

# Predictive Models Based on Personal, Indoor and Outdoor Air Pollution Exposure

Roland Hosein<sup>1</sup>, Paul Corey<sup>2</sup>, Frances Silverman<sup>3</sup>, Anthony Ayiomamitis<sup>4</sup>,  
R. Bruce Urch<sup>5</sup> and Neil Alexis<sup>6</sup>  
Gage Research Institute, Toronto, Ontario M5T 1R4

## Abstract

Portable air pollution samplers were used to measure sulphur dioxide (SO<sub>2</sub>), nitrogen dioxide (NO<sub>2</sub>) and respirable suspended particulates (RSP) in a study of a group of nineteen asthmatics during two periods in the winter and summer respectively. One sampler was carried by each subject, one was placed in the home indoors, and one outdoors by the home. In addition, similar pollutants were measured at a central stationary site within a 15 km radius during the same time periods. Samplers were not placed, however, in other indoor spaces where subjects spent part(s) of the day. We used the data from all the sampling sites to develop predictive models for personal exposure. With 330 person-days of exposure data, multiple regression of these "fixed site" measures of pollution against the personal exposure measures revealed a predictive relationship whose power increased proportionally to the time the subjects spent indoors. This relationship was limited, however, since samplers were not placed at other in-

door spaces, thus leaving the predictive model incomplete. A pollution index in which these indoor and outdoor pollutant measures were weighted by the time spent at home indoors and outdoors was predictive of personal exposure for NO<sub>2</sub> and RSP ( $R=0.78, 0.44$  respectively); the SO<sub>2</sub> levels were too low to be used in the comparative analyses ( $R=0.19$ ).

## Introduction

The health consequences of exposure to low levels of air pollution remain unclear. The association of lung function and symptom reporting with pollution levels measured at stationary outdoor sites is very weak, or probably non-existent, because the outdoor levels may not reflect true exposure. The most valid method for assessing true exposure is by personal sampling, but in epidemiological studies in which large populations are used, this method may be impractical if the exposure pattern of each subject has to be characterized.

The earliest models for assessing exposure were based on coal consumption rates (Pattle et al., 1979; Holland and Reid, 1965), and by comparing populations with urban or rural residence. As better measurement methods were developed, a smoke index was used that referred to suspended particulates only (Holland et al., 1969; Lawther et al., 1970). Later on a gravimetric method for total suspended particulates (TSP) was used, and in addition, liquid absorption methods were developed for sulphur dioxide (SO<sub>2</sub>) and nitrogen diox-

## KEY WORDS:

Multi-pollutant sampler, Nitrogen dioxide (NO<sub>2</sub>), Sulphur dioxide (SO<sub>2</sub>), Respirable suspended particulate (RSP), Pollution index

Manuscript received: 22 October 1990

Accepted for publication: 7 November 1991

<sup>1</sup> Associate Professor, Occupational and Environmental Health Unit, Faculty of Medicine, University of Toronto.

<sup>2</sup> Professor, Department of Preventive Medicine and Biostatistics, Faculty of Medicine, University of Toronto.

<sup>3</sup> Assistant Professor, Department of Medicine, Faculty of Medicine, University of Toronto.

<sup>4</sup> Research Assistant, The Gage Research Institute.

<sup>5</sup> Charge Technologist, The Gage Research Institute

<sup>6</sup> Graduate Student, The Gage Research Institute.

ide ( $\text{NO}_2$ ). All these methods were designed primarily to assess air pollution in communities and were related to monitoring the contribution from emission sources, but were not designed for examining correlations with health effects (Ferris et al., 1973; Shy et al., 1973).

