A METHOD FOR ASSESSING THE HEALTH RISKS ASSOCIATED WITH ALTERNATIVE AIR QUALITY STANDARDS FOR OZONE

Strategies and Air Standards Division
Office of Air Quality Planning and Standards
U.S. Environmental Protection Agency
Research Triangle Park, N.C. 27711

Table of Contents

1.0 Introduction

The National Ambient Air Quality Standard (NAAQS) for photochemical oxidants is being reviewed in conjunction with the reissuance of the criteria document.[1] This review affords an opportunity not only to examine the standard in the light of a more extensive data base, but also to develop a general method for setting NAAQS's.

Although the scientific data base is now more extensive than it was when the original criteria document was published, many uncertainties relevant to standard setting remain. These include uncertainties about the concentrations, exposure times, and patterns of exposure which contribute to each category of health effect associated with oxidants in general and ozone in particular. Due in large part to the unpredictable nature of meteorological conditions, there are also uncertainties about the maximum ambient ozone concentrations that will occur in a given period of time, whatever the precursor emission situation. Dealing with these uncertainties requires setting a standard with an adequate margin of safety. The method described in this report provides a framework and suggests a quantitative approach to accomplish this end.

The National Academy of Sciences has recommended that EPA make use of some of the principles and techniques developed in the discipline of decision analysis which are helpful to rational decision-making under uncertainity.[2,3] The method discussed below incorporates some of these principles and techniques. For example, the technique of "probability encoding", which enables optimal use of the quantitative judgments of health experts, plays an important role. The decision analysis principle of reducing complex judgments to smaller,more manageable subjudgments whose logical implications can be determined mathematically is employed.

The kernel of the suggested approach is a systematic assessment of the risks associated with alternative standards in a carefully defined sense. Unfortunately, there are secondary uncertainties about how to represent the primary uncertainties which give rise to the risks. Yet, an ambient air quality standard specifies a precise averaging time, level, and expected number of exceedances of the level. Another principle of decision analysis and the spirit of yet another are applied in dealing with this situation. First, the method attempts to define terms precisely so that uncertainty about what a quantity means is not added to the inherent uncertainty about its value. Second, the output of the method clearly displays how a calculated risk varies with the particular choice made from a reasonably comprehensive set of representations of the primary uncertainties which give rise to the risk. In other words the "softness" of the risk calculations which results from the secondary uncertainties is dealt with directly in such a way as to give the decision-maker(s) a conception of its degree.

2.0 Underlying Principles of the Method

2.1 Risk and Margins of Safety

2.1.1 Legislative Guidance

Guidance for setting a primary NAAQS is given by the Clean Air Act in the following passage:

> National primary ambient air quality standards . . . shall be ambient air quality standards the attainment and maintenance of which in the judgment of the Administrator, based on [air quality] criteria and allowing an adequate margin of safety, are requisite to protect the public health. (4)

A fundamental motivation for the development of a method for assessing the health risks associated with possible primary air quality standards is the premise that in order for EPA to make the most meaningful judgment on whether a possible standard provides an adequate margin of safety it needs as clear a conception of the risks associated with the possible standard as it is feasible to get at the given time. This important premise is supported in Appendix A. The point that in general safety can only meaningfully be interpreted in terms of risk is made in reference (5).

The provision in the Act for an adequate margin of safety and the following clarification of the intent of the Act given in its Legislative History make clear the sense in which the term 'risk' must be interpreted to capture the intent of the Act:

> ... In setting such air quality standards the Secretary should consider and incorporate not only the results of research summarized in air quality criteria documents, but also the need for margins of safety. Margins of safety are essential to any health-related environmental standards if a reasonable degree of protection is to be provided against hazards which research has not yet identified.
> ... the Committee emphasizes that included among those persons whose health should be protected by the ambient standard are particularly sensitive citizens such as bronchial asthmatics and emphysematics who in the normal course of daily activity are exposed to the ambient

> environment. In establishing an ambient standard necessary to protect the health of these persons, reference should be made to a representative sample of persons comprising the sensitive group rather than a single person in such a group.
>
> Ambient air quality is sufficient to protect the health of such persons whenever there is an absence of adverse effect on the health of a statistically related sample of persons in sensitive groups from exposure to the ambient air. An ambient air quality standard, therefore, should be the maximum permissible ambient air level of an air pollution agent or class of such agents (related to a period of time) which will protect the health of any group of the population.
>
> For purposes of this description a statistically related sample is the number of persons necessary to test in order to detect a deviation in the health of any person within such sensitive group which is attributable to the condition of the ambient air. (6)

The passage explicitly states that the intent of the Act is to protect the most susceptible group in the general population, implicitly assumes there is an adverse health effects threshold concentration of any NAAQS pollutant, and implicitly acknowledges that the threshold concentration will be unknown. If there were no uncertainty the objective would be to set standards at the maximum level (related to a period of time) for which peak pollutant concentrations would not exceed the health effect threshold when the standard is met. Since there is uncertainty, standards must provide a margin of safety. Even if the assumption that there is an adverse health effects threshold concentration is true, there is no positive concentration which the threshold is known to be above. Hence, standards cannot be set so that there is no risk that peak pollutant concentrations will exceed the health effect threshold when the standard is met.

In order to make a meaningful judgment on whether a possible standard provides an adequate margin of safety, a conception is needed of the threshold risk associated with the possible standard. The threshold risk associated with

a possible standard is the risk that ambient concentrations of the pollutant will exceed the health effects threshold concentration for the most sensitive group in the general population when air quality just achieves that standard. If the threshold risk associated with the possible standard is deemed to be acceptable in view of the circumstances then that standard is judged in a meaningful way to allow an adequate margin of safety.

The part of the passage quoted above from the Legislative History of the Act which says, "Margins of Safety are essential ... if a reasonable degree of protection is to be provided against hazards which research has not yet identified," is ambiguous. 'Hazards which research has not yet identified', could refer to: (1) types of health effects of the pollutant that research has not yet identified which are caused by lower concentrations than cause known or suspected health effects of the pollutant; or (2) lower concentrations of the pollutant than have been shown by research to date to contribute to health effects that research has shown are effects of the pollutant at higher concentrations; or (3) both (1) and (2).

There is a way to assess the threshold risk associated with a standard for health effects on which there is evidence, but there is no way to assess the risk for health effects on which there is no evidence. Hence, the approach taken here is to assess the risk for effects on which there is evidence. It is suggested that the primary concern in standard setting should be with effects for which there is evidence. Truisms are the most that can be said about effects for which there is no evidence: there may not be any; if there are any their threshold concentration may be either higher or lower than the threshold concentration of effects for which there is evidence; and, the lower the risk of effects for which there is evidence, the lower the risk of effects for which there is no evidence.

Decision-makers may wish to keep these truisms in mind when deciding what constitutes an acceptable risk for those effects for which there is evidence.

2.1.2 Health Effect Threshold

It is obviously important to give the health effect threshold concept a precise definition. Uncertainty about what the concept means in general would hinder attempts to deal with the uncertainty about particular health effect threshold concentrations. Defining the health effect threshold concept precisely must be distinguished from measuring particular health effect threshold concentrations accurately. The latter cannot be done, but the former can. The health effect threshold concept is a very useful one, despite the inherent limitations in the accuracy with which particular threshold concentrations can be determined.

One task involved in giving the concept of a health effects threshold a precise definition is defining precisely what is to be regarded as a health effect. A health effect threshold concentration must be distinguished from a physiological response threshold concentration. It is well known that one molecule of a pollutant can react biochemically with the human body. However, not every physiological effect need be classified as a health effect. Effects such as a disease or increased susceptibility to a disease are clearly health effects, but there are other effects whose classification is not clear.

There is no precise and general technical definition of the health effect concept which is in accord with common usage where common usage is clear, but also guides classification of the unclear cases as health effects or non-health effects. This is because there are inherent difficulties in trying to draw the line in a non-arbitrary and general way between physiological changes which

are health effects and those which are not. Many physiological responses are on a continuum from trivial to obviously adverse effects with no obvious point at which the one leaves off and the other begins. In such cases, pragmatic judgments have to be made about how much of the given response is to be considered a health effect; i.e., how much of the given response to protect against.

A second aspect of defining the health effects threshold concept precisely requires distinguishing the health effects threshold concentration for a group from the health effects threshold concentration for a person. People within a group, even a sensitive group, will vary in their sensitivity. One person can be said to be more sensitive than another if he has a lower health effect threshold concentration. From a strictly theoretical point of view a natural definition of the health effects threshold concentration for a group would be that concentration which is the health effect threshold concentration for the most sensitive member of the group. For practical reasons addressed in section 2.2, a different definition has been adopted in the initial application of this method, even though one reading of the above passage from the Legislative History of the Clean Air Act is that this natural definition is the one intended. The definition used is: the health effect threshold for the most sensitive group is that concentration which is the health effect threshold concentration for the least sensitive member of the most sensitive 1% of the most sensitive group.

A third aspect to making the concept of the health effect threshold for the most sensitive group precise is the stipulation of the conditions of exposure, such as: the concentrations of co-pollutants with which the given pollutant may be additive in causing an effect; the concentration - time patterns the exposure may have; and, the state of the subjects with regard to possible stresses such as exercise. This aspect will also be addressed in Section 2.2.

