right) \\
&\quad + \exp\left(\Lambda + \frac{1}{2}V\right) \\
&\quad \cdot \left(1 - \Phi\left(\frac{\Lambda + V - \log(R_{\text{remed}})}{\sqrt{V}}\right)\right)\Bigg],
\end{aligned} \tag{9}
$$

where Λ and V are the posterior mean and variance, from equation (4).

(b) *Do not remediate.* Expected loss is

$$
\begin{aligned}
L_{3b} = \$50 &+ D_r \frac{1}{30} \exp\left(M + \frac{1}{2}S^2\right) \\
&+ D_r \exp\left(\Lambda + \frac{1}{2}V\right).
\end{aligned} \tag{10}
$$

4.3.1 *Decision of whether to remediate given a measurement.* To evaluate the decision tree, we must first consider the inner decision between 3(a) and 3(b), conditional on the measurement y. Let y_0 be the point (in log space) at which you will choose to remediate if $y > y_0$, or do nothing if $y < y_0$. (Because of measurement error, $y \neq \theta$, so $e^{y_0} \neq R_{\text{action}}$.) We shall solve for y_0 in terms of the prior mean M, the prior standard deviation S, and the measurement standard deviation σ, by solving the implicit equation

$$
L_{3a} = L_{3b} \quad \text{at } y = y_0. \tag{11}
$$

The expected losses L_{3a} and L_{3b} depend on y_0 only through $\Lambda = (M/S^2 + y/\sigma^2)/(1/S^2 + 1/\sigma^2)$, and so we can solve for y_0 by first solving for Λ_0 in (11), then setting

$$
y_0 = \left(1 + \frac{\sigma^2}{S^2}\right)\Lambda_0 - \frac{\sigma^2}{S^2}M. \tag{12}
$$

Thus the relation between y_0 and M is linear, with the slope depending only on the variance ratio σ^2/S^2.

Given σ^2/S^2 and D_r, we solve for Λ_0 numerically, using the bisection method to converge on the value of Λ that satisfies (11). Figure 1 shows the measurement action level e^{y_0} as a function of the perfect-information action level R_{action}, evaluated at values of the prior GM radon level e^M ranging from 0.5 to 4.0. For this example, we have assumed that $\sigma = \log(1.2)$, and that $S = \log(2.3)$ for all counties.

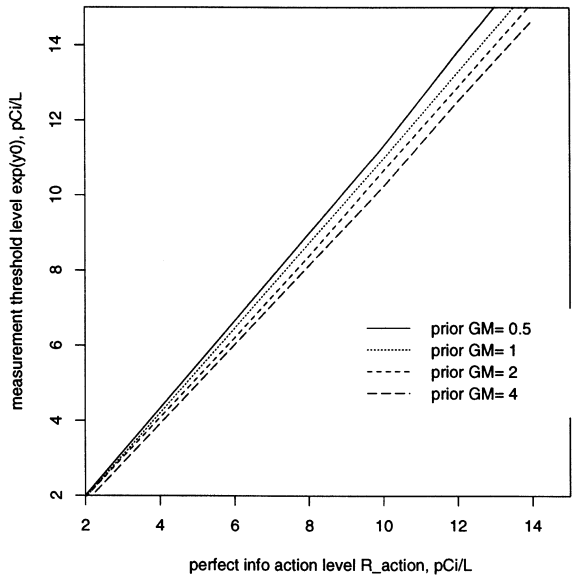

FIG. 1. *Measurement action levels e^{y_0} as a function of the perfect-information action level R_{action}, evaluated at values of the prior GM radon level e^M ranging from 0.5 pCi/L to 4 pCi/L. For a given R_{action} and prior GM, find the "measurement threshold"; if the measured value exceeds this threshold, then remediation is recommended. The threshold differs substantially from R_{action} only for low prior GM and high R_{action}.*

4.3.2 *Deciding whether to measure.* We determine the expected loss for branch 3 of the decision tree by averaging over the prior uncertainty in the measurement y,

$$
L_3 = \text{E}(\min(L_{3a}, L_{3b})). \tag{13}
$$

Given (M, S, σ, D_r), we evaluate this expression as follows.

1. Simulate 5000 draws of $y \sim N(M, S^2 + \sigma^2)$.
2. For each draw of y, compute $\min(L_{3a}, L_{3b})$ from (9) and (10).
3. Estimate L_3 as the average of these 5000 values.

Of course, this expected loss is valid only if we assume that you will make the recommended optimal decision once the measurement is taken.

We can now compare the expected losses L_1, L_2, L_3, and choose among the three decisions. Figure 2 displays the expected losses as a function of the perfect-information action level R_{action} for several values of e^M. As with Figure 1, we illustrate with $\sigma = \log(1.2)$ and $S = \log(2.3)$. For any value of M and R_{action}, the recommended decision is the one with the lowest expected loss.

For any R_{action}, we can summarize the decision recommendations as the cut-off levels M_{low}

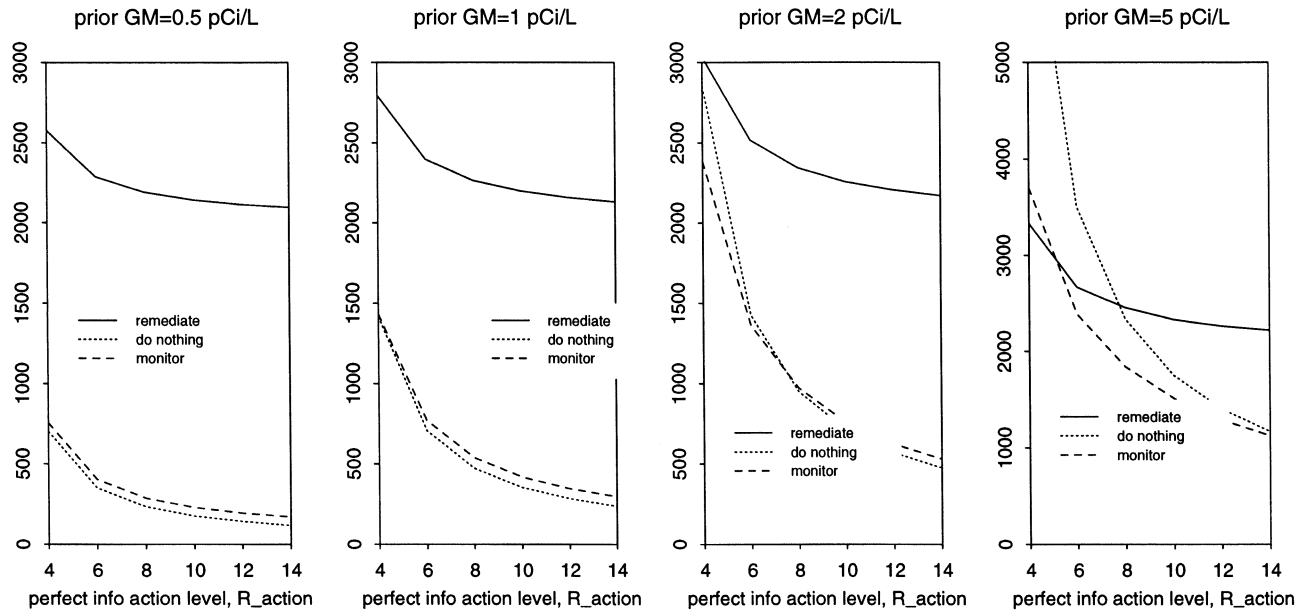

FIG. 2. *Expected losses in dollars (including the dollar value of the expected reductions in radon levels) of the three decisions*: (1) *remediate,* (2) *do nothing*, (3) *monitor (take a measurement), as a function of the perfect-information action level* R_{action}. *The four plots correspond to four different values of the prior geometric mean radon level* e^M.

and M_{high} for which decision 1 is preferred if $M > M_{\text{high}}$, decision 2 is preferred if $M < M_{\text{low}}$, and decision 3 is preferred if $M \in [M_{\text{low}}, M_{\text{high}}]$. Figure 3 displays these cut-offs as a function of R_{action}, and thus displays the recommended decision as a function of (R_{action}, e^M), once again under the simplifying assumption that $\sigma = \log(1.2)$ and $S = \log(2.3)$ for all counties. For example, setting $R_{\text{action}} = 4$ pCi/L leads to the following recommendation based on e^M, the prior GM of your home

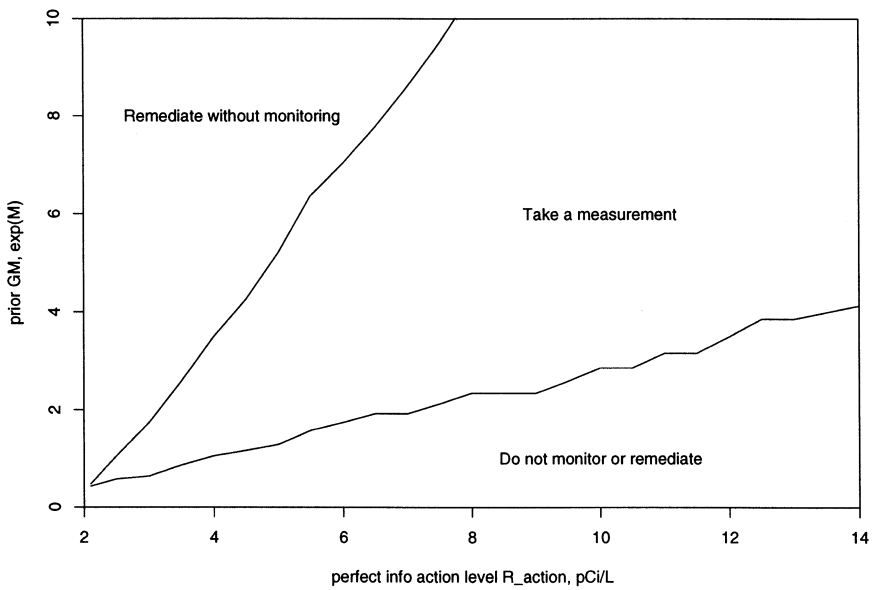

FIG. 3. *Recommended decisions as a function of the perfect-information action level* R_{action} *and the prior geometric mean radon level* e^M, *under the simplifying assumption that* $e^S = 2.3$. *You can read off your recommended decision from this graph and, if the recommendation is "take a measurement," you can do so and then use Figure* 1 *(with interpolation or extrapolation if necessary) to tell you whether to remediate. The horizontal axis of this figure begins at* 2 pCi/L *because remediation is assumed to reduce ALAA radon level to* 2 pCi/L, *so it makes no sense for* R_{action} *to be lower than that value. Wiggles in the lines are due to simulation variability.*

radon based on your county and house type:

- If e^M is less than 1.0 pCi/L (which corresponds to 68% of U.S. houses), do nothing.
- If e^M is between 1.0 and 3.5 pCi/L (27% of U.S. houses), perform a long-term measurement (and then decide whether to remediate).
- If e^M is greater than 3.5 pCi/L (5% of U.S. houses), remediate immediately without measuring. Actually, in this circumstance, short-term monitoring turns out to be (barely) cost-efficient: the reason for the recommendation of immediate remediation is that the excess risk associated with occupying the home for a year while a long-term measurement is made is not worth bearing, given the high likelihood that the home will eventually be remediated anyway. However, if a short-term measurement is made and is sufficiently low, then the home is unlikely to have such an exceptionally high level that one additional year of exposure carries a large risk. In this case, long-term monitoring can be performed to determine whether remediation is really indicated. We will ignore this additional complexity to the decision tree, since it occurs rarely and has very little impact on the overall cost-benefit analysis.

4.4 Decision-Making If a Short-term Measurement Has Been Taken

We do not in general recommend taking short-term measurements, because long-term measurements are much superior in terms of both bias and variance. However, short-term measurements are quite popular (partly because these are often taken as a condition of sale of a house), and so it is worth considering the decision problem in this situation.

In fact, the above decision framework is immediately adaptable to a homeowner who has already taken a short-term measurement. The only change that needs to be made is that the prior distribution (2) needs to be updated given the information from the short-term measurement. We thus replace M and S^2 in the above formulas by

$$M_{\text{new}} = \frac{M/S^2 + (y_{\text{st}} - \log b)/\sigma_{\text{st}}^2}{1/S^2 + 1/\sigma_{\text{st}}^2}, \qquad S_{\text{new}}^2 = \frac{1}{1/S^2 + 1/\sigma_{\text{st}}^2}, \tag{14}$$

where y_{st} is the logarithm of the short-term measurement, b is the correction factor derived from Table 2 and $\sigma_{\text{st}} = \log(1.8)$. If the short-term measurement was not made in winter, then a seasonal correction factor will also apply; see, for example, Mose and Mushrush (1997) and Pinel et al. (1995). At this point, we can return to the procedure described in the previous sections.

