SURVEY METHODOLOGY FOR STUDYING RADON LEVELS IN HOMES

Brenda G. Cox and Shelton M. Jones, Research Triangle Institute
Brenda G. Cox, P. O. Box 12194, Research Triangle Park, NC 27709

KEY WORDS: radon, environmental surveys

For a sizable fraction of the total public, recent studies indicate that the inhalation of radon decay products in the indoor environment results in exposure to ionizing radiation, with some sources estimating that indoor radon exposure may be responsible for more than 10% of the United States incidence of lung cancer (Bundnetz, 1978). Radon is a radioactive gas which is produced through the decay of radium-226, a naturally occurring element present in trace amounts throughout the earth's crust. Radon and its decay products rarely reach elevated levels outdoors because of continual dispersion and dilution. However, the reduced ventilation in enclosed buildings may enable radon and its decay products to reach levels indoors that are orders of magnitude above the outdoor level.

Elevated radon levels have been found in homes in many states (Alter and Oswald, 1987). In this paper we present issues in the design of surveys to measure radon concentrations in indoor air using the New York State Radon Survey as an example. Sponsored by the New York State Energy Research and Development Authority (NYSERDA), the survey was conducted in 1985-86 to determine the extent to which high radon concentrations occurred in New York State homes and the relationship between indoor radon levels and local geology and house construction characteristics (Sheldon, et al., 1988).

1. Establishing the Survey Objectives

The primary objective for the New York State Radon Survey was to describe the overall frequency distribution of radon concentrations within residential structures in New York State. Because human health effects are related to cumulative exposure, measurements over time were desired rather than grab samples collected at a single point in time. Moreover, indoor radon concentrations vary greatly over time depending upon ventilation (open versus closed windows and doors) and heating and cooling practices. Hence, an average exposure measurement over at least one heating season, but preferably one year, was clearly needed to satisfy survey objectives. The alpha track detector was chosen as the measurement device. To satisfy these objectives, both a three-month heating season and an annual radon measurement was made in each sample home.

Note that a different set of survey objectives could have led to a different choice of monitoring instruments. For instance, a number of states are participating in EPA-supported investigations whose purpose is to identify radon "hot spots," geographic areas where radon levels may be elevated in homes (Ronca-Battista, et al., 1987). These surveys are made using a three-day charcoal canister measurement taken during the winter. Because the measurement is for a short period of time, charcoal canisters cannot provide a good assessment of human health risk. But they can be used to identify areas where radon may be a potential problem in indoor air.

2. Defining the Target Population

The target population for a survey is the entire set of elements for which survey data will be used to make inferences (Cox and Cohen, 1985). For the New York State Radon Survey, the target population was defined as all year-round, single-unit, owner-occupied housing units in New York State that were continuously occupied during the monitoring time period. This definition is more inclusive than the target population definitions used for past studies of indoor radon concentrations. The rationale behind the choices made in developing the definition follows.

Housing units versus households. A household is the set of individuals who reside in a housing unit. Since radon concentrations are predominantly the result of housing unit rather than household characteristics, analyses tend to occur at the housing unit level. This suggested that the target population was best defined in terms of housing units.

Group quarters. Whether or not group quarters should be included was another issue. Examples of group quarters include mental hospitals, homes for the aged, correctional facilities, military barracks, dormitories, and rooming houses. Since their residents tend to be transient and the health effects of radon are related to exposure over time, it appeared reasonable to exclude group quarters from the target population.

Usual home elsewhere or vacant. The population of housing units can essentially be divided into three categories: (1) units occupied as principal residences, (2) seasonal, migratory, and occasional use units, and (3) unoccupied units. For the New York State Radon Survey, long term exposure to radon was of interest. For this reason, the study was restricted to housing units that were occupied as principal residences.

Households who move. A practical constraint imposed by an exposure measurement is that the housing unit should be continuously occupied. Leaving the radon detector in a vacant home after the household moves presents a problem since the resultant measurement does not reflect exposure of humans to health risk. Keeping a detector in place when one household moves out and another moves in results in operational problems. For these reasons, the target population for radon monitoring was restricted to nonmovers. To insure that the characteristics of homes occupied by nonmovers and movers could be contrasted, the target population for the telephone interview included movers.