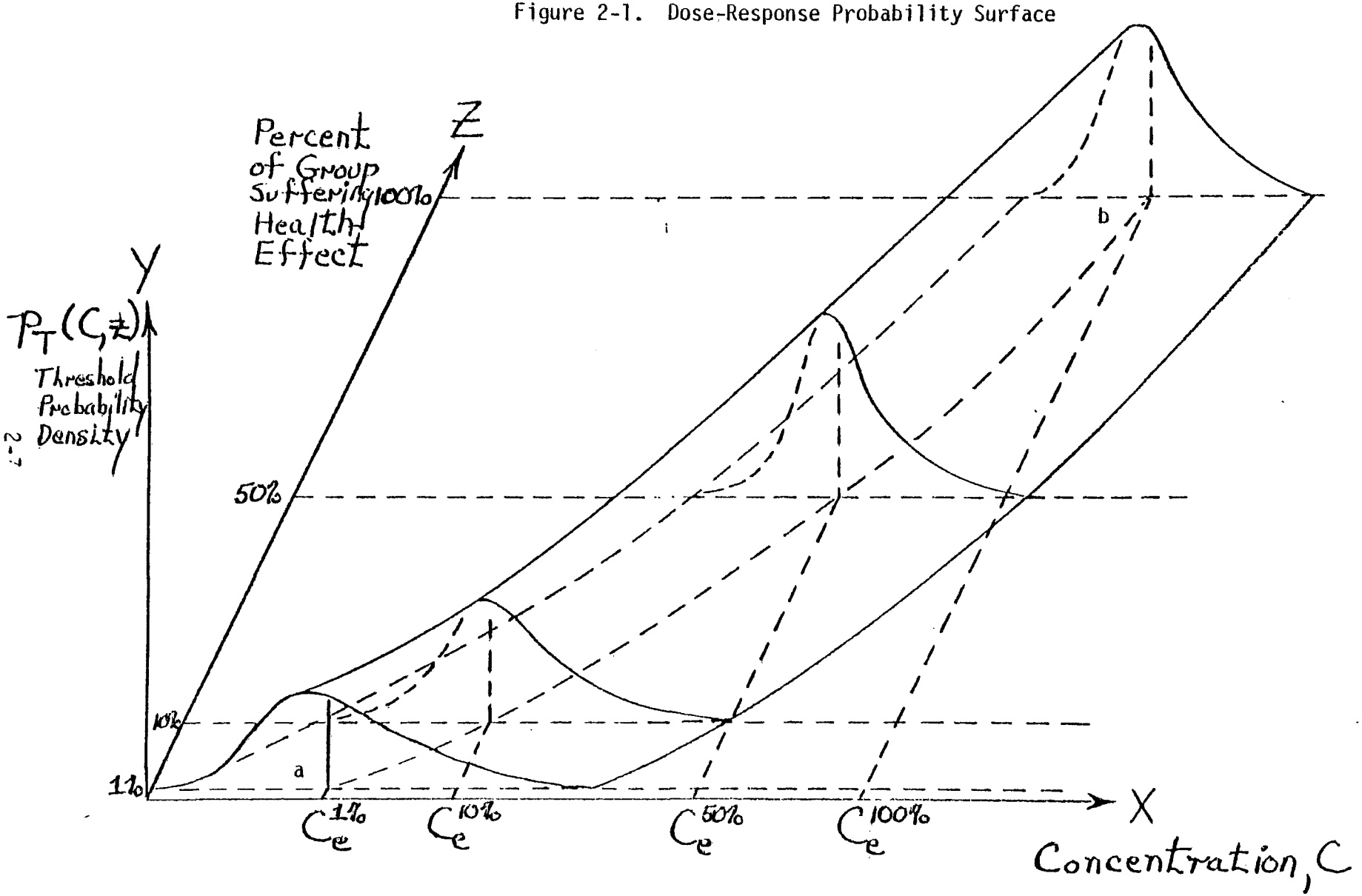
2.1.3 Idealized ~~Risk Surfaces~~ Dose-Response Probability Surfaces

Suppose the conditions of exposure are uniform throughout a hypothetical geographical region of concern, and constant over time; suppose for all persons in the region under the conditions of exposure which obtain, the amount of physiological response contributed to by pollutant X is a non-decreasing function of its maximum concentration; suppose all necessary judgments have been made on what is to count as a health effect; and, suppose a most sensitive group has been identified, all of whose members will suffer a health effect due to the presence of X if its concentration is high enough. Then, if pollutant X reaches a certain concentration it will contribute to a health effect in 1 percent of the people in the most sensitive group; if it reaches a certain higher concentration it will contribute to a health effect in 10 percent of the people in the most sensitive group; and so on.

Suppose that this somewhat idealized situation is realistic in the sense that the relationship between the peak concentration and the percentage of the most susceptible group affected is unknown. Suppose that the risk surface depicted in Figure 2-1 represents this uncertainty. Then, if the health effect threshold for the most sensitive group is defined to be that concentration which is the health effect threshold concentration for the least sensitive member of the most sensitive 1 percent of the most sensitive group, the intersection of the plane which is perpendicular to the XZ - plane and to the Z-axis at the 1 percent point with the risk surface gives a probability density distribution which represents the uncertainty as to what concentration is the health effect threshold for the most sensitive group.

In Figure 2-1, line ab in the XZ-plane is the set of medians of the probability distributions obtained by intersecting the risk surface with planes perpendicular to the XZ-plane and to the Z-axis at the full range of points on the Z-axis. Concentration $C_e^{1\%}$ is the expected concentration at

Figure 2-1. Dose-Response Probability Surface



which 1 percent of the group would be affected, $C_e^{10\%}$ is t
tration at which 10 percent of the group would be affected, and so on.

2.1.4 "Risk" Nomenclature

In the public health literature where dose-response relationships have been estimated without constructing a risk surface, concentrations corresponding to $C_e^{1\%}$ have been said to have a risk of 1 percent. "Risk" in this sense is an estimate of the percentage of a group which will suffer a health effect; this percentage is called risk because it is interpreted to be the probability that a generic member of the group will suffer a health effect. In the context at hand, it is better terminology to refer to the risk that n percent of the members of the group will suffer a health effect at various concentrations. These risks can be determined from the n% probability distribution, where the n% probability distribution is the distribution which is the intersection of the risk surface with the plane that is perpendicular to the XZ-plane and to the Z-axis at the n% point.

Obviously, a risk surface such as the one shown in Figure 2-1 would be a very valuable input to a real-world attempt to determine the expected concentration at which a certain percentage of a group would suffer health effects, the expected percentage of people suffering health effects at a certain concentration (slice the surface perpendicular to the x-axis), the risk that a certain percentage of the group would suffer health effects at a certain concentration, and so on. However, the idealized nature of the situation described above must be kept in mind. Conditions of exposure are not uniform throughout an area and are not constant over time: pollutant mixes vary, concentration-time patterns vary, some people are sitting, some people are walking, some people are jogging, etc. Different people react differently to different concentration-time patterns. How many people are being exposed varies over time. Within a given control area concentrations of a pollutant vary at any given time.

In view of the above complexities, the initial application of this method (to ozone) concentrates on determining the risk that the health effects threshold for the most sensitive group would be exceeded. This is a risk that some sensitive people would suffer adverse effects. This approach, as well as seeming to capture the sense of the Clean Air Act, avoids many formidable problems the above complexities pose for attempts to make various types of "head count" estimates such as those mentioned in the preceding paragraph. Of course, these estimates give a more complete risk picture, and so the capability to make meaningful estimates of this type is being developed.

2.1.5 Basic Model

In estimating the risk that a health effects threshold will be exceeded it is necessary to take into account, in addition to the uncertainties concerning the location of health effect thresholds, another major source of uncertainty. This is the uncertainty concerning the maximum concentration of the pollutant which will occur in the time period over which the risk is to be estimated and when the general air quality just meets the NAAQS for the pollutant. This uncertainty comes about because the ambient concentrations of a pollutant are subject to the random changes in meteorological conditions and in the emissions of the pollutant or its precursors into the atmosphere. As a result, even though pollutant or precursor levels in an area have been brought to a general level at which the overall ambient air quality is meeting the NAAQS, the highest concentration levels observed over the area in a given time period (e.g. 1, 2, or 5 years) will vary over a succession of time periods. The uncertainty in the maximum concentration can be accounted for by a probability distribution which can be estimated from aerometric data.

Figure 2-2 broadly represents the situation that must be dealt with. It shows two probability distributions in the form of probability density functions. The curve drawn in the upward direction is the probability density function for a health effects threshold. The curve drawn in the downward direction is the probability density function for the maximum hourly average ozone

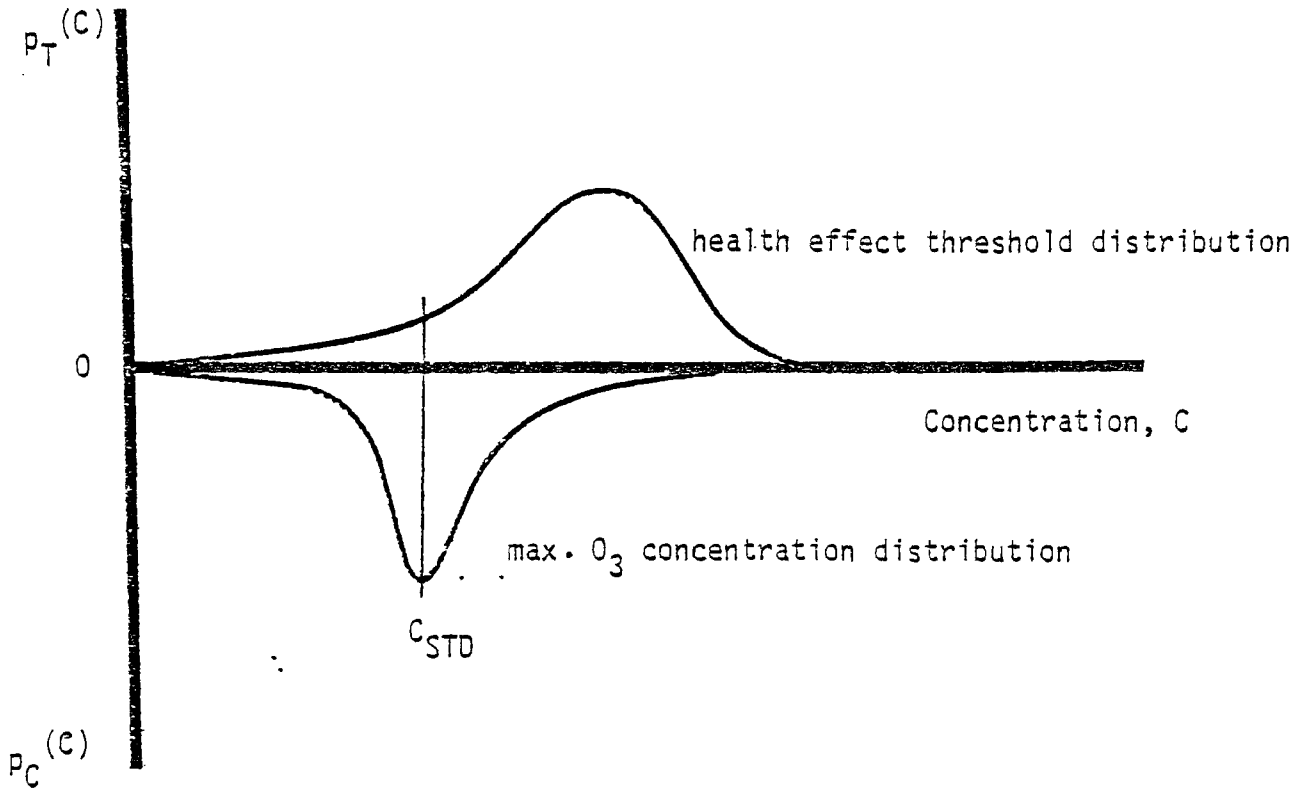


Figure 2-2. Probability Distributions on the Two Major Sources of Uncertainty

concentration observed over some time period. This distribution is drawn in the downward direction so that both distributions can be conveniently displayed on the same axis. The location on the concentration axis of the distribution representing the uncertainty about the maximum ozone value depends on which potential standard is under consideration. The more stringent the potential standard, the farther it is to the left, that is, shifted toward lower concentrations. The farther it is to the left, the lower the risk that the maximum ozone value will exceed the overall health effect threshold concentration.

