A RAND NOTE

MODELS FOR HUMAN EXPOSURE TO AIR POLLUTION

Naihua Duan

July 1982

N-1884-HHS/RC

Prepared for

The U.S. Department of Health and Human Services

Rand
SANTA MONICA, CA. 90406

The research reported herein was performed pursuant to Grant No. 016B80 from the U.S. Department of Health and Human Services, Washington, D.C. and in part by The Rand Corporation in accordance with its program of public service.

Published by The Rand Corporation

A RAND NOTE

MODELS FOR HUMAN EXPOSURE TO AIR POLLUTION

Naihua Duan

July 1982

N-1884-HHS/RC

Prepared for

The U.S. Department of Health and Human Services

Rand
SANTA MONICA, CA. 90406

PREFACE

This Note reviews and compares various models for human exposure to air pollution. The work was undertaken as part of the methodological research effort of the Rand Statistical Research and Consulting Group. As such, the research was supported in part by a grant from the U.S. Department of Health and Human Services, and in part by Rand corporate research funds.

This Note is also being published in a special issue of *Environment International*.

SUMMARY

Four models for human exposure to air pollution are discussed and compared. *Simple microenvironment monitoring* measures pollutant concentrations at fixed locations, regarded as proxies for similar locations or microenvironments. This model does not require pollutant measurements on the individual level, therefore is easy to implement. However, the model can be used only to estimate the average exposure in a population and does not provide any estimate of the variability and distribution of individual exposures. *Replicated microenvironment monitoring* provides some estimates of the variability and distribution. However, because of the possible discrepancy between the microenvironment concentration distribution and the individual concentration distribution, some adjustment might be necessary.

Integrated personal monitoring allows direct estimation of the average exposure as well as the variability and distribution of individual exposures. Coupled with the appropriate time budget data, a regression analysis can be applied to estimate the contribution from each microenvironment type. However, possible collinearity problems might result in low precision in those estimates. Moreover, it might be difficult to adjust for a possible Hawthorne effect.

Continuous personal monitoring has the advantage of recording exposure in each microenvironment type separately, allowing direct estimation of the average exposure as well as the variability and distribution of exposures in each microenvironment type. Moreover, it can also be conducted in conjunction with a two stage sampling scheme, using information from a large data

base on activity patterns, thereby making more efficient use of the monitoring data. It is also easier to adjust for a possible Hawthorne effect in this design.

CONTENTS

1. INTRODUCTION

In recent years, a substantial amount of field work has been done on human exposure to air pollution. This Note presents a theoretical comparison of several models that have been used or proposed in the literature.

2. MICROENVIRONMENTS AND MICROENVIRONMENT TYPES

During the daily activities of a human subject, he will likely pass through many different "locations" with different pollutant concentrations. We define a *microenvironment* to be a chunk of air space with homogeneous pollutant concentration. Examples of microenvironments can be "room 106," "room 107," "sidewalk along 1700 block of Main Street," etc. Since the pollutant concentration of the same air space might also vary over time, a more rigorous definition of microenvironment should be a four dimensional space × time concept, such as "sidewalk along 1700 block of Main Street during rush hour," "room 106 at night time," etc.

If we keep track of all microenvironments that a human subject passes through during his daily activity, we can represent his total (integrated) exposure (E) during the day (or any other fixed time period) as a linear combination of concentration in distinct microenvironments (γ_j, *microenvironment concentration*), weighted by the time (τ_j) the person spent in the corresponding microenvironment. The relationship can be expressed as the following formula:

$$E = \sum_j \gamma_j \tau_j \tag{2.1}$$

where the summation extends over all possible microenvironments.

Apparently there are far too many microenvironments for us to keep track of; therefore we have to group them, grouping all microenvironments of a similar nature together. For example, we might group all the indoor microenvironments together, and all the outdoor microenvironments together. We will refer to such groups of microenvironments as *microenvironment types*. Duan (1981, §8) derives a criterion for grouping microenvironments into microenvironment types.

