J. ENVIRON. SYS., Vol. 1(2), pp. 111-132, June, 1971

Probabilistic Models for Calculating Air Pollution Damage

SEYMOUR SCHWARTZ
Division of Environmental Studies
University of California, Davis

ABSTRACT

A decision process that weighs the real benefits of pollution-control—the reduced incidence of undesirable effects—against the costs of control requires the prediction of future damage. An alternative to a cost-benefit evaluation is the use of damage statistics which display the effects of pollution on different segments of the population. Probabilistic models are developed to yield both the expected value statistic for use in a cost-benefit evaluation and "percentage-frequency" relationships, which show the distribution of damage as well as the amount. The probabilistic response characteristic, which is the key to the transformation of air quality information to damage information, is introduced. Its construction is illustrated with actual eye irritation data. The submodels required for the complete damage calculation are identified and their interconnections shown. Time and geographical variations in pollutant concentration, as well as variability in human response to pollution, are accounted for in this model.

Introduction

Improvement in environmental management can be achieved on two levels: in the making of specific policy, and in the process of how decisions are made. The objective of this paper is to develop an idealized model for the calculation of the real benefits of air pollution control—the reduction in the incidence of damage. From such a model, improvement in both levels of decision-making is possible. For specific policy, the basis for a comprehensive evaluation of costs and reduced damage is provided. With realistic shortcuts and simplifications imposed by the existing lack of

doi: 10.2190/YHUT-NEGJ-D0RM-M4R7
http://baywood.com

information, a more complete assessment of costs and benefits than is currently performed can be undertaken. For the decision-process, the ultimate information requirements of the idealized model may serve as a basis for the planning of research, data-gathering, and modeling activities for increasing the quality of future decision-making.

An ideal model for predicting damage must be probabilistic to account for the time and geographical variation of pollutant concentration and the variability of human responses to pollution. Three critical submodels are required:

1. one which calculates (predicts) the probability density function of pollutant concentration from emission data and meteorologic statistics;
2. one which gives the incidence of an effect (response) at different concentrations;
3. one which performs statistical transformations to yield damage information.

The response model and the statistical transformations are treated in detail in this paper. It is assumed that suitable air quality models will be available for calculating pollutant concentration statistics.

The details of the response submodel depend upon the decision-models to be used in selecting from among alternative control policies. A cost-benefit approach is favored by many economists as an approximation to the rational choice model, which instructs us to invest in a project up to the point at which marginal benefits are equal to marginal costs.[1] The cost-benefit model requires the expected value of damage. However, under conditions of uncertainty, the decision-maker may wish to consider the variance of damage as well as the expected value.[2] More to the point, practical defects in the cost-benefit approach leads to the search for more realistic and flexible, though still quantitative, models.[3] Damage may be quantified in terms of incidence of occurrence and its distribution among different fractions of the population. Once the distribution of damage is of interest to the decision-maker, the representation of the response submodel by regression methods—the conventional approach[4]—becomes inadequate. Thus, the more general probabilistic response characteristic is introduced along with the statistical transformations that are needed to provide complete damage information.

Overview of the Model for the Calculation of Pollution Damage

A simplified flow diagram of the required steps in the calculation of pollution damage is shown in Figure 1. Construction of each of the

