Atmospheric Environment Vol. 22, No. 10, pp. 2115–2125, 1988.
Printed in Great Britain.

0004–6981/88 $3.00 + 0.00

THE BOSTON RESIDENTIAL NO_2 CHARACTERIZATION STUDY—II. SURVEY METHODOLOGY AND POPULATION CONCENTRATION ESTIMATES

P. BARRY RYAN, MARY LOU SOCZEK, ROBERT D. TREITMAN and JOHN D. SPENGLER
Department of Environmental Science and Physiology, Harvard School of Public Health, 655 Huntington Avenue, Boston, MA 02115, U.S.A.

IRWIN H. BILLICK
Gas Research Institute, 8600 West Bryn Mawr Avenue, Chicago, IL 60631, U.S.A.

(*First received* 10 *October* 1986 *and in final form* 30 *November* 1987)

Abstract—A recently completed year-long study of NO_2 and air exchange rates in over 500 homes in the Boston Metropolitan area provides data to quantify the component of total NO_2 exposures attributable to indoor sources, especially to gas-fired appliances. The approach of this work was to provide field data for validation or refinement of exposure models developed in previous, related work. For an indoor characterization field study, sample sizes of 450 gas- and 150 electric-range-equipped housing units were selected based on: (1) modeled estimation of precision and stability of parameter estimates using various sample sizes; and (2) calculations including anticipated attrition from one monitoring period to the next. The sample was selected using standard area probability sampling to allow extrapolations of survey sample results to the larger population. The survey design included stratification by range fuel and area clustering for sampling and logistical efficiency. This paper presents the sampling results and field work progress through the year, with a discussion of response rates typical for exposure monitoring investigations. Monitoring results provide NO_2 concentration data to evaluate the overall success of the survey implementation. A series of analyses isolate and quantify the standard errors of distribution estimates. Using sample data weighted for stratification, population exposure distribution parameter estimates are presented. Overall, analyses indicate that key model assumptions are valid. The relatively low standard errors of exposure parameters indicate that the design used in the study was relatively efficient. This illustrates the utility of standard survey research methodology in exposure assessment problems.

Key word index: NO_2, survey research, indoor air quality, exposure assessment, Boston.

INTRODUCTION

Collection of data on the indoor home microenvironment is essential to examine the dynamics of the relationship between indoor concentrations and total personal exposures because of the large amount of time individuals spend in their homes (Szalai, 1972; Chapin, 1974) and the documented importance of this indoor environment on total personal exposures (Spengler and Soczek, 1985). The Harvard School of Public Health Indoor Air Quality Group conducted a large field study of NO_2 concentrations and air exchange rates in over 500 households in the Boston, Massachusetts, Standard Metropolitan Statistical Area (SMSA). The overall objective of this work was to quantify the component of total NO_2 exposures attributable to indoor sources, especially unvented gas-fired appliances (Soczek *et al.*, 1985; Ryan *et al.*, 1988). We adopted an approach in which empirically-based, microenvironmental exposure models were developed (Ryan *et al.*, 1986) and data collected to validate or refine the models. Prior work has shown that the microenvironment contributing the most to personal exposure variance is the indoor home microenvironment. (Quackenboss *et al.*, 1985; Spengler *et al.*, 1985; Ryan *et al.*, 1988.) Thus, we directed our efforts towards characterization of indoor environments to improve model estimates.

The microenvironmental approach, while offering a useful avenue for analysis of environmental exposure, presents some difficulty in designing a sampling strategy. Selection of a purely random sample of elements of the population, in this case specific microenvironments, offers the advantage of ensuring that such elements would occur in the sample as they occur in the population as a whole. But this strategy suffers in that elements which have more intrinsic variability in any particular parameter would be relatively poorly-described. A stratified-random sample, allowing separation of the sample into strata, affords a more efficient design. Previous work (Spengler *et al.*, 1983; Letz *et al.*, 1983; Quackenboss *et al.*, 1982; Quackenboss *et al.*, 1985) had shown that homes equipped with gas ranges display higher NO_2 concentrations and are more variable than homes without these appliances. This suggests an efficient design which separates the sample according to the presence or absence of a gas range and samples these residences at a higher rate. This represents a novel application of survey sampling methodology to indoor air quality research.

Optimum sample sizes of 275 gas-range-equipped (Gas) housing units (HUs) and 110 electric-range-equipped (Electric) HUs were chosen based on modeling work in which the stability of exposure distribution parameters at various sample sizes was estimated (Soczek *et al.*, 1985; Ryan *et al.*, 1988). With samples of these sizes, we expected to estimate the mean NO_2 concentration in Gas HUs within 2.5 ppb; in Electric HUs within 1.5 ppb. In addition, other distribution parameters can be estimated adequately with these sample sizes.

To investigate the anticipated seasonal component of the distribution, we needed to replicate our monitoring during consecutive seasonal periods. An analysis of data from the year-long indoor monitoring studies conducted in Portage, Wisconsin (Spengler *et al.*, 1983) and Topeka, Kansas (Letz *et al.*, 1983) showed that 75–95 % of the variance in the annual mean NO_2 concentration could be recovered with measurements from as few as three or four well-chosen periods. According to these results, the maximum variance would be recovered by monitoring within the January–February, July–August and October–November periods. Additional information could be provided with data from a fourth, April–May period. Four monitoring periods were planned. Initial samples of 450 Gas and 150 Electric HUs were specified to allow for an expected 10 % attrition from period to period (Lopez, personal communication, 1985).