To better take into account the statistical behavior of ambient pollutant concentrations, the proposed standard for ozone has been given the following statistical form.

> C_{STD} ppm hourly average concentration with an expected number of exceedances per year less than or equal to E.

Through the use of an expected exceedance rate an area is allowed to occasionally experience a measured exceedance rate of greater than E so long as the expected rate, estimated by averaging the rate over a number of years, is not exceeded.

The lower the numbers C_{STD} and E the more stringent the standard and the more the lower curve in Figure 2-2 will be displaced to the left. Raising one of these two numbers and lowering the other sets up a tradeoff in which the particular numbers and the particular context determine whether the standard is more or less stringent. For any C_{STD} or E a standard of any desired stringency can be obtained by making the other number high or low enough. However, for reasons which will be discussed more fully in a later section it is desirable to maintain the value of E in the vicinity of one or lower.

It should be clear that the standard level and the health effect threshold concentration are not (in general) the same. Hence, a standard which allows one exceedance of a standard level may have a very low threshold risk associated with it. The standard, the standard level, and the health effect threshold are all distinct concepts.

The shape of the maximum ozone concentration distribution is a function of the time period considered. For example, a one-year distribution and a three-year distribution for the same standard will have different shapes. Hence, the risk is a function of the time period; the longer the period the greater the risk. It is required by law that the scientific criteria for NAAQS's be reviewed and reissued every five years. Hence, NAAQS decisions apply to at least a five-year period. At the end of five years when the new criteria are reissued the risks can be reassessed. So, the time period chosen over which to assess the risks is five years.

2.2 Uncertainty Concerning Health Effect Threshold Concentrations

An important part of assessing threshold risk is representing the uncertainty about the health effect threshold concentration for the most sensitive group. Representing this uncertainty requires subjective judgments that go beyond strict scientific interpretation. However, these judgments are best made by members of the set of health scientists who are most familiar with the health science information reviewed and assessed in the criteria document. For, although some of the judgments required go beyond strict scientific interpretation, they can best be made on the basis of the scientific information presented in the criteria document, and by those with the expertise to understand the scientific implications of that information.

2.2.1 Subjective Probability

Since an integral step in arriving at the desired representation of the uncertainty about the health effect threshold concentration for the most sensitive group is the elicitation of subjective probability distributions for individual categories of health effects, the required subjective judgments are made in probability encoding sessions. "Probability encoding" is a term used in the management science and decision analysis literature to refer to the elicitation of subjective probability distributions from experts. It is an explicit, precise, and formal technique for quantifying the probability judgments of experts. The technique has been pioneered by Stanford Research Institute (SRI), amongst others; two SRI-related publications describe the technique and the history of its use. (7) (8)

Several distinctions regarding "subjective" and "objective" probability need to be understood. The first distinction is between:

(a) the method by which a probability distribution is derived; and

(b) the decision to accept a probability distribution as the representation of uncertainty about the value of a quantity for some practical purpose.

In the sense of (a), probability distributions are either objective or subjective; that is, either the procedure by which a probability distribution is derived from a body of information is completely determinate, so that two different people following the procedure exactly will arrive at the same distribution, or probability judgments are required in going from the information to the distribution, so that in general different probability distributions will result when different people make the judgments.

In the sense of (b), all probability distributions are subjective. The decision to use a particular probability distribution to represent the uncertainty about the value of a quantity is a subjective decision, no matter whether the distribution is derived objectively or subjectively.* The uncertainty about the health effect threshold concentration for a pollutant could be represented by any number of objectively derived distributions about which any health scientist well informed on the topic would presumably make the (subjective) judgment that he could contribute a subjectively derived probability distribution which would represent the uncertainty better. In some situations, a health expert might feel that a particular objectively derived distribution represents the uncertainty best. However, as will be seen, the situation is generally so complex that this will generally not be the case.

Certainly the situation is too complex for the type of objectively derived probability distributions which have historically played a large role in statistical inferences in the empirical sciences. Another distinction needed here is between subjective and objective interpretations

*Hence, Quinn and Matheson suggest the term "judgmental probability" be used rather than the term "subjective probability". (8)

of probability. The mathematical relationships between various probability statements, distributions, etc. can be treated as part of an abstract axiomatic system in which the terms remain uninterpreted. But an important question concerns what these statements mean. Two basic interpretations of probability statements dominate the history of the probability concept. (9) One is the subjective (or epistemic) interpretation in which a probability is interpreted as a measure of degree of belief, or as the quantified judgment of a particular individual. Degree of belief consists of a disposition to make certain specific kinds of choices in well-defined choice situations (10). The second interpretation is "the frequentist (or aleatory) interpretation according to which the probability of an event is the relative frequency of occurrence of that event to be expected in the long run (11).

The relative frequency interpretation has been the foundation of sampling - theory (or classical) statistical inference in the empirical sciences. Often in the empirical sciences experiments can be designed such that random samples can be taken from a well-defined population or process. If, by agreement of those in a position to know, such an experiment satisfies certain conditions, a probability calculated on the basis of a sample relative frequency can be the only rational probability to assign to an event. Such a probability is not only objective in its derivation, but also gains general acceptance because of the compelling nature of its empirical content. Such general agreement brought about by empirical content is the trademark of good empirical science.

In such situations, subjectivity has been reduced to a minimum. But, there is still the subjective element of agreeing that the relevant conditions obtain, and the subjective interpretation of what the probability means applies.

Since the sample relative frequency is the only rational probability to assign the event, it is the one assigned. Within the confines of such contexts whether the probability is given the relative frequency or the subjective interpretation makes little practical difference. However, the concept of probability has always applied to a much wider range of uncertain situations than those that can be investigated directly and empirically through random sampling. Many of these situations are like the situation at hand; a probability or probability distribution based on the available information is needed for the most rational approach to decision-making, but it cannot be determined on the basis of scientific data alone what probability or probability distribution best represents the available knowledge and the remaining uncertainty. The subjective interpretation of probability can be extended to these situations and the relative frequency interpretation cannot. Hence, the subjective interpretation of probability is the one that is used in decision analysis.

The subjective interpretation of probability has come to be identified in some minds with the Bayesian approach to statistical inference, since in that approach sample information in the form of a likelihood ratio is often combined with prior information in the form of subjective probability distributions. But, in the Bayesian approach, which is an extension of the classical approach, enough sample information will "swamp" the subjective prior distribution in formally arriving at the posterior distribution. So, the only constraint to "objectifying" the posterior distribution to any degree desired is a cost constraint on the amount of sample information it is rational to obtain under the circumstances. In the situation at hand it is not a resource constraint that leads to the use of subjective probability distributions, but rather an inherent inability to get sample information which is a <u>direct</u> input to the distribution in question.

Of course, either the Bayesian or the classical approach to statistical inference can be used in fitting models to data outside the range of interest in order to extrapolate from the range in which there is data to the range of interest (12) (13). Such models can be used by an expert as an aid in arriving at the needed distribution. But, unless there is wide agreement that the model holds within the range of interest, it should not be felt that a distribution arrived at using such a model is necessarily a better representation of the uncertainty than a distribution arrived at without using the model.

To sum up, the representation of the uncertainty about the health effect threshold concentration cannot be determined by solely scientific means. The goals of empirical science and the goals of practical decision-making must be distinguished. For excellent reasons, whereof empirical science cannot speak it remains silent. The goal of practical decision-making is to use mathematics and science as far as they go, but if there remains a gap to adopt whatever means seem at the time to be best for bridging the gap. The question is not <u>whether</u> subjective judgments will be made, but rather <u>how</u> and <u>by whom</u>. It is argued here that these judgments are best made explicitly, under well-defined choice situations, by the best-informed health scientists; rather than implicitly, under ill-defined circumstances, and/or by non-experts.

2.2.2 Independence of Health Effects

The criteria document for the typical pollutant discusses several health effects contributed to by the pollutant. A risk surface similar to the one in Figure 2-1 of section 2.1 could be constructed for each of these effects. However, if the risk of exceeding a health effect threshold is made the focus of the risk assessment, several simplifications are possible. First, the health effects can be grouped into independent categories according to the following definition of independence: If for two health effects the judgments of an expert, about the probability that the threshold for either effect is below various concentrations, would be affected by knowing the true health

effect threshold concentration for the other, then for that expert the two health effects are threshold dependent; otherwise the two health effects are threshold independent. Threshold-dependent health effects are grouped into the same health effect category.

One point of grouping the health effects into threshold-independent categories is to enable the health experts to concentrate on one category of health effects at a time in making their judgments.* Once subjective probability distributions are obtained for each individual health effect category, a composite health effect threshold distribution can be determined mathematically from the individual distributions (see section 3.0). The mathematical laws of probability used to obtain the composite distribution can only be applied to independent distributions.

Although the grouping of the health effects into threshold-independent categories reduces the number of effects that must be addressed by the health expert at one time, he still may have to address several effects within a given category. Just as in making his judgments about threshold independence the expert focuses on whether knowledge about the threshold concentration for one effect would affect his judgments about the threshold concentration for another, and does not concern himself with whether he believes there is any relationship between the two effects at higher concentrations, so he focuses on the effect within the category that he believes has the lowest threshold concentration. One reason two given effects may be in the same category is that the expert is sure or thinks it likely that one of the effects occurs at a lower concentration than the other.

*See subsection 2.2.6 for two other reasons.