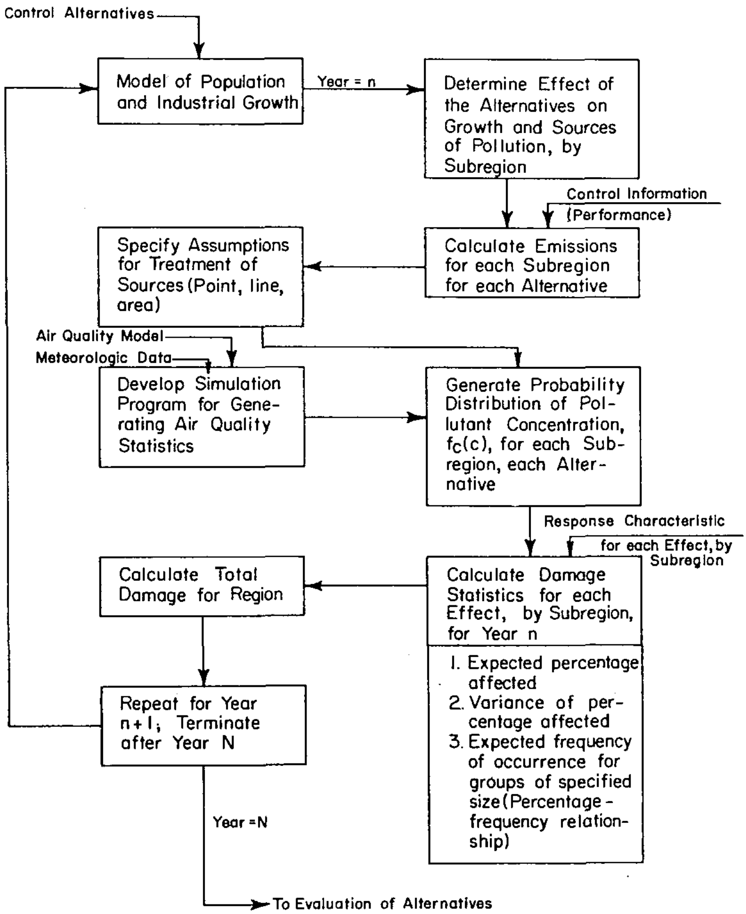

Figure 1. Flow Diagram of Damage Calculation

submodels will involve numerous tasks and decisions, which are not shown here. We proceed under the assumption that the information required to build each of the models is available or will be available. Since the required information includes knowledge of the behavior of complex physical and social systems, this is an optimistic assumption.*

First, the air quality region is divided into K subregions. The damage calculation is performed for each subregion. The selection of the number

* Detailed descriptions of the tasks and information requirements and an assessment of the prospects for building the various submodels are given in Reference 3.

and shape of the subregions depends on several factors, the most prominent of which is the cost versus resolution trade-off of running the air quality model.

The control alternative is an input to the growth model as industrial and vehicle growth will, in general, be a function of the type of control. Only a rough guess at the effect of the control policy on growth will be possible. The output of the growth model, in addition to population of each subregion, is industrial source type and location, and number of vehicular sources, suitably aggregated by subregion. The emission model converts the source information to a time profile of rate of pollutant emission, for each pollutant type, for each control alternative. The emission information is input to the air quality model, which yields as an output pollutant concentration at a receptor point in each subregion. It is assumed that either by direct use of the air quality model or by the construction of a transfer function, the probability distribution for pollutant concentration, $f_C(c)$ is generated for each subregion and pollutant. This distribution is assumed to apply to a one-year period, and is calculated from yearly statistics of meteorologic and climatologic parameters. As $f_C(c)$ is a function of source growth and the control program, it changes over time. For practical purposes, a given distribution may be accurate for a period of several years.

It is assumed, for generality, that response characteristics are available for each subarea, since they may differ from one subarea to another. Except in special cases, where an unusually high density of sensitive individuals live in a given subarea—e.g., in a retirement community—a single response characteristic may be assumed to apply.

Response Characteristics

The vital link in the damage model is the response characteristic, which transforms air quality to pollution damage—i.e., the number of people suffering from a particular effect of pollution during a given time period. It is a difficult transformation to obtain because: it applies to very inhomogenous populations; some effects are dependent upon the history of exposure to pollution of the population; and the specific biochemical and physiological mechanisms which cause many effects are unknown. Because of this complexity, a probabilistic view of the response characteristic is essential. It has not been adequately presented in the past, although various statistical attempts have been made to unravel the "ill-behaved" data. Nowhere has it been suggested that the response characteristic should consist of the conditional probability density (frequency) function of percentage of the population affected given pollutant concentration (for one or several pollutants).

THEORY AND DESCRIPTION OF RESPONSE CHARACTERISTICS

The response characteristic (or response curve) appears in the air pollution literature as a plot of the percentage of the population suffering from a particular effect versus pollutant concentration. Figure 2 is typical.[5]

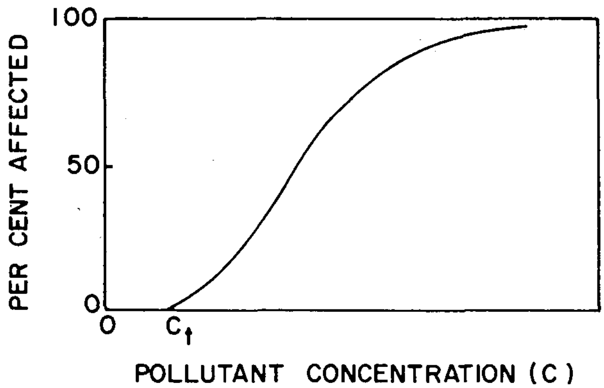

Figure 2. Typical Response Characteristic in Literature

This curve implies that either: a deterministic (cause-effect) relationship exists between the pollutant and the effect of interest; or, a point estimate of the percentage affected for every value of pollutant concentration should be used. If a deterministic relationship were to exist, then the percentage of a population suffering from an effect would indeed be the same each time a given concentration of pollutant is observed. Cassell,[6] Amdur,[7] and other investigators have stated the case against the possibility of isolating deterministic relationships between effect and cause in the real atmosphere. That is, the precise contribution of each pollutant to an effect cannot be determined because of "complicating factors": other, perhaps unknown, pollutants, temperature, humidity, wind, air ions, and the mental and physical condition of each individual. The percentage affected at a given pollutant concentration should, therefore, be viewed as a random variable.