For sampling and logistical efficiency, a 2-stage sampling scheme incorporating stratification by range fuel and clustering was used. First, each city and town in the SMSA was assigned to one of three gas-density strata depending upon the expected percentage of HUs using gas as the primary cooking fuel, as reported in the 1980 Census. To increase the likelihood of clusters containing both Gas and Electric HUs, the interval of selection of clusters (Census blocks) and the number of HUs selected within the blocks was adjusted for each density stratum. Next, Census blocks were selected within each stratum. Individual HUs within each block were then identified, and a set designated as a cluster. A 1/900 sample of 1091 housing units in 63 clusters was identified. Finally, to produce the desired number of Gas and Electric HUs, all Gas units in the sample were automatically eligible for inclusion whereas 40 % of Electric units in the sample would be randomly selected. We expected to find 800 eligible housing units and obtain a response rate of 75%, to yield a final sample of 600 HUs. For a more complete description of the sampling methods, see Ryan *et al.* (1988).

Results from a pilot study are presented elsewhere (Soczek *et al.*, 1985; Ryan *et al.*, 1988). In this paper we present data from the full study. The main purpose of this analysis is to test for design effects in the data, assess the success of the survey approach and present weighted population exposure estimates. As in our preliminary work with the pilot study data, we will limit this discussion to NO_2 concentrations.

SURVEY METHODOLOGY

Sampling results

Sample screening and household recruitment began with a pilot study conducted on 10 % of the sample in the fall of 1984 (Soczek *et al.*, 1985; Ryan *et al.*, 1988). The observed 69 % response rate, defined as the percentage of eligible individuals contacted who agreed to participate in the monitoring, while somewhat lower than anticipated, suggested an adequate number of households would be recruited into the full sample. Full-scale sampling began in January 1985. In February, it became evident that a response rate of approximately 60 % would result, necessitating the expansion of the selection process to ensure sufficient participation.

The expansion of the selection process was carried out in a manner identical to the selection of the original sample. The addition of housing units in 18 new clusters and continued recruiting in the original 63, bringing the total to 1498 units, changed the overall probability of selection to 1/700. Figure 1 shows the location of the 81 clusters in the SMSA. The selection and interview process with the additional housing units required an extended monitoring period; as a result, sampling and monitoring actually spanned the planned winter and spring monitoring periods and reduced the number of actual monitoring periods to three. A total of 973 eligible HUs were identified, with 581 agreeing to participate in the monitoring. Thus, the final response rate was 59.7 %.

Monitoring progress

We completed three full periods of indoor monitoring; the numbers of HUs monitored each period are presented in Fig. 2. Although we anticipated attrition from one monitoring period to the next, we expected to monitor all 581 HUs in the sample in first monitoring period (WIN/SPR). In actuality, we set up only 501 (86.2 %) of these. Technicians received refusals from 45 HUs (7.7 %) and were not able to schedule field visits for the remaining 35 HUs (6.0 %).

In the subsequent monitoring periods, the pool of eligible HUs was designated as the HUs monitored in the previous period, plus the 'potentials' for which no final disposition had been reached. Only HUs whose residents had *refused* monitoring were dropped from the pool.

In the second period (SUMMER), 465 HUs (86.7 %) of the eligibles were monitored; in the last period (FALL), 439 HUs (87.1 %) were monitored. The success of efforts to maintain the sample of 501 monitored in WIN/SPR to provide year-long data, plus recruit the problematic 'potential' HUs to provide additional

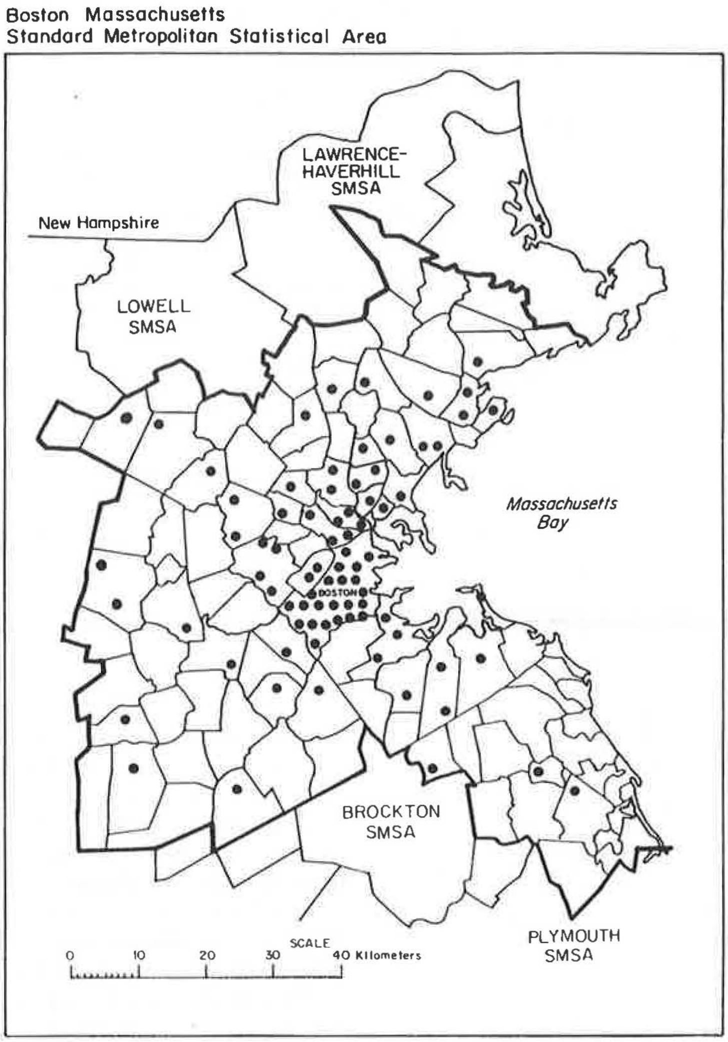

Fig. 1. Map of Boston Standard Metropolitan Statistical Area (SMSA) showing location of clusters monitored.

numbers for cross-sectional analyses are shown in Fig. 3. Of the original 501 monitored, 42 dropped out prior to the SUMMER monitoring period. An additional 34 dropped out before the FALL monitoring period. Eight homes monitored in the WIN/SPR period, missed the SUMMER period but were monitored again in the FALL period. Fourteen new HUs were monitored in the SUMMER period. Twelve of those were also included in the FALL sample. Two residences were monitored only in the FALL period.