The levels of air pollution measured at outdoor sites were of low predictive power of either acute or chronic respiratory illnesses and diseases. Indoor air pollution levels were different from outdoor levels in many studies undertaken in the period after 1970 (Binder et al., 1976; Ott and Eliason, 1973; Dockery et al., 1981; Ott and Flachsbar, 1982; Quackenboss et al., 1986; Spengler and Soczek, 1984; Fischer et al., 1986) and the sources of such indoor pollution were established (Quackenboss et al., 1986; Fischer et al., 1986; Harlos et al., 1988). The availability of small motors, pumps and battery packs allowed the development of portable samplers to measure personal exposure to air pollution (Ott and Eliason, 1973; Fischer et al., 1986; Harlos et al., 1988; Spengler et al., 1987; Sexton et al., 1984; Sega and Fugas, 1982). We designed a multi-pollutant personal sampler which was portable and was used to sample simultaneously  $\text{SO}_2$ ,  $\text{NO}_2$ , and RSP for up to 8 hours (Mintz et al., 1981; Silverman et al., 1982). In this study, our subjects carried the personal samplers and, in addition, similar samplers were placed inside and outside the home. A stationary government site that was central in the city but within a 15 km radius of each subject's home was also used to generate data as part of the study. We used the data to develop predictive models of personal exposure using the other samplers, as personal sampling may be impractical for large-scale population studies. Duan (1982) felt that advancing technology on passive sampling may allow personal sampling of large groups. Some researchers have tried to develop predictive models for personal exposure based on exposure data that may be easier to obtain; for example, Spengler et al. (1987)

used indoor home, indoor school and outdoor levels to develop a model for  $\text{NO}_2$  exposure in children; Sega and Fugas (1982) used a model of personal and outdoor levels for RSP; Horie and Eldon (1979) postulated that component monitoring will be practical; and Ott (1983-84) stated that the amount of time spent in each location may be predictive of real personal exposure.

## Methods

The subjects selected for this study came from a group of asthmatics at the Gage Research Institute Clinic, Toronto, who lived in homes without gas stoves. Nineteen asthmatics were chosen to participate in the study and to carry the personal sampler 8 hours per day for 2 weeks in the heating season and 2 weeks in the non-heating season. During the same time period, a similar sampler was placed within each subject's home in an area most frequently used by the subject. Another similar sampler was placed outside of the subject's home at least 15 meters from any street, in the backyard but not under a tree. Both indoor and outdoor sampler probes were kept at about 1 meter above ground. Another identical sampler was placed at a central location in the city that was within a 15 km radius of all of the subjects' homes.

The sampler (Mintz, et al. 1981; Silverman, et al. 1982) was designed to sample simultaneously for respirable suspended particulates,  $\text{NO}_2$ , and  $\text{SO}_2$ . It consisted of a battery-operated sampling pump system connected to a filter assembly for RSP, and to two spill-proof impingers, each containing appropriate absorbing reagents for  $\text{NO}_2$  and  $\text{SO}_2$ . The samplers were calibrated daily for flow. The systems for  $\text{NO}_2$  and  $\text{SO}_2$  were calibrated against standard permeation tube techniques and real time monitors (Mintz et al., 1981). Respirable suspended particulates were collected (flow rate 1.7 l/min) using a 10 mm nylon cyclone assembly and filter cassette

or school and out-  
model for NO<sub>2</sub> expo-  
d Fugas (1982) used  
outdoor levels for  
(1979) postulated that  
will be practical; and  
the amount of time  
may be predictive of

r this study came  
cs at the Gage Re-  
ronto, who lived in  
es. Nineteen asth-  
participate in the  
personal sampler 8  
eks in the heating  
a the non-heating  
time period, a simi-  
within each subject's  
quently used by the  
sampler was placed  
ome at least 15 me-  
e backyard but not  
and outdoor sam-  
about 1 meter above  
ical sampler was  
on in the city that  
is of all of the sub-

al. 1981; Silverman,  
o sample simultan-  
ended particulates,  
d of a battery-oper-  
em connected to a  
d to two spill-proof  
ng appropriate ab-  
and SO<sub>2</sub>. The sam-  
y for flow. The sys-  
<sub>2</sub> were calibrated  
on tube techniques  
Mintz et al., 1981).  
rticulates were col-  
using a 10 mm ny-  
and filter cassette

which effectively collected particles with an aerodynamic diameter of less than 10  $\mu\text{m}$ .

Subjects were instructed in the care and use of their assigned personal sampler since the gas collection impingers contained liquid reagents. Our technicians attended to the other samplers. At the start, flow rate was measured, and a built-in timer recorded the start-up time; at the end of the day the elapsed time was again recorded in addition to the flow rates. The filters were transferred to petri dishes for conditioning and weighing. The NO<sub>2</sub> concentration was assessed using the TGS-ANSA method (Mulik et al., 1974), the SO<sub>2</sub> by the West-Gaeke method (West and Gaeke, 1956).