The second possible implication of Figure 2 is that it is recognized that a deterministic relationship does not exist, and a point estimate is sought. Regression analysis provides a convenient means for fitting a smooth function to the data, thereby eliminating its variability.* The view that the variability of data points represents an error to be minimized is inappropriate in our situation. Here we have a complex combination of factors interacting to produce variations in response. The decision-maker may be most interested in the extreme portion of the distribution which describes the responses of particularly sensitive people and the effect of very high pollution days. These people and these days exist because of unusual combinations of circumstances. Their presence should not be obscured by regression, which yields a point estimate, if it is possible to describe response in terms of the conditional density function. From the latter, more information can be extracted, particularly, how often specified fractions of the population are affected. This is referred to as a *percentage-frequency* description of pollution effects.

DEFINITION OF PROBABILISTIC RESPONSE CHARACTERISTICS

The probabilistic response characteristic is defined as the set of conditional probability density functions $f(x|c)$ for all values of C, where x is the value of the random variable X—the percentage of the population affected or responding, and c is the value of the random variable C—the pollutant concentration at the receptor point of interest. (C may be used to represent dosage of pollution if the effect is dependent upon dosage.) For each discrete value of C ($C = c_j$), we determine the conditional probability function.**

$$f(x|c_j) = P(X = x|C = c_j)$$

GENERALITY

The above definition suggests a single pollutant, but it can be extended to any number of pollutants when necessary. The multidimensional response characteristic is then defined as

$$f(x|c_1, c_2, \ldots c_q, \ldots) = P(X = x|C_1 = c_{1j} \cap C_2 = c_{2j} \cap \ldots) \tag{1}$$

where the first subscript refers to pollutant type.

* This attitude is illustrated in Reference 5. Other early surveys and laboratory experiments averaged the eye irritation scores for each of the participating groups, thereby obscuring individual differences, e.g., Schuck.[8]

** The notation for random variables and probability functions follows closely that of Lindgren.[9] Lindgren indicates that it might be clearer to write the conditional distribution as $f_{X|Y} = y_j(x_i)$ rather than $f(x_i|y_j)$ but that the notation is too cumbersome. For the probability function of a single random variable, we exhibit the random variable, as $f_X(x)$. $f(x)$ is used for discrete distributions, whereas some authors prefer $p(x)$.

A further generalization is necessary for particulate pollution—particle size. Letting R be the random variable denoting particle size, we write the response characteristic as

$$f(x|c_1, c_2, \ldots, r) = P(X = x|C_1 = c_{1j} \cap \ldots \cap R = r) \tag{2}$$

CHOICE OF POLLUTANT

The choice of the appropriate pollutant or combination of pollutants for constructing the response characteristic depends upon prior knowledge of relationships between effects and specific pollutants. This knowledge may result from laboratory experiments in a controlled (usually one-pollutant) environment or from statistical analysis of effects in the ambient atmosphere. We seek high positive correlation between effect and pollutant in choosing the pollutants. It is evident from reviews of effects of pollution that adequate knowledge exists to begin the task of constructing response characteristics. The best quality information is available for sensory irritation, visibility reduction, and vegetation damage.

THE TIME DIMENSION

The time dimension appears implicitly in the response characteristics. We need to answer the questions: "What averaging time should be used for the measurement of pollutant concentration?" and "When should the population be sampled for response, with respect to the time of the pollution measurement—i.e., should there be any delay?" The answer to these questions lies in the nature of the physical mechanism. Schuck[8] shows that eye irritation occurs almost immediately after exposure to irradiated auto exhaust in a test chamber and reaches peak intensity within 10 minutes of exposure. Analysis of Los Angeles County Air Pollution Control District (LAAPCD) data for ambient air showed that the maximum percentage affected generally occurs within one hour of peak oxidant concentration. For eye irritation, a one-hour or shorter averaging time is used and there is no time delay. If an effect depends on dosage rather than short-term maximum concentration, an averaging time of 4, 8, or 24 hours may be needed. As a hypothetical example, we assume that pollution-caused headaches are due to cumulative exposure of 4 to 6 hours, on a smoggy day. We could use 4-hour average values of pollutant concentration or 4-hour dosage

$$\int_{t=0}^{t=4\text{ hours}} C(t)\, dt.$$

They differ only by a scale factor, and will give equivalent damage information if treated properly. McCarrol, et al.,[10] have investigated the possibility of effects being delayed in time for typical pollution patterns—not only for pollution "incidents." For delayed effect, the response characteristic will show the percentage affected at pollutant concentrations (or dosage) which were observed some time in the past.

Construction of a Response Characteristic

A sample response characteristic is constructed from eye irritation and oxidant concentration data obtained from the LAAPCD. Employees reported hourly on whether or not they were experiencing eye irritation. Oxidant concentration averaged over the previous hour was recorded each hour as was the daily peak concentration. Based on Schuck's results, peak oxidant and the maximum percentage affected during the day were used. Usually, the two peaks were within one hour of each other. Oxidant was chosen as the pollutant because of a high correlation with eye irritation.*

The sample consists of 231 points, representing those days between July 1, 1962, and June 30, 1963, on which a complete record was available. Oxidant concentration was divided into 12 categories of 0.05 ppm. each and the percentage reporting eye irritation was divided into 10 categories of 10 per cent each. Table 1 is a summary of the frequency of occurrence of each of the 120 combinations.