2.2.3 Responses of Concern

Once it has been determined how many independent health effect categories an expert judges there to be for which much or most of the evidence is from his field of expertise (such as epidemiology), there are several types of important judgments he makes for each category. First, he describes what he considers the health effect to be that has the lowest threshold concentration for the category. In those cases where the effect is a continuum of physiological response, he judges how much response should be regarded as a health effect; EPA representatives involved in the probability encoding session clarify as much as possible for the expert the sense of the Act on what is to be regarded as a serious enough response to be considered an adverse health effect. But, since neither the Act nor its Legislative History is explicit on this point, much is left to the expert's judgment. Since experts naturally differ some in their judgments in cases where drawing the line involves some inevitable arbitrariness, the definition of a health effect in such cases usually varies some among experts. So, differences in subjective probability distributions on the health effect threshold concentration for such categories can be due both to differences in judgment about what concentration (for the averaging time) will cause what response and how much response should be considered an adverse effect on health. Once one or more applications of the method have been made in which various health experts have addressed the question of how much response of a given type is an adverse health effect, the option of designating a given amount of response as the amount EPA regards as a health effect and wants to protect against may become more attractive.

2.2.4 Sensitive Population

Another judgment the expert makes for each category he addresses is what group, characterizable in general terms, is the most sensitive group for that category. The most sensitive group for a category is the most

sensitive group for the health effect of that category which is judged to have the lowest threshold concentration. Since experts may differ on what group they judge to be the most sensitive group, the number of people in the most sensitive group may differ for different experts. This fact means that in specifying the health effect threshold concentration for a group it may be that either the "target individual", whose threshold concentration is used to define the threshold concentration for the group (above in section 2.1.2, the least sensitive member of the most sensitive 1 percent of the most sensitive group), will be different for the two experts, or the number of people more sensitive than the target individual will be different.

2.2.5 Seriousness of Effect

There are two aspects to the risk of exceeding a health effect threshold: the probability of exceeding the threshold and the seriousness of the adverse effect. Determining the probability is the more complex task, but giving EPA decision-makers and interested parties as good a conception as feasible of the seriousness of the effect is also very important. Everything else being equal, less risk is acceptable for serious effects than for effects which are not serious. Hence, each expert is asked to describe how serious he believes the given effect to be and to say anything he feels would help clarify for a decision-maker how much concern there should be about the effect.

2.2.6 Uncertainty About Causality

For some categories of health effects there is no uncertainty about the existence of a causal relationship between the pollutant and the category of health effect. For others, the existence of a causal relationship is uncertain. This uncertainty often arises in cases where there is toxicological evidence from animal studies that the pollutant causes the

effect in one or more species of animals, but there is no conclusive evidence from clinical and/or epidemiological studies that a causal relationship exists between the pollutant and the effect in humans. In those cases where the existence of a causal relationship for a category is in doubt, experts addressing that category are asked to judge the probability that a causal relationship exists between the pollutant and the effect in humans.

Distributions which represent the expert's uncertainty about the health effect threshold concentration if there exists a causal relationship between the pollutant and the effect in humans are encoded for such categories in the same way as the distributions for categories about which there is no doubt about the existence of such a causal relationship. The existence probability is used to apportion the encoded distribution between the range of concentrations the expert believes contains the threshold concentration if a causal relationship exists and a range of concentrations much higher than would ever be found in the ambient atmosphere. The latter portion must be maintained somewhere on the concentration axis for mathematical reasons (see Appendix B); its particular shape does not affect the risk calculations. The fraction of the distribution in the lower concentration range is equal to the existence probability. The part of the distribution in this range has the same shape as the encoded distribution. Figure 2-3 depicts a probability density function for a health effect whose threshold is thought to be in the vicinity of 0.20 ppm, but which has only a 20 percent probability of being real.

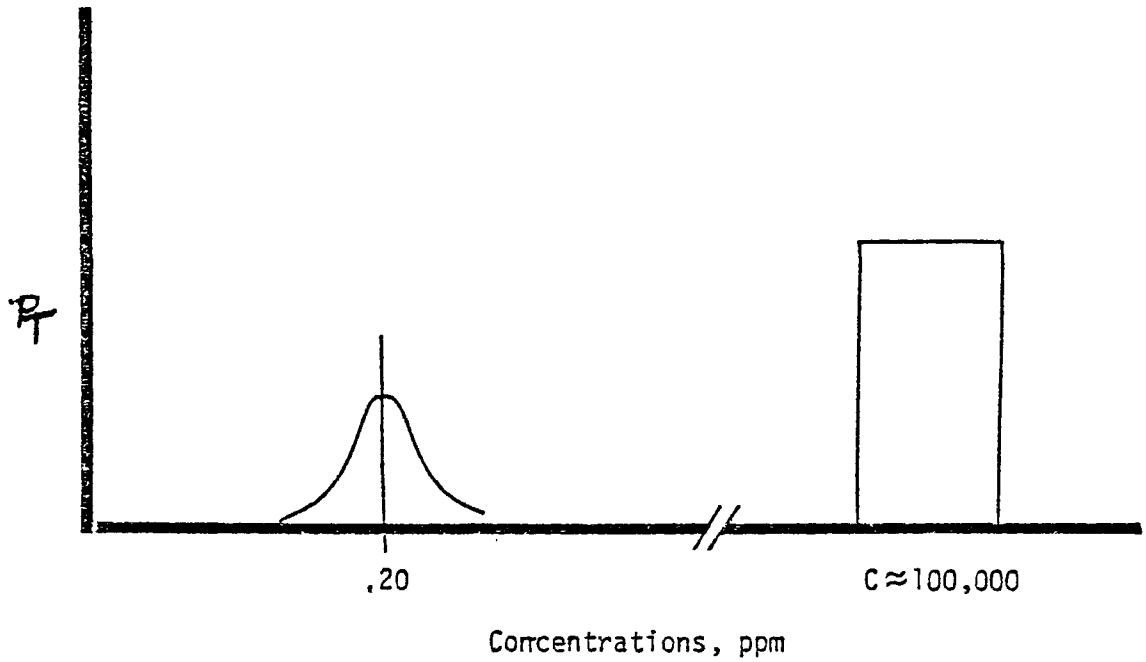


Figure 2-3 Hypothetical Threshold Probability Density Function for a Health Effect with a 20% Chance of Being Real

Having the expert address the question of the probability of existence separate from the question of the most probable concentration of the health effect threshold if a causal relationship exists is a second simplification of the judgments which he has to make.* There can be a tendency when such a breakdown is not made to either assume a causal relationship exists in risk calculations if the existence probability is thought to be high, or neglect the possible existence of a causal relationship in risk calculations if the existence probability is not thought to be high. Unless the existence probability is negligibly different from 1 or 0, this tendency amounts to the mistake of using outcomes rather than probabilities of outcomes in calculating probabilities (risks). The formal incorporation of an existence probability into the mathematics used to make the risk calculations, as is done in Appendix B, avoids this mistake.

The third reason for grouping the health effects into threshold-independent categories is that the probability that at least one of several unknown and independently distributed health effect thresholds is less than a given concentration is greater than the probability that any selected one of the thresholds is less than the given concentration. To put the point another way, the composite health effect threshold distribution calculated from the threshold distributions for single health effect categories will always be displaced toward lower concentrations from all of the single category distributions. For any given concentration the total area under the composite density curve to the left of that concentration will be greater than the total area under any one of the individual density curves to the left of that concentration.

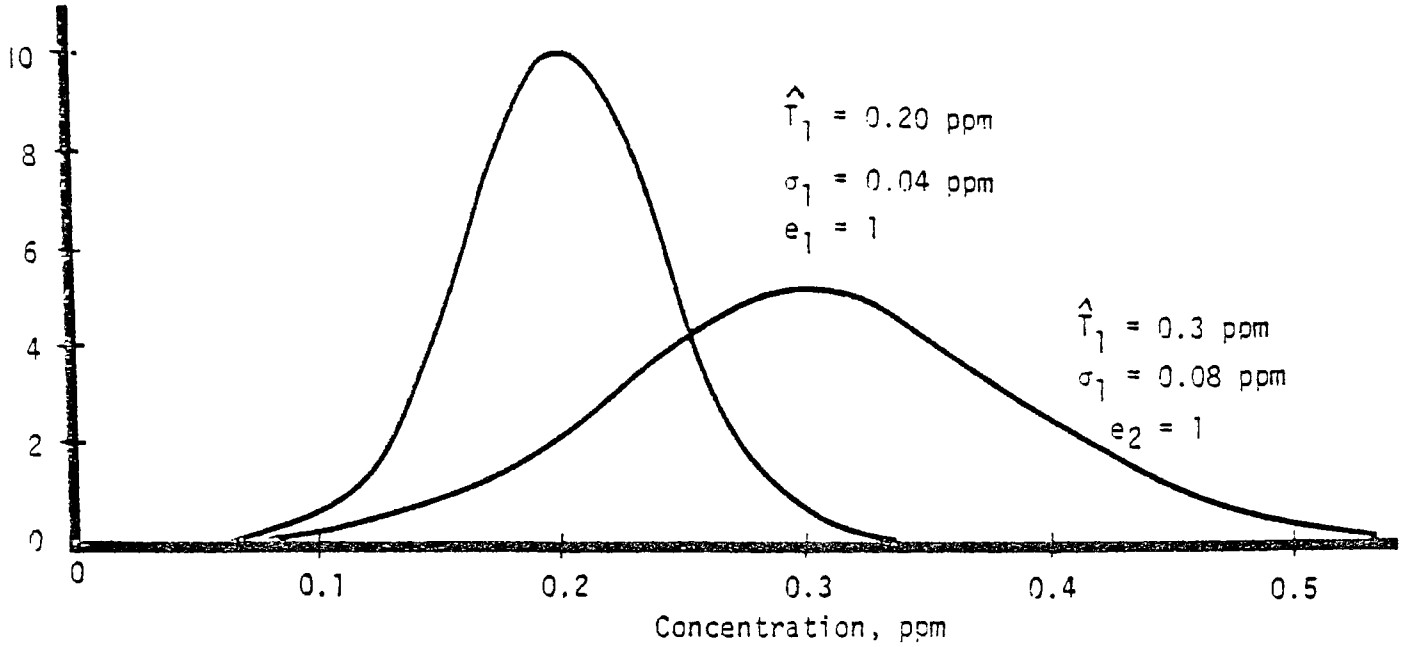
The leftmost individual health effect distribution will in general have the most influence on the composite distribution. The one exception to this generalization can be a case in which the leftmost distribution is a distribution for a health effect category about which the existence of a causal

*See p. 2-17.