Overall, 417 HUs were monitored all three periods. The breakdown by range type for these HUs (298 Gas; 119 Electric) compares favorably to our original target of 275 Gas and 110 Electric HUs. Only 64 HUs were never monitored; we monitored at least once in 517 (89.0 %) of the original 581.

In addition to the basic cross-seasonal analyses, the monitoring strategy was planned to allow estimation of concentration differences associated with usage of a home by different residents. Census records for the Boston area indicated that approximately 20 % of our sample would move during the one-year study period (U.S. Bureau of the Census, 1982). As we encountered HUs with new residents, an attempt was made to recruit them into the study. In SUMMER we found 44 (8.2 %) of the 536 potential HUs in the sample had new residents. We recruited 31 of these new-resident HUs into the study. Of these 31 residences, 23 had been monitored previously through the former residents. In FALL, 31 (6.2%) of the remaining 504 potential HUs had new residents. We recruited 14 of these, 12 of which had been monitored in the previous period. One HU had different residents in all three periods, but was monitored only during WIN/SPR period.

Sample retrieval

To eliminate return visits to retrieve the samplers, we implemented a participant sampler capping and mailing system (Soczek *et al.*, 1985; Ryan *et al.*, 1988).

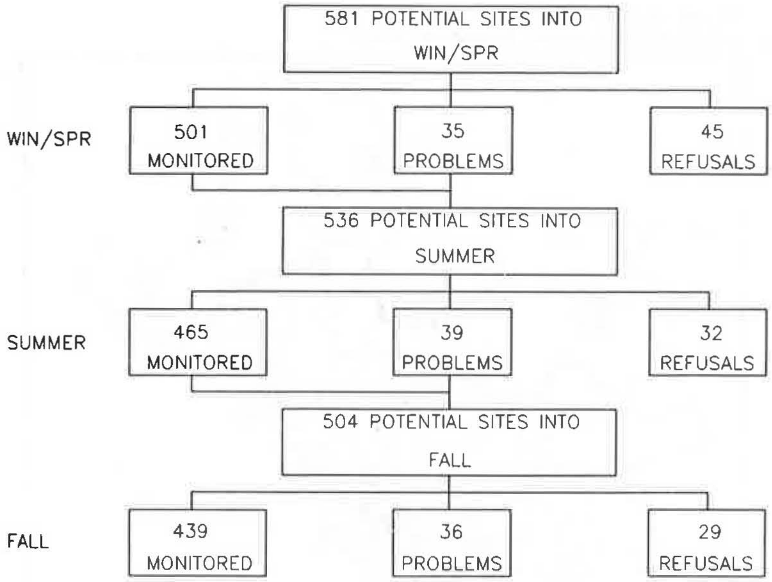

Fig. 2. Sample monitoring and attrition throughout the year.

Fig. 3. Individual site participation through the year.

Mailers from over 94 % of the monitored HUs were returned each period. Of these, data loss due to missing or improperly capped samplers averaged 9 %. Overall, the percentage of HUs with valid data, relative to the number of HUs monitored, averaged 88 % over the three monitoring periods.

ANALYSIS

Quality assurance

An extensive quality assurance effort was instituted in this study to ensure the validity of the data collection process. NO_2 samplers were prepared in batches with approximately 5 % of the samplers in each batch held in the laboratory for laboratory blanks. Additionally, 5 % of all samples taken in the field were replicate samples and an additional 5 % were field blanks. Laboratory blanks were used to correct the observed absorbances measured in field samples to account for any contamination in manufacturing or storage. The field blanks were not used to correct field data but rather as a check on the quality assurance of the system. The absorbances of the field blanks, once corrected for the laboratory background blanks, showed non-significant increases equivalent to less than 1 ppb for a 2-week sample.

The analysis of all replicate samples is depicted in Fig. 4. The strong correlation and relatively small scatter about the 1:1 line suggests a precise measurement. Precision estimates are: 10.3 % for WIN/SPR; 4.7 % for SUMMER; and 8.0 % for FALL presented as relative standard deviation for replicate pairs.

Data summaries

NO_2 concentrations measured inside and outside of Gas and Electric HUs for each period are shown in Table 1. For both groups, measured values were somewhat lower than predicted with the exposure models. In all periods NO_2 concentrations in the Gas HUs were higher and showed more variance than concentrations in Electric HUs. Indoor levels in Gas HUs were higher than the outdoor levels, whereas in Electric HUs the opposite was true.