The response characteristic $f(x|c_j)$ is then the set of relative frequencies of each element (i,j) of the table, conditioned by the event—the occurrence of the j^{th} concentration category. The meaning of $f(x_i|c_j)$ is—the probability that the random variable X is in the i^{th} category of percentage affected, given that the random variable C is in the j^{th} concentration category. x_i and c_j can be assigned the midpoint values of their respective categories.

$$f(x_i|c_j) = P(X \text{ in } i^{th} \text{ category} | C \text{ in } j^{th} \text{ category})$$

$$= \frac{\dfrac{N(i,j)}{N_{total}}}{\dfrac{N(j)}{N_{total}}} = \frac{N(i,j)}{\sum_i N(i,j)}$$

where N(i,j) is the number of occurrences of the $(i,j)^{th}$ combination. The conditional probabilities are tabulated in Table 2.

* Based on results reported by the California State Department of Public Health.[5] A partial correlation coefficient of 0.89 is shown for eye irritation and oxidant (other variables held constant).

Table 1. Number of Days on Which Each Combination of Pollutant Concentration and Percentage Reporting Eye Irritation Occurred

Oxidant Concentration (C)	*Percentage Reporting Eye Irritation (x)*									
(ppm)	*0-10*	*11-20*	*21-31*	*31-40*	*41-50*	*51-60*	*61-70*	*71-80*	*81-90*	*91-100*
.01-.05	42	2								
.06-.10	41	6	2	2						
.11-.15	15	12	2	2	6	4				
.16-.20	4	6	6	6	16	4	4	1	2	0
.21-.25	1	1	0	3	5	1	2	4	4	1
.26-.30	1	0	1	2	0	0	2	4	3	1
.31-.35									1	2
.36-.40										2
.41-.45										
.46-.50										
.51-.55										1
.56-.60										

Source: Derived from LA APCD data.

For example, we consider the pollutant concentration category j = 2 ($0.06 \leqslant C \leqslant 0.10$). There are a total of 51 points in this category: $\sum_i N(i, 2) = 51$. For i = 1 ($0 \leqslant X \leqslant 0.10$), $N(1,2) = 41$ points. Therefore

$$f(x_1 | c_2) = \frac{41}{51} = 0.805$$

Similarly,

$$f(x_2 | c_2) = \frac{6}{51} = 0.118$$

$$f(x_3 | c_2) \text{ and } f(x_4 | c_2) = \frac{2}{51} = 0.038$$

These conditional probabilities define the response characteristic for the concentration range $0.06 \leqslant C \leqslant 0.10$. For eye irritation, the percentage affected refers to a single day, therefore x has units of 1/day. This must be kept in mind in calculating total annual damage.

Calculation of Damage Statistics

It is assumed that $f_C(c)$ has been calculated for each control alternative, for each year in the time horizon of the study; and, the probabilistic response characteristic, $f(x|c)$, or the point-estimate characteristic, $\hat{x}(c)$, is available for each effect of interest. Where the probabilistic response characteristic is available, we are able to calculate:

1. The expected percentage and number affected in each year
2. The variance of the number affected
3. The expected frequency of occurrence of an effect for groups of specified size
4. The percentage and number affected at different pollutant concentrations for a specified frequency of occurrence

EXPECTED PERCENTAGE AND NUMBER AFFECTED

For the calculation of benefits in a cost-benefit analysis, the expected value of the total number of occurrences of an effect, per year, is desired. The reduction in occurrences (damage) from one alternative to a null alternative (reference level) is taken as the net benefit of the first. To calculate total damage, the expected value of the percentage affected in each of the K subregions is calculated first. The expected value is first expressed in terms of the joint density function:

Table 2. Calculated Conditional Probabilities: $f(x_i|c_j)$

Oxidant Concentration (C) (ppm)	*Percentage Reporting Eye Irritation (x)*									
	0-10	*11-20*	*21-30*	*31-40*	*41-50*	*51-60*	*61-70*	*71-80*	*81-90*	*91-100*
.01-.05	.955	.045								
.06-.10	.805	.118	.038	.038						
.11-.15	.333	.267	.044	.044	.133	.089	.000	.000	.000	.022
.16-.20	.082	.122	.122	.122	.326	.082	.082	.020	.041	.000
.21-.25	.045	.045	.000	.136	.227	.045	.091	.182	.182	.045
.26-.30	.071	.000	.071	.143	.000	.000	.143	.286	.216	.071
.31-.35									.333	.667
.36-.40									.000	1.000
.41-.45										
.46-.50										
.51-.55										1.000
.56-.60										

$$E(X) = \sum_x \sum_c x f(x,c) \tag{4}$$

Substituting

$$f(x,c) = f(x|c) f_C(c)$$

$$E(X) = \sum_x \sum_c x f(x|c) f_C(c) \tag{5}$$

Since we have available the conditional probability function, i.e., the response characteristic, and the pollutant concentration distribution, the expected value can be calculated. Recasting the equation, we have*