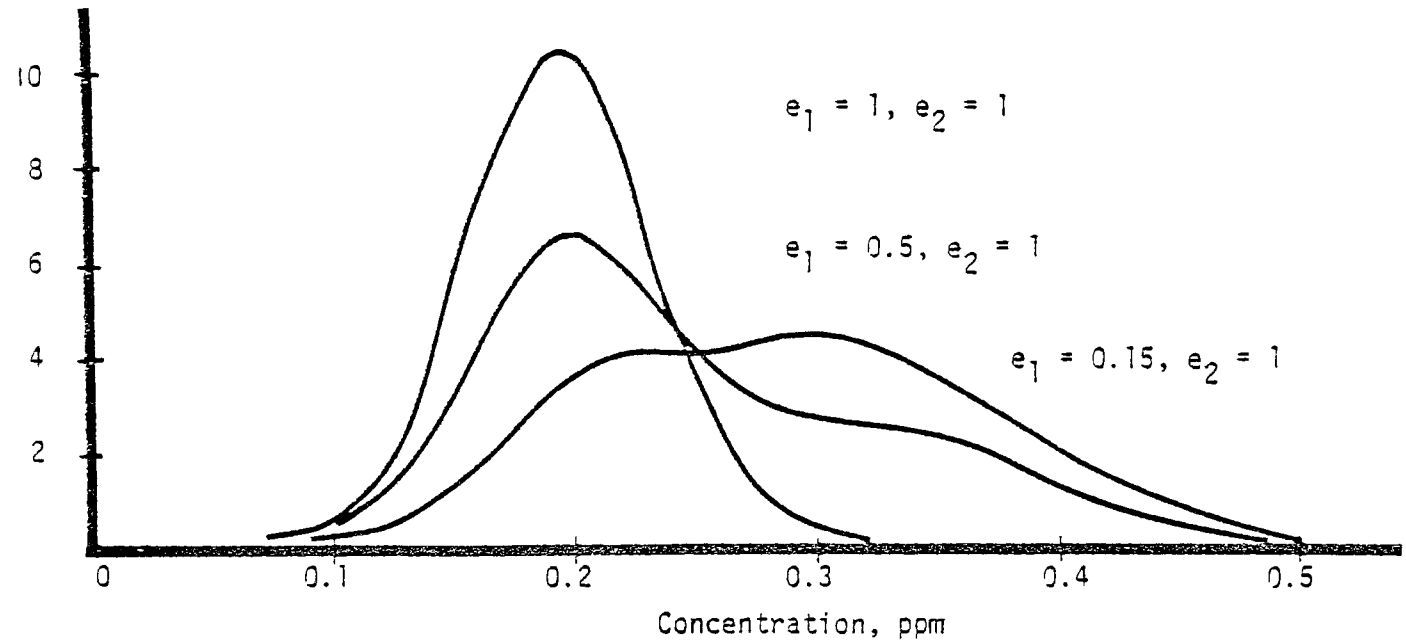
relationship between the pollutant and it is uncertain. Figure 2-4 illustrates this point. Figure 2-4(b) shows three examples of composite functions derived from two individual health effect threshold probability density functions (Figure 2-4(a)). The two hypothetical individual functions used in the example are both normal distributions. As can be seen in section 4.0, real subjective probability distributions will tend to be neither symmetrical nor a well-known distribution.

2.2.7 Defining the Encoding Variable

The probability encoding technique is designed to minimize the motivational and cognitive biases which psychologists have found can arise in the elicitation of subjective probability distributions. Care can be taken to guard against two possible "motivational" biases: the building of "safety margins" into judgments, and the favoring of what has been scientifically demonstrated over what is viewed to be most likely in view of the current evidence. Judgments about the degree to which a health effect threshold concentration has been scientifically demonstrated to be below certain concentrations could also be elicited in the probability encoding sessions. In fact, these would be useful judgments that would best be elicited in such a session where the nature of the judgments could be carefully defined; belief functions (14) on degrees of confirmation (15) could be elicited. But, whether this is done or not, such judgments should not be confused with the probability judgments made in the encoding of a subjective probability distribution. (16) It is the latter that must be used to calculate risk.



(a) Individual Health Effect Threshold Probability Density Functions



(b) Composite Probability Density Functions for Various Probabilities of Existence are Assigned to Function with $\hat{T}$ = 0.2 ppm

Figure 2-4 Variation in Composite Health Effect Threshold Probability Density Functions of Two Independent Density Functions as Different Probabilities of Existence are Assigned to the Lower Threshold Function

Defining well the uncertain quantity about which probability judgments are elicited is a normal part of a probability encoding session. As was discussed in section 2.1, there are three aspects to defining the health effects threshold for the most sensitive group precisely:

(1) definition of what is to be regarded as a health effect;

(2) stipulation by his place in the distribution of sensitivities within the most sensitive group the member of the group whose threshold concentration is to serve as the threshold concentration for the group;

(3) stipulation of the conditions of exposure.

In the initial application of the method (to ozone) the definition of a health effect for each expert was based on his own best judgment of how much response should be regarded as a health effect. The expert's judgments concerned the concentration of ozone that would cause the effect in 1 percent of the group he judged to be most sensitive for that effect <u>if the whole group were exposed under the stipulated conditions of exposure</u>. The selection of the 1 percent figure unavoidably involves <u>some</u> arbitrariness, but was selected for several reasons. To use the health effect threshold of the most sensitive group rather than the threshold of the least sensitive member of the most sensitive 1 percent of the most sensitive group would be impractical for at least one reason. For the types of effects contributed to by ozone, the most sensitive member of the most sensitive group is an unknown type of person who would be extremely difficult for the health expert to make judgments about. Subjective judgments grade from well-informed judgments to sheer guesses; judgments about the most sensitive member of the most sensitive group would be

at the latter end of the scale. The health expert is applying his own expertise to make judgments on the basis of the scientific information in the criteria document. While judgments about what dose of a pollutant would affect a small percentage of the people in the most sensitive group under certain conditions on the basis of the data in a criteria document are very difficult and extrapolative, they are not as uninformed as would be judgments about the most sensitive person.

As stated previously, the legislative history of the act indicates that EPA is to protect the most sensitive group as a whole, rather than the most sensitive person. The approach chosen satisfies this intent, for the threshold risk determined on the basis of this definition of the health effect threshold can reasonably be interpreted as the risk that the pollutant will affect the group as a whole. It would be incorrect to interpret this definition as an indication of a utilitarian judgment to trade off the interests of 1 percent of a sensitive group against the interests of society as a whole. Threshold risk so defined is not the risk that 1 percent of the most sensitive group will be affected by the pollutant. How many people would be affected if the ambient air level of a pollutant exceeds the health effects threshold concentration as it is defined here is unknown; the point is that it would affect some people.

By redefining the threshold in terms of 0.005 or 0.05 of the most sensitive group we can say that roughly half as many or five times as many people will be affected. So the somewhat arbitrary choice of 1 percent is not unimportant. But this kind of choice must be made. The threshold risk can be calculated for more than one definition to give the decision-maker an idea of how much it varies with the definition chosen.

The decision-makers can also keep in mind that the threshold risk for individuals more sensitive than the least sensitive member of the most sensitive 1 percent of the most sensitive group, including the most sensitive member of the most sensitive group, is greater than the threshold risk for the most sensitive group. The question of whether the health effect threshold concentration for the most sensitive member of the most sensitive group is greater than zero is left unanswered and unaddressed. In general, for any low concentration the lower the probability that the health effect threshold concentration of the most sensitive group is less than the given concentration, the lower the (higher) probability that the health effect threshold for the most sensitive member of the most sensitive group is less than the given concentration. A relationship could be developed to extrapolate from the most sensitive group threshold probability (which is itself extrapolated from the available scientific data) to a probability for the most sensitive member of the most sensitive group.

The approach taken in stipulating the conditions of exposure is to have the expert make his judgments for an idealized situation in which the conditions of exposure are assumed to be the same for the whole sensitive group and the whole group is assumed to be exposed. One condition of exposure is that it is in the ambient air of an "average" United States city. If the expert's best judgment is that there are additive effects from other pollutants in the ambient air he is instructed to take them into account. He is to assume, however, that the ambient concentration of other NAAQS pollu-

tants is their standard level. This stipulation is to avoid double-counting, yet deal with the fact that some NAAQS pollutants may be additive with one another in causing some health effects. Non-NAAQS pollutants are to be assumed to be at levels found in an average U.S. city.

Another condition of exposure for short-term effects concerns the concentration-time patterns the exposure may have. If one exposure can cause a health effect, the expert is instructed to make his judgments for the following exposure pattern: the peak concentration lasts the length of the averaging time with a "typical" build-up and drop-off in concentration before and after the peak concentration. If possible, the averaging time is selected before the elicitation of subjective probability distributions from experts.

The experts are instructed to consider the members of the most sensitive group to be under any normally occurring stress such as exercise; mild exercise such as tennis or jogging is interpreted to be normally occurring, but heroic exercise such as marathon running is not. To the extent that adaptation is considered to be an important factor in making the judgments, the history of the group is that they have all lived in that environment in which the expert feels it most likely the pollutant would cause exactly 1 percent of the group to suffer adverse health effects if they were all exposed on the worse days.

The point of having the health expert make judgments for an idealized situation is to simplify the judgments. Even taking one health effect category at a time, the expert is faced with portions of the criteria document which review, critique, and interpret studies of different types, ages, and scientific validity; some of these studies report positive results, some negative results,

and possibly some contradict each other; each of the three types of studies, clinical, epidemiological, and toxicological, has inherent limitations, although they often complement each other. To take this scientific work and the interpretation of it given in the criteria document and determine its implications for even an idealized real world is a task only the flexibility of the human mind can deal with, but a very complex task.

The degree of idealization which makes the judgments simplest is an open question. If the evidence is mainly epidemiological the degree of idealization suggested here is probably too great. When the evidence is mixed among the three types of studies, as it usually is, it is hard to judge what degree of idealization would be best. When risk surfaces are estimated for the purpose of making the various kinds of head-count risk estimates which would fill out the risk picture, there is another consideration. This consideration is the type of risk surface which would be most convenient for the second step from the risk surface to the real world — in that case a quantitative step. The problem becomes that of designing the risk surface optimally for serving as a pier between the information in the criteria document and a full risk picture; the first span being the health expert judgments and the second span being analysis, exposure pattern data, air quality data, etc.

2.3 Uncertainty in Peak Air Quality Levels

As indicated in Section 2.1.5 the second of the two primary uncertainties which must be accounted for in assessing the risk of exceeding a health effects threshold is the uncertainty as to the peak levels of oxidant concentration which may be encountered in the time period over which the risk is assessed.