With the microenvironment types, we can reformulate the earlier relationship (2.1) as follows:

$$E = \sum_k c_k \, t_k \tag{2.2}$$

where the individual's *time allocation* (t_k) to the k-th microenvironment type is the sum of times spent in microenvironments of the k-th type:

$$t_k = \sum_j \delta_{jk} \, \tau_j \tag{2.3}$$

$$\delta_{jk} = \begin{cases} 1 & \text{if j-th microenvironment belongs to type k} \\ 0 & \text{otherwise} \end{cases}$$

and the individual's average concentration (c_k, *individual concentration*) in the k-th microenvironment type is the average of the microenvironment concentrations γ_j in microenvironments of type k, weighted by the time spent in the corresponding microenvironment:

$$c_k = \sum_j \delta_{jk} \, w_{jk} \, \gamma_j \tag{2.4}$$

where the proportions

$$w_{jk} = \tau_j / t_k$$

are the individual's *visit frequencies*.

It should be noted that the microenvironment concentrations γ_j do not depend on the individual, whereas the individual concentrations c_k might vary from individual to individual because each individual might have his own unique visit frequencies.

With the microenvironment types representation (2.2), there are two possible approaches to study exposure. One is the *direct approach:* we are interested in exposure, therefore we simply measure exposure directly. The other is the *indirect approach*, in which we measure the time allocations and concentrations separately and then combine them to reconstruct exposures. It is the goal of this Note to study the various models based on these two approaches.

3. SIMPLE MICROENVIRONMENT MONITORING

Fugas (1975) proposed a model that follows the indirect approach and does not require the use of personal monitoring. For each microenvironment type, a fixed microenvironment is taken as representative of all microenvironments of the same type. Microenvironment concentrations (γ_j) are measured at these representative microenvironments and are taken as estimates for the individual concentrations (c_k):

(3.1) $\hat{c}_k = \gamma_j$, the j-th microenvironment is the representative of the k-th microenvironment type

While Fugas (1975) did not specify explicitly how the time allocations t_k were obtained, presumably we can use either a diary or appropriate recollection.

The individual's total exposure (E) can be estimated by the time weighted average:

$$\hat{E} = \sum_k \hat{c}_k \, t_k \tag{3.2}$$

For many goals, we might also be interested in the average exposure $\bar{E}$, averaged over a target population of interest, which can be estimated by the average-time weighted average:

$$\hat{\bar{E}} = \sum_k \hat{c}_k \, \bar{t}_k \tag{3.3}$$

where $\bar{t}_k$ is the average time allocation for the k-th microenvironment type averaged over the target population.

If representative microenvironments are chosen appropriately, (3.2) and (3.3) provide reasonable estimates of the individual exposure and the average exposure. For example, if the representative microenvironments are chosen to be unbiased for the corresponding microenvironment type, the estimates (3.2) and (3.3) are also unbiased if certain independence assumptions are met, namely, that the individual concentrations and time allocations are stochastically independent. If the independence assumption is not satisfied, the average total exposure can be expressed as

$$\begin{aligned}
\bar{E} &= \overline{\sum_k c_k \, t_k} \\
&= \sum_k \overline{c_k \, t_k} \\
&= \sum_k [\bar{c}_k \, \bar{t}_k + \mathrm{Cov}\,(c_k, t_k)] \\
&= \sum_k \bar{c}_k \, \bar{t}_k + \sum_k \mathrm{Cov}\,(c_k, t_k)
\end{aligned}$$

On the other hand, if $\hat{c}_k$ is an unbiased estimate for $\bar{c}_k$, the expectation of average-time weighted average (3.3) is

$$\sum_k \bar{c}_k \, \bar{t}_k \, ,$$

which differs from the average exposure E in that the covariance term

$$\sum_k \text{Cov}\ (c_k,\ t_k)$$

has been lost.

Simple microenvironment monitoring does not account for individual differences in individual concentrations c_k, therefore it cannot be used to estimate the variability and distribution of individual exposures. (The model does account for individual differences in time allocations t_k, and therefore can partially account for individual differences in exposure E.) While the microenvironments of the same type are similar in nature, the microenvironment concentrations γ_j in distinct microenvironments are likely to be somewhat different. Moreover, different individuals will likely have different visit frequencies w_{jk} to the various microenvironments of the same type. As a result, the individual concentrations c_k in the same microenvironment type are likely to be different for different individuals. However, simple microenvironment monitoring represents all microenvironments of the same type by the same representative microenvironment, which dictates that individual concentrations c_k in the same microenvironment type are estimated to be the *same* for different individuals.