$$E(X) = \sum_j \left[\sum_i x_i f(x_i|c_j)\right] f_C(c_j) \tag{6}$$

Letting $POP_{k,n}$ = population in region k in year n
$TOT_{l,m,n}$ = total damage for effect in year n for alternative m
and introducing subregion, effect, control alternative, and year subscripts on E(X), total damage for effect 1, year n, over the entire region is

$$TOT_{l,m,n} = \sum_{k=1}^{K} A_t E(X_{k,l,m,n}) POP_{k,n} \tag{7}$$

The multiplier A_t is 365 when X has dimensions 1/day, as for eye irritation and other short-term effects. $f_C(c)$ is obtained for each alternative, for each year, from the air quality model and should bear subscripts m,n.

For the second type of response characteristic—the point estimate, $\hat{x}(c)$—the expected value is given by

$$E(X) = \sum_j \hat{x}(c_j) f_C(c_j) \tag{8}$$

It is seen from Eqs. 6 and 8 that if $\hat{x}(c_j)$ is the conditional expectation $E(X|C = c)$, the results are identical. If $\hat{x}(c)$ is obtained in any other way, e.g., by regression analysis, they differ.

If only E(X) is desired, observed values of X may be averaged for each value of C, and used in Eq. 8 as $\hat{x}(c_j)$. That is, letting $x'_{ii}|c$ denote sample values of percentage affected at concentration c, we obtain $\hat{x}(c_j)$ as

$$x(c_j) = \frac{1}{NS} \sum_{ii=1}^{NS} x'_{ii}|c_j \qquad ii = 1,2,3,\ldots, NS \tag{9}$$

Thus, the conditional sample mean is used as an estimate of $E(X|C) = \sum_i x_i f(x_i|c_j)$—the conditional mean of the random variable X given C.

* Although the calculation of E(X) proceeds from this form, the right side can be expressed more compactly. The term in brackets is the conditional expectation of X: $E(X|C = c_j)$. Since this is a function of c, the outer summation is the expectation, over C, of the conditional expectation of X: $E(X) = E_C[E(X|C)]$.

ILLUSTRATIVE CALCULATION

$$E(X) = \sum_j [\sum_i x_i f(x_i|c_j)]\, f_C(c)$$

will be calculated from the data in Table 2, using midpoint values for x_i (e.g., 0.05, 0.15, . . . , 0.95). First $E(X|C = c_j) = \sum_i x_i|c_j)$ is calculated.

$$\begin{aligned} E(X|C = c_1) &= 0.05f(x_1|c_1) + 0.15f(x_2|c_1) + \ldots + 0.95f(x_{10}|c_1) \\ &= 0.05(.955) + 0.15(.045) + 0 \\ &= 0.055 \end{aligned}$$

$$\begin{aligned} E(X|C = c_2) &= 0.05f(x_1|c_2) + 0.15f(x_2|c_2) + \ldots + 0.95f(x_{10}|c_2) \\ &= 0.05(.805) + 0.15(.118) + 0.25(.038) + 0.35(.038) \\ &= 0.082 \end{aligned}$$

The remaining values of $E(X|C = c)$, along with $f_C(c)$ for the year in which the response data were recorded, is given in Table 3.

Thus, the expected value of the percentage affected, per day, is 27 per cent. The approximate total damage per year for Los Angeles County, assuming the same concentration distribution throughout the county,* is

$$TOT = 0.27(365)(6.8 \text{ million}) = 670 \text{ million man-days.}$$

The variance of the damage statistics should be calculated along with the expected value. Then, a decision based on the expected value and the variance is possible. For example, an alternative with a higher expected value of damage may be preferred to one with a lower expected value if the variance of the latter alternative is higher. The variance of the percentage affected may be calculated as:

$$Var(X) = \sum_x (x-\bar{x})^2 f_X(x) \tag{10}$$

where the marginal density function $f_X(x)$ is obtained from:

$$f_X(x) = \sum_c f(x|c) f_C(c)$$

and $\bar{x}$ is the expected value, E(X). If the point-estimate $\hat{x}(c_j)$ is used, the variance may be calculated as:

$$Var(X) = \sum_c [\hat{x}(c_j) - \bar{x}]^2 f_C(c) \tag{11}$$

* Since pollutant levels are higher in downtown Los Angeles than the average for the county, the calculated figure is somewhat high. It is used merely for illustration and for a rough idea of the magnitude of the eye-irritation effect. Population is for 1967 for the standard metropolitan statistical area.