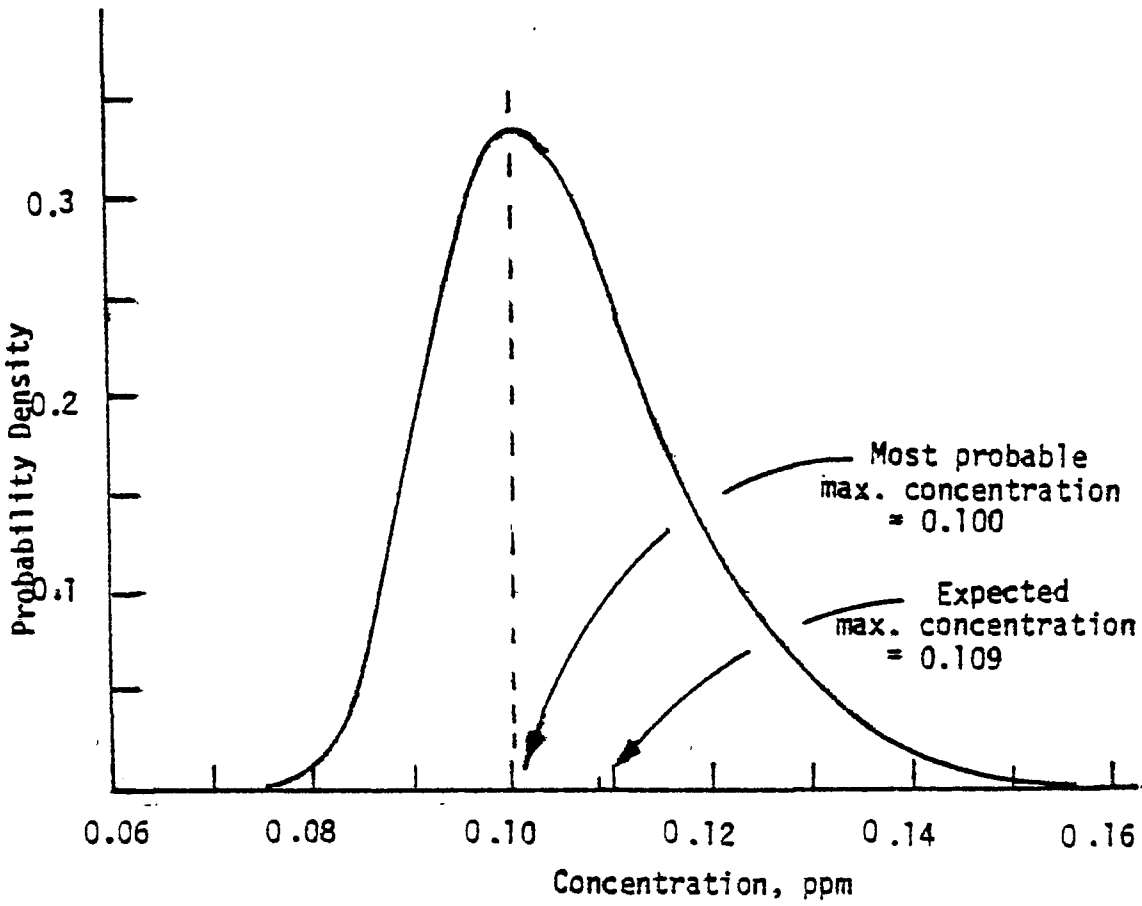
There exists a certain amount of confusion concerning peak pollutant levels in that there is a tendency to think that a geographic area at a given level of precursor emissions will experience a characteristic highest concentration of pollutant in any calendar year or period of several calendar years. This tendency is more evident in the belief that there is a characteristic second-highest concentration which may be related to the level of precursor emissions. In reality the highest concentration or second highest (or third or fourth, etc.) are statistical variables which can change significantly from one calendar year to the next even though the average level of precursor emissions remains constant. These changes are largely of a random nature and come about because of random fluctuations in weather and in emission levels. It is precisely because of this random behavior that the uncertainty in the highest concentrations experienced in a given period of time needs to be accounted for in the risk assessment. If the peak concentration for a given time period was a constant for a given area at a given emission level, only the uncertainty in the position of the health effect threshold would need to be of concern.

The uncertainty in the value of the highest hourly average concentration is represented in Fig 2-5(a) by a probability density function. In the

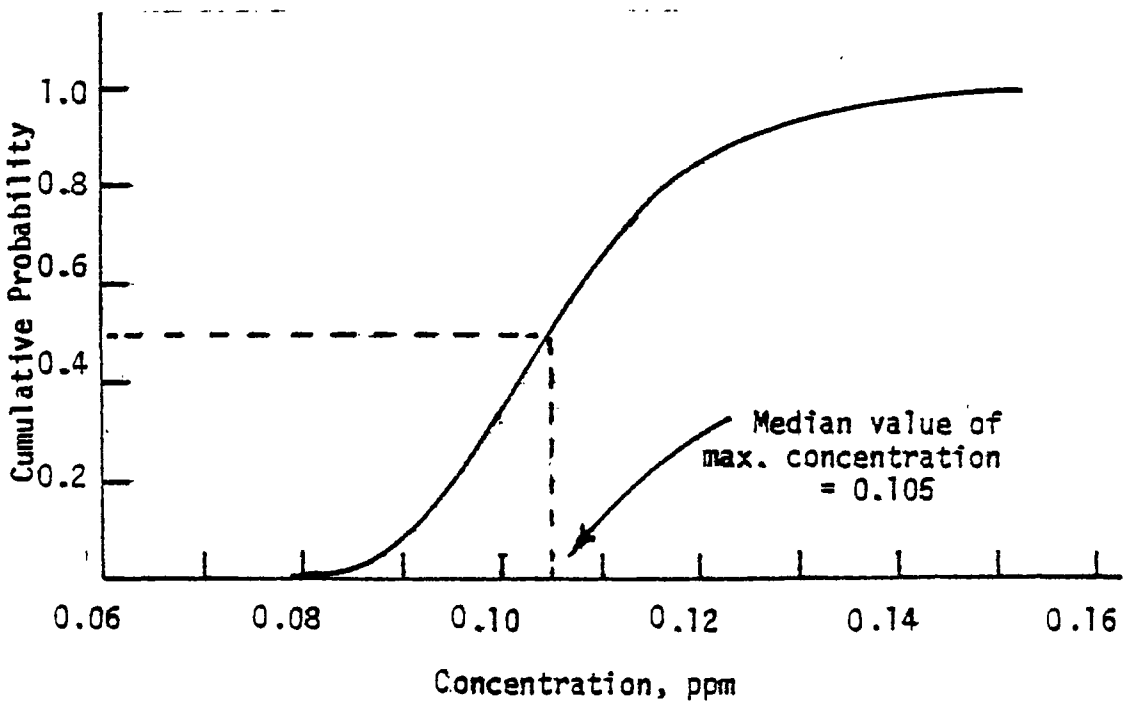
actual calculation of risk, as discussed in Section 3, it is more convenient to use the cumulative distribution function which is just an alternate way of expressing the same uncertainty. Figure 2-5 shows a hypothetical probability density function and the corresponding cumulative distribution of the annual maximum hourly average ozone concentrations. For the hypothetical situation depicted in the figure, the most probable value of the maximum concentration in a calendar year is 0.10 ppm. The cumulative distribution, customarily called the distribution, is such that the value of a point on the distribution read from the vertical axis is the probability that the observed maximum hourly average concentration for a given time period is less than the concentration read off the horizontal axis. For example, Figure 2-5(b) shows that there is a 50 percent chance that the maximum ozone concentration in any year is less than or equal to 0.105 ppm. This concentration is usually referred to as the median value. The expected or mean value of the annual maximum concentration is 0.109 ppm. This distribution is referred to in the following sections as the P_C function.

It may be desired to assess the risk of exceeding a health effect threshold "m" or more times in a given period, where m is an integral number such as 1, 2, 3,$\cdots$ etc. In this case there is no change in the threshold probability density functions, but there is a change in the P_C function. As is discussed more fully in Section 3.2, the P_C function for "m" or more exceedances is the distribution of the mth highest concentration observed in the time period. The distribution in this case is referred to as $P_C^{(m)}$. Using this notation, the distribution P_C of the highest concentration, would be designated as $P_C^{(1)}$.

Figure 2-5. Probability Density Functions and Cumulative Distribution (P_C) of Annual Highest Hourly Average Concentration



(a) Density Function



(b) Cumulative Distribution Function, P_C

To use the P_C function in estimating risk it is necessary to be able to estimate the P_C functions corresponding to potential or alternative levels of the ambient air quality standard. The averaging time, the concentration level and the overall form of the standard are all factors in making this connection. The averaging time of the concentrations to which the P_C function is applied is the same as the averaging time specified in the standard. Theoretically, there is no limitation in the averaging time. It could be one hour, one day, one month or even one year. In the case of ozone the one-hour average is most appropriate because protection is desired against short-term peak concentrations. The remaining discussion will assume hourly average concentrations, but it should be kept in mind that the approach is not limited to this time period.

The concentration level of the standard determines the general position of the P_C function along the concentration axis. The manner in which it does is dependent upon the form of the standard. For example, if the standard were to have the following form:

(A) C_{STD} ppm, expected maximum hourly average concentration in one year,

where C_{STD} is the concentration level of the standard, then the P_C function is placed so that its mean concentration value corresponds to C_{STD} (Refer to Figure 2-5). If the standard level had corresponded to the median (0.5 fractile) of the maximum concentration instead of the expected maximum value, then the 50% point of the P_C function would be set at C_{STD}.