While the average exposure is important for many research goals, individual difference is also important for many goals. For example, for many policy decisions, it is important to know the proportion of the population under exposure in excess of a certain threshold. Therefore we need also estimate the *distribution* of exposures in the target population of interest. Apparently the simple microenvironment monitoring model is insufficient for this goal.

4. REPLICATED MICROENVIRONMENT MONITORING

While simple microenvironment monitoring cannot be used to estimate the distributions of concentrations c_k and exposures E because only one microenvironment is chosen as representative of all microenvironments of the same type, a plausible modification is to use more than one representative microenvironment of the same type. For example, Flachsbart and Ott (1981) measured the carbon monoxide concentrations in a large number of microenvironments in commercial districts of several California cities and studied the distribution of microenvironment concentrations γ_j.

The author agrees with the implicit assumption in Flachsbart and Ott (1981) that distinct microenvironment types can be distinguished by examining the microenvironment concentrations γ_j. However, it should be noted that the distribution of microenvironment concentrations γ_j can be *different* from the distribution of individual concentrations c_k. Recall from section 2 that individual concentration c_k in the k-th microenvironment type is the weighted average

$$c_k = \sum_{j=1}^{J} w_j \gamma_j \tag{4.1}$$

where the summation extends over the J microenvironments of the k-th type, γ_j is the concentration in the j-th microenvironment, and w_j is the individual's visit frequency

$$w_j = \tau_j / t_k$$

where τ_j is the time the individual spent in the j-th microenvironment, and the t_k is the total time spent in microenvironments of the k-th type. Note that for simplicity of expressions, the notations are slightly different from (2.4).

There is one and only one special case in which the distribution of microenvironment concentrations (γ_j) and the distribution of individual concentrations (c_k) are the same. This special case is that each individual chooses to visit only one microenvironment of the k-th type, and each microenvironment has the same probability to be chosen. In this case the individual concentrations (c_k) are actually a random sample from the distribution of microenvironment concentrations (γ_j).

The conditions for this special case are rarely satisfied. First, people are likely to pass through more than one microenvironment of the same type during their daily activities. Second, even if they are restricted to only one microenvironment, it is unlikely that the microenvironments are chosen with equal probability.

When the conditions for the special case are not satisfied, the distribution of individual concentrations (c_k) is *different* from the distribution of microenvironment concentrations (γ_j). The following theorem shows that the individual concentrations (c_k) are usually *less variable* than the microenvironment concentrations (γ_j).

Theorem 1

Let $c_k = \sum_{j=1}^{J} w_j \gamma_j$ as in (4.1).

Assume (1) $E(\gamma_j) = \mu$

$\text{Var}(\gamma_j) = \tau^2$

$\text{Cov}(\gamma_j, \gamma_\ell) = 0 \qquad j \neq \ell$

$$(2)\quad E(w_j) = \omega_j$$

$$\mathrm{Var}(w_j) = \sigma_{jj}$$

$$\mathrm{Cov}(w_j, w_\ell) = \sigma_{j\ell} \qquad j \neq \ell$$

$$(3)\quad \mathrm{Cov}(\gamma_j, w_j) = 0$$

$$\mathrm{Cov}(\gamma_j, w_\ell) = 0 \qquad j \neq \ell$$

Then

$$(4.2)\qquad E\left[\mathrm{Var}(c_k|\gamma) - \frac{1}{J}\sum_j (\gamma_j - \bar{\gamma})^2\right] \leq 0$$

where equality holds if and only if

$$(4.3)\qquad w_j = \begin{cases} 1 & \text{with probability } 1/J \quad \text{for some } j \\ 0 & \text{with probability } (J-1)/J \quad \text{for all other } j\text{'s} \end{cases}$$

In theorem 1, we assume that the microenvironment concentrations γ_j are generated for each specific time period and are common to all individuals. The individual visit frequencies w_j are generated for each individual for specific time periods. For a fixed time period, the distribution of individual concentrations c_k is conditioned on the common microenvironment concentrations γ_j.