Table 3. Calculation of E(X)

j	*Concentration Category*	$E(X \mid C = c_j) = \sum_i x_i f(x_i \mid c_j)$	$f_C(c)$	$[E(X \mid C = c_j)] f_C(c)$
1	0.01-0.05	0.055	0.190	0.0104
2	0.06-0.10	0.082	0.221	0.0181
3	0.11-0.15	0.213	0.195	0.0416
4	0.16-0.20	0.388	0.212	0.0824
5	0.21-0.25	0.577	0.095	0.0548
6	0.26-0.30	0.630	0.061	0.0384
7	0.31-0.35	0.907	0.013	0.0118
8	0.36-0.40	0.950	0.009	0.0085
9	0.41-0.45	–	0.000	–
10	0.46-0.50	–	0.000	–
11	0.51-0.55	0.950	0.004	0.0038
12	0.56-0.60	–	0.000	–
				$\sum_j = 0.270$

CALCULATION OF "PERCENTAGE-FREQUENCY" DAMAGE CURVES

A serious defect of the cost-benefit approach to decision-making is that it aggregates costs and benefits, but does not consider how they are distributed among individuals. The impact on particularly sensitive individuals could easily be overlooked. For example, if an average of 5% is affected per day over the year, it could mean that a sensitive group of 5% is affected every day of the year, or that 10% is affected on 180 days per year, or that 100% is affected on 18 days per year. The ability to calculate *percentage-frequency* curves, which show how many days per year specified percentages of the population suffer from an effect of pollution, enables the decision-maker to observe the existence of particularly sensitive groups and to determine their size. It also provides him with a sounder basis for establishing air quality goals and standards than now exists.

The specified percentage of the population, for which we calculate the number of days per year on which an effect occurs, is referred to as the *population of concern* and denoted by x_0. The number of days per year on which x_0 is affected is denoted by $N(x_0)$. Population of concern x_0 is affected when $X \geqslant x_0$ for $x_0 > 0$. We calculate the probability of this event as:

$$P(X \geqslant x_0) = \sum_{X=x_0}^{X=1.0} \sum_{\text{all } c} f_{X,C}(x,c) \tag{12}$$

Substituting for the joint distribution, we obtain

$$P(X \geqslant x_0) = \sum_{X=x_0}^{X=1.0} \sum_{\text{all } c} f(x|c)\ f_C(c) \tag{13}$$

This probability has a well-defined physical meaning in terms of frequency of occurrence: *it is the relative frequency of days on which x_0 is affected*. That is, if $P(X \geqslant x_0) = P_0$, we say that, in the long run, the group x_0 is affected $100P_0$ per cent of the time. Therefore, the number of days per year x_0 is affected is given by:

$$N(x_0) = 365P_0 = 365P(X \geqslant x_0) \tag{14}$$

which is obtained from Eq. 13. Thus, we have the theoretical basis for constructing percentage-frequency curves for each subregion. But, decisions cannot be made as if subregions were independent. The distribution of pollutant concentration in a subregion depends upon sources in other subregions and the meteorology of the region. Therefore, the percentage-frequency curve for the entire region is needed as a basis for a decision. It is obtained as follows: (See Figure 3.)

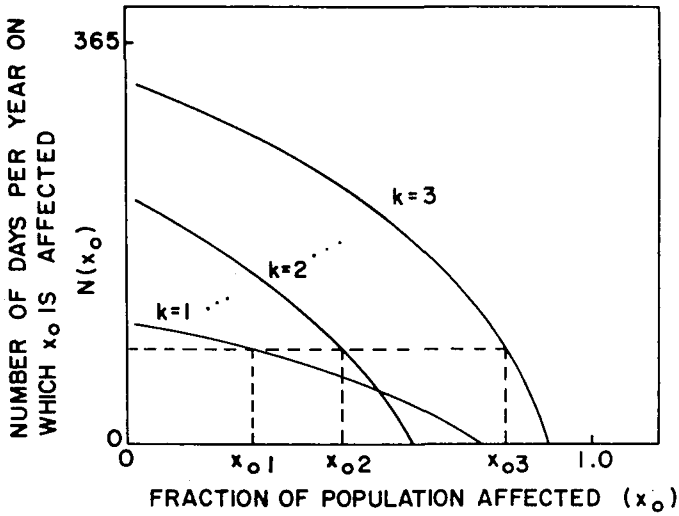

Figure 3: Use of Subregion Percentage-Frequency Curves to Construct Regional Curve

First, a value of N is specified, and the corresponding value of x is obtained for each of the K subregions (k = 1,2 . . . , K). The population affected in each subregion is then $x_{0k}POP_k$. The value of x_0 for the region, x_{0t}, for the specified N is:

$$x_{0t} = \frac{1}{POP_t} \sum_{k=1}^{K} x_{0k}POP_k \tag{15}$$

The calculation is repeated for several values of N to construct the regional curve.