Units in structure. The New York State Radon Survey also restricted attention to structures that contained only one housing unit (i.e., single-unit homes). Single unit homes include single-unit detached homes, attached homes such as town houses and row houses, and mobile homes and trailers. Previous studies have also excluded multi-unit dwellings such as apartment complexes since many of the homes are above ground and can be expected to have lower radon concentrations.

Owners versus renters. The New York State Radon Survey also restricted the target population to owner-occupied homes. The rationale behind the restriction was twofold: (1) a perception that owners must explicitly authorize the monitoring, and (2) the assumption that owner authorization would be difficult and costly to obtain for renter-occupied dwellings. It was felt that owner authorization was needed since hazardous levels of radon could reduce the market value of the dwelling or necessitate expensive mitigation remodeling.

3. Choosing the Mode of Data Collection

Two data collection modes were considered for use in administering the questionnaire and placing the detectors. The two options were a face-to-face interview design where detectors are placed by interviewers versus a telephone interview design where detectors are mailed to respondents after the questionnaire has been administered by telephone. Face-to-face methods were expected to produce higher response rates and better quality data. However, the telephone interview design was thought to be less costly. Even more important, greater geographic dispersion of the sample is possible with telephone data collection, an important characteristic in studying radon.

A telephone interview design was chosen for the New York State Radon Survey. The procedure consisted of sampling from the set of all potential telephone numbers in New York State. Interviewers dialed each sample number to determine whether the number was associated with an eligible residence. For year-round, owner-occupied, single-unit residences, the interviewer administered the questionnaire and then requested name and address information. Responding households were sent a package of detectors with instructions for placement and a postcard to return indicating the date they installed the detectors. After the monitoring period was complete, the household received a letter requesting return of the detectors. Prompting post cards and telephone calls were also used to encourage timely return of the detectors.

4. Developing a Stratification Scheme

During sample selection, ancillary data can be used to increase the precision of study estimates and to control the distribution of the sample by partitioning the target population into groups or strata and sampling independently within each stratum. Current research indicates that the principal source of radon in indoor air is radon in soil gas beneath the structure. The amount of radon in soil gas is related to the bedrock geology and surficial soil structure (including permeability). Therefore, geologic characteristics are the most important variables for use in stratifying the target population prior to sample selection.

For this study, New York State was partitioned into seven explicitly defined regions or strata based on geological characteristics and cost of interviewing. These strata were defined as the result of discussions between project staff, NYSERDA representatives and a New York State geologic consultant. Exhibit 1 characterizes these seven strata.

5. Constructing the Sampling Frame

The sampling frame for a survey is the list or mechanism used to enumerate population elements for sample selection purposes. The June 1985 AT&T list of telephone exchange codes was used to identify all telephone exchange codes used in New York State. An exchange code is defined by the area code and first three digits of the ten-digit telephone number. All possible last four digits could conceptually be added to these New York State exchange codes to create a frame of all telephone numbers allocated to New York State.

To stratify the frame prior to sample selection, New York State telephone exchange codes had to be linked to counties. The rate center city corresponding to each exchange code was used to identify the county associated with each telephone number. Since exchange codes do not exactly follow county boundaries, the classification of telephone numbers into counties was approximate. Our experience with this classification system indicates that the actual county of residence will differ from the county of classification for about ten percent of the telephone numbers. This approximate classification into counties is quite acceptable for stratification purposes. For analysis of the survey data, the actual county of residence must be obtained in the telephone interview.

6. Selecting the Sample

Much of the radon research prior to 1985 was based upon subjectively selected samples. These studies tended to be more exploratory in nature attempting to determine if certain geologic conditions could produce high radon levels in homes. For the New York State Radon Survey, the objectives were much broader requiring both a defensible quantification of the extent to which elevated radon levels were found in New York State homes and a capacity to explore the correlates of high radon levels. For this reason, probability sampling was indicated rather than the haphazard methods used in past studies. When probability sampling is used, every unit on the frame is given a known, nonzero probability of inclusion in the study (Cox and Cohen, 1985).

To reduce data collection costs, the Mitofsky-Waksberg variation of random digit dialing was used to select sample telephone numbers from each stratum (Waksberg, 1978). This method defines clusters of random digits as blocks of 100 numbers formed by taking the most recent AT&T national list of telephone area codes and prefix numbers (often referred to as central offices exchanges or NNX codes) and then adding all possible seventh and eighth digits from 00 to 99, thus constructing a sampling frame of all possible first eight digits associated with the universe of ten-digit telephone numbers.