The first term in (4.2) is the variance of the individual concentrations c_k for a fixed realization of microenvironment concentrations γ_j. The second term is the population variance of the realized microenvironment concentrations γ_j. The inequality implies that on the average, the individual concentrations c_k are less variable than the microenvironment concentrations γ_j.

The conditions (4.3) are the conditions cited earlier for the "special case."

As a result of the theorem, knowledge about the visit frequencies w_j is necessary for estimating the distribution of individual concentrations c_k, as well as the distribution of exposures E.

If each individual visits only one microenvironment of a specific type in a specific time period, a possible remedy of the problem is to count or estimate the number n_j of people present at each microenvironment j sampled for microenvironment monitoring. The counts n_j can then be used as the weights for the corresponding microenvironment concentrations γ_j in the analysis. For example, instead of using the unweighted average and variance of γ_j, the weighted average and variance of γ_j (weighted by the counts n_j) can be used to estimate the average and variance of c_k. Similarly, the weighted sample distribution of γ_j can be used to estimate the distribution of c_k.

The problem is much harder if individuals visit more than one microenvironment of the same type. If continuous personal monitoring is conducted together with replicated microenvironment monitoring, it might be possible to use the directly measured individual concentrations c_k to adjust the microenvironment concentration distribution. However, the validity of such adjustment has to be tested in future field work.

5. INTEGRATED PERSONAL MONITORING

While the adjustment for visit frequencies in replicated microenvironment monitoring is a difficult problem, personal monitoring can be regarded as an automatic adjustment for visit frequencies. An appropriate sample of human subjects will sample the microenvironments according to their individual visit frequencies.

Recent field works have employed integrated personal dosimeters such as passive badge samplers to measure the total exposure E directly. If such direct measurements of exposure are taken on a sample of human subjects representative of the target population, we can estimate the average exposure, the variance of exposures, and the distribution of exposures in the target population by the corresponding sample average, sample variance, and sample distribution of the exposures measured in the sample. Furthermore, if a diary is used in conjunction with the integrated measurements to measure the time allocation t_k, Tosteson and Spengler (1981) used an ingenious regression analysis to estimate the average concentration in each microenvironment type:

$$E = \sum_k \bar{c}_k \, t_k + \varepsilon$$

The integrated exposures E are regressed on the time allocations t_k to estimate the unknown coefficients $\bar{c}_k$. (Even if our main interest is in the total exposure, appropriate estimation of the concentrations in each microenvironment type is very useful for identifying subpopulations at risk.)

While the regression analysis has the potential for estimating the average concentration in each microenvironment type without measuring it, it also has the built-in collinearity problem that the sum of explanatory variables (time allocations) is a constant, namely, the length of the total measurement period. Moreover, if we restrict to specific subpopulations such as urban workers, some of the explanatory variables might be highly correlated. For example, the amount of time spent at work is likely to be nearly a constant. Such collinearity problems can result in poor precision in the estimated regression coefficients.

A more crucial problem with integrated personal monitoring is that the inference about exposure based on direct measurements might not be efficient. We will discuss the efficiency comparison with indirect measurements in section 6.

Another crucial problem with the integrated personal monitoring model is that there might be a Hawthorne effect, namely, measuring an individual might change his behavior (Rossi et al., 1979). In the nine person pilot CO study (Ziskind et al., 1981), one of the subjects was overly cooperative in varying his daily activities: on one day, he biked to work; on another day, he took the bus. On a third day, he drove to work. Such activity pattern during a field study period might not represent a normal activity pattern, and as a result, the exposures during the study period might not represent normal exposures, either.

If the study period is short, i.e., exposure is measured on each individual for a few times only, there does not seem to be any way to test for the existence of the Hawthorne effect or to eliminate it. If the study period is long as is the case with the nine person study, it is possibly reasonable to assume that the Hawthorne effect is strongest during a short initial portion of the study period ("break-in period"), and will likely drop off after the subject is more accustomed to the measurement. Therefore a possible way to test for the existence of the Hawthorne effect is to divide the study period appropriately into two portions (we might want to choose the cut-off point empirically), and test for any possible difference in the activity patterns and exposures in the two periods. If a substantial amount of difference is found on some subjects, we might want to reject the data collected during the break-in period as biased and restrict

the final analysis to the data collected after the break-in period. However, a substantial amount of monitoring data has to be sacrificed with such adjustment.