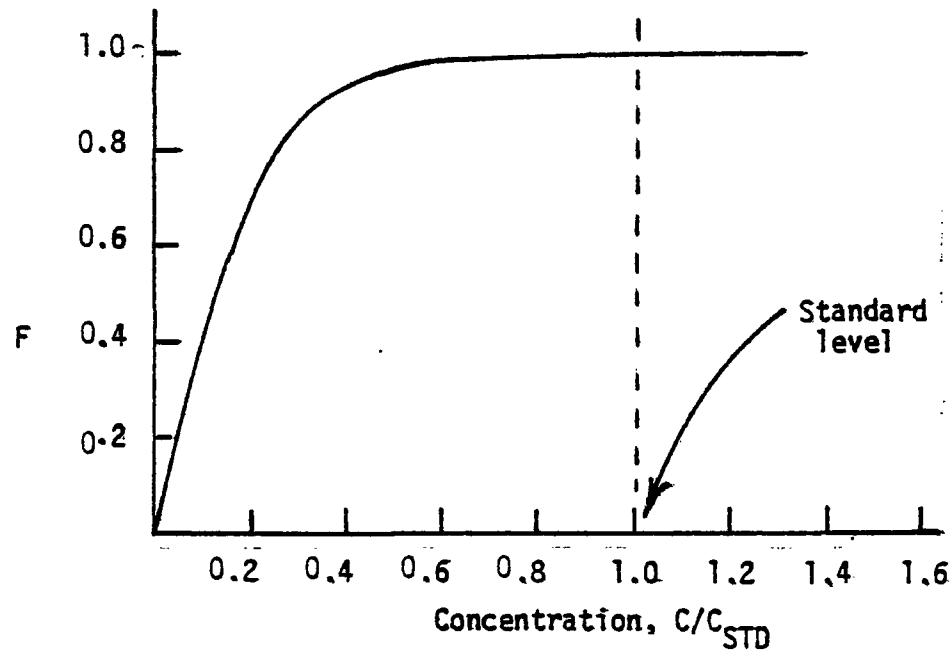
In both of the above-mentioned forms, the standard deals directly with the behavior of the annual maximum hourly average concentration and, therefore, there is a direct connection between the standard level and the P_C function. In other forms the connection may be less direct. For example, the form of the ozone standard which has been proposed is:

> (B) C_{STD} ppm hourly average concentration with an expected number of exceedances per year less than or equal to one.

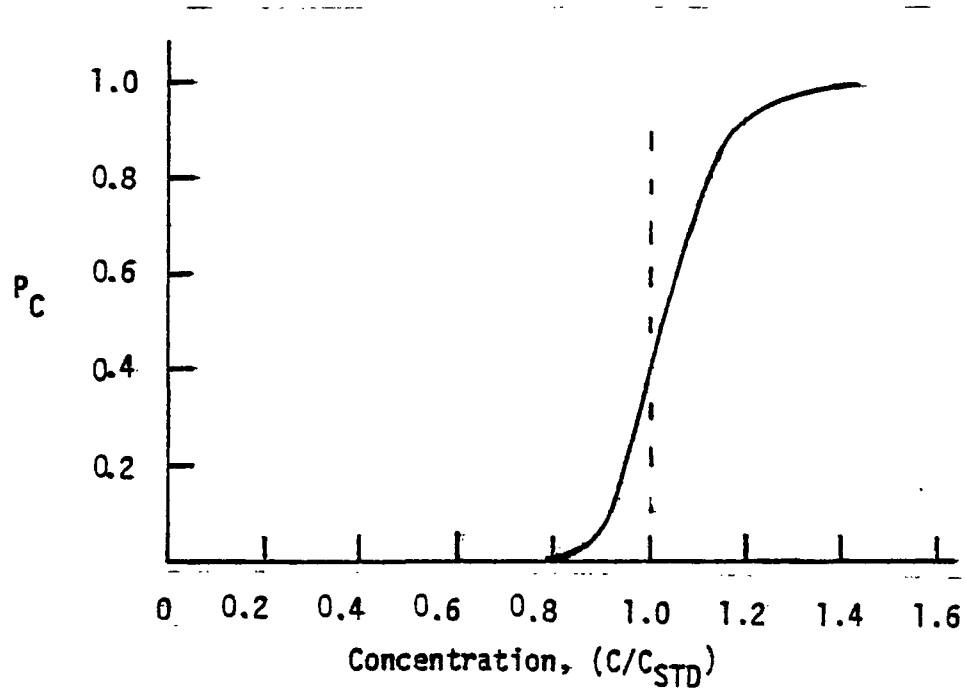
In this case the standard is not based on the annual maximum hourly average concentration but on the concentration which is expected to be exceeded once a year on the average. That is, the proposed form is directly connected to the distribution of hourly average concentrations rather than to the distribution of the annual maximum hourly average concentrations. In fact, it can be shown that C_{STD} is that concentration which corresponds to the 1 - 1/8760 = 0.999886 (where 8760 is the number of hours in one year) fractile on the distribution function for the hourly averages.

The P_C distribution and the hourly average concentration distribution are distinctly different, as shown in Figure 2-6. The hypothetical distribution for the hourly averages shown is a Weibull distribution which has been demonstrated to provide a good fit to ambient hourly average ozone concentrations (17). The P_C function in the figure corresponds to a time period of one year. The concentration axis is in units relative to the standard level as defined by (B) above; that is, the relative concentration 1.0 corresponds to the level of the standard. As would be expected, the effective range

Figure 2-6. Hypothetical Distribution of Hourly Average Concentrations and the Corresponding Distribution of the Annual Maximum Hourly Average Concentration



(a) Distribution of Hourly Average Concentration



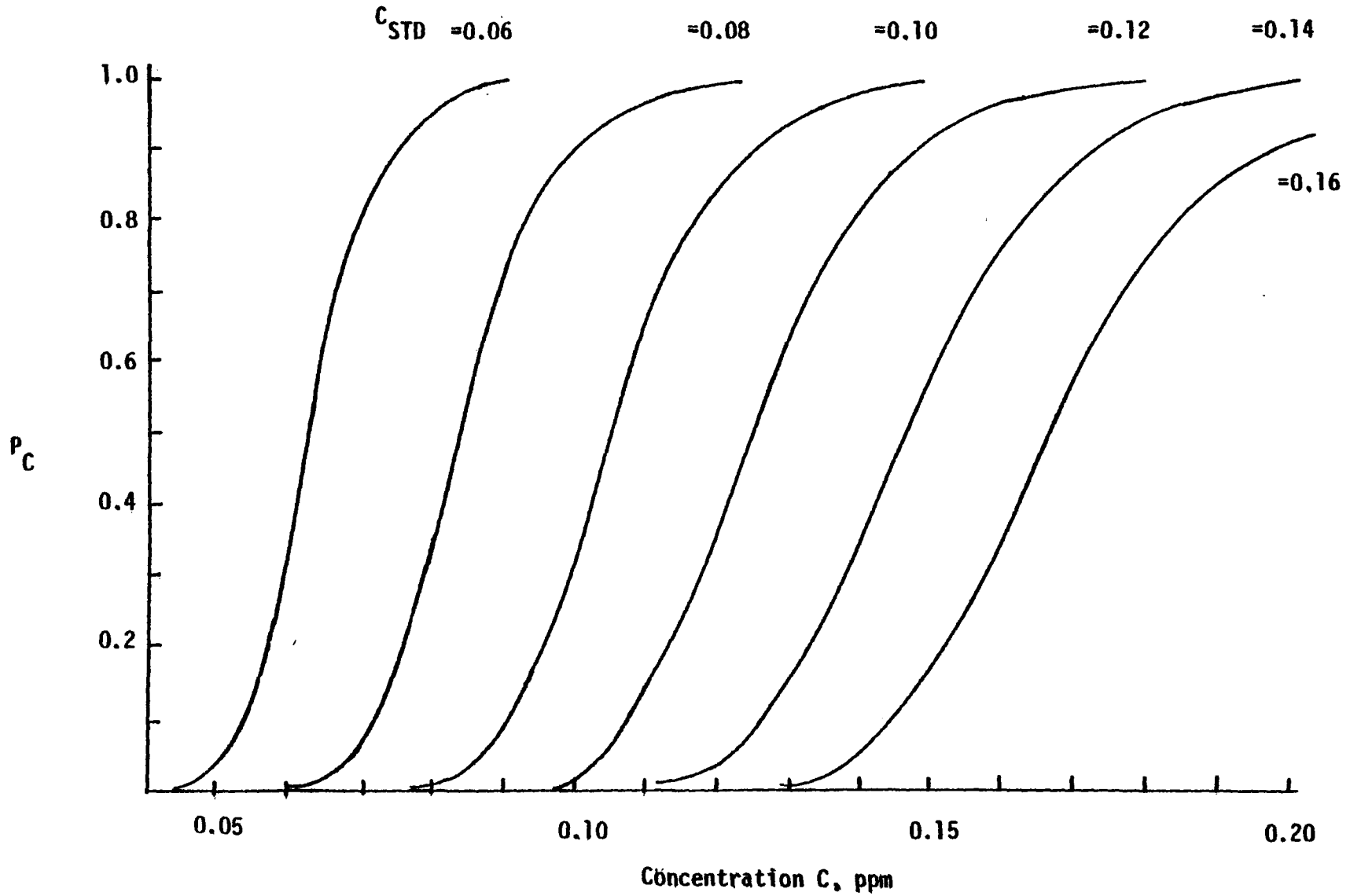
(b) Distribution of Annual Maximum Hourly Average Concentration

(values sensibly above zero and below one) of the P_C function is much narrower than that of the hourly average distribution and encompasses the standard level. On the other hand, the hourly average distribution is very close to its limiting value of 1 at the standard level. Under the circumstances depicted in the figure, the value of the P_C function at the standard level is 0.368 and will be this value at all levels of the standard. Figure 2-7 shows a series of P_C functions for various alternative levels of the standard. It is seen that the chief effect of the standard level is to establish the position of the P_C function along the concentration axis. Note, also, there is a tendency for the P_C function to spread as the standard level is increased.

To determine the corresponding P_C function it is necessary to know the distribution function for the hourly average concentrations. However, the exact connection is influenced by the degree to which the hourly average concentrations in neighboring hours are correlated and/or are affected by the time of day, of the week and of the year. The concentrations of pollutants tend to show such dependencies and ozone concentrations in particular show strong autocorrelation and time dependence. The P_C functions shown in the figures of this section are all based on the assumption of no correlation and no time dependence. As discussed in Section 3.2 this case is important in the development of approaches to taking these interactions into account.