Design effect analysis

In air quality research and in particular indoor air quality research, it is often necessary to cluster samples for logistical reasons: proximity to laboratories; convenience for field technicians, etc. The clustering may be temporal, spatial or both. In this study, we chose a clustered design with the belief that such a selection would prove the most efficient for sample

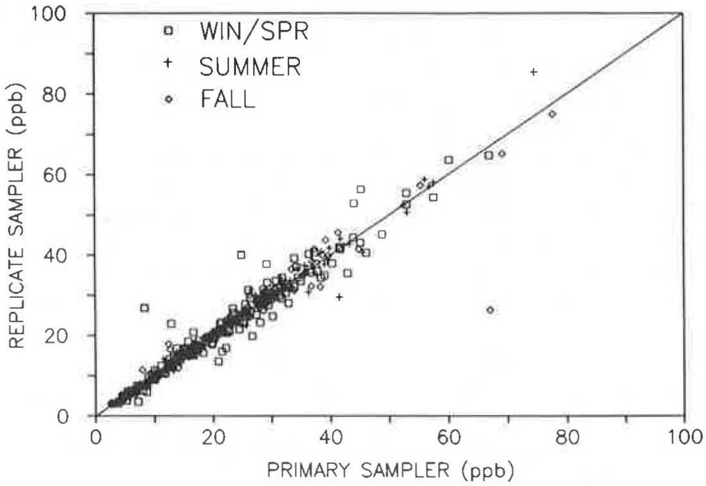

Fig. 4. Palmes tube replicate data.

Table 1. Summary statistics for Gas and Electric populations for each monitoring period

Period	Location	Gas			Electric		
		N	Mean	Std	N	Mean	Std
1	Kitchen	298	38.7	18.8	121	9.4	7.0
WIN/SPR	Living room	298	29.0	14.5	124	9.3	5.0
	Bedroom	291	25.6	14.0	115	8.2	7.8
	Outdoor	242	21.2	7.7	103	17.6	7.3
		N	Mean	Std	N	Mean	Std
2	Kitchen	301	34.7	16.1	117	13.0	5.4
SUMMER	Living room	298	26.8	10.5	117	12.9	5.5
	Bedroom	283	24.0	8.7	114	12.1	5.0
	Outdoor	251	22.0	8.9	103	16.7	7.3
		N	Mean	Std	N	Mean	Std
3	Kitchen	277	39.1	21.0	117	8.6	4.4
FALL	Living room	276	27.8	15.2	116	8.6	4.6
	Bedroom	269	25.0	14.7	115	8.0	5.1
	Outdoor	242	23.0	6.9	101	19.9	6.0

Concentrations in [ppb].

placement. This design included spatial clustering in that the homes in a given cluster were geographically proximate to one another, and the potential for temporal clustering as technicians had the option of setting up homes at their convenience. No attempt has been made to quantify the temporal and spatial effects separately. Both are considered part of the clustering.

In a survey design relying on a clustered approach to sampling, there always exists a possibility of decreased power of prediction when compared to a fully random design. This so-called 'design effect' results in an effective loss of sample size due to the fact that elements taken from within a cluster are, generally, more alike than elements taken randomly from the population as a whole. To test whether such effects are significant in that they may compromise the ability to extrapolate from the sample to a larger population, one may take a phased approach to analysis in which a preliminary analysis of variance is performed and the cluster variable itself is introduced as a potential predictor of, in this case, NO_2 concentration (Ryan *et al*., 1988). If the cluster variable is not a significant predictor, no significant design effects exist and analysis may proceed. If such analysis of variance suggests the cluster variable as a significant predictor of variability, then further analyses are warranted in which this effect is quantified.

The approach in this work is to perform the analyses of variance prior to the more sophisticated approaches. The results of such analyses then can be used as a guide to components in need of further study.

Analysis of variance

To test for design effects, the influence of stratification and clustering must be quantified. Stratification increases the efficiency of the design and a complete or near-complete separation of sample elements by the stratification variable is desirable. However, clustering can reduce the power of prediction if the elements within each cluster are fairly homogeneous. A further problem is that a correlation between the stratification variable and the cluster variable would produce an apparent design effect where none actually existed, as the cluster variable would act as a surrogate for the true effect of the stratification variable. If significant design effects exist, analysis must first adjust for these effects to ensure that subsequent findings are not invalidated by the potential correlation with a design element (Kish, 1965).

Accordingly, a series of analyses were performed to test for the presence of design effects in the data. With each measurement, using the NO_2 concentration as the dependent variable, we parameterized a general linear model with the cluster variable (Cluster) and the stratification variable (range type, Range) as main effects. Results are presented separately for each monitoring period in Tables 2–4. The Type I sum of squares (SS) corresponds to the fraction of the variance explained by the variable if it were the first or only effect in the model. The Type III SS corresponds to the marginal variance that this effect explains after the other effect is evaluated.

In all analyses of *indoor* measurements Cluster as a Type I effect alone accounts for a significant proportion of the total variability of the NO_2 concentrations, suggesting a very strong design effect. Analysis of the explained variance associated with the Type III SS, however, shows that Cluster is acting *partly* as an incomplete surrogate for Range. This observation stands in contrast to the results of the pilot phase of this project (Ryan *et al.*, 1988), which showed no

Table 2. Results of analysis of variance for WIN/SPR nitrogen dioxide concentrations accounting for effects of Cluster and Range

Nitrogen dioxide	Model effect	df	Type SS	% Variance explained	Significance test F	p
Kitchen	Cluster	78	I	43.8	4.35	<0.001
			III	17.1	1.69	0.001
$N-419$	Rangc	1	I	39.9	312.53	<0.001
			III	13.1	104.41	<0.001
Living room	Cluster	78	I	43.3	4.10	<0.001
			III	20.5	1.94	<0.001
	Range	1	I	34.0	255.01	<0.001
$N=422$			III	11.2	84.07	<0.001
Bedroom	Cluster	78	I	39.1	3.11	<0.001
			III	20.0	1.58	0.003
$N=407$	Range	1	I	28.2	177.05	<0.001
			III	9.1	56.84	<0.001
Outdoor	Cluster	78	I	54.6	4.16	<0.001
			III	50.3	3.83	<0.001
$N=345$	Range	1	I	4.3	25.39	<0.001
			III	0.0	0.02	0.888

Table 3. Results of analysis of variance for SUMMER nitrogen dioxide concentrations accounting for effects of Cluster and Range