4.5 Summary of the Individual Decision Process

Ideally, an individual homeowner in the United States can now make a remediation decision using the following process:

1. Determine the radon level R_{action} above which you would remediate, if you knew your home radon level exactly. This value can be chosen in its own right or by choosing a value of D_r based on the perceived gains from lowering radon level or by assigning a dollar value D_d to a millionth of a life and computing based on the number of ever-smokers S and never-smokers N in the house. As discussed in Section 4.2, current understanding of the risks of radon and the effects of remediation suggest that the EPA's recommendation of 4 pCi/L is a reasonable catch-all value, with 8 pCi/L being a more reasonable value for nonsmokers.
2. Look up e^M and e^S, the GM and GSD of the posterior predictive distribution for your home's radon level, as estimated from the hierarchical model described in Section 3.
3. If a short-term measurement has been taken, update the prior distribution using (14) and the bias correction from Table 2 (and possibly an additional seasonal correction).
4. Calculate the expected losses of decisions (1), (2) and (3) from the formulas in Section 4.3 and, if decision (3) is chosen, the recommended measurement action level e^{y_0}. The recommended decision—that with the lowest expected loss, corresponds to that indicated in Figure 4.3.2 (with slight alterations depending on the exact value of S).
5. If decision (3) is chosen, perform a long-term measurement. In one year, the measurement e^y is available. Remediate if $e^y > e^{y_0}$.

We are in the process of setting up a website at `http://www.stat.columbia.edu/radon` to automate the steps listed above and supply other information about decision-making for radon hazards.

5. AGGREGATE CONSEQUENCE OF DECISION STRATEGY

Now that we have made idealized recommendations, we consider their aggregate effects if followed by all homeowners in the United States. In particular, how much better are the consequences compared to other policies such as the current one, implicitly endorsed by the EPA, of taking a short-

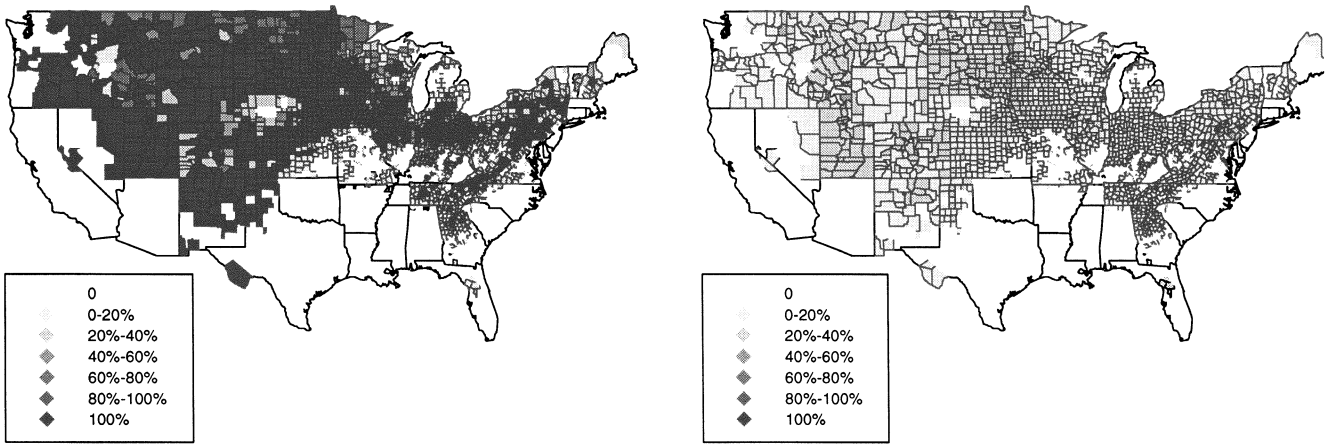

FIG. 4. *Map (a) showing fraction of houses in each county for which measurement is recommended, given the perfect-information action level of* $R_{\text{action}} = 4\ pCi/L$; *(b) expected fraction of houses in each county for which remediation will be recommended, once the measurement y has been taken. For the present radon model, within any county the recommendations on whether to measure and whether to remediate depend only on the house type: whether the house has a basement and whether the basement is used as living space. Apparent discontinuities across the boundaries of Utah and South Carolina arise from irregularities in the radon measurements from the radon surveys conducted by those states, an issue we ignore in the present paper.*

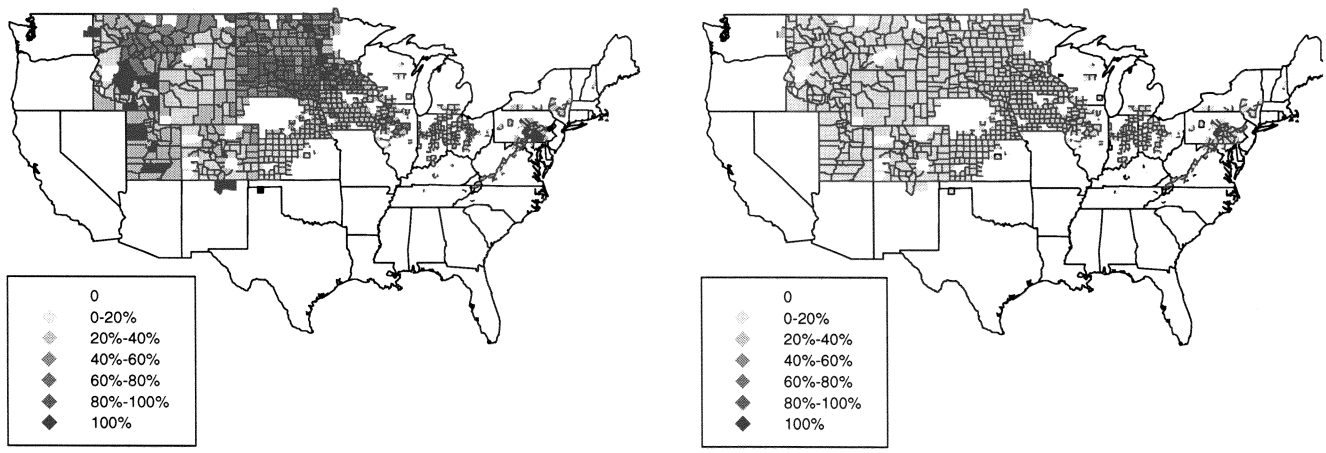

FIG. 5. *Map (a) showing fraction of houses in each county for which measurement is recommended, given the perfect-information action level of* $R_{\text{action}} = 8\ pCi/L$; *(b) expected fraction of houses in each county for which remediation will be recommended, once the measurement y has been taken. As with the previous figure, the decision recommendations depend only on county and house type.*

term measurement as a condition of a home sale and performing remediation if the measurement is higher than 4 pCi/L?

5.1 Estimated Consequences of Applying the Recommended Decision Strategy to the Entire United States

Figures 4 and 5 display the geographic pattern of recommended measurements (and, after one year, recommended remediations), based on action levels R_{action} of 4 and 8 pCi/L, respectively. These recommendations incorporate the effects of parameter uncertainties in the models that predict radon distributions within counties, so these maps would be expected to change somewhat as better predictions become available. Note that these maps are *not* based on a single estimated parameter such as "the probability that a home's concentration exceeds 4 pCi/L." Although a discrete action level does play a role in the decision process (after all, each home must either monitor or not, and remediate or not) the benefit of remediation is a continuous function of the initial radon concentration, and that concentration is assumed to be drawn from a continuous distribution. It is the confluence of these continuous distributions and the

discrete willingness-to-remediate point that give rise to the fairly complex expressions for expected loss in Section 4.3.

From a policy standpoint, perhaps the most significant feature of the maps is that even if the EPA's recommended action level of 4 pCi/L is assumed to be correct—and, as we have discussed, it does lead to a reasonable value of D_d, under standard dose-response assumptions—monitoring is still not recommended in most U.S. homes. Indeed, only 28% of U.S. homes would perform radon monitoring. A higher action level of 8 pCi/L, a reasonable value for nonsmokers under the standard assumptions, would lead to even more restricted monitoring and remediation: only about 5% of homes would perform monitoring.

5.2 Decision Strategies Considered and Evaluation Criteria

In this section, we shall consider various decision strategies.

1. Follow the recommended strategy from Section 4.3 (that is, monitor homes with prior mean estimates above a given level, and remediate those with high measurements).
2. Perform long-term measurements on all houses and then remediate those for which the measurement exceeds a specified level: $e^y > R_{\text{action}}$.
3. Perform short-term measurements on all houses and then remediate those for which the bias-corrected measurement exceeds a specified level: $e^{y_{\text{st}}}/b > R_{\text{action}}$ (with b defined as described in Section 4.4).
4. Perform short-term measurements on all houses and then remediate those for which the uncorrected measurement exceeds a specified level: $e^{y_{\text{st}}} > R_{\text{action}}$.

We evaluate each of the above strategies in terms of aggregate lives saved and dollars cost, with these outcomes parameterized by the radon action level R_{action}. Both lives saved and costs are considered for a thirty-year period. For each strategy, we assume that the level R_{action} is the same for all houses (this would correspond to a uniform national recommendation) and that 0.30 male and 0.27 female ever-smokers and 1.07 male and 1.16 female never-smokers live in each house (or, rather, that these are the averages over the thirty-year period).

We also evaluate strategies based on the estimated cost per life saved. This aggregate cost per life is different from the marginal cost per life used to set the action level R_{action} in Section 4.2. For example, as discussed previously, an action level of $R_{\text{action}} = 4$ pCi/L approximately corresponds to a net present value of \$0.21 per microlife, which corresponds to a marginal cost of \$210,000 per life saved. However, if the optimal recommendation is followed for the entire country, the estimated aggregate cost per life saved is only \$87,000: the aggregate cost averages over the whole population, ranging from mitigations that are barely cost-effective through mitigations that are highly efficient in terms of risk reduction for a given cost. See also Figure 10 for a comparison of aggregate and marginal costs per life saved.

5.3 Modeling the Variation in the Population of U.S. Homes

Because we use inferences from a hierarchical model, we are able to give different recommendations for different houses in the population as characterized by location as well as continuous covariates.

Thus, aggregate effects are determined by adding up the individual decisions over all the ground-contact homes in the country. Considering 3078 counties with three house types within each, we have 3078×3 pairs of (M, S) obtained from the hierarchical model fit to the national and state radon survey data as described in Section 3. Given $(M, S, R_{\text{action}})$, the decisions of whether to monitor and whether to measure are made as described in Section 4.5, and expected number of lives saved and cost spent are assessed if remediation is implemented.

For any of the decision strategies, in any given house, we evaluate the total cost,

$$\begin{aligned}\text{Expected cost} &= \$50\ \Pr(\text{measurement}) \\ &\quad + \$2000\ \Pr(\text{remediation}),\end{aligned} \tag{15}$$

where

$$\Pr(\text{measurement}) = \begin{cases} 1_{\{M_{\text{low}} < M < M_{\text{high}}\}}, & \text{for strategy 1,} \\ 1, & \text{for strategies 2, 3 and 4,} \end{cases}$$

and

$$\Pr(\text{remediation}) = \begin{cases} \Pr(M > M_{\text{high}}) \\ \quad + \Pr\big((M_{\text{low}} < M < M_{\text{high}}) \text{ and } (y > y_0)\big) \\ \quad = 1_{\{M > M_{\text{high}}\}} + 1_{\{M_{\text{low}} < M < M_{\text{high}}\}} \\ \qquad \cdot \left(1 - \Phi\left(\dfrac{\Lambda - y_0}{\sqrt{V}}\right)\right), & \text{for strategy 1,} \\ \Pr(y > \log(R_{\text{action}})) \\ \quad = 1 - \Phi\left(\dfrac{\Lambda - \log(R_{\text{action}})}{\sqrt{V}}\right), \\ & \text{for strategy 2,} \\ \Pr(y_{\text{st}} - \log b > \log(R_{\text{action}})) \\ \quad = 1 - \Phi\left(\dfrac{M_{\text{new}} - \log(R_{\text{action}})}{S_{\text{new}}}\right), \\ & \text{for strategy 3,} \\ \Pr(y_{st} > \log(R_{\text{action}})) \\ \quad = 1 - \Phi\left(\dfrac{M_{\text{new}} - \log(R_{\text{action}} + \log b}{S_{\text{new}}}\right), \\ & \text{for strategy 4} \end{cases}$$

(with \$50 replaced by \$15 for strategies 3 and 4 in which short-term measurements are used), and we evaluate the expected lives saved,

$$\begin{aligned} &\text{Expected lives saved} \\ &\quad = A \cdot \mathrm{E}(\max(e^{\theta} - R_{\text{remed}}, 0) | \text{remediation}) \\ &\qquad \cdot \Pr(\text{remediation}) \\ &\quad = A \cdot R_{\text{reduced}} \Pr(\text{remediation}), \end{aligned} \tag{16}$$

where

$$R_{\text{reduced}} = \exp(\Lambda + V/2)\Phi\left(\frac{\Lambda + V - \log(R_{\text{remed}})}{\sqrt{V}}\right) - R_{\text{remed}}\Phi\left(\frac{\Lambda - \log(R_{\text{remed}})}{\sqrt{V}}\right),$$

and A is the expected lives lost in a thirty-year period per pCi/L of home radon exposure, given by D_r/D_d from equation (5) for any home, and equal to 0.0113 for the "average household" of 0.3 male ever-smokers, 1.07 male never-smokers, 0.27 female ever-smokers and 1.16 female never-smokers. In the above formulas, Λ and V are given by (4), and M_{new} and S_{new} are given by (14).