In addition to carrying the personal sampler, subjects filled in a daily diary concerning activities which may affect their exposure or health (Silverman et al., 1982). A more detailed questionnaire was also administered by the technician who attended to the samplers. Lung function tests (using a Vitalograph spirometer) were performed at the start and end of each day by the technician.

All data management and statistical analyses were carried out using the statistical package SAS (SAS Users Guide, 1979). Linear correlation analyses were conducted comparing the four methods of sampling for each pollutant to determine whether the levels obtained from the indoor, outdoor or outdoor central samplers bore any relationship to the results obtained using a personal sampler. The p values associated with the tests of significance of correlation and regression coefficients serve only as a guide since no account was taken of the auto-correlation between repeated measurements on the same subjects over time.

On each day, subjects were asked to record the amount of time spent in one of four categories: home indoors, home outdoors, away from home outdoors (other outdoors), and away from home indoors (other indoors). Because of the logistics involved we were not able to have a sampler for this last category

(other indoors), hence any predictive model based on a combination of the other three categories would be strongest for the days on which the percentage of time spent indoors away from home was low. Multiple regression analyses were performed between the three exposure categories on five mutually exclusive subsets of data based on the time spent indoors away from home. The subsets were developed arbitrarily and the ranges chosen for each subset were for analytical convenience only.

Another index was used which weighted the indoor and outdoor central estimates by the percent of time spent in these areas. This index consisted of multiplying the indoor pollutant level by the percentage of time spent indoors added to the outdoor central level multiplied by the time spent outdoors i.e.:

$$\text{Index} = \text{Percent of Time Indoors (Indoor Level)} + \text{Percent of Time Outdoors (Outdoor Level)}$$

## Results

The highest correlations were observed between personal and indoor sampling for NO<sub>2</sub>, SO<sub>2</sub> and RSP, but they explained less than 32% of the variation (Table 1). The relationship between the personal and outdoor home samples and that between personal and outdoor central samples, explained a much smaller percentage of the variation for NO<sub>2</sub> and RSP; the personal SO<sub>2</sub> levels showed a good relationship with outdoor home levels. As expected, outdoor home and outdoor central levels for NO<sub>2</sub> and SO<sub>2</sub> showed a strong relationship explaining 34% and 45% of the variation respectively.

Because of this low predictive power of the various measures, a linear combination was explored since subjects spent some time away from home (Table 2). In an attempt to improve the prediction of the personal exposure estimate the predictive ability of a linear

**Table 1.** Linear correlation between personal, home indoor, home outdoor and outdoor central sampling for NO<sub>2</sub>, SO<sub>2</sub> and RSP

	Correlation Coefficient			Mean ( $\mu\text{g}/\text{m}^3$ ) N = 330	S.D. +/- ( $\mu\text{g}/\text{m}^3$ )	Range min-max ( $\mu\text{g}/\text{m}^3$ )
	Indoor home	Outdoor home	Outdoor central			
<b>NO<sub>2</sub></b>						
Personal	0.57*	0.36	0.31	40	22	10-152
Indoor home		0.43	0.29	36	21	2-118
Outdoor home			0.58*	42	22	1-115
Outdoor central				52	22	2-120
<b>SO<sub>2</sub></b>						
Personal	0.45*	0.53*	0.36	4	8	0-92
Indoor home		0.43	0.34	2	4	0-37
Outdoor home			0.67*	16	23	0-138
Outdoor central				24	28	0-147
<b>RSP</b>						
Personal	0.50*	0.31	0.16	51	23	11-130
Indoor home		0.25	0.19	43	25	0-197
Outdoor home			0.42	32	18	0-161
Outdoor central				38	16	2-91

\*Significant at the 5% level of a conventional test of a correlation coefficient.