6. CONTINUOUS PERSONAL MONITORING

Recent development in hardware (Ott, 1981) has made it feasible to conduct personal monitoring with continuous dosimeters and record the disaggregated measurements automatically on miniature memory chips. It is then possible to reconstruct the exposures from each microenvironment type either using a diary or a built-in timer.

Duan (1981) proposed that such a dosimeter be used in conjunction with a two stage sampling design, consisting of a diary stage and a monitoring stage. In the diary stage, extensive measurements of time allocation are taken on a large sample of subjects using an activity diary. In the monitoring stage, concentration measurements are taken on a subsample of subjects or a subsample of subject-days. For example, we can continue the diaries throughout the study period and sample a few days for each individual for concentration measurements. Duan (1981, §6) describes a convolution method to combine the two sets of measurements to estimate the distribution of exposures, as well as the average exposure and the variance of exposures. (The convolution method also assumes stochastic independence between the concentrations and time allocations.)

Compared with integrated personal monitoring, continuous personal monitoring (in conjunction with the two stage sampling scheme) can achieve better efficiency. The reason for the efficiency gain is that with extended diary measurements, we can reduce the variation in our estimates that is due to the variation in time allocation. A formula for the efficiency gain is given in Duan (1981, §7).

Further efficiency gain can be achieved by using an optimal sampling design in the monitoring stage. After some diary measurements have been taken, the subjects in the sample can be stratified according to their observed activity patterns. If some auxiliary information on the concentrations in the various microenvironment types like Flachsbart and Ott (1981) is available, we can estimate the variability of measurements in each strata. The sample design then should follow the general sampling theory (e.g., Cochran 1962) that the more variable strata should be over-sampled.

With the continuous personal monitoring model, we can adjust for any possible Hawthorne effect without loss in the monitoring data. For example, if we suspect that the activity patterns during the monitoring stage might not be representative of the subject's normal activity pattern, we can restrict the time allocation data to the diary stage and reject the time allocation data during the monitoring stage. (Therefore we use only the concentration data from the monitoring stage.) Furthermore, if we suspect that the activity patterns during the break-in period of the diary stage might be biased, we can further restrict to the time allocation data collected after the break-in period. While we lose some of the time allocation data, we do not lose any monitoring data. Furthermore, we can also test for the existence of the Hawthorne effect on either the monitoring stage or the diary stage.

Compared to the integrated personal monitoring, the treatment of the Hawthorne effect in this model is more efficient in that we retain the use of all monitoring data.

REFERENCES

COCHRAN, W.G., *Sampling Techniques*, J. Wiley and Sons, New York (1962).

DUAN, Naihua, "Microenvironment Types: A Model for Human Exposure to Air Pollution," *Technical Report No. 47*, SIMS, Department of Statistics, Stanford University, Stanford, CA (1981).

FLACHSBART, P., and W. OTT, "Field Surveys of Carbon Monoxide in Commercial Settings Using Personal Exposure Monitors," *Technical Report* (in preparation), U.S. Environmental Protection Agency, Washington, DC (1981).

FUGAŠ, Mirka, "Assessment of Total Exposure to an Air Pollutant," Proceedings of the International Conference on Environmental Sensing and Assessment, Paper No. 38-5, Vol. 2, *IEEE #75 - CH 1004-1 ICESA* (1975).

OTT, W., Private Communication (1981).

ROSSI, P., H. FREEMAN, and S. WRIGHT, *Evaluation of a Systematic Approach*, p. 176, SAGE, Beverly Hills, CA (1979).

TOSTESON, T.D., and J.D. SPENGLER, "Statistical Models for Personal Exposure Data," presented at Environmetrics 81, Alexandria, VA (1981).

ZISKIND, R.A., R. FITE, and T. MAGE, "Personal Exposure to Carbon Monoxide: Nine Person Study Data Analysis and Conclusions," presented at the International Conference on Indoor Air Pollution, Health and Energy Conservation, Amherst, MA (1981).