Yet, another form of response information may be of value to the decision-maker, especially if a short-cut approach to goal- or standard-setting is desired. This is a display of the percentage affected vs. concentration for a specified frequency of occurrence. We start with the statement:

$$P(X \geqslant x|C = c) = \sum_{X=x}^{1.0} f(x|c) \tag{16}$$

If we set this probability of occurrence, which is the relative frequency fraction x of the population is affected at concentration c, we can obtain x. That is, we desire to determine the size of the group that is affected $100P_v$ per cent of the time at concentration c, where P_v denotes the specified relative frequency. We solve for x

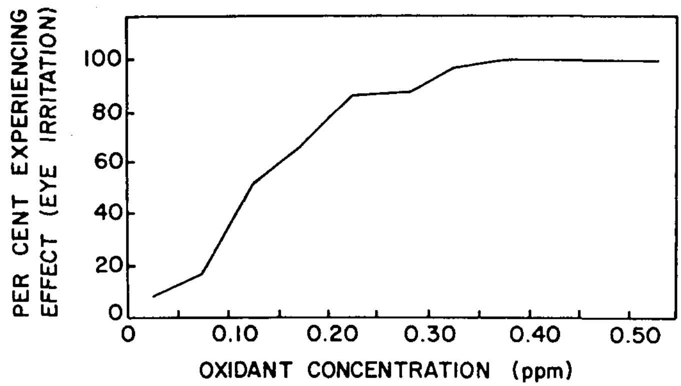

Figure 4: Response Curve for Effect Occurring 10 Per Cent of the Time

$$\sum_{X=x}^{1.0} f(x|c_j) = P_v, \text{ for all } c_j \tag{17}$$

For discrete data x is found by interpolation. To illustrate, the data of Table 2 is used to plot Figure 4 with $P_v = 0.1$. The curve shows that at an oxidant concentration of 0.15 ppm., about 60 per cent of the population experiences eye irritation on 10 per cent of the days per year. This curve is in the nature of a point-estimate response characteristic. It is derived only from response information and does not have the predictive capability of the percentage-frequency curves, which are functions of pollutant distribution, $f_C(c)$, as well as the response characteristic.

A Note on The Modeling Process

The practicability of constructing a probabilistic response characteristic depends upon several factors: prior knowledge of cause-effect relationships; data availability; cost of data acquisition and analysis; and, the potential impact of the information on the control decision. Where necessary conditions are not met, the less informative point estimate response characteristic can be obtained by regression methods or by other forms of data analysis. We examine the modeling process with a look at the breakdown mechanism and the role of statistical analysis. A simple block diagram (Figure 5) for a single effect helps to illustrate the process.

Pollution plus complicating personal and environmental variables are considered to be inputs to a population. Output is the number suffering from a particular effect in a given time period. The transformation or "black-box" is the breakdown mechanism that causes some individuals to

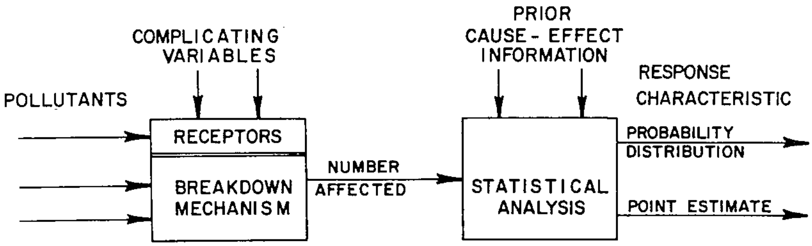

Figure 5: Construction of Response Characteristics

be affected by pollution. Although the breakdown (physical) mechanism may be known, that knowledge is not enough to predict the number affected. We must also know how the population varies in the level (exposure to pollution) at which breakdown occurs. In constructing the desired response characteristics, we consider the approaches of the epidemiologist and the toxicologist.

The epidemiologist is not primarily concerned with the nature of the transformation, i.e., the "why" of the breakdown mechanism and the variability among individuals. Given sufficient input and output data and some evidence that they are related, he can construct a probability density function (response characteristic). This approach appears to be promising for sensory irritation, and mild, short-term illness. Its use is implied in the prior discussion. For long-term effects the prospects are poor.

The toxicologist attempts to unravel the breakdown mechanism. For our task, however, he is severely limited in method and point of view. He deals mainly with a few individuals in a one- or two-pollutant laboratory setting. His methods are valuable for ascertaining what the effects of pollution may be, but they are inadequate for constructing response characteristics in a real setting.[7]

A synthesis of the approaches of the epidemiologist and the toxicologist holds the promise of better models. Recent work by Friedlander[11] illustrates how knowledge of (or hypotheses about) the breakdown mechanism is combined with assumptions about the variability of response of the population. Friedlander's model for mortality during high pollution periods is based on hypotheses regarding the effect of pollution in speeding up the normal gene-controlled aging process, which results in a fatal decline in activity of the respiratory system of some individuals. Here, statistical analysis is applied at the subsystem level, i.e., the respiratory function, to estimate the percentage of the population that will reach critical values under different durations and concentrations of exposure.