**Table 2.** Percentage of time spent indoors and outdoors

Ranges of time spent indoors other(%)	N <sup>a</sup>	Percentage			
		Indoors		Outdoors	
		home	other	home	other
0	84	94.6	0	2.7	2.7
1 - 19	52	82.9	11.2	0.6	5.3
20 - 79	80	51.0	46.7	0.5	1.8
80 - 99	55	8.1	90.8	0.2	0.9
100	59	0	100	0	0
Total	330	50.9	46.1	0.9	2.1

<sup>a</sup>N = total number of samples collected.

combination of the three exposure measurements was investigated using multiple linear regression. It was anticipated that the regression model would have the most predictive power for those days on which individuals spent less time indoors away from home for which no corresponding exposure measurements were available. Therefore these multiple regression analyses were performed

on five mutually exclusive subsets of the data, based on the time spent indoors away from home. The five subsets of data are given in Table 2. Twenty-five percent (84) of the samples came from subjects who spent 95% of their time indoors at home and 5% outdoors. The findings also revealed that our subjects spent between 94-99% of their time indoors over the sampling period at home or indoors elsewhere. It is important to note that one third of the subjects (114) spent more than 90% of their time away from home indoors. Since we did not assess the levels in other indoor spaces, the predictive model development is somewhat incomplete.

The multiple linear regression coefficients between personal, indoor home, outdoor home, and outdoor fixed station (central location) for NO<sub>2</sub> are shown in Table 3. For those subjects who spent no time indoors away from home, the estimates were well predicted by the indoor home and outdoor central levels, i.e. R = 0.81. As the propor-

central sampling for NO<sub>2</sub>,

S.D. +/- (µg/m<sup>3</sup>)  
Range min-max (µg/m<sup>3</sup>)

22 10-152  
21 2-118  
22 1-115  
22 2-120

8 0-92  
4 0-37  
23 0-138  
28 0-147

23 11-130  
25 0-197  
18 0-161  
16 2-91

...sive subsets of the  
...e spent indoors away  
...subsets of data are  
...y-five percent (84) of  
...subjects who spent  
...ors at home and 5%  
...also revealed that our  
...94-99% of their time  
...ng period at home or  
...s important to note  
...subjects (114) spent  
...eir time away from  
...did not assess the le-  
...aces, the predictive  
...mewhat incomplete.  
...egression coefficients  
...oor home, outdoor  
...d station (central lo-  
...own in Table 3. For  
...nt no time indoors  
...estimates were well  
...home and outdoor  
...0.81. As the propor-

**Table 3.** Regression coefficients using three fixed sites to predict personal exposure for NO<sub>2</sub>

Ranges of time spent indoors other (%)	N	Indoors home		Outdoors home		Outdoors central		R <sup>c</sup>
		B <sup>a</sup>	p <sup>b</sup>	B	p	B	p	
0	84	0.89	0.0001	-0.03	0.82	0.23	0.03	0.81
1 - 19	52	0.83	0.0001	0.14	0.19	-0.01	0.95	0.85
20 - 79	80	0.43	0.0001	0.07	0.48	0.22	0.02	0.64
80 - 99	55	-0.02	0.85	0.43	0.007	-0.03	0.79	0.42
100	59	0.13	0.51	-0.21	0.21	0.45	0.03	0.35
combined	330	0.52	0.0001	0.08	0.18	0.12	0.03	0.59

<sup>a</sup>B is the regression coefficient

<sup>b</sup>p is the statistical probability

<sup>c</sup>R is the multiple correlation coefficient of the predictive equation

tion of time spent away from home indoors increased, the predictive power decreased. The indoor home and outdoor central levels jointly made a significant contribution to the regression model.

Because of the relationship between the personal samples and the indoor home and outdoor central samples, we decided to undertake multiple regression analyses for NO<sub>2</sub>, SO<sub>2</sub> and RSP on those subjects who spent no time indoors away from home (Table 4). For each pollutant the resulting regression equation explained over 50% of the variation in the measurements made by the personal sampler. Both the indoor home and outdoor central made separate contributions to the regression for NO<sub>2</sub> and SO<sub>2</sub> respectively.