These eight-digit numbers are treated as primary sampling units (PSUs). Each PSU represents a cluster of telephone numbers which may contain from 0 to 100 working telephone numbers. Using the Mitofsky-Waksberg procedure, m PSU clusters are selected and interviews are attempted with k+1 residential telephone numbers within each PSU. The PSUs are selected for the sample in the following manner. With-replacement sampling is used to select an NNX code and then the two remaining digits are randomly generated. If this telephone number corresponds to a residential address, the PSU is retained in the sample. Additional random last two digits are generated and dialed within this same eight-digit group of telephone numbers until a total of k+1 residences are identified. If the initial number is either a nonworking number or a nonresidential number, then the PSU is not included in the sample and no further calls are made to telephone numbers with these eight leading digits.

The Mitofsky-Waksberg design was created to take advantage of the fact that although only one in five telephone numbers in this country is residential, a rather high proportion of these eight-digit groups of telephone numbers contain no residential numbers at all. Groves (1978) presents empirical evidence that nationally 65 percent of the eight-digit groups contain no residential numbers. Thus, for those eight-digit clusters containing at least one residential number, the proportion of residential numbers would be much higher than for all telephone numbers (approximately 60 percent). A clustered approach to selecting telephone numbers for interviewing can be more cost effective than simple random sampling under these circumstances.

7. Allocating the Sample to Strata

Except when oversampling of population subgroups is needed, allocation of the sample to strata is usually made proportional to the size of the strata. For the New York State Radon Survey, the objectives required that estimates of reasonable precision be guaranteed for the State as a whole and for the first three strata since these three strata were thought to be potential trouble spots with regard to radon. (Initially the precision for Long Island estimates was to be controlled as well. This provision was dropped when exploratory field studies failed to find evidence of elevated radon levels on Long Island.) The first and third strata were sufficiently small that separate estimation capability could not be assured for these strata without oversampling.

In addition, data collection costs could be expected to vary substantially across strata with the New York City interviews the most expensive due to the low incidence of single-unit, owner-occupied homes there. This situation made optimal allocation of the sample an attractive alternative since it could result in substantial cost savings. Following customary practice, the optimal allocation was designed to minimize total survey cost subject to specified variance constraints (Cox and Cohen, 1985).

The following design constraints were involved in allocating the sample size:

- Population proportions statewide and for each stratum. This is the proportion of eligible homes having radon levels that exceed a specified value of interest. The allocation was developed to give precise estimates for proportions in the neighborhood of 0.05.

- Relative stratum sizes. The relative size of a stratum is the ratio of the number of eligible homes in the stratum to the number of eligible homes in the State. These values were computed using 1980 county-level Census information on the total number of housing units and the proportion of single-unit structures.

- Interviewing costs. This includes the cost of unproductive telephone numbers (i.e., ones not yielding a single-unit home), and the cost of completed telephone interviews with eligible households. These costs were estimated to be $2 and $20, respectively, and were assumed to be the same for all strata.

- The probability of identifying a cluster from a randomly selected telephone number. This probability varied from stratum to stratum. Its value was estimated for each stratum based on the number of exchange codes, the number of households, and the percentage of households that were single-unit.

- The conditional probability of obtaining an eligible home within an identified cluster. This probability also varied from stratum to stratum. The values used were taken from previous RTI experience in random digit dialing for the respective locations and adjusted to account for the restriction to single-unit homes.

- The intracluster correlation. This design constant describes the tendency for homes in the same cluster to have similar radon levels. Its value was assumed to be 0.05 for all strata. The value was chosen based on previous survey experience and on certain features of this study.

These design constraints are summarized in Exhibit 2.

Having specified the design constants, sample size solutions were obtained in two steps using Chromy's optimal allocation procedure (Chromy, 1979). The optimal second-stage sample sizes for each stratum were computed first. (These cluster sizes were later adjusted upwards to compensate for nonresponse.)

Given the second-stage sample sizes, the first-stage stratum sample sizes required to meet the variance constraints were computed. The variance constraints were specified in terms of the coefficient of variation (CV), which is defined as

$$CV(\hat{P}) = \frac{[Var(\hat{P})]^{1/2}}{P} ,$$

where P is the population proportion being estimated, $\hat{P}$ is the estimator based on the sample data, and $Var(\hat{P})$ is the variance of $\hat{P}$. The variance constraints were imposed at P = 0.05 for this design, as discussed earlier.