Although an attempt to model the breakdown mechanism promises more accurate results, the obstacles to unraveling the complex physical processes are formidable. Beyond our ability to develop breakdown models, the data requirements for long-term effects, such as emphysema and bronchitis, are massive. We may not be able to acquire pollution-exposure histories for a large enough number of individuals to construct probability density functions for the different exposure dosages. Even if we were able to construct the distributions, we cannot determine what role pollution played in causing these effects and, consequently, what fraction to attribute to pollution. The difficulty in disaggregating the role of pollution from other factors is a major shortcoming of the epidemiological (or input–output) approach to response models. Lave and Seskin[4] show that regression methods can be helpful in uncovering statistical relationships between effects and concentration of different pollutants (singly or in combination). However, considerable care must be exercised in the application and interpretation of regression analysis. Watt[12] summarizes:

> At best, this procedure will yield the researcher a rough idea as to which variables are accounting for a significant proportion of the variance in the system. At worst, blind application of multiple regression analysis leads to erroneous conclusions. A significant regression coefficient may not mean that a factor is important; rather, the factor may be highly correlated with some important factor that has not been included in the regression analysis. Furthermore, a factor may not appear to be important when in fact it is, because the wrong model may have been postulated.

Concluding Remarks

The damage model indicates what information is required to approach a rational-choice decision process, i.e., a complete evaluation of costs and benefits. The requirements are staggering. Much of the information will be costly to obtain and some may not be obtainable at all. Achievement of a rational ideal is clearly beyond reach for the short-term and must be doubted for the long-term. This does not mean that an assessment of costs and benefits is worthless. Where damage information is obtainable, it may clarify the choices for the decision-maker and the public. But, even if we had better damage models available we could not expect optimum decisions—there is still no magic answer to resolving differences in the value placed on reduced damage by different people.

We have stressed those aspects of the model that apply to sensory irritation and aesthetic insult in the belief that these effects motivate people to action. As Dales[13] points out, in his perceptive book, people who argue for reduced pollution because their senses and sensibilities are offended cannot be proved wrong by logical argument. When they seek morbidity and mortality data for their case, they encounter poor information and conflicting evidence. Opponents of control can argue that pollution does not *cause* illness or death, it merely predisposes, hastens, or complicates. Proof is difficult. We believe it is appropriate to rely more heavily upon sensory and aesthetic damage in decision-making. The potential exists for obtaining better quality information with a smaller investment of time and money as compared to morbidity and mortality information.

Further research in, and modeling of, pollution processes is needed to make an informed judgment as to the role complex models should play in the decision-process. We strongly warn against requiring, by law or administrative decree, the use of complex models or techniques in decision-making before they are fully tested. Such a warning may seem too obvious to merit statement, but we need only look to air quality management to see that it is not. An elaborate decision process was required by the Air Quality Act of 1967 and by subsequent administrative guidelines.[14] Clearcut methods for translating air quality criteria to ambient air standards to a plan for implementation—especially the last stage—were not available. The result has been great confusion and a delay of several years over what could have been achieved had a simpler strategy for implementation been adopted. The desire for a high degree of rationality and scientific support can create large costs by delaying action. The potential benefits of elaborate scientific support for a decision should be weighed against the costs of delay.

ACKNOWLEDGMENTS

I wish to thank Gerald A. Fleischer and Theodore C. Foin, Jr., for their valuable comments. I am grateful to the U.S. Public Health Service, National Air Pollution Control Administration, for the financial support which made this research possible (Fellowship No. 5 F3 AP 38,732002).

REFERENCES

1. Harold Wolozin, "Setting Criteria for Public Expenditures on Air Pollution Abatement: Theoretical Foundations and Limitations," in *The Economics of Air Pollution* ed. by Wolozin, H., W. W. Norton & Co., New York (1966) p. 168.
2. Gerald A. Fleischer, "Capital Allocation Theory," Appleton-Century-Crofts, New York (1969) pp. 156-157.

3. Seymour I. Schwartz, "Models for Decision-Making in Air Quality Management," Ph.D. dissertation, University of Southern California, Los Angeles, August 1970.
4. Lester B. Lave and Eugene P. Seskin, *Science*, **169**, 723-733 (August 21, 1970).
5. California State Department of Public Health, "Technical Report of California Standards for Ambient Air Quality and Motor Vehicle Exhaust," Berkeley, 1960.
6. Eric J. Cassell, *Law and Contemporary Problems*, **33**, 197-217 (1968).
7. Mary Amdur, *J. Air Pollution Control Assoc.*, **19**, 638-643 (1969).
8. E. A. Schuck, "Eye Irritation from Irradiated Auto Exhaust," Air Pollution Foundation, Los Angeles (1957).
9. B. W. Lindgren, "Statistical Theory," 2nd ed., Macmillan, New York (1968).
10. James McCarrol, et al., *Archives of Environmental Health*, **14** 178-184 (1967).
11. S. K. Friedlander, *Environmental Science and Technology*, 2, 1101-1108 (1968).
12. Kenneth E. F. Watt, "Ecology and Resource Management," McGraw-Hill, New York (1968), p. 241.
13. J. H. Dales, "Pollution, Property and Prices," University of Toronto Press, Toronto (1968). p. 103.
14. National Air Pollution Control Administration (NAPCA), "Guidelines for the Development of Air Quality Standards and Implementation Plans," Washington, D.C. (1969).