Although the multiple regression was sat-

**Table 4.** Multiple linear regression between personal, indoor home and outdoor central exposure for which subjects spent no time in other indoor locations (N = 84)

Pollutant	Indoor home		Outdoor central		R <sup>c</sup>
	B <sup>a</sup>	p <sup>b</sup>	B	p	
NO <sub>2</sub>	0.88	0.0001	0.21	0.02	0.81
SO <sub>2</sub>	0.72	0.0001	0.08	0.0001	0.75
RSP	0.62	0.0001	-0.09	0.50	0.70

<sup>a</sup>B is the regression coefficient.

<sup>b</sup>p is the statistical probability.

<sup>c</sup>R is the multiple correlation coefficient of the predictive equation.

isfactory in explaining a large proportion of the variation, a weighting index was developed which was based on time spent indoors at home and outdoor central (Table 5). The correlation between the index and the personal sample was significant for NO<sub>2</sub> and RSP but not for SO<sub>2</sub>. This low correlation with SO<sub>2</sub> was probably due to its relatively low level indoors and overall.

**Table 5.** Relationship between personal sample and index<sup>a</sup> (N = 84)<sup>b</sup>

Index for	B	p	R
NO <sub>2</sub>	0.94	0.00001	0.78
SO <sub>2</sub>	0.27	0.084	0.19
RSP	0.62	0.0003	0.44

<sup>a</sup> Weighted combination of indoor and outdoor central estimates.

<sup>b</sup> 84 days sampled on which individuals spent no time in other indoor locations.

## Discussion

Most of our subjects spent a very large proportion of their time at home, yet the predictive power of the indoor sampling was low when compared with the personal sampling. This finding may be explained in part by the fact that judgement was used in placing the indoor sampler in the room that we thought was most frequently used by the subjects.

This measure gave an average level for that particular room, and would not reflect the exposure of the subject to pollutants that may be higher in other rooms. Hosein et al. (1986) reported mean NO<sub>2</sub>, SO<sub>2</sub> and TSP levels inside asthmatic homes of 27 µg/m<sup>3</sup>, 2 µg/m<sup>3</sup>, and 67 µg/m<sup>3</sup> respectively. The same study also reported mean NO<sub>2</sub> levels inside homes (including non asthmatics) without gas stoves of 76 µg/m<sup>3</sup> (summer) and 80 µg/m<sup>3</sup> (winter). Spengler et al. (1979) reported mean NO<sub>2</sub> levels ranging from 7 µg/m<sup>3</sup> to 45 µg/m<sup>3</sup> inside homes (living rooms) without gas cooking. Mean RSP levels of 35 µg/m<sup>3</sup> and 40 µg/m<sup>3</sup> (Hosein et al., 1986) have been reported and ranges of 50 - 124 µg/m<sup>3</sup> (Hosein and Corey, 1987) have been reported, both inside non-smoking homes. The mean levels found in this study are for the most part consistent with levels reported by prior investigators. However, matching and controlling for many source variables were not done as in previous work, due to the modeling focus of this paper.

The potential sources of air pollution and/or influencing factors in this study were: outside diffusion, heating systems, air conditioners, number of occupants, furniture, floor and wall coverings, cooking with flour, indoor hobbies (sanding/grinding) and fireplaces. Homes with gas stoves were excluded from this study and since all sampled homes were occupied by asthmatics, smoking and pets were not identifiable sources. Perhaps the relatively low levels of NO<sub>2</sub> measured were due to the absence of the major contributing source to high indoor NO<sub>2</sub> levels, gas stoves (Hosein et al., 1986). In addition, smoking would have been expected to account for a high RSP level (Hosein et al., 1986) yet smoking was absent here, thereby implicating other pollution sources. Prior work (Hosein et al., 1986) has demonstrated that the number of occupants, outside infiltration (air tightness) and pets are inconsistent effectors on indoor NO<sub>2</sub>, SO<sub>2</sub> and RSP levels inside homes.