Nitrogen dioxide	Model effect	df	Type SS	% Variance explained	Significance test F	p
Kitchen	Cluster	78	I	60.7	8.40	<0.001
			III	36.1	4.99	<0.001
$N=418$	Range	1	I	33.1	360.98	<0.001
			III	8.4	91.70	<0.001
Living room	Cluster	78	I	58.7	7.47	<0.001
			III	35.5	4.51	<0.001
	Range	1	I	31.3	314.77	<0.001
$N=418$			III	8.1	81.22	<0.001
Bedroom	Cluster	78	I	58.1	6.71	<0.001
			III	32.9	3.81	<0.001
$N=397$	Range	1	I	32.5	296.56	<0.001
			III	7.3	66.94	<0.001
Outdoor	Cluster	78	I	75.8	11.40	<0.001
			III	68.6	10.31	<0.001
$N=354$	Range	1	I	7.7	89.09	<0.001
			III	0.4	5.11	0.02

Table 4. Results of analysis of variance for FALL nitrogen dioxide concentrations accounting for effects of Cluster and Range

Nitrogen dioxide	Model effect	df	Type SS	% Variance explained	Significance test F	p
Kitchen	Cluster	78	I	42.3	4.67	<0.001
			III	16.0	1.40	0.024
$N=394$	Range	1	I	38.1	260.06	<0.001
			III	11.7	79.97	<0.001
Living room	Cluster	78	I	40.3	3.87	<0.001
			III	18.2	1.44	0.016
	Range	1	I	31.3	193.22	<0.001
$N=392$			III	9.2	56.71	<0.001
Bedroom	Cluster	78	I	33.8	2.91	<0.001
			III	15.3	1.05	0.382
$N=384$	Range	1	I	27.8	148.42	<0.001
			III	9.3	49.65	<0.001
Outdoor	Cluster	78	I	50.6	3.52	<0.001
			III	46.3	3.26	<0.001
$N=343$	Range	1	I	4.3	23.21	<0.001
			III	0.0	0.00	0.998

significant effects of Cluster after accounting for Range. The statistically significant marginal variance explained by Cluster *after* accounting for Range indicates that a significant cluster effect exists and that further analysis of these data to determine the strength of the effect is necessary. The effect of clustering of sample elements varies from season to season, with Cluster explaining more variability in SUMMER measurements but only producing marginal significance in FALL.

Analysis of the *outdoor* NO_2 concentration data, on the other hand, shows Range to be an incomplete surrogate for the true cluster effect, as Cluster explains all of the variability associated with Range. Range acts as a surrogate for Cluster as Gas HUs are found more frequently in the urban areas of the SMSA and because the urban and rural clusters are of small geographic extent and, hence, small relative variability in NO_2 concentrations. We speculate the cluster effect observed in the indoor concentrations may well reflect the infiltration of different ambient levels of NO_2 concentrations across the population.

In further analyses, the effect of Cluster on prediction of NO_2 concentrations was isolated with parallel analyses of the effect of Cluster in the separate Gas and Electric samples, as shown in Table 5. In WIN/SPR, Cluster explains over 72 % of the variability in NO_2 concentrations in Electric HUs, but only 32–34 % of the variance in Gas HUs. We suggest the variability *not* explained in the Gas HUs is due to the differences in the usage of gas fuel, air exchange rates and volumes among HUs within each cluster.

In SUMMER, the power of Cluster in predicting NO_2 levels in Gas HUs is more pronounced than in WIN/SPR. During FALL, the predictive power of Cluster for indoor NO_2 concentrations falls below the level of statistical significance. This is due, in part, to the modestly decreasing sample sizes and increasing standard deviations in the measurements as suggested by data presented in Table 1. One may speculate that the modest increase in variability among the indoor measurements may be due to varied times over which housing units are 'closed-up' for the winter season. The SUMMER and FALL monitoring periods suggest Cluster to be an accurate predictor of Electric HU variability during these periods with 72–82% of the variability accounted for by this variable.

Intra-class correlation and design effect calculation

Analysis of variance procedures test for the presence of a relationship, but give no indication of the strength of the association between a dependent and an independent variable. A measure of the strength of association which also lends itself easily to direct design effect analysis is the intra-class correlation coefficient, *rho*. *Rho* is a measure of the homogeneity within sample clusters; it is essentially a characteristic of a clustered sample and affects the variance component estimation. It is the portion of the total element variance that is due to clustering.

The calculation, after Blalock (1960), is as follows:

$$rho=\frac{(MS_c-MS_e)}{(MS_c+(n-1)MS_e)} \tag{1}$$

where MS_c = Mean Square attributable to clustering
MS_e = Mean Square attributable to error
n = average number of elements per cluster.

Rho is normally positive, showing some homogeneity of elements, with values from 0 to +1. (A negative *rho* results when elements within clusters are extremely heterogeneous.) *Rho* equals 0 when the Mean Square attributable to cluster equals the Mean Square attributable to error, that is, when elements drawn from within a cluster are no more similar than

Table 5. Results of one-way analyses of variance of nitrogen dioxide concentrations accounting for effects of cluster