The following variance constraints were used to obtain the allocation:

$CV(\hat{P}) \leq 0.13$ (statewide);

$CV[\hat{P}(1)] \leq 0.35$ (Stratum 1);

$CV[\hat{P}(h)] \leq 0.25$ h = 2, 3; and

$CV[\hat{P}(h)] \leq 1.00$ h = 4, 5, 6, 7.

The strongest constraint was placed on the variance of the statewide estimate. Constraints were also imposed for Strata 1, 2, and 3 estimates, because these strata were of special interest to the study, beyond their contribution to statewide estimates. Variances for the remaining stratum-level estimates were only minimally constrained. The sample allocation solutions to the above constraints are shown in Exhibit 2.

8. Results of Sampling

Use of the Mitofsky-Waksberg design resulted in calls to a total of 21,813 telephone numbers. These numbers comprised the bank of telephone numbers for which calls were made by telephone interviewers. The data collection experience is given in Exhibit 3 for different stages of data collection. Of the 21,813 telephone numbers dialed, 35% were working residential numbers. Fifty-four percent of these working residential numbers were determined to be eligible households and 92% of these yielded a completed interview. Of these completed interviews, 131 households were ineligible for radon monitoring due to the respondents' plans to move within the next 12 months. This meant that 97% of the respondents satisfied the movement status criteria (at this stage). Each respondent's participation was solicited, which was determined by whether or not the respondent would agree to radon monitoring and provide residential address information. Of those who satisfied the movement criteria, 89% of the households agreed to radon monitoring. In an effort not to exceed desired sample design yields, a probability subsample of 3,115 households was selected for radon monitoring. The sample design assumed that 65% of the households mailed three-month detectors would yield valid radon measurements. In fact, the response rate realized was 73% or 2,267 homes with valid heating-season measurements for the primary living area. The twelve-month monitors required a longer detector period, therefore, response was not as high as for the short-term heating season. At a rate of 62% of the nonmoving participants, a total of 1,930 households provided valid twelve-month measurements.

Having to call 21,813 telephone numbers to get valid heating-season measurements from 2,267 homes and annual measurements from 1,930 homes seems excessive at first glance. Some explanations are in order.

Unfortunately, only a small fraction of the sample telephone numbers corresponded to an eligible residence. The following different types of survey ineligibles were identified in the sample telephone numbers:

- 14,135 nonresidential or nonworking numbers,
- 3,531 residences that were multi-unit, renter-occupied or vacation homes, and
- 131 households who indicated they intended to move within 12 months.

An additional 149 households were not selected for radon monitoring. Removing the above cases from the total sample results in 3,867 homes being solicited to obtain valid measurements from 59% for the heating season and 50% for the year.

Unfortunately, low response rates such as these are characteristic of environmental surveys (Cox, Mage and Immerman, 1988). The telephone approach with its lack of personal contact may make the problem worse since respondents may find it easier to refuse over the telephone than in a face-to-face contact. In addition, valid radon measurements were not obtained for some cooperating household because of mistakes they made in installing or returning the detectors, such as failing to note the beginning or end date of monitoring or improper return packaging. Another problem that caused poor response was the fact that most respondents had never heard of radon, which led many to question our motives. The wide publicity attached to the hazards of radon gas in homes since 1985 may have eliminated this problem for future studies.

REFERENCES

Alter, H.W. and R.A. Oswald (1987), Nationwide Distribution of Indoor Radon Measurements: A Preliminary Data Base, Journal of the Air Pollution Control Association 37, 227-231.

Bundnetz, R.J., J.V. Beck, C.D. Hallowell, W.W., Nazerwoff, A.T. Nero, and A.H. Rosenfeld (1978), Human Disease from Exposure. The Impact of Energy Conservation in Buildings, Laurence Berkeley Laboratory, Publication 7809, University of California at Berkeley.

Chromy, J.R. (1979), Sequential Sample Selection Methods. Proceedings of the American Statistical Association, Survey Research Methods Section, 401-406.

Cox, B.G. and S.B. Cohen (1985,, Methodological Issues for Health Care Surveys. New York: Marcel Dekker, Inc.

Cox, B.G., D.T. Mage and F.W. Immerman (1988), Sample Design Considerations for Indoor Air Exposure Surveys, Journal of the Air Pollution Control Association Vol. 38, No. 10.

Geomet Technologics (1981), Indoor Air Monitoring Protocol Guidelines for Residential Weatherization Program, G.T.I. Report No. ES-923, Washington, DC: U.S. Environmental Protection Agency.