We evaluate the expectations in (15) and (16) by simulation. First, we simulate 5000 draws of $y \sim N(M, S^2+\sigma^2)$ (for strategies 1 and 2) or $y \sim N(M+\log b,\ S^2+ = \sigma_{\text{st}}^2)$ (for strategies 3 and 4). Second, for each draw of y, we compute $A \cdot R_{\text{reduced}}$ under the constraints of $M > M_{\text{high}}$ or $((M_{\text{low}} < M < M_{\text{high}})$ and $(y > y_0))$ or $y > \log(R_{\text{action}})$, and then estimate (16) and (15) as the average of these 5000 draws. Simulations average over uncertainties in home radon levels R and variability in measurements e^y (or $e^{y_{\text{st}}}$). For these calculations we used the actual model estimates of S, rather than setting them all equal to a single value as was done for illustrative purposes in the previous section.

We then multiply by the total number of ground contact houses for each (M, S), that is, for each house type and for each county, and sum them up to get expected total costs and lives saved over a thirty-year period in the United States.

5.4 Results

For the present county-level radon model, within each county monitoring is recommended for some subset of homes: for all homes, for all homes with basements, for all homes with living-area basements or for no homes. The maps in Figure 4 display, for each county, the fraction of houses that would measure and the estimated fraction of houses that would remediate if the recommended decision strategy were followed everywhere with $R_{\text{action}} = 4$ pCi/L. About 26% of the 70 million ground-contact houses in the United States would monitor. This would result in detection of and remediation of 2.8 million homes above 4 pCi/L (74% of all such homes), and 840,000 of the homes above 8 pCi/L (91% of all such homes). Some additional estimates of the program's effectiveness are presented in Tables 3 and 4, and Figure 5 displays similar maps for an 8 pCi/L action level.

In order to understand the effects of the different decision strategies on aggregate outcomes, we have developed a series of graphs. Figures 6 and 7 illustrate the efficiency of the recommended remediation strategy by showing the overall distributions of radon levels (and total radon exposures) and the distributions of homes to be monitored and remediated. As is apparent in the figures, even with the large uncertainties in individual county distributional parameters the recommended program is quite effective at focusing on the homes with the highest indoor radon concentrations.

Figure 8 displays the trade-off between expected cost and expected lives saved over a thirty-year period for the four strategies listed in Section 5.2. The numbers on the curves are action levels R_{action}. This figure allows us to compare the effectiveness of alternative strategies of equal expected cost or equal expected lives saved. For example, the recommended strategy (the solid line on the graph) at $R_{\text{action}} = 4$ pCi/L would result in an expected 83,000

TABLE 3
*Some summary statistics on the effectiveness of various home radon measurement and remediation strategies for the 4 pCi/L action level**

	Strategy			
	1	**2**	**3**	**4**
Fraction of all U.S. homes that would measure	26%	100%	100%	100%
Fraction of all U.S. homes that would remediate	5%	6%	8%	17%
Fraction of all homes over 4 pCi/L that would remediate	74%	89%	74%	92%
Fraction of all homes over 8 pCi/L that would remediate	91%	100%	95%	99%
Total cost ($ billion)	7.32	11.56	12.20	25.06
total cost of measuring	0.97	3.50	1.06	1.06
total cost of remediation	6.35	8.06	11.14	24.00
Expected lives saved	84,000	97,000	88,000	110,000
ever-smokers	49,000	57,000	51,000	64,000
never-smokers	35,000	40,000	37,000	46,000
Aggregate $ cost per life saved	87,000	119,000	138,000	228,000

*(1) recommended strategy based on decision analysis using the hierarchical model, (2) long-term measurements on all houses, (3) bias-corrected short-term measurements on all houses, (4) uncorrected short-term measurements on all houses. All are based on an action level of $R_{\text{action}} = 4$ pCi/L. Costs and lives saved cover 30 years.

TABLE 4
*Some summary statistics on the effectiveness of various home radon measurement and remediation strategies for the 8 pCi/L action level**

	Strategy			
	1	**2**	**3**	**4**
Fraction of all U.S. homes that would measure	5%	100%	100%	100%
Fraction of all U.S. homes that would remediate	0.7%	1.4%	2.5%	7.0%
Fraction of all homes over 4 pCi/L that would remediate	12%	26%	37%	67%
Fraction of all homes over 8 pCi/L that would remediate	44%	87%	70%	91%
Total cost ($ billion)	1.11	5.54	4.63	10.94
total cost of measuring	0.19	3.50	1.06	1.06
total cost of remediation	0.92	2.04	3.53	9.84
Expected lives saved	27,000	50,000	51,000	82,000
ever-smokers	16,000	29,000	30,000	48,000
never-smokers	11,000	21,000	21,000	34,000
Aggregate $ cost per life saved	42,000	110,000	90,000	133,000

*(1) recommended strategy based on decision analysis using the hierarchical model, (2) long-term measurements on all houses, (3) bias-corrected short-term measurements on all houses, (4) uncorrected short-term measurements on all houses. All are based on an action level of $R_{\text{action}} = 8$ pCi/L. Costs and lives saved cover 30 years.

lives saved at an expected cost of $7.3 billion. Let us compare this to the EPA's implicitly recommended strategy based on uncorrected short-term measurements (the dashed line on the figure). For the same cost of $7.3 billion, the uncorrected short-term strategy is expected to save only 32,000 lives; to achieve the same expected savings of 83,000 lives, the uncorrected short-term strategy would cost about $19 billion.

Figure 9 displays these results in another way, as estimated cost per life saved, as a function of expected cost, for the four strategies. Finally, Figure 10 displays the estimates for both marginal and average cost per life saved, for the recommended decision strategy, as a function of the radon action level R_{action}. The average cost per life saved is estimated as described above, and the marginal cost per life saved is simply $10^6 D_d$ (as defined in Section 4.1). Average cost per life saved is always lower than marginal cost because, for any action level, the average includes all houses at or above that level, and remediations are more efficient (in terms of lives saved per dollar) in the higher-radon houses.

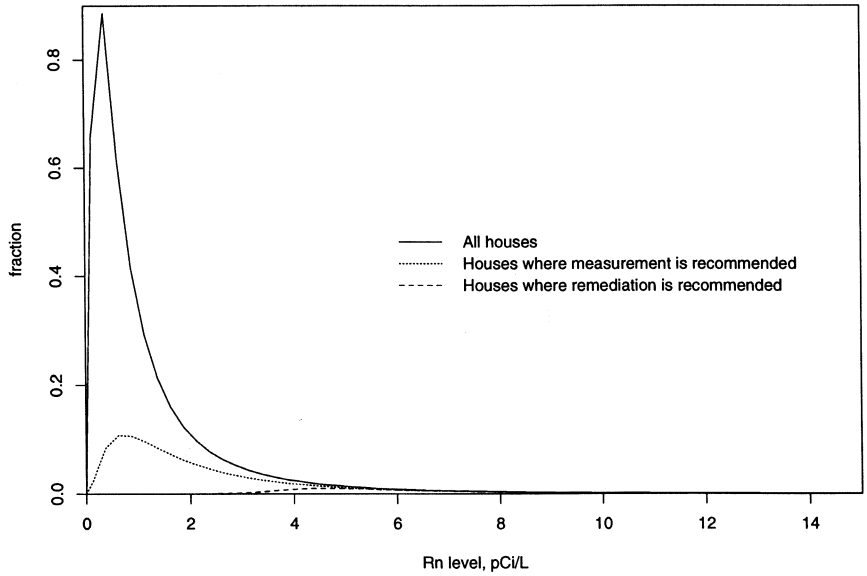

FIG. 6. *Estimated distributions of annual living area average radon concentrations in (upper, solid line) all U.S. houses, (middle, dotted line) all houses where measurement is recommended under the optimal strategy (with* $R_{\text{action}} = 4\ pCi/L$*), and (lower, dashed line) all houses where remediation is recommended immediately or after measurement.*

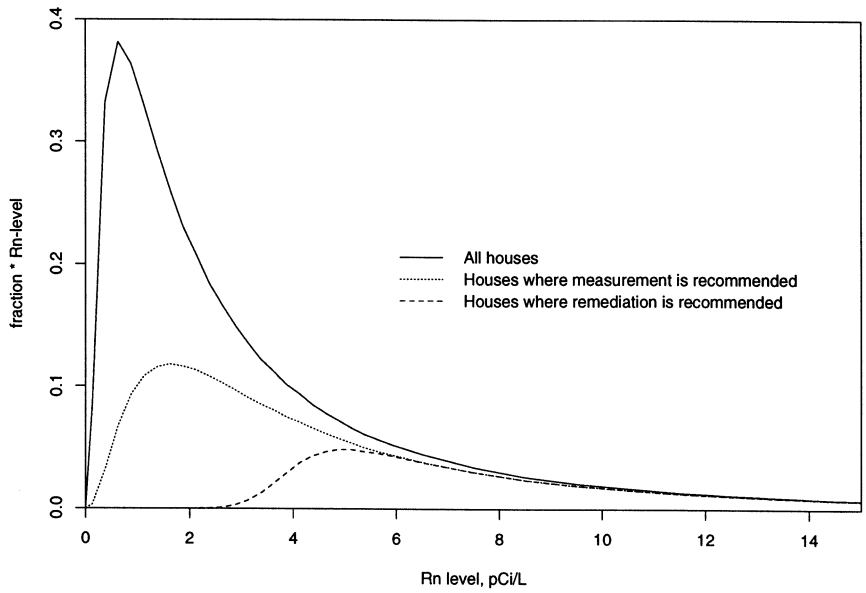

FIG. 7. *Fraction of total radon exposure, as a function of indoor radon concentration. Where the previous plot shows* $f(\theta)$*, this plot shows* $\theta f(\theta)$*. Curves are shown for (upper, solid line) all U.S. houses, (middle, dotted line), all houses where measurement is recommended under the optimal strategy (with* $R_{\text{action}} = 4\ pCi/L$*) and (lower, dashed line) all houses where remediation is recommended immediately or after measurement.*

6. SENSITIVITY TO ASSUMPTIONS

Our results are subject to potential error in:

- Estimates of annual average living area radon exposure (and its variation) from home radon measurements and the hierarchical model (including basement information and geographic predictors).
- The magnitude of cancer risk from a given radon concentration (including the assumed linearity of cancer risk as a function of radon level).
- The effects of remediation.

In this section, we consider each of these issues in turn and then discuss other factors, involving indi-

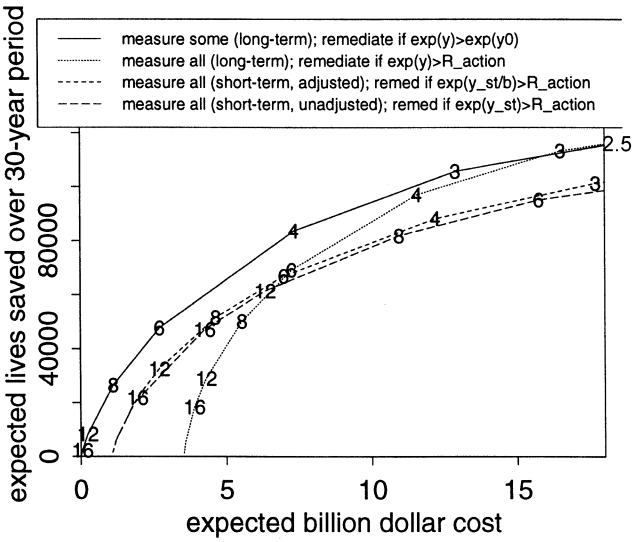

FIG. 8. *Expected lives saved versus expected cost, for various radon measurement–remediation strategies discussed in Section* 5.2. *Numbers indicate values of* R_{action}. *The solid line is for the recommended strategy of measuring only certain homes; the others assume that all homes are measured. All results are estimated totals for the United States over a thirty-year period.*

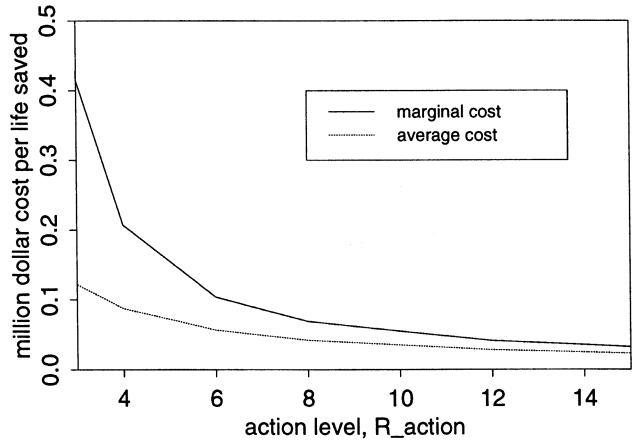

FIG. 10. *Estimated average and marginal costs per life saved versus action level* R_{action} *for the recommended decision strategy. Average cost per life saved is computed averaging over the distribution of U.S. houses, as displayed in Figure* 8. *Marginal cost per life saved is* $10^6 D$ *(as defined in Section* 4.1*) based on a household with* 0.3 *male and* 0.27 *female ever-smokers and* 1.07 *male and* 1.16 *female never-smokers. Marginal cost is always higher than average cost because the marginal houses are those for which it is just barely cost-effective to remediate.*

vidual preferences and behavior, that might affect the decisions.