Period	Sample	Location	N	df	Variance explained (%)	F	p
	Gas	Kitchen	298	66	32.3	1.66	0.003
1		Living room	298	66	33.7	1.78	0.001
		Bedroom	291	66	32.5	1.64	0.004
		Outdoor	242	66	49.5	2.62	<0.001
	Electric	Kitchen	121	50	83.0	6.85	<0.001
WIN/SPR		Living room	124	51	72.0	3.62	<0.001
		Bedroom	115	50	91.0	12.96	<0.001
		Outdoor	103	47	84.5	6.40	<0.001
	Gas	Kitchen	301	68	56.6	4.46	<0.001
2		Living room	298	68	52.3	3.70	<0.001
		Bedroom	283	68	51.4	3.32	<0.001
		Outdoor	251	66	80.6	11.60	<0.001
	Electric	Kitchen	117	47	76.4	4.76	<0.001
SUMMER		Living room	117	47	78.8	5.46	<0.001
		Bedroom	114	46	72.8	3.90	<0.001
		Outdoor	103	42	86.4	9.09	<0.001
	Gas	Kitchen	277	66	28.8	1.29	0.094
3		Living room	276	66	30.6	1.40	0.039
		Bedroom	269	65	24.7	1.02	0.361
		Outdoor	242	64	51.4	2.93	<0.001
	Electric	Kitchen	117	49	82.1	6.28	<0.001
FALL		Living room	116	49	79.0	5.06	<0.001
		Bedroom	115	47	80.7	5.97	<0.001
		Outdoor	101	44	77.7	4.44	<0.001

elements drawn randomly from the population as a whole.

A statistic directly related to *rho* is the design effect (*Deff*). *Deff* is calculated from *rho* and the average number of elements per cluster (n):

$$Deff = 1 + rho(n-1). \qquad (2)$$

Deff represents the increase in sample variance due to clustering. The overall sample variance is greater in a clustered sample than in a simple random sample because the within-cluster variance is generally smaller than if the set of elements were drawn from a simple random sample. A design effect of unity, equivalent to a *rho* of zero, would indicate that elements within a clustered sample are no different from elements in a simple random sample. A design effect greater than unity reflects the degree to which sample elements are clustered or are more similar than randomly chosen elements. For example, a design effect of 2 would mean a sample of 200 would provide only as much information as a simple random sample of 100. Thus, as the design effect *increases*, the effective sample size *decreases*.

A third statistic, the standard error of the mean, is a measure of the variability of the mean and an indicator of how much error may exist in a given estimate. The usual formula for the standard error gives values appropriate only for estimates made with simple random samples or cluster samples with design effects of unity. In samples with design effects greater than 1, an appropriate correction factor is the square root of the *Deff* associated with that measurement. Thus the formula:

$$SE = \sqrt{\left(\frac{(Deff\, s_c^2)}{n}\right)} \qquad (3)$$

where s_c^2 = observed standard deviation for clustered sample

is used for calculating standard errors in clustered samples. Using adjusted standard errors, accurate 95 % confidence intervals for the means (95 % CI) can be calculated.

The *rho*, *Deff* and 95 % CI associated with each concentration in the separate Gas and Electric samples are presented in Table 6. In all monitoring periods, the *rhos* for indoor NO_2 concentrations are greater for the Electric than for the Gas samples. In the WIN/SPR and FALL monitoring periods, each Electric HU in the sample provides less new information than each Gas HU. In SUMMER data, a different pattern is found. The Gas measurements are more highly inter-correlated than for Gas residences in the other periods, with design effects of averaging 2.3, larger than any other monitoring period for Gas or Electric residences. One may speculate that this is a reflection of the highly-correlated outdoor measurements ($rho = 0.74$) found for Gas residences in this monitoring period. Overall, the effective sample size of the Electric sample is approximately half of the actual

Table 6. Intra-class correlation coefficients, design effects and 95% confidence limits, for each location by monitoring period

Period	Location	Gas			Electric		
		Rho	*Deff*	95% CI	*Rho*	*Deff*	95% CI
1	Kitchen	0.13	1.45	2.57	0.71	1.98	1.75
	Living room	0.15	1.51	2.02	0.52	1.73	1.16
WIN/SPR	Bedroom	0.13	1.43	1.92	0.84	2.06	2.04
	Outdoor	0.31	1.81	1.30	0.72	1.82	1.90
		Rho	*Deff*	95% CI	*Rho*	*Deff*	95% CI
2	Kitchen	0.44	2.49	2.87	0.61	1.87	1.34
	Living room	0.38	2.28	1.80	0.65	1.93	1.38
SUMMER	Bedroom	0.36	2.12	1.48	0.54	1.78	1.22
	Outdoor	0.74	3.03	1.92	0.77	2.08	2.03
		Rho	*Deff*	95% CI	*Rho*	*Deff*	95% CI
3	Kitchen	0.06	1.20	2.71	0.69	1.93	1.11
	Living room	0.09	1.28	2.03	0.64	1.84	1.14
FALL	Bedroom	0.01	1.02	1.77	0.67	1.94	1.30
	Outdoor	0.34	1.93	1.21	0.61	1.75	1.55

Concentrations in [ppb].

sample size as indicated by *Deff* values of about 2.0. The effective sample size of the Gas sample, of primary interest in our study, is reduced by a factor of only 1.5. Design effects for outdoor concentrations are similar across seasons and for both samples.

The 95 % CI show that the accuracy of means estimated for the Gas population are approximately 2.0 ppb in each monitoring period. For the Electric population, the precision varies from slightly greater than 1.0 ppb to about 2.0 ppb over the year. These observed error estimates agree quite closely with the errors specified in the initial sample size calculations, 2.5 for Gas and 1.5 for Electric samples. Thus the design effects present in the data do not compromise the desired level of precision in prediction.

ESTIMATING POPULATION EXPOSURES

Weighting

All the previous analyses have been based on the actual sample data. To estimate population parameters, the sample elements must be weighted to account for the differential probabilities of selection introduced by the stratified design. All Gas HUs were assigned a stratification weight of 1.0. Since only a percentage of Electric HUs were included in the final sample, each Electric element was assigned a stratification weight of 2.0 or 2.5, determined by which of two selection protocols was implemented at the second stage of sampling.