It is well known that in industrial environments, exposure measurements usually show a poor correlation between area sampler levels and levels measured on workers carrying personal samplers. The ratios of measurements from personal exposure monitors to those made by area exposure monitors are summarized to be typically 3 to 10 for occupational settings and 1.2 to 3.3 for residential settings (Rodes et al., 1991). The relatively lower correlation between personal levels and indoor levels for SO<sub>2</sub> may in part be due to the extremely low indoor levels; the high correlation between personal and outdoor home levels for SO<sub>2</sub> was probably due to the fact that the main contribution of indoor SO<sub>2</sub> was from infiltration of outdoor air. The correlation between the outdoor home and outdoor central samples for all pollutants was significant as might be expected, but the differences between the means confirm that sites close to each other still experience concentrations that are only partially related (Ott and Eliason, 1973; Hosein et al., 1977). The low predictive power of indoor exposure and health responses, as shown elsewhere (Binder et al., 1976; Hosein and Corey, 1987), may be explained by the poor relationship between indoor and personal exposure levels.

In the total sample of 330 person days an average of 46.1% of the time was spent indoors away from home, an activity for which there existed no estimate of exposure. For this reason the multiple correlation coefficient of 0.59 for the regression model that utilized all 330 person days of NO<sub>2</sub> exposure was only slightly larger than the simple linear correlation coefficient of 0.57 between the personal and indoor home samplers. Hence if we are not able to control for the time spent away from home, the other types of sampling provide no better level of prediction of true exposure than indoor sampling alone.

The high predictive power of the multiple regression analyses between personal samp-



- Ott, W. and Flachsbart, P. (1982) "Measurement of CO concentration in indoor and outdoor locations using personal exposure monitors", *Environment International*, 8, 295-304.
- Ott, W. (1983-84) "Exposure estimates based on computer generated activity patterns", *Journal of Toxicology and Clinical Toxicology*, 21(42), 97-128.
- Pattle, R.E., Webb, C.O. and Burgess, F. (1979) "The acute toxic effects of black smoke", *British Journal of Industrial Medicine*, 16, 216.
- Quackenboss, J.J., Spengler, J.D., Kanavck, M.S., Letz, R. and Duffy, C.P. (1986) "Personal exposure to NO<sub>2</sub>: relationship to indoor/outdoor air quality and activity patterns", *Environmental Science and Technology*, 20(8), 775-783.
- Rodes, C.E., Kamens, R.M. and Wiener, R.W. (1991) "The significance and characteristics of the personal activity cloud on exposure assessment measurements for indoor contaminants", *Indoor Air*, 2, 123-145.
- Sega, K. and Fugas, M. (1982) "Personal exposure versus monitoring station data for respirable particles", *Atmospheric Environment*, 8, 259-263.
- Sexton, K., Spengler, J.D. and Treitman, R. (1984) "Personal exposure to respirable particles: a case study in Waterbury, Vermont", *Atmospheric Environment*, 18(7), 1385-1398.
- Shy, C., Hasselblad, V. and Burton, R.M. (1973) "Air pollution effect on ventilatory function of U.S. school children", *Archives of Environmental Health*, 27, 124.
- Silverman, F., Corey, P., Mintz, S., Olver, P. and Hosein, H.R. (1982) "A study of effects of ambient air pollution using personal samplers", *Environmental Monitoring and Assessment*, 2, 233.
- Spengler, J.D., Ferris, B.G., Jr., Dockery, D.W. and Speizer, F.E. (1979) "Sulphur dioxide and nitrogen dioxide levels inside and outside homes and the implications on health effects research", *Environmental Science and Technology*, 13, 1276.
- Spengler, J.D. and Soczek, M.L. (1984) "Evidence for improved ambient air quality and the need for personal exposure research", *Environmental Science and Technology*, 18(9), 268A-280A.
- Spengler, J.D., Ware, J., Speizer, F., Ferris, B., Dockery, D., Lebrecht, E. and Brunnekreef, B. (1987) "Harvard's indoor air quality respiratory health study". In: Seifert, B., Esdorn, H., Fischer, M., Ruden, H. and Wegner, J. (eds) *Proceedings of Indoor Air '87*, Berlin (West), Institute of Water, Soil and Air Hygiene, Vol. 2, pp. 218-223.
- Statistical Analysis System (1979) *SAS Users Guide*, Box 8000, Cary, North Carolina, SAS Institute.
- West, P.W. and Gaeke, G.C. (1956) "Fixation of sulphur dioxide as sulfitemercurate III and subsequent colorimetric determination", *Analytical Chemistry*, 28, 1816.