Statistical model of home radon levels. The model has been extensively validated (see Price, Nero and Gelman, 1996; Price and Nero, 1996; Price, 1997).

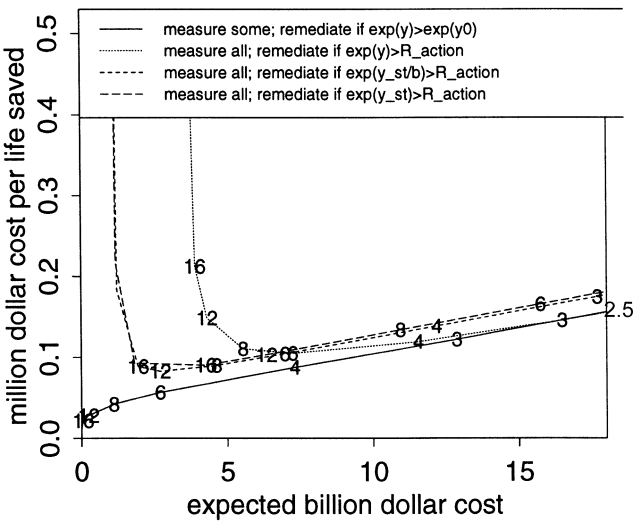

FIG. 9. *Estimated cost per life saved versus expected cost (over a thirty-year period), for various radon measurement–remediation strategies discussed in Section* 5.2. *The solid line is for the recommended strategy of measuring only certain homes; the others assume that all homes are measured. For comparison, remediating every house without making any measurements has an estimated cost per life saved of* $2.2 *million.*

In general, the model behaves well; cross-validation indicates that the uncertainty intervals are approximately correct, for example. However, it is likely that the lognormality assumption (for homes in a given county, with a given set of explanatory variables) underestimates the number of homes in the high tail of radon concentrations for some counties. For instance, Hobbes and Maeda (1997) suggest that some counties in southern California might be better fit as a mixture of two lognormals, one with a low geometric mean for most of the homes and one with a high geometric mean for the small fraction of homes on a particular geologic deposit. Similar high-radon pockets or exceptionally high within-county variability are known to occur in a few counties in Florida, New York, Washington and elsewhere.

From the standpoint of individual decisions, an underestimate of the size of the very high tail of radon concentrations would generally have a small effect: as long as the cumulative exposure for homes exceeding the action level is not seriously in error, the recommendation of whether or not to monitor will not be affected, so if the fraction of homes over 4 pCi/L or 8 pCi/L is fairly accurately estimated using the lognormal approximation, the exact distribution of a small number of very high homes is not critical. The fraction of homes over 4 or 8 pCi/L is fairly well estimated under the lognormal approximation for most of the counties with GM's over 1 or 2 pCi/L, respectively, and most counties with GM's lower than that have such low numbers of homes

over the action level that even a large relative error in their prevalence would probably not change the monitoring recommendations. Given the large number of counties in the United States (over 3000) it is very likely that there are at least a few for which nonlognormality is a significant issue, but it is unlikely to seriously affect most of our results.

This whole issue becomes more important, though, if the action level is set very high (e.g., for a female nonsmoker living alone); we would not trust the model's exact predictions when estimating the frequency of rare cases such as homes over 20 pCi/L.

On a different model-related topic, it is possible that the model can be improved by including more spatial or geological information (see, e.g., Boscardin, Price and Gelman, 1996; Geiger and Barnes, 1994; Mose and Mushrush, 1997; Miles and Ball, 1996), which would cause predictions for individual homes to become more precise and the prior standard deviations S to decrease. For instance, radon mapping within counties would allow recommendations to discriminate more precisely among houses and thus increase expected lives saved for any given dollar expenditure. Indeed, such targeted recommendations may already be possible in localized (subcounty) areas that can be confidently identified as having disproportionate numbers of high-radon homes.

Magnitude of cancer risk from radon exposure. There is disagreement as to the estimate of lung cancer risk attributable to radon exposure, despite the efforts of the BEIR committee to thoroughly review the data available. The main issue is whether the results from the analysis of the data for miners could be generalized and applied to the progeny of radon in homes (Lubin and Boice, 1997). Even if a linear no-threshold model is appropriate, the coefficients (the risk per unit exposure) are uncertain by at least a factor of 1.4. In addition, the model of discrete risks for smokers and nonsmokers is a simplification since smoking levels vary and many nonsmokers are exposed to secondhand smoke.

Linearity of the dose-response function. Experiments on animals, plus epidemiological studies with miners and others exposed at very high doses, suggest that at high doses the dose-response is approximately linear (see Nazaroff and Nero, 1988, Chapters 8–9). However, there are really no good data at low doses. The Environmental Protection Agency assumes the function is linear all the way to zero, but others have suggested that there are a threshold (an exposure below which there are no effects) or even a protective effect at low concentrations (Cohen, 1995; Bogen, 1997).

Case-control studies suggest that, if there is a protective effect at low levels, it cannot be large, but mild protective effects, or a threshold so that levels below 5 pCi/L or so have no effect, cannot be ruled out. However, in spite of claims to the contrary by Cohen (1995), we are confident that long-term exposure to 2 pCi/L is safer than exposure to, say, 10 pCi/L (see, e.g., Lubin and Boice, 1997).

Moreover, our results are less sensitive than one might suppose to nonlinearities in the dose-response function at low concentrations. This is because we assume that remediation reduces radon levels to 2 pCi/L, so the dose-response below that concentration is irrelevant. For instance, if long-term exposure at 2 pCi/L were actually safer than no exposure at all, that would have no effect on our analysis under the present assumptions.

To get some idea of the sensitivity of our results to the details of the dose-response relationship at low doses, we consider the effects of a relationship with a threshold at 4 pCi/L, so that exposure below that level has no health effect. One might examine this issue in several ways. For instance, we could ask what the optimal strategy would be under this modified dose-response relationship and see how the recommended actions (e.g., which homes should monitor and which should remediate) would change compared to the recommendations based on the linear dose-response. Instead, we look at how the number of lives saved would change if the strategy based on the linear dose-response were implemented; that is, if all of the same homes monitor or remediate as for the linear dose-response, but if the dose-response actually has a threshold. This seems to us to be the more relevant question, since our goal is to understand the robustness of the present analysis rather than to seriously propose analyses under alternative dose-response functions. Also, alternative recommendations would merely entail further restrictions on which homes are candidates for monitoring, so determining exactly which homes those are is not likely to be particularly instructive.

Given a threshold at 4 pCi/L, remediations in homes close to that threshold are mostly wasted (and all remediations are less beneficial), so we expect a reduction in lives saved. Some summary statistics are given in the columns labeled (b) in Table 5. As expected, the resulting number of lives saved substantially changes according to this assumption; compared to the situation with a linear dose-response, 37% fewer lives are saved for $R_{\text{action}} = 4$ pCi/L, and 23% fewer are saved for $R_{\text{action}} = 8$ pCi/L. Costs per life saved are still lowest under the recommended strategy 1.

TABLE 5

*Sensitivity analysis: expected total lives saved and cost per life saved under four strategies for a grid of R_{action}, under three different models**

Action Level	Strategy	Total lives saved (30 years) (a)	(b)	(c)	Total cost ($ billion)	Aggregate dollars per life saved (a)	(b)	(c)
2.5 pCi/L	1	116,000	59,000	108,000	18.4	158,000	310,000	169,000
	2	119,000	60,000	113,000	20.6	172,000	342,000	183,000
	3	108,000	57,000	105,000	22.0	204,000	380,000	210,000
	4	119,000	60,000	126,000	39.7	332,000	664,000	317,000
3 pCi/L	1	105,000	57,000	94,000	12.9	122,000	222,000	136,000
	2	113,000	60,000	102,000	16.5	145,000	273,000	161,000
	3	102,000	56,000	95,000	17.7	175,000	314,000	185,000
	4	116,000	59,000	119,000	33.6	288,000	565,000	282,000
4 pCi/L	1	84,000	53,000	73,000	7.3	87,000	138,000	101,000
	2	97,000	59,000	84,000	11.6	119,000	195,000	136,000
	3	88,000	53,000	80,000	12.2	138,000	231,000	151,000
	4	110,000	57,000	108,000	25.1	228,000	430,000	232,000
6 pCi/L	1	48,000	37,000	40,000	2.7	57,000	74,000	66,000
	2	68,000	51,000	59,000	7.2	105,000	141,000	123,000
	3	67,000	45,000	59,000	7.0	104,000	157,000	119,000
	4	95,000	54,000	89,000	15.8	165,000	289,000	177,000
8 pCi/L	1	27,000	22,000	22,000	1.1	42,000	51,000	49,000
	2	50,000	40,000	43,000	5.5	110,000	137,000	129,000
	3	51,000	37,000	45,000	4.6	90,000	126,000	104,000
	4	82,000	49,000	73,000	10.9	133,000	219,000	148,000
12 pCi/L	1	8,000	7,000	7,000	0.2	28,000	32,000	33,000
	2	29,000	25,000	25,000	4.3	148,000	171,000	173,000
	3	32,000	25,000	28,000	2.7	83,000	107,000	97,000
	4	61,000	41,000	54,000	6.3	103,000	153,000	117,000

*(a) cancer risk without threshold and postremediation radon level 2 pCi/L; (b) cancer risk with threshold 4 pCi/L and postremediation radon level 2 pCi/L; (c) cancer risk without threshold and postremediation radon level with a lognormal distribution with GM equal to the square root of the preremediation radon level and GSD of 1.3. The four strategies are (1) recommended strategy based on decision analysis, (2) long-term measurements on all houses, (3) short-term measurements on all houses, adjusted for bias, (4) unadjusted short-term measurements on all houses, uncorrected.

Effect of remediation. We have assumed that remediation reduces a home radon level to 2 pCi/L. This cannot be accurate for several reasons. First, the postremediation radon level must, in reality, vary among houses. In the context of our linear dose-response model, we can account for variation by considering the assumed postremediation level as an expected radon level, averaging over houses. Second, the assumed reduction level of 2 pCi/L is a rough estimate from sparse data on remediation effects. Raising or lowering this postremediation level would correspondingly raise or lower the recommended action level R_{action} and raise or lower the estimated costs per life saved. Third, the postremediation level must certainly, in reality, depend on the initial radon level in a more complex way than simply $\mathrm{E}(\text{postremediation level}|R) = \min(R, R_{remed})$. In particular, we would expect that, for some houses with initially low radon levels (below 2 pCi/L), remediation might still have an effect. Unfortunately, available data on remediation effectiveness have been collected only for houses with fairly high preremediation levels; see Henschel, 1993, for examples.

For a sensitivity analysis, we consider a model in which the postremediation radon level is lognormally distributed with GM equal to the square root of the preremediation radon level (in pCi/L) and GSD of 1.3, further constrained to not exceed the preremediation level. This rule is arbitrary, of course, but it behaves reasonably in that postremediation radon levels are variable and are sometimes above 2 pCi/L for houses originally above 4 pCi/L. Under this model, high-radon houses are typically not remediated all the way down to 2 pCi/L, so it is not surprising that the effects of the measurement–remediation strategy are less, with reductions of 12% and 17% of estimated total lives saved for $R_{action} = 4$ and 8 pCi/L, respectively (see columns (c) of Table 5).