Parameter estimation

Using weighted data, population exposure distribution parameters were estimated. Means, standard deviations and selected percentiles for the overall population and the Gas and Electric subgroups are presented in Tables 7–9. As a point of reference, the National Ambient Air Quality Standard (NAAQS) for NO_2 is 53 ppb based on an annual average. Examination of kitchen estimates shows that 10 % of the population as a whole have values exceeding 53 ppb, for a 2-week average, during WIN/SPR and in FALL while 5 % show similar concentrations during SUMMER. Analysis of the values for the two subgroups shows that this group is composed almost entirely of Gas HUs. In the Gas population, nearly 20 % exceed 53 ppb in WIN/SPR and FALL, 10 % in SUMMER; whereas in the Electric subgroup, one exceedance, a WIN/SPR bedroom, is found. Outdoor estimates show that neither in the population as a whole nor in the subgroups are any exceedances found. This is consistent with historical monitoring data for the Boston Metropolitan area (Spengler *et al.*, 1979).

CONCLUSION

Data from this survey allow examination of two key areas: (1) the validity of assumptions underlying the monitoring and modeling approach used in this study and (2) the advantages and limitations of applying standard survey methods to exposure assessment problems. Analyses of response rates, design effects and standard errors of the parameter estimates provide the needed information for each evaluation.

In this study, a lower than anticipated original sampling response rate was offset by our success in maintaining the sample through the year; the final sample sizes are adequate for accurate population parameter estimation. Our previous analyses indicate the modification to the monitoring schedule resulting in the loss of the spring season of data will not seriously impact on estimates of annual averages.

The design effect analyses show that the efficiency of the clustered design did result in less predictive power than would have been obtained with a simple random

Table 7. Concentration percentile estimates for total population [ppb]

Period/ location	Percentiles								
	1	5	10	25	50	75	90	95	99
WIN/SPR									
Kitchen	1.2	3.8	5.0	8.1	16.3	36.3	53.4	65.4	87.4
Living room	2.8	4.1	4.9	8.1	14.4	26.6	37.5	48.4	71.2
Bedroom	1.5	2.8	4.3	6.4	13.1	23.4	36.4	44.2	75.8
Outdoor	0.4	8.9	10.5	13.9	18.8	23.9	28.0	32.5	42.5
SUMMER									
Kitchen	3.5	6.4	8.8	12.4	19.8	32.7	43.4	53.9	78.9
Living room	2.6	6.4	8.1	11.7	17.7	26.0	34.0	39.8	56.1
Bedroom	3.1	6.1	7.6	10.8	16.3	23.5	30.6	35.2	46.4
Outdoor	6.2	7.1	9.4	13.1	18.3	23.5	29.2	33.2	49.8
FALL									
Kitchen	2.8	3.4	4.9	7.8	15.5	35.3	52.3	63.1	95.1
Living room	2.5	3.4	4.9	7.5	13.1	25.4	36.2	44.0	70.5
Bedroom	1.7	2.9	3.8	6.3	12.4	23.1	32.8	40.1	73.6
Outdoor	10.5	13.0	14.6	16.8	20.6	25.0	28.8	30.7	42.9

Table 8. Concentration percentile estimates for Gas population [ppb]

Period/ location	Percentiles								
	1	5	10	25	50	75	90	95	99
WIN/SPR									
Kitchen	5.5	13.3	17.0	26.2	35.0	48.7	65.2	74.6	103.6
Living room	5.6	10.3	13.9	19.0	26.5	34.8	48.5	57.6	84.1
Bedroom	4.3	9.2	11.6	16.6	22.4	32.3	42.3	51.7	82.7
Outdoor	0.4	9.7	12.0	16.4	21.5	25.4	29.8	33.2	44.1
SUMMER									
Kitchen	10.5	15.5	18.3	24.6	31.6	41.4	53.7	62.1	92.8
Living room	6.4	12.7	14.4	20.4	25.4	31.7	39.8	46.2	58.6
Bedroom	7.0	11.6	13.5	17.9	23.2	29.1	35.2	40.7	49.4
Outdoor	5.4	10.1	12.1	16.5	21.2	26.4	31.5	36.5	54.0
FALL									
Kitchen	8.5	14.0	16.8	24.1	35.6	48.3	63.1	75.9	115.7
Living room	6.4	10.7	12.6	17.0	25.4	34.1	44.2	57.6	84.4
Bedroon	4.4	8.7	11.0	14.5	22.7	31.2	40.3	53.2	87.4
Outdoor	10.9	14.7	15.9	18.9	22.4	26.3	29.7	31.3	57.8

Table 9. Concentration percentile estimates for electric population [ppb]

Period/ location	Percentiles								
	1	5	10	25	50	75	90	95	99
WIN/SPR									
Kitchen	0.9	2.6	3.8	5.4	8.2	11.8	14.6	19.0	50.7
Living room	2.0	3.7	4.1	5.4	8.2	11.3	15.4	17.8	22.9
Bedroom	1.0	2.1	2.8	4.5	6.6	9.6	14.4	17.4	60.0
Outdoor	0.3	8.4	9.7	12.8	16.8	21.7	26.0	28.7	42.4
SUMMER									
Kitchen	2.5	5.5	6.4	9.4	12.4	15.7	20.3	22.8	27.8
Living room	2.6	4.6	6.5	9.3	11.9	15.8	20.0	22.9	28.6
Bedroom	2.2	4.7	6.4	8.9	11.0	14.8	20.2	21.8	23.8
Outdoor	6.3	6.7	7.8	11.8	15.7	20.4	27.8	29.1	49.2
FALL									
Kitchen	2.1	3.0	3.5	6.0	8.0	10.2	15.2	16.7	25.6
Living room	2.0	2.8	3.5	5.4	7.8	10.0	15.1	16.9	26.2
Bedroom	1.2	2.5	3.0	4.6	6.5	9.8	14.8	19.3	29.6
Outdoor	7.6	12.3	13.7	15.9	18.6	23.1	27.5	30.4	40.2

sample of the same size. The confidence intervals for the NO_2 concentration measurements, however, show that the desired degree of precision was indeed obtained. This is due to the fact that the design effects are small and that the concentration measurements show somewhat less variance than originally predicted.