Glasser, G.J. and G.D. Metzger (1975),"National Estimates of Nonlisted Telephone Households and Their Characteristics, Journal of Marketing Research XII, 359-61.

Health, R.C. (1984), Ground-Water Regions of the United States. U.S. Geological Survey, Water Supply Paper 2242, 78 pp.

Kish, L. (1965). Survey Sampling. New York: John Wiley and Sons, Inc., 334-339.

Ronca-Battista, M., M. Moon, J. Bergsten, S.B. White, B. Alexander, and N. Holt (1987), Radon-222 Concentrations in the United States - Results of Sample Surveys in Five States, Proceedings of the Natural Radiation Environment IV, Radiation Protection Dosimetry.

Sheldon, L.S., T.D. Hartwell, B.G. Cox, J.E. Sickles II, E.D. Pellizzari, M.L. Smith, R.L. Perritt and S.M. Jones (1988), An Investigation of Infiltration and Indoor Air Quality, Contract No. 736-CON-BCS-85, Albany, NY: New York State Energy Research and Development Authority.

U.S. Bureau of the Census (1985), Statistical Abstract of the United States: 1986. Washington, DC: U.S. Government Printing Office.

Waksberg, J. (1978), Sampling Methods for Random Digit Dialing. Journal of the American Statistical Association, 73, Number 361.

Exhibit 1. Characterization of the Strata Used in the New York State Radon Survey

Stratum Number and Title	Description	Occupied Housing Units
1. Binghamton	This is an area of special interest based upon data from the National Uranium Resource Evaluation (NURE) Program. High levels of radon have also been observed in water.	127,872
2. Undeformed Sediments	Relatively undeformed Paleozoic sediments dominate central and southern New York, from Albany to Buffalo, and south to Binghamton. High radon concentrations have been found in this region.	1,766,039
3. Metamorphic Rock	The Adirondack metamorphic/igneous rocks contain facies analogous to those associated with high radon concentrations in water in Maine.	182,230
4. Deformed Sediments and Rock	East of Albany and extending down to New York City, this is belt of complex deformed, orogenic sediments and metamorphic rocks.	645,578
5. Staten Island	Staten Island is in part underlain by Triassic sediments similar to those associated with unusually high concentrations of radon in unpublished "grab" sample studies.	114,551
6. Long Island	Much of Long Island is underlain by glacial sediments. Radon levels should be low in this area containing predominantly sand deposits.	792,521
7. New York City	This is not a separate geological region, but the size of the city warrants its individual consideration since single-unit, owner-occupied homes are less frequent here than elsewhere in the State.	2,674,215
TOTAL		6,303,006

Exhibit 2. Stratum-Level Design Constants

Stratum (h)	Relative Stratum Size (Single-Unit Homes) N(h)/N	Probability of Identifying a Cluster	Probability of Identifying a Single-Unit Home, Given a Cluster	First-Stage Sample Size (Number of Clusters)	Second-Stage Sample Size (Housing Units per Cluster)	Total Sample Size (Housing Units)
1. Binghamton	3.1%	16.7%	35.7%	151	6	906
2. Undeformed Sediments	41.8%	13.4%	29.0%	410	6	2,460
3. Metamorphic Rock	4.5%	11.0%	31.9%	181	8	1,448
4. Deformed Sediments and Rock	14.0%	13.8%	24.0%	149	6	894
5. Staten Island	2.1%	25.9%	27.2%	49	5	245
6. Long Island	2.4%	21.8%	36.4%	246	4	984
7. New York City	10.1%	3.1%	4.5%	54	30	1,620

Exhibit 3. Telephone Interviewing Results

Stratum Number	Telephone Numbers Called	Eligible Households	Completed Interviews	Nonmovers Identified	Cooperating Nonmovers	Nonmovers Receiving Monitors	Valid 3-Month Measurements	Valid 12-Month Measurements
1	2,128	552	518	507	448	448	339	302
2	7,074	1,419	1,307	1,245	1,107	958	730	624
3	4,029	855	795	770	697	697	516	440
4	2,357	457	419	405	353	353	246	212
5	543	102	92	90	78	78	46	37
6	2,920	625	565	549	485	485	326	265
7	2,762	137	117	116	96	96	64	50
Total	21,813	4,147	3,813	3,682	3,264	3,115	2,267	1,930