Additional modeling and decision issues. We have made several simplifying assumptions and choices regarding what parameters to calculate. These include the following:

1. Examining benefits in terms of "lives saved" rather than, say, "quality-adjusted life-years saved."
2. Ignoring the influence of age and latency on personal risk: it takes several to many years for lung cancer to develop and to kill, once it has been initiated, so there is little benefit of remediation for, say, 70-year-old persons. If a cancer has already been initiated then remediation is too late, whereas if they don't yet have cancer then they are likely to die of another cause before a cancer can kill them.
3. Ignoring possible interactions of radon exposure and age (e.g., children may have a different dose-response from adults).
4. Assuming risk is a function of cumulative thirty-year exposure: if risk per dose is highly nonlinear then details of the temporal variation in radon exposure become important (so, for example, the effect of people moving from home to home must be considered).
5. Ignoring the distinction between remediating to reduce personal risk and remediating a house that is to be sold (thus reducing the risk to future occupants).
6. Implicitly assigning zero cost to the hassle and stress of performing radon testing and remediation (a simplification that could be handled by adjusting the associated dollar costs).
7. Ignoring the pattern of residential mobility: if people currently living in high-radon homes remediate their houses, the majority of the resulting health benefits will accrue to future occupants of their homes (Warner, Courant and Mendez, 1995; Warner, Mendez and Courant, 1996).

All of these issues, and more, could in principle be addressed by adding additional parameters to the overall model of risks and values. We chose instead to keep the model relatively simple, since our main goals are to illustrate how the hierarchical radon model can feed into a cost-benefit analysis and to begin to bridge the gap between modeling of hazards and recommendations of actions; for both of these goals our conceptually straightforward model seemed appropriate. In particular, we think the EPA might be persuaded to make geographically specific recommendations and possibly even to make recommendations that vary for smokers and nonsmokers, but is very unlikely to make recommendations that vary by household size, age and so on, in part because more complicated recommendations might lead to considerable confusion.

7. DISCUSSION

We have used a Bayesian hierarchical model to analyze radon data in the United States, thereby generating estimated distributional information, and uncertainties, for different types of homes in every state of the conterminous United States. We used these results, along with estimates of radon risk taken from epidemiological data, to construct a formalism by which monitoring and remediation programs can be evaluated, allowing for individual variation in risk tolerance. To illustrate the use of this formalism, we examined the implications of a policy derived from the current EPA recommendation that sets 4 pCi/L as a remediation level, but that takes account of the wide variation in radon levels among counties. This sort of analysis can in principle be used by individuals trying to decide what actions to take but, more importantly, can be used by policy-makers to decide what actions to recommend or legislate.

As for the results themselves, under the assumptions used in this paper, radon is indeed a major cause of lung cancer in the United States, associated with thousands of extra lung cancers per year, and yet, we recommend monitoring only for 26% of the population (or less, if separate action levels are to be used for smokers versus nonsmokers), and remediation is recommended only for homes in the highest few percent of all homes in the United States. Our baseline recommended strategy (based on $D_d = 0.21$, equivalent to a marginal cost per life saved of \$210,000), would save only approximately 3,000 lives per year out of the estimated 15,000 radon-related deaths per year at an average cost per life saved of \$140,000. The problem is that because of the lognormality of the radon distribution, most of the total exposure (and thus, most of the expected radon deaths) is in people exposed at low levels of radon that cannot be substantially reduced by remediation (see Figure 7). That is unfortunate from the standpoint of cancer prevention but fortunate from the standpoint of our analysis since it renders our recommendations relatively insensitive to the dose-response at very low concentrations. However, if cancer risk is a strongly nonlinear function of radon concentration for concentrations in the range of 2–10 pCi/L, then both the details of the dose-response and the effects of remediation for low-radon homes are crucial unknown quantities in the decision. Unfortunately, we see lit-

tle hope for clarification of the dose-response issue for many years to come.

7.1 Policy Implications

As discussed in this paper, smokers are thought to be at much higher risk of radon-induced lung cancer than are nonsmokers. This makes radon a peculiar issue from the standpoint of public policy, as noted by Ford et al. (1998). Under the assumptions made in this paper, a large majority of remediations *should* be performed by smokers, but smokers might be willing to accept more risk for lung cancer than are never-smokers. As Nazaroff and Teichman (1990) comment in an article that touches on many issues of radon risk reduction, "it seems unlikely that most smokers would make the necessary investment to reduce the radon-related risk of lung cancer when the dominant cause of their risk is smoking."

The results presented above incorporate uncertainties in the county radon distributions and explicitly allow for estimation using different assumptions about risk tolerance. As illustrated in the discussion of sensitivity analysis, it is also possible to tinker with the dose-response function and the assumptions about remediation effectiveness. An optimal decision strategy, within the framework of the model, can be determined for any choice of these parameters, but any such strategy is optimal only in the simplified world of the model. Reality differs from the model in many ways: not all people will act rationally or follow the recommendations of the model; it may be politically difficult to call for radon testing in some areas and not others, since doing so may lower property values; similarly, it may be difficult to call for different action levels for smokers and nonsmokers, though in some sense it clearly makes sense to do so; people are impatient and may vastly prefer short-term tests to long-term ones and so on. In the policy world, psychological, political and economic considerations can be at least as important as the scientific and statistical issues considered in this paper. Moreover, even some scientific issues (most notably, uncertainty in the dose-response relation) are not fully addressed in our results.

However, this is not to say that our scientific and statistical results are useless. To the contrary, some conclusions are so clear that we think that policy can and should be changed to reflect them. Even considering possible nonlognormality within counties and variation in risk tolerance, there is no plausible scenario in which it makes sense to monitor every house in the country with a short-term measurement. The fact that high-radon homes (i.e., over 4 pCi/L) have been found in every area of the country, which the EPA states when recommending universal testing, is true but irrelevant. Someone living in a nonbasement home in Louisiana surely has many risk-reduction options that are vastly more efficient uses of money and time than is performing a radon test (e.g., buy a smoke detector, get the car's brakes checked, visit a doctor, etc.). This is true even if we use the EPA's recommended remediation level of 4 pCi/L. As we have seen, that action level itself is not unreasonable, but it does not justify monitoring every home. Of course, we say this with the luxury of having a great deal more information on the geographical distribution of radon than was available when the EPA's recommendations were first promulgated.

7.2 Generalizations to Other Decision Problems

In the Bayesian approach to decision analysis, decision options are evaluated in terms of their expected outcomes, averaging over a probability distribution that is assigned jointly to all unknown quantities. The probability distribution is typically obtained by elicitation from experts, literature review, and sometimes data analysis (in which case it is identified as a posterior rather than a prior distribution). However, it is not yet common for decision analyses to use the sorts of hierarchical models that are becoming standard in Bayesian statistics (see, e.g, Carlin and Louis, 1996; Gelman et al., 1995), as we have done in the present paper. Indeed, we are unaware of any other case in which spatially varying recommendations have been made based on the output of a hierarchical model, with correct incorporation of spatially varying statistical uncertainties.

A nonhierarchical model that has geographic variation would allow spatially varying recommendations, but, given the (inevitable) existence of spatial variation unexplained by the model, would yield less accurate predictions and thus yield decision recommendations that were not as well calibrated locally. The hierarchical model, in contrast, allows parameter estimates and uncertainties to vary by area, so that location-specific recommendations can be made, and the influence of recommended actions within local areas can be assessed.

More generally, we suspect that hierarchical modeling can be combined with decision analysis in a wide variety of problems, which we hope will make the data analysis more useful and the decision-making more individually focused. We also anticipate more sophisticated methods for computation (since, in general, the hierarchical posterior distributions that are input to these decision

analyses will be summarized by simulation) and graphical display of the varying decision recommendations, continuing on the work developed in this case study.

ACKNOWLEDGMENTS

We thank the Editors and reviewers for helpful comments, the NSF for Grants DMS-94-04305 and SBR-97-08424 and Young Investigator Award DMS-97-96129 and the Department of Energy for grant DE-AC03-76SF00098.

REFERENCES

ALEXANDER, B., RODMAN, N., WHITE, S. B. and PHILLIPS, J. (1994). Areas of the United States with elevated screening levels of Rn222. *Health Physics* **66** 50–54.

BOGEN, K. T. (1997). Do U.S. county data disprove linear no-threshold predictions of lung cancer risk for residential radon? A preliminary assessment of biological plausibility. *Human and Ecological Risk Assessment* **3** 157–186.

BOSCARDIN, W. J., PRICE, P. N. and GELMAN, A. (1996). A spatial random effects model for indoor radon concentrations in Washington State. Technical report, Energy and Environment Div., Lawrence Berkeley National Lab.

CARLIN, B. P. and LOUIS, T. (1996). *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman and Hall, London.

COHEN, B. L. (1995). Test of the linear-no threshold theory of radiation carcinogenesis for inhaled radon. *Health Physics* **68** 157–174.

COLE, L. (1993). *Element of Risk: The Politics of Radon*. Oxford Univ. Press.

DAKINS, M. E., TOLL, J. E., SMALL, M. J. and BRAND, K. P. (1996). Risk-based environmental remediation: Bayesian Monte Carlo analysis and the expected value of sample information. *Risk Analysis* **16** 67–79.

EDDY, D. (1989). Screening for breast cancer. *Annals of Internal Medicine* **111** 389–399.

EDDY, D. (1990). Screening for cervical cancer. *Annals of Internal Medicine* **113** 214–226.

ENGLEHARDT, J. D. and PENG, C. (1996). A Bayesian benefit-risk model applied to the South Florida Building Code. *Risk Analysis* **16** 81–91.

EVANS, J. S., HAWKINS, N. C. and GRAHAM, J. D. (1988). The value of monitoring for radon in the home: a decision analysis. *JAPCA* **38** 1380–1385.

FORD, E. S., KELLY, A. E., TEUTSCH, S. M., THACKER, S. B. and GARBE, P. L. (1998). Radon and lung cancer: a cost-effectiveness analysis. *American Journal of Public Health*. To appear.

GEIGER, C. and BARNES, K. B. (1994). Indoor radon hazard: a geographical assessment and case study. *Applied Geology* **14** 350–371.

GELMAN, A., CARLIN, J. B., STERN, H. S. and RUBIN, D. B. (1995). *Bayesian Data Analysis*. Chapman and Hall, London.

HENSCHEL, D. B. (1993). Radon reduction techniques for existing detached houses. U.S. Environmental Protection Agency, EPA/625/R-93/011, October.

HENSCHEL, D. B. and SCOTT, A. G. (1987). Testing of indoor radon reduction techniques in eastern Pennsylvania: an update. In *Proceedings of the Second APCA Indoor Radon II, International Specialty Conference* 146–159. Air Pollution Control Association, Pittsburgh, PA.

HOBBS, W. E. and MAEDA, L. Y. (1996). Identification and assessment of a small, geologically localized radon hot spot. *Environment International* **22** S809–817.

KLOTZ, J. B., SCHOENBERG, J. B. and WILCOX, H. B. (1993). Relationship among short- and long-term radon measurements within dwellings: influence of radon concentrations. *Health Physics* **65** 367–374.

LAGARDE, F., PERSHAGEN, G., AKERBLOM, G., AXELSON, O., BAVERSTAM, U., DAMBER, L., ENFLO, A., SVARTENGREN, M. and SWEDJEMARK, G. A. (1997). Residential radon and lung cancer in Sweden: risk analysis accounting for random errors in the exposure assessment. *Health Physics* **72** 269–276.

LUBIN, J. H. and BOICE, J. D. (1997). Lung cancer risk from residential radon: meta-analysis of eight epidemiologic studies. *J. National Cancer Institute* **89** 49–57.

LUBIN, J. H. and STEINDORF, K. (1995). Cigarette use and the estimation of lung cancer attributable to radon in the United States. *Radiation Research* **141** 79–85.

MARCINOWSKI, F., LUCAS, R. M. and YEAGER, W. M. (1994). National and regional distributions of airborne radon concentrations in U.S. homes. *Health Physics* **66** 699–706.

MILES, J. and BALL, K. (1996). Mapping radon-prone areas using house radon data and geological boundaries. *Environment International* **22** S779–S782.

MOSE, D. G. and MUSHRUSH, G. W. (1997). Variable spacial and seasonal hazards of airborne radon. *Atmospheric Environment* **21** 3523–3530.

NATIONAL RESEARCH COUNCIL, COMMITTEE ON THE BIOLOGICAL EFFECTS OF IONIZING RADIATION (BEIR VI) (1998). *Health Risks of Radon and Other Internally Deposited Alpha-emitters*. National Academy Press, Washington, D.C.

NATIONAL RESEARCH COUNCIL, PANEL ON DOSIMETRIC ASSUMPTIONS AFFECTING THE APPLICATION OF RADON RISK ESTIMATES (1991). *Comparative Dosimetry of Radon in Mines and Homes*. National Academy Press, Washington, D.C.