Several important conclusions can be drawn from this study. While response rates of 80–90 % for general HU samples in metropolitan areas are typical (Fowler, 1984; Marquis, 1979), this work supports evidence that, in studies requiring in-house monitoring, response rates somewhat lower can be expected (Mage *et al.*, 1986). It has been suggested that the limiting factor is quite likely the larger respondent burden produced by the introduction of the physical measurement process (Whitmore, 1985). Future work should explore potential bias due to non-response and possibly develop methods for obtaining higher initial response rates.

Stratification and clustering generally increase the efficiency of probability-based survey designs, and data from this study strongly support its use in exposure assessment studies. Specifically, generalizations can be made from our findings with NO_2 data that while design effects in samples with clusters averaging 5–10 elements each on the order of 2.0 for Electric and 1.5 for Gas populations can be expected, this does not seriously affect the standard error of key parameter estimates if adequate sample sizes are maintained.

Acknowledgements—Work reported in this paper was conducted under Gas Research Institute Contract No. 5082-251-0739. The authors acknowledge the contribution of Floyd J. Fowler, Jr, of the Center for Survey Research, and Jennifer L. Porter, Research Assistant. We thank an anonymous reviewer for numerous useful comments.

REFERENCES

Blalock H. M. (1960) *Social Statistics.* McGraw Hill, New York.

Chapin F. S. (1974) *Human Activity Patterns in the City.* Wiley-Interscience, New York.

Fowler F. J., Jr (1984) *Survey Research Methods, Applied Social Research Methods Series, Vol.* 1. Sage Publications, Beverly Hills.

Kish L. (1965) *Survey Sampling.* John Wiley and Sons, New York.

Letz R., Spengler J. D. and Miller D. P. (1983) Relationships of measured NO_2 concentrations at discrete sampling locations in residences in Topeka, Kansas. Paper presented at the Third Annual National Symposium on Recent Advances in Pollutant Monitoring of Ambient Air and Stationary Sources, Raleigh, NC, May 1983.

Mage D. T., Cox B. G. and Immerman F. W. (1986) Sample design considerations for indoor air exposure surveys. Presented at the 79th Annual Meeting of the Air Pollution Control Association, Minneapolis, MN, June 1986.

Marquis K. H. (1979) Survey response rates: some trends, causes, and correlates. *Health Survey Research Methods: Second Biennial Conference*, National Center for Health Services Research. Hyattsville, MD. DHEW Pub. No. (PHS) 79-3207,: 3–12.

Quackenboss J. J., Kanarek M. S., Spengler J. D. and Letz R. (1982) Personal monitoring for nitrogen dioxide exposure: methodological considerations for a community study. *Envir. Int.* **8,** 249–258.

Quackenboss J. J., Spengler J. D., Kanarek M. S., Letz R. and Duffy C. P. (1985) Personal exposure to nitrogen dioxide: relationship to indoor/outdoor air quality and activity patterns. *Envir. Sci. Technol.* **20(6),** 775–783.

Ryan P. B., Soczek M. L., Spengler J. D. and Billick I. H. (1988) The Boston residential NO_2 characterization study—I. A preliminary evaluation of the survey methodology. *J. Air Pollut. Control Ass.* **38,** 22–27.

Ryan P. B., Spengler J. D. and Letz R. (1986) Estimating personal exposures to NO_2. *Envir. Int.* **12,** 395–400.

Soczek M. L., Ryan P. B., Spengler J. D., Fowler F. J. and Billick I. H. (1985) A survey methodology for characterization of residential NO_2 concentrations. Presented at the 78th Annual Meeting of the Air Pollution Control Association, Detroit, MI, June 1985.

Spengler J. D., Duffy C. P., Letz R., Tibbitts T. W. and Ferris Jr, B. G. (1983) Nitrogen dioxide inside and outside 137 homes and implications for ambient air quality standards and health effects research. *Envir. Sci. Technol.* **17,** 164–168.

Spengler J. D., Ferris B. G., Jr., Dockery D. W. and Speizer F. E. (1979) Sulfur dioxide and nitrogen dioxide levels inside and outside homes and the implications on health effects research. *Envir. Sci. Technol.* **13,** 1276–1280.

Spengler J. D. and Soczek M. L. (1985) Evidence for improved ambient air quality and the need for personal exposure research. *Envir. Sci. Technol.* **18,** 268A–280A.

Spengler J. D., Treitman R. D., Tosteson T. D., Mage D. T. and Soczek M. L. (1985) Personal exposures to respirable particles and implications for air pollution epidemiology. *Envir. Sci. Technol.* **19,** 700–707.

Szalai A. (ed) (1972) *The Use of Time: Daily Activities of Urban and Suburban Populations in Twelve Countries.* Mouton, The Hague.

U.S. Bureau of the Census (1982) *Census of Population and Housing, Supplementary Report: Provisional Estimates of Social, Economic and Housing Characteristics—States and Selected Standard Metropolitan Statistical Areas.* U.S. Government Printing Office, Washington, DC.

Whitmore R. W. (1985) Design of surveys for residential and personal monitoring of hazardous substances. Presented at the Workshop on Human Exposure Assessment: Monitoring and Modeling, Harvard University, Boston, MA.