NAZAROFF, W. W. and NERO, A. V., eds. (1988). *Radon and Its Decay Products in Indoor Air*. Wiley, New York.

NAZAROFF, W. W. and TEICHMAN, K. (1990). Indoor radon-exploring United States federal policy for controlling human exposures. *Environmental Science and Technology* **24** 774–782.

NERO, A. V., GADGIL, A. J., NAZAROFF, W. W. and REVZAN, K. L. (1990). Indoor radon and decay products: concentrations, causes, and control strategies. Technical Report LBL-27798, Lawrence Berkeley Lab.

NERO, A. V., SCHWEHR, M. B., NAZAROFF, W. W. and REVZAN, K. L. (1986). Distribution of airborne Radon-222 concentrations in U.S. homes. *Science* **234** 992–997.

OWENS, D. K., HARRIS, R. A., SCOTT, P. M. and NEASE, R. F. (1995). Screening surgeons for HIV infection. *Annals of Internal Medicine* **122** 641–652.

PINEL, J., FEARN, T., DARBY, S. C. and MILES, J. C. H. (1995). Seasonal correction factors for indoor radon measurements in the United Kingdom. *Radiation Protection Dosimetry* **58** 127–132.

PRICE, P. N. (1997). Predictions and maps of county mean indoor radon concentrations in the Mid-Atlantic States. *Health Physics* **72** 893–906.

PRICE, P. N. and NERO, A. V. (1996). Joint analysis of long- and short-term radon monitoring data from the Northern U.S. *Environment International* **22** S699–S714.

PRICE, P. N., NERO, A. V. and GELMAN, A. (1996). Bayesian prediction of mean indoor radon concentrations for Minnesota counties. *Health Physics* **71** 922–936.

PRILL, R. J., FISK, W. J. and TURK, B. H. (1990). Evaluation of

radon mitigation systems in 14 houses over a 2-year period. *Journal of the Air and Waste Management Association* **40** 740–746.

REVZAN, K. L., PRICE, P. N., NERO, A. V., GUNDERSEN, L. C. S. and SCHUMANN, R. R. (1998). Bayesian analysis of the relationship between indoor radon concentrations and predictive variables in U.S. houses. Technical Report, Energy and Environment Division, Lawrence Berkeley National Lab.

RHOADS, S. E. (1980). *Valuing Life: Public Policy Dilemmas*. Westview Press, Boulder, CO.

STIDLEY, C. A. and SAMET, J. M. (1993). A review of ecologic studies of lung cancer and indoor radon. *Health Physics* **65** 234–251.

TURK, B. H., HARRISON, J. and SEXTRO, R. G. (1991). Performance of radon control systems. *Energy and Buildings* **17** 157–175.

UPFAL, M., DIVINE, G. and SIEMIATYCHI, J. (1995). Design issues in studies of radon and lung cancer: implications of the joint effect of smoking and radon. *Environ. Health Perspect.* **103** 58–63.

U.S. CONGRESS, OFFICE OF TECHNOLOGY ASSESSMENT. Researching Health Risks. U.S. Government Printing Office, OTA-BBS-570, Washington, D.C.

VISCUSI, W. K. (1992). *Fatal Tradeoffs: Public and Private Responsibilities for Risk*. Oxford Univ. Press.

WARNER, K. E., COURANT, P. N. and MENDEZ, D. (1995). Effects of residential mobility on individual versus population risk of radon-related lung cancer. *Environmental Health Perspectives* **103** 1144–1149.

WARNER, K. E., MENDEZ, D. and COURANT, P. N. (1996). Toward a more realistic appraisal of the lung cancer risk from radon: the effects of residential mobility. *American Journal of Public Health* **86** 1222–1227.

WATSON, S. R. and BUEDE, D. M. (1987). *Decision Synthesis: The Principles and Practice of Decision Analysis*. Cambridge Univ. Press.

WHITE, S. B., CLAYTON, G. A., ALEXANDER, B. V. and CLIFFORD, M. A. (1990). A statistical analysis: predicting annual Rn-222 concentration from 2-day screening tests. In *Proceedings of the 1990 Symposium on Radon and Radon Reduction Technology* **1** 3-117–3-118. U.S. EPA, Washington, D.C.

WHITE, S. B., BERGSTEN, J. W., ALEXANDER, B. V., RODMAN, N. F. and PHILLIPS, J. L. (1992). Indoor Rn222 concentrations in a probability sample of 43000 houses across 30 states. *Health Physics* **62** 41–50.

Comment

Bradley P. Carlin

First, congratulations to the authors on a fine paper, which shows quite clearly how formal Bayesian decision-theoretic tools may be combined with modern hierarchical modeling techniques to produce clear and sensible guidelines in an important environmental health problem setting. Of course, Bayesians have long argued that their techniques offer significant advantages over the traditional, more informal analytic procedures often used by decision makers, but only with the advent of modern Markov chain Monte Carlo (MCMC) computing methods in the last decade or so can these benefits be fully realized. As seen in the present paper, the Bayesian engine does not obviate the need for a variety of (potentially highly influential) assumptions in the analysis, but it does provide a framework in which these assumptions can be carefully structured, and their impact assessed.

Before commenting on specific aspects of the authors' work, it is worth mentioning a possible confusion in the use of the term "Monte Carlo analysis" by the risk assessment and applied Bayesian camps. As already mentioned, to the latter group this typically refers simply to the integration methods used to evaluate the "denominator integral" in Bayes' Rule; that is,

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\theta})\,p(\boldsymbol{\theta})}{\int p(\mathbf{y}|\boldsymbol{\theta})\,p(\boldsymbol{\theta})\,d\boldsymbol{\theta}}, \tag{1}$$

where $\mathbf{y}$ denotes the observed data and $\boldsymbol{\theta}$ the vector of unknown parameters. However, to risk assessors, "Monte Carlo analysis" is an approach by which a sample of potential risk or exposure values is obtained by first specifying distributions relating the various observed and unobserved quantities in the model and then simulating values from the resulting hierarchical risk model. While my reading of this literature is admittedly only cursory, this seems to be the approach taken in several risk assessment textbooks (e.g., Vose, 1996), and one that has been recently codified by an EPA panel assembled to "promote scientific consensus on risk assessment issues and to ensure that this consensus is incorporated into appropriate risk assessment guidance" (Environmental Protection Agency, 1997). But to Bayesians, this approach is tantamount to "sampling from the prior"; the Monte Carlo method is being used only to simulate values from assumed

Bradley P. Carlin is Associate Professor, Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, Minnesota 55455 (*e-mail: brad@muskie.biostat.umn.edu*).

distributions, not to assist in the formal prior-to-posterior updating of (1) above. To be fair, the usual approach does encourage using observed data when determining distributions for unknown quantities whenever possible, and the literature is beginning to distinguish the variability in unknown quantities (which is sometimes called "uncertainty") from the variability of observed quantities given the unknowns (which instead is called "variability"; see, e.g., Rai and Krewski, 1998). Still, with a few notable exceptions (Taylor, Evans and McKone, 1993; Brand and Small, 1995; Dakins, Toll, Small and Brand, 1996), the risk assessment literature seems in need of more formal Bayesian thinking, for which the present work (and the earlier work of Wolpert, Steinberg and Reckhow, 1993) may well serve as a blueprint.

One way in which the risk assessment literature is ahead of that in statistics is in its willingness to discuss the value of human life on a dollar (or some other meaningful quantitative) scale. In clinical trials, for instance, real advances in decision theoretic solutions to the interim monitoring and final analysis problems have been stymied by the unwillingness of most statisticians, epidemiologists and clinicians to even contemplate such a mapping (though a few brave first attempts have been made by Berry and Ho, 1988; Stangl, 1995). An important feature of the present paper is the authors' description of how already established government guidelines for what constitutes a radon exposure level worthy of remediation *implicitly* determines dollar values per microlife (Section 4.2). Clearly such a linear scale is not appropriate when we move far from the origin (no reasonable person would surrender one million of his own microlives for *any* dollar amount), but discussions of this sort may well have beneficial impact in risk assessment strategies far beyond environmental settings, if in no other way but informing decision-makers as to what implicit values their recommendations are placing on fractions of lives.

Turning then to specific comments on the authors' approach, given the power of modern MCMC techniques I was surprised that the model components considered in Section 3 were essentially confined to normal distributions. The model apparently treats the variance parameters τ^2 and σ^2 (as well as a variety of tuning parameters in Section 4) as constants, instead of more plausibly assuming distributions for them. Indeed, some of the modeling is not even being shown: (2) is written as a prior (or a "predictive" in the authors' nomenclature), but in fact it must be the result of a preliminary prior-to-posterior calculation, combining some prior on the regression parameters $\boldsymbol{\beta}$ with some preliminary data $\mathbf{y}^*$ on typical radon concentrations in U.S. homes. What is this preliminary data and model? Do its residuals suggest any evidence of lingering spatial correlation? Also, the two-stage implementation of the preliminary (Section 3.1) and house-specific (Section 3.2) models is odd, since it forfeits the usual Bayesian advantage of a single unifying model that enables all sources of variability and uncertainty to correctly propagate throughout its levels.

As the authors mention, the paper's main focus is on the decision analysis in Section 4. Here there are any number of assumptions with which one could quibble (the flat \$2000 to remediate any home regardless of location, the 70-year life expectancy for every occupant, etc.); one could either place distributions on these quantities as well, or simply undertake a variety of sensitivity analyses (as the authors describe in some detail in Section 6). While I don't wish to nitpick further here, I did find the approach for "discounting the value of a life," described near the beginning of Section 4.1, to be somewhat confusing. At first blush, if D_d is the amount we are willing to pay to save one microlife now, then since lives saved 20 years in the future are worth less, it seems the revised D_d should be *decreased* (not increased) by a factor of 1.05^{20}. However, recall that the paper does not really specify D_d from first principles, but rather "backs it out" by viewing the \$2000 remediation cost as fixed. Thus if the value of the lives saved decreases, our cost per life saved must go up. Yet even here, it seems that the appropriate increased cost must be backed out from (5) and (6) as well, discounting each future year's risk separately in the thirty-year decision period rather than applying a single inflation factor to D_d.

Of course, the actual dollar amount any given person would spend per microlife saved is probably more a function of their own financial resources and aversion to risk than any governmentally recommended remediation levels. I am personally acquainted with a suburban couple with three children who, after reading an early report on the alleged dangers of living near high-voltage electrical lines like the ones near their home, immediately sold the place and moved. Because they did this at just the time when popular concern over this potential risk was at its zenith, their total financial loss in the transaction (including moving expenses and remodeling their new home) was in the neighborhood of \$80–100,000: in the light of more recent data on the subject, a colossal amount spent per expected microlife saved. In the language of (6), for this couple D_d (hence D_r) was essentially infinity for this perceived risk, and thus $R_{\text{action}} \approx R_{\text{remed}}$.

Looking again at Figure 3, this means the couple would not consider taking a new risk measurement, and apparently the decision minimizing their expected loss was to remediate (i.e., move) immediately. While this is obviously an extreme example, I would venture to guess that many middle- to upper-class persons would adopt a similar strategy for radon remediation, especially given the relatively small cost involved here (only $2000, though this does not include the headaches of getting the work done and then maintaining the fan system once installed).

Finally, in Section 4.3.1, the authors correctly recognize that the "inner decision" of whether to remediate or not given the future observation y must be considered first; only then can we decide the broader question of whether to take this extra measurement at all (versus simply deciding based on the prior) by integrating over the possible values of y we might see. This approach is a special case of the general Bayesian decision-theoretic approach to sequential analysis problems, *backward induction* (see, e.g., DeGroot, 1970, Chapter 12), in which multistage decision problems are decided by "working backward" through the potential future observation stages, alternately minimizing expected loss and integrating over the as-yet-unseen data values. It is easy to show (and intuitively clear) that it is always better to continue sampling if there is no cost associated with obtaining these new samples. Still, backward induction is seldom used in practice, due to the explosion in analytical and bookkeeping complexity as stages are added to the model. Recently, however, Carlin, Kadane and Gelfand (1998) have proposed a "forward sampling" algorithm that substantially eases the analytic and computational burden and can be used to identify the best member of a plausible class of strategies when backward induction is infeasible. Such an approach has obvious appeal in clinical trial monitoring (where a monitoring board wishes to check the trial's progress at various intervals and stop the trial as soon as the best decision is clear) and might also be useful in environmental settings where a series of measurements is anticipated, with remediation an available option at each interval.

In summary, this paper makes important methodological contributions to the field of environmental decision analysis and similarly important contributions to the substantive problem of radon remediation (indeed, homeowners would do well to consult this paper and its Web site rather than simply rely on any of the "one-size-fits-all" government guidelines). I look forward to future developments in fully Bayesian decision analysis and its further incorporation into the practice of risk assessment.

ACKNOWLEDGMENTS

The author is grateful to Ms. Li Zhu and Mr. Cong Han for helpful discussions during the preparation of this comment. The author's work was supported in part by National Institute of Allergy and Infectious Diseases (NIAID) Grant R01-AI41966, and by National Institute of Environmental Health Sciences (NIEHS) Grant 1-R01-ES07750.

Comment: In Praise of Decision Analysis in Environmental Health

Carl V. Phillips

If we knew all the right decisions ex ante, then making public health policy (or advising individuals making decisions) would be relatively easy. If we had no information to inform a certain decision, then any new test or study could be useful and it would be difficult to predict what new study would be the most useful. For most public health questions, of course, our knowledge lies somewhere between these extremes: we do not know enough to make a definitive decision or recommendation, but we know (or would know if we looked carefully) what further studies or tests would help make the decision. Yet somehow, most of the public health literature fails to recognize the implications of this.

Carl V. Phillips, Division of Environmental and Occupational Health and Center for Environmental and Health Policy, University of Minnesota School of Public Health, Minneapolis, Minnesota 55455.

The failures include implicit assumptions that no further data is available (making policy recommendations prematurely), the next study will inevitably clear up all uncertainty (ignoring the need to figure out what the next study should focus on), and further data gathering cannot be targeted to subpopulations based on existing data. (Sadly, to the extent that targeted data collection is recommended, it is often focused on the subpopulation that is judged to be most "at risk." This is often the population that is least in need of further study, since existing data is sufficient to warrant intervention with a high degree of confidence.)

Lin, Gelman, Price and Krantz (1999) do a great service by reminding us, using a wonderfully clear example, how ex ante knowledge helps us target further data gathering and thereby helps assess when the benefits exceed the costs. They observe that our priors about the risk from household radon exposure vary by geography and household characteristics, and that we can use this convenient data to dramatically improve the efficiency of decisions about testing and remediation. The addition of previous measurements further improves the precision of the geographic estimates, and thus the value of the recommendations. (Decision analysis could still be used without this addition, just using the predicted mean county radon concentrations and household characteristics, since they vary enough to produce different recommendations. It might be, however, that the precision would be so lowered that the benefits of decision analysis would be diluted.)

Social scientists and policy analysts have long discussed how to use existing information to decide about gathering further information or to make different recommendations for different individuals (e.g., Stokey and Zeckhauser, 1978). But the critical importance of quantifying underlying probability distributions and sources of uncertainty—indeed, the understanding that decision-makers can even use such information—has been overlooked by other health researchers. This has created the sad situation where policy analysts are waiting for probability data for use in decision analysis, while health researchers are providing only the point estimates that they apparently assume to be preferable and useful.

An issue like radon testing provides a useful way to introduce the social value of data gathering into public health without the medical arena's limits of allowable practice. Medical testing is functionally equivalent to other areas of public health, such as radon testing or finding out how safe a highway is by building it and watching what happens. But the culture of medicine and the legal climate complicate things. While allowing the gathering of patient data for the social good, the current culture makes it difficult to perform tests with expected social benefits but a net expected cost to the particular patient, or to withhold tests that have negative expected value but have any chance of improving a diagnosis. Indeed, it is sometimes difficult to even discuss making more efficient decisions in the medical arena. An environmental health issue like radon allows socially optimal recommendations because most people are amenable to persuasion about the right choice, given their underlying lack of knowledge, the relatively low individual costs and risks and the noninvasiveness of most actions. At the same time, the social costs and risks are fairly high, and it is worth the effort to try to minimize them.

One major policy advantage of the situation described by Lin et al. (one which should probably be given more attention in the research literature and policy process) is that it allows individuals with different tastes for risk to take different actions. Unlike public health decisions that must be made by a central authority for everyone (cases ranging from effluent regulation to airplane safety features), the decision about radon parallels the decisions about medical care and consumption. If someone is more willing to risk disease or less willing to spend money to avoid it (or does not believe that the risk is actually real), then he has the option of making a different decision than the official recommendation. This would be particularly reasonable if, as suggested by Lin et al., EPA made recommendations that ignored household composition. Single nonsmoking assistant professors could rationally choose to ignore the radon risk in their homes.

An extension of these principles in a different direction (and into more controversial areas) is to assess what population-level research would be most useful given our current data and priors (Phillips and Maldonado, 1999). Epidemiologic studies, along with most quantitative health research, tend to conclude by calling for more research, but very seldom assess exactly what the further research should do. Further research can be used to eliminate some of the measurement error, simply assess the level of measurement error, eliminate confounders, measure confounders or just increase the sample size. Within all these choices, there are continuous ranges of choices along multiple dimensions. Yet the decision analysis principles are still the same as those in Lin et al. By fully assessing what we know and what more we would be likely to know following future research, we can better determine when to act, when to walk away or what more we want to know.

Given the clear policy prescriptions in an analysis like Lin et al., it is crucial that authors are careful to identify what policy question they are answering and get certain key social-science-based parameters right. It is here that I take some issue with the article. (It should be emphasized that this criticism is an indication that Lin et al. are victims of their own success. Criticism at this level would not be worthwhile if the article did not have such great potential for guiding policy.) The key number relating to human preference, the value of a life saved, is too low. There is, of course, no clear number for the value of a life saved, and willingness to pay for a probabilistic saved life (by a consumer or regulatory agency) varies wildly. Nonetheless, the implicit values of $100,000 to $500,000 per life saved preferred by Lin et al. are low by almost an order of magnitude compared to the typical discussion in the economics-of-health or economics-of-regulation literature. This has a substantial impact on the recommended policy. The choice of discount rates also affects the decision through the same pathway and is always controversial in policy discussions. The use of 5% for discounting lives saved will be criticized by some as being high (though others have used even higher values).

Lin et al. allude to the difference between public and private health impacts of remediation, particularly in their acknowledgment of simplifying assumptions. But they understate the importance of this issue, which is probably the most important challenge to actual decision making, dwarfing the practical implications of their other assumptions or sensitivity analyses. Most houses change residents many times over two or three decades. Israeli and Nelson (1992) report that the mean total residence time for U.S. households is less than five years, and for people who own their home it is about eleven years. The medians are considerably lower. It is also likely that tenure is lower among those who would be most inclined to remediate (the relatively young and affluent). If the radon exposure were perfectly capitalized into the value of a house, then the private and public decision would be the same (setting aside the different impacts of radon based on the number, age, and smoking habits of the residents). The remaining benefit from a remediation expenditure would be captured by the current owner when the house was sold (or rented) and so the private optimization decision about remediation would achieve the social optimum.

However, there is little chance that this perfect capitalization will occur. Currently, most consumers ignore radon risk when making housing choices, while a few overreact to it. Greater attention to the risk in the media or by the government would tend to increase the level of concern. But there is no reason to believe that it will get to the "correct" level of concern (or that if it does, that it will not shoot past it into widespread overreaction). In general, the vaunted invisible hand of microeconomics can only promise that prices will be set correctly when there is room for someone to make an arbitrage profit from someone else's miscues. If you underprice the wheat you are selling, I can make a profit by buying it. But if you underprice the value of protecting yourself from radon, there is no way for me to make a profit, and thus no market pressure for the price to rise to its proper level. Thus, the major decision variable for individual homeowners, perhaps more important than even the concentration of radon, is likely to be how long they plan to stay in the house, a variable which is omitted from the analysis.

Given the failure of perfect capitalization, the optimal public health result could still be achieved by requiring all property owners to take the steps recommended by the decision analysis, regardless of personal taste or plans to leave the house. However, this would be such an implementation nightmare that it is not even worth discussing. The analysis in Lin et al. does not clearly position itself as either a recommendation to the homeowner (in which case it should consider expected tenure in the house and level of capitalization from the risk) or for some national public health initiative (in which case, issues of implementation, social attitudes toward risk, and politics will probably dominate the rational assessment).

Despite addressing some of the uncertainty between households through the hierarchical model, Lin et al. cannot do much to deal with the great underlying uncertainty of how risky low-dose radon really is, especially for nonsmokers. In many cases, policy should be made based on the best-available estimates of important values. However, there are limits to this approach. On a practical level, even with an impeccably flawless decision analysis, it may be difficult to get consensus on a precise action point for a decision tree when there is huge disagreement about the value of the central parameter. Apart from this, when certain actions have irreversible costs (such as $1500 to install a radon remediation system), it may pay to wait for more information, in case the new information might suggest a different optimal action. (In economics, this concept is known as *option value*.) This, in turn, creates another layer of optimization decision based on our priors about what new information will emerge and when.

Before paralysis sets in about the enormity of the task of optimizing our decisions under risk and uncertainty, we should remember the value of a high-quality analysis that covers most of the important options and sources of uncertainty. Such an analysis is much more likely to produce good decisions than is taking rhetorical refuge in the complexity of the decision and resorting to rules of thumb, political pressure or simple inertia. Lin et al. show how to conduct such an analysis, and (modulo our disagreement about economic parameters and the complication of individual versus social decisions) carry it out. Their analysis is part of an important trend in health research toward considering all costs and benefits of an action and seriously analyzing what we would learn from further data gathering. This is a huge improvement over the standard practices of either making policy recommendations as if a given study were the last word in an area or assuming that we will have all the answers soon and refusing to make a recommendation until then.

Rejoinder

Chia-yu Lin, Andrew Gelman, Phillip N. Price and David H. Krantz

1. INTRODUCTION

We thank both discussants for their comments, especially for their explorations of the connections between statistical modeling, decision analysis and public health outcomes. Our paper has two main goals: to illustrate the benefits of hierarchical modeling in probabilistic decision analysis and to determine an improved radon recommendation that would have a chance of actually influencing the government's radon policy. In this rejoinder, we respond to the specific comments of the discussants in the context of our major concerns about radon and decision analysis in general.

Phillips points out that we never stated exactly what policy question we are attempting to answer, and that it is unclear whether we are making recommendations to individuals or to policy-makers. The answer to the latter question is that we are hoping to influence government radon policy; we do not expect our recommendations to reach a substantial number of individual homeowners. We do think that the government, that is, the Environmental Protection Agency, or perhaps state health departments, could recommend use of some house- and occupant-specific information in making radon decisions, but, as discussed below, the complexity of the decisions would probably be kept very low. Of course, even though it is the government's radon policy that we are attempting to influence, the eventual costs and benefits (and decisions) would still be up to individual homeowners (except for a few government-owned buildings such as schools and military base housing, some of which have already been monitored and remediated).

As to our not explicitly stating the policy question that we were trying to answer, that's certainly a valid point. Rather than claim that what we did is "right," we will explain why we did the analysis the way we did, since we think the same issues will apply to many other decision analyses.

2. RADON POLICY

We did not set out to answer the open-ended question of what the U.S. government's radon policy should be, which is such a complicated question that it is hard to see how to directly address it with a decision analysis. For instance, the question of whether radon policy should be set by the federal government, state health departments or local zoning boards, is both a political question and a matter of organizational efficiency that we have no clear way to analyze. In practice, the radon policy can only be chosen from within a universe of possible policies, and we don't even know what that universe encompasses.

However, we have had several years of experience analyzing radon data with the goal of identifying high-radon homes, as part of a project sponsored by the EPA, the Department of Energy, and the U.S. Geological Survey. During this time we grew dissatisfied with several aspects of the current U.S. radon program. For instance, many people make decisions about radon remediation decisions (and even decisions about what house to buy) based on short-term monitoring in the basement, which bothers us since

we know that this protocol leads to a very large fraction of "false positive" results. Also, when we found that some large areas of the country have nearly no high-radon homes, it seemed silly to recommend monitoring everywhere. Finally, there's the risk difference between smokers and nonsmokers, which certainly seemed as though it ought to influence decisions.

2.1 Why Not Look for the Optimal Policy for Individuals?

Once we begin to consider the factors that can affect people's decisions, it seems appropriate to analyze the whole problem from the perspective of individual homeowners, but this approach quickly becomes extremely complicated. To fully model risk, one should consider, in addition to the predicted indoor radon concentration, at least the number of occupants, and the sex, age and smoking status of each occupant (or in the case of young nonsmokers, the probability that they will become smokers). Of course, all of this is in addition to the very large uncertainty in the dose-response relation, particularly at low doses.

Also, in contrast to our simplified model, both the costs of remediation and the postremediation radon concentration are variable (and unknown). As far as costs are concerned, some unknown fraction of the remediation cost is, or ought to be, recovered upon sale of the home, which takes place after a variable and uncertain time period, with the recovered cost dependent on both the risk tolerance of the new buyers and the makeup of their household.

In short, attempting to determine the best course of action for a particular person or home is a mess. We could, of course, create models and distributions for all of the factors listed above and generate individual recommendations, but to what end? There's no chance that any sizable fraction of homeowners would actually do the work needed to determine what we recommend they should do, and we also can't picture radon policy-makers sifting through the resulting reams of analysis in order to try to formulate a new radon policy. And perhaps they shouldn't; would the benefits of having a better-targeted, but much more complicated, radon policy outweigh the ill effects of added confusion and complexity?

2.2 What Do Governments Think a Radon Policy Should Look Like?

Most governments that have official radon policies (this includes most of the "developed" nations) have a single recommended action level, or sometimes two or three separate target levels, for existing, rebuilt and new buildings (Cole, 1993). In the United States, EPA and state health department officials have told us that when people ask them for radon advice, they don't want to have to think about a lot of different issues; they just want to know what a "safe" radon level is. Whether or not the policymakers are right about the need for simplicity, it is clear that official radon recommendations will in fact be based on quite simple monitoring and remediation criteria. So when it came to deciding what policies to analyze, we decided to restrict ourselves to fairly simple variations on the EPA's current recommendations. We make no claim that the resulting policies are the best of all possible ones; we only claim that they would be improvements to the current recommendations.

3. VALUE OF A MICROLIFE

3.1 Why We Used the Values We Used

The discussion of which simple policies should be considered leads us to one of the specific issues raised by the discussants. Phillips objects to the low "value of a microlife" that we used, and Carlin says that our derivation of it from the EPA's recommendations is confusing.

We considered three parameterizations (value of a microlife, or a radon action level, or a dollar value associated with reduction of 1 pCi/L), and it is possible to perform the analysis conditional on any of these. One reason for allowing decisions to be analyzed in terms other than "cost per life saved" was to allow more direct treatment in, say, a regulatory framework in which decisions are made based on an action level. Of course, a value for any of the three parameters determines values for the others, so there is no escaping that any decision implies a marginal cost per life, but, as a practical matter, radon decision-makers might prefer to work with one of the other parameterizations.

Our analytical results were worked out with the parameter values unspecified, but when we plugged in example values we chose them for a reason. As a practical matter, retaining the action level is important. There is probably no stone tablet in Washington that reads "Thou shalt use a 4 pCi/L action level," but there might as well be: the EPA has already faced significant heat from nonbelievers who think that breathing radioactive gas is good for you, or at worst harmless, and there is no chance that the action level will be decreased. On the other hand, there is also little chance that the EPA will completely abandon its long-standing 4 pCi/L threshold in favor of a higher threshold, particularly in view of

the fact that under conventional dose-response models it is not a very protective standard; as we show in the paper, the implied marginal cost per life saved is only around $200,000, which, as Phillips notes, is very low compared to the values people usually use.

However, interpreting this implied marginal cost is a bit tricky. The radon decision is about paying now to reduce the risk of lung cancer death between 5 and 35 years in the future (under our assumption that remediation is effective for 30 years and assuming 5 years between the cancer initiation event and death). On average, ignoring the age distribution of the population for the moment, performing remediation now can be thought of as saving a fraction of a statistical life at the middle of that period, or 20 years. This delay can also be interpreted in terms of "discounting" of the future; in our paper, we suggested that a discount rate of 5% leads to a "net present value" of a life of $1.05^{20} D_d = 2.7 D_d$, but that is actually slightly wrong. We should have used $\int_5^{35} 1.05^t D_d \, dt = 2.9 D_d$. However, we prefer to avoid this discounting formulation and rather simply recognize that the lives saved by radon remediation will be on the order of 20 years in the future. Another way of analyzing the decision, consistent with the medical decision analysis literature, is to look at years of life saved. If lung cancer deaths from radon are approximately uniformly distributed across the population of smokers, this gives on the order of 25 years of life saved, so that, for example, $200,000 per life corresponds to about $8,000 per undiscounted year of life.

Getting back to Phillips' comment that these dollar amounts are an order of magnitude lower than typical values in the economics-of-health or economics-of-regulation literature, we note that in addition to the discounting issue discussed above, in many health and regulatory decisions there is a potential for prescribing the action to be taken: requiring (by law) that insurance companies pay for a particular procedure, or that companies reduce their emissions below a certain level. In contrast, with a few exceptions radon recommendations are just that—recommendations—and the fact that most people do not now remediate, or even monitor for radon, suggests that there would be little point to determining a radon policy based on a higher value per microlife with its correspondingly lower recommended action level. We hope that an improved radon policy will also meet with better compliance with the recommendations of that policy, and if so then the action level could be revisited, but for now a lower action level would probably simply not be respected.

3.2 Variation in the Value of a Microlife

Carlin suggests that the amount people are actually willing to spend is "probably more a function of their own financial resources and aversion to risk than any governmentally recommended remediation level." Actually, we think all three of these factors are important. In practice, based on ample anecdotal evidence, many people *do* take the recommended action level of 4 pCi/L into account when deciding on whether to remediate (though they do not necessarily follow the EPA's recommendations; see Evdokimoff and Ozonoff, 1992), and many radon mitigation companies guarantee that the long-term postremediation concentration will be below 4 pCi/L and will perform additional remediation if that standard is not met.

However, Carlin's point is well taken. If "risk-aversion" is measured in dollar terms, then it may be more a measure of financial resources than of psychological attitude toward risk. Consider the most extreme case of avoiding certain risks. Carlin points out that a linear scale of value per microlife is "not appropriate far from the origin since no reasonable person would surrender one million of his own microlives for *any* dollar amount." Although true, the question that is more relevant to our analyses is not how much someone would have to pay you in order for you to tolerate a given risk, but rather how much *you* would pay to avoid a given risk. There is clearly a finite answer to this latter question: you cannot pay money that you cannot raise. The theoretical "ability to pay" to save a million of your own microlives might be a couple of million dollars for a reader of this journal, up to many tens of billions of dollars for, say, Bill Gates.

4. THE ROLE OF FORMAL "DECISION ANALYSIS" IN DECISION-MAKING

Formal decision analysis requires setting up a decision–uncertainty tree, estimating the costs and probabilities associated with the potential outcomes, setting up a value–utility function for the outcomes and evaluating the tree using averaging and maximization. It is often said that the most important parts of formal decision analysis are (a) explicitly setting down the possible decisions and outcomes and (b) revealing possible incoherence in existing decision procedures.

The key difficulty of using a decision analysis to make an actual recommendation is that the inputs to the analysis may be more controversial than the outputs. For example, in our analysis, any homeowner can obtain a recommendation to remediate simply by increasing the value of a microlife past

a certain value. To put it another way, there may be as much arbitrariness in choosing the relative values of money and life as there is in setting a perfect-information remediation threshold, or even in making a measurement–remediation decision. As with the medical decision-making literature, we can use decision analysis as a tool to produce best estimates of costs in dollars and lives, which then must be balanced by policymakers and individuals.

In the radon example, the clearest benefit of the decision analysis is in allowing us to construct a spatially varying family of decision recommendations that we expect would save more lives at a lower cost than a uniform national recommendation. In addition, the formal analysis allows us to calibrate decision thresholds (in pCi/L) in terms of dollars per microlife. Both these benefits require a realistic statistical model of measurements and home radon levels.

5. DECISION ANALYSIS AND HIERARCHICAL MODELING

To see the connection between statistical models and decision analysis, we consider how decision recommendations change as the underlying statistical models become more complex.

The simplest decision analyses are uniform across the population and are expressed as, for example, Should a patient with a certain medical condition undergo the risk of a certain diagnostic test? The recommendation might then be in terms of dollars per life saved, or even simply as positive or negative expected lives saved. More sophisticated analyses allow relevant probabilities to depend on known covariates, so that the question of undergoing the diagnostic test might be conditional on the age, sex, and some assessment of the health status of the patient.

From a statistical standpoint, conditional decision recommendations correspond to interactions between treatment effect and covariates, and they can have important practical considerations. For example, optimal recommendations for cancer screening depend on age, with the particular age recommendations depending on the pattern of cancer onset (see, e.g., Eddy, 1990), and in economics, a program that has a negative effect on the population can be estimated to have a highly positive effect if targeted on the individuals with covariates that predict a highly positive interaction with treatment (Dehejia, 1998).

In response to Carlin's comments about the underlying statistical model: yes, we previously fit a fully Bayesian model to a large set of short- and long-term radon measurements, along with other information on houses in the dataset and counties in the United States. We used the posterior distribution of that analysis as the prior distribution for the analysis in this paper. For each county and house type, we used the posterior simulation draws from our previous analysis to compute a prior mean and standard deviation for the mean log radon level for those counties and house types. We then assumed that the standard deviations of the measurements within that county and house type were estimated to a high precision (and could thus be summarized by posterior point estimates?), which is not too bad an approximation given the large datasets used to construct that posterior distribution. If we were less confident that the posterior distribution was close to normal, then we would have worked with the simulation draws themselves, but in this case, we wanted the convenience of the normal approximation, which allowed some of the steps of the decision analysis to be performed analytically. Another approach would be to use the normal approximation, but then check it (or correct for it) at the end of the analysis, using importance sampling.

Finally, we believe it is important to link the concerns of statistical modeling to those of decision analysis. Sensitivity analysis is already recognized as a crucial step in any practical decision analysis. In addition, the iterative steps of modeling, fitting and model-checking are as relevant for decision analysis as for inference. In particular, in a decision problem, it makes sense to check that the decision recommendations for the model applied to the data are consistent with what would be expected under the model; that is, decision recommendations can be used as test variables in predictive checks as in Gelman, Meng and Stern (1996). In the radon example, other natural predictive checks arise from concerns expressed with the model; for example, are there pockets of high-radon homes in otherwise low-radon counties, beyond that predicted by the model? More generally, the decision analysis should guide the model-checking as well as the inference and the modeling itself.

ADDITIONAL REFERENCES

BERRY, D. A. and HO, C.-H. (1988). One-sided sequential stopping boundaries for clinical trials: a decision-theoretic approach. *Biometrics* **44** 219–227.

BRAND, K. P. and SMALL, M. J. (1995). Updating uncertainty in an integrated risk assessment: conceptual framework and methods. *Risk Analysis* **15** 719–731.

CARLIN, B. P., KADANE, J. B. and GELFAND, A. E. (1998). Approaches for optimal sequential decision analysis in clinical trials. *Biometrics* **54** 964–975.

DAKINS, M. E., TOLL, J. E., SMALL, M. J. and BRAND, K. P. (1996). Risk-based environmental remediation: Bayesian Monte Carlo analysis and the expected value of sample information. *Risk Analysis* **16** 67–79.

DEGROOT, M. H. (1970). *Optimal Statistical Decisions*. McGraw-Hill, New York.

DEHEJIA, R. (1998). Program evaluation as a decision problem. Technical report, Dept. Economics, Columbia Univ.

ENVIRONMENTAL PROTECTION AGENCY (1997). *Guiding Principles for Monte Carlo Analysis*. Technical Report EPA/630/R-97/001, Risk Assessment Forum, U.S. Environmental Protection Agency, Washington, D.C.

EVDOKIMOFF, V. and OZONOFF, D. (1992). Compliance with EPA guidelines for follow-up testing and mitigation after radon screening measurements. *Health Physics* **63** 215–217.

GELMAN, A., MENG X. L. and STERN, H. S. (1996). Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Statist. Sinica* **6** 733–807.

ISRAELI, M. and NELSON, C. B. (1992). Distribution and expected time of residence for U.S. households. *Risk Analysis* **12** 65–72.

LIN, C., GELMAN, A., PRICE, P. and KRANTZ, D. (1999). Analysis of local decisions using hierarchical modeling, applied to home radon measurement and remediation. *Statist. Sci.* **14** 305–337.

PHILLIPS, C. V. and MALDONADO, G. (1999). Is More Research Really Needed? Unpublished manuscript, Univ. Minnesota School of Public Health, Division of Environmental and Occupational Health.

RAI, S. N. and KREWSKI, D. (1998). Uncertainty and variability analysis in multiplicative risk models. *Risk Analysis* **18** 37–45.

STANGL, D. K. (1995). Prediction and decision making using Bayesian hierarchical models. *Statistics in Medicine* **14** 2173–2190.

STOKEY, E. and ZECKHAUSER, R. (1978). *A Primer for Policy Analysis*. Norton, New York.

TAYLOR, A. C., EVANS, J. S. and MCKONE, T. E. (1993). The value of animal test information in environmental control decisions. *Risk Analysis* **13** 403–412.

VOSE, D. (1996). *Quantitative Risk Analysis: A Guide to Monte Carlo Simulation Modelling*. Wiley, Chichester.

WOLPERT, R. L., STEINBERG, L. J. and RECKHOW, K. H. (1993). Bayesian decision support using environmental transport-and-fate models (with discussion). *Case Studies in Bayesian Statistics. Lecture Notes in Statist.* **83** 241–293. Springer, New York.