National Ambient Air Quality Standards are, of course, set for the entire U.S. While the level of the standard tends to locate the position of the corresponding P_C function, the degree of spread or effective range of the function can vary from one geographic area to the next. The shape of the function can also vary. This, of course, means that the risk of exceeding

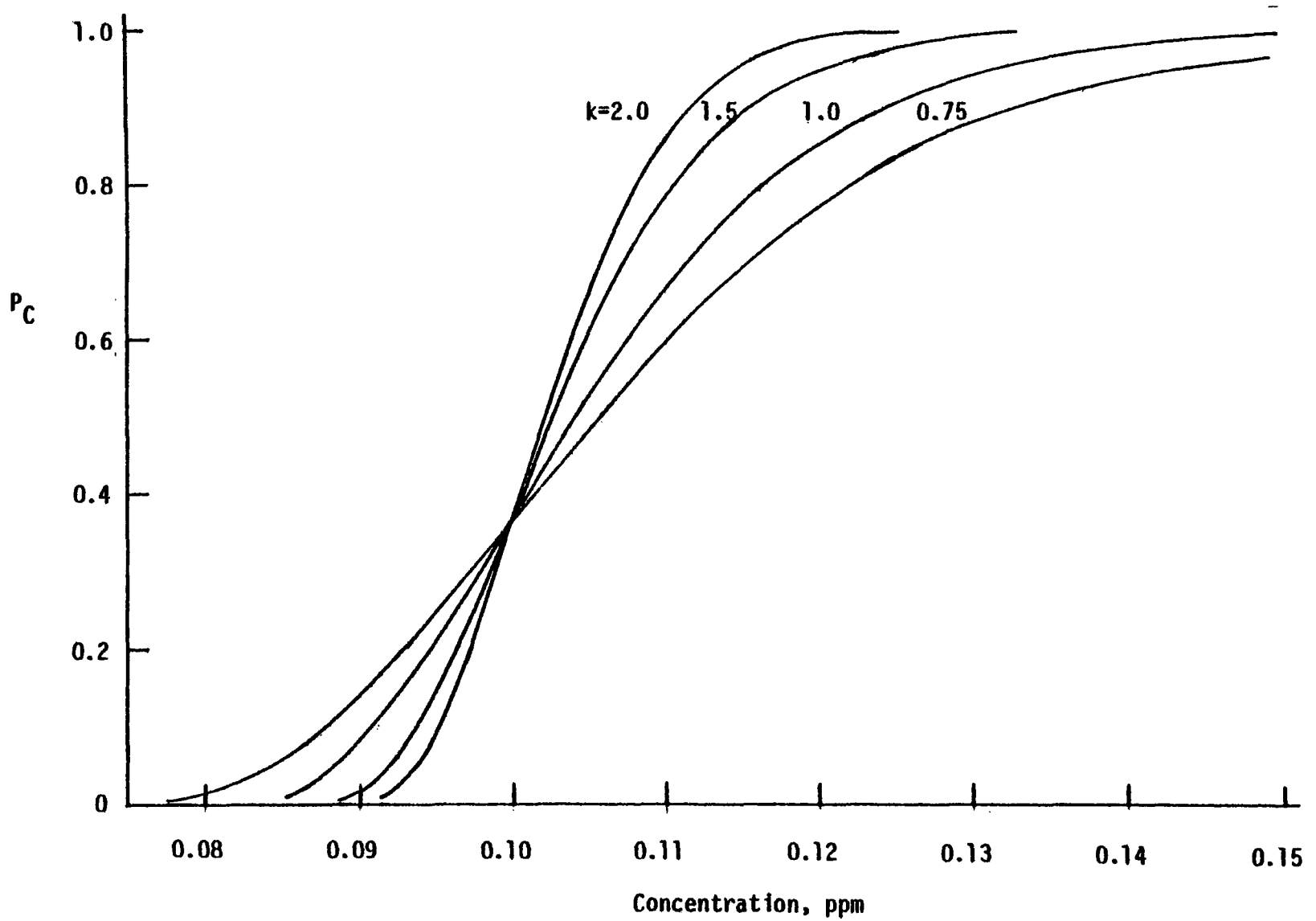
Figure 2-7. Change in P_C Function With Change in Standard Concentration Level



a health effects threshold is not the same with a given standard level for all parts of the country. This is not a limitation of the proposed risk assessment method but rather a limitation in the procedure of setting a single standard level applicable to all of the U.S. Therefore, in setting a single national ambient air quality standard it is important to determine the range of risks associated with each level of the standard. As indicated, the standard level sets the position of the P_C function and tends to be the major factor determining the risk. The variation from one geographic area to the next primarily affects the spread of the P_C function and this effect has less impact on the risk than the placement of the P_C function. Variations in shape of the function as distinct from changes in spread are expected to have a distinctly lower effect on risk. There may also be some influence of geographic area on the placement of the P_C function if the nature of the correlation between hourly average concentrations and time dependence were to vary markedly from one area to the next. This possibility has not been investigated sufficiently as yet but is presently judged to introduce less variation than the change in spread since the broad aspects of the correlation and time dependence probably do not vary substantially with location, at least for urban areas. Figure 2-8 shows the possible variations in the spread of a hypothetical P_C function. Each of the P_C functions in the figure passes through $P_C = 0.368$ at the level of the standard. Section 3.2 treats P_C functions in more detail and describes methods for deriving them from the distribution of hourly average concentrations.

One further difficulty in estimating the appropriate P_C functions for a pollutant is that available air quality data may be at a significantly

Figure 2-8. P_C Function for One-Year Period for Different Values of Weibull Shape Factor, k



higher level than would exist if the alternate levels of the standard under consideration were being attained. In this case it will be necessary to make an extrapolation to the more stringent air quality levels, making the most of whatever information is available.

2.4 Secondary Uncertainties and Public Probability

In section 2.1, it has been pointed out that the two primary uncertainties which give rise to the threshold risk associated with NAAQS's are the uncertainty about the health effect threshold concentration and the uncertainty about the maximum pollutant concentration which will be reached in a given period of time if a given standard is just attained. Section 2.2 discussed the factors involved in obtaining a representation of the first primary uncertainty, a composite health effects threshold probability distribution. Section 2.3 discussed the factors involved in obtaining a representation of the second primary uncertainty, a P_c probability distribution. Once a composite threshold distribution and a P_c probability distribution have been determined, the mathematical formulas presented in section 3.1 can be used to calculate a threshold risk estimate.

If the only uncertainties that needed to be addressed were primary uncertainties, it would only be necessary to calculate one threshold risk estimate for each alternative standard under consideration. However, there are also secondary uncertainties about how to best represent the primary uncertainties; these secondary uncertainties cannot be ignored because they can greatly affect the calculated risk values, which in turn can affect the decision on which alternative standard to adopt.

In saying that there is secondary uncertainty about how to best represent the primary uncertainty about the health effect threshold concentration, the possibility is left open that there may be no best representation of this

uncertainty in one sense; that is, in the sense of there being a probability distribution which most rationally represents the available evidence. If for each health effect category there exists a threshold distribution which most rationally represents the available evidence and if there exist probabilities which most rationally represent the available evidence on the existence of a causal relationship between the pollutant and each category of effect about which the existence of such a causal relationship is in doubt, then the composite distribution obtained from these probability distributions and these probabilities is the most rational representation of the available evidence; otherwise, no such most rational representation exists. It is an open philosophical question whether there exist most rational representations of evidence in the two required senses. The situation is similar with regard to the P_c distribution.

Subjective probability distributions on threshold concentrations for individual categories tend to be subjectively derived and hence different for each health expert; likewise probability-of-existence judgments. One of the primary aims in standard setting is to minimize arbitrariness; to minimize arbitrariness in standard-setting decisions, arbitrariness must be minimized in the generation of the information on which the decision is made.

From the public's point of view as a whole, it would be arbitrary to let one health expert make the required subjective judgments. Ideally, all of the fully qualified health scientists who are best informed about the information in the criteria document would contribute judgments. But resource constraints and the inadvisability of attempting to determine the exact membership of the set of best-informed scientists dictate that only a representative sample of the best-informed experts contribute judgments. Selection of the experts is obviously an important juncture that deserves close attention, both in terms of the particular experts to be selected for a particular assessment and the criteria and process by which experts are selected in general.

Presumably, one criteria which deserves to play a role in the selection of experts is diversity in viewpoint. The whole idea of having more than one expert make the necessary judgments is to avoid arbitrarily basing a standard on one point of view when more than one rational point of view may exist about the most probable implications of the same evidence. To the extent that diverse, well-informed viewpoints are recognizable a priori, this criteria could either enter in selecting the sample of experts or in selecting a subset of the best-informed scientists from which a truly random sample could be selected.

Perhaps the most widely known technique for the use of expert judgment in decision-making is the Delphi Technique.(18) (19) Diverse approaches have been included under the Delphi rubric, but at least three themes seem to be common to most: (1) systematicness, (2) sharing of information and perspective among experts, (3) convergence of judgment among experts. The method being suggested here subscribes to (1) and (2). It also subscribes to (3) to the extent that convergence of judgment is brought about by the sharing of information and perspective. However, to the

extent that convergence of expert opinion is made a goal in itself in the Delphi approach, the philosophy of the approach being suggested is different. In some formal work, distinct from the Delphi approach, convergence of judgment for the sake of convergence is a formal goal. (20) The method described here does not accept the idea that convergence of judgment in representing the implications of inconclusive evidence for the sake of convergence is a legitimate goal; at least not for public decision-making. Rather, the goal is to represent to the decision-maker, as much as is feasible, the implications of the actual diversity of well-informed judgments that can remain after information and perspective have been shared; ideally the diverse judgments would be representative of the diversity of judgments that would be made by the whole set of well-informed scientists were it feasible to have them all participate.

Consider as an example a case in which there are two threshold-independent health effect categories, one of which requires a probability-of-existence judgment: suppose that three subjective probability distributions are elicited on the health effect threshold concentration of each category and three probability-of-existence judgments are made for the category about which the existence of a causal relationship is uncertain. Then, there are 3 x 3 x 3 = 27 different combinations of distributions and existence judgments which give rise to 27 different composite health effect distributions. Letting 'd_i^j' represent the jth distribution for the ith category, 'e_i^j' represent the jth existence judgment for the ith category, 'I' represent the compositing function, and D_i represent the ith composite distribution, the situation can be represented formally as follows:

$$I\ (d_1^1\ ,\ e_1^1\ ,\ d_2^1) = D_1$$

$$I\ (d_1^2,\ e_1^1\ ,\ d_2^1\ ) = D_2$$

$$\vdots \qquad\qquad \vdots$$

$$I\ (d_1^3\ ,\ e_1^3,\ d_2^3\ ) = D_{27}$$

On the air quality side assume there exists a two-parameter distribution that best fits available air quality data, that can be used to derive an approximation of the true P_c distribution (P_c^t), and that has a scale parameter which varies with the standard level and a shape parameter, k. Suppose that for any given alternative standard level the value of k which gives the best approximation to P_c^t for that standard level is unknown, but 5 equiprobable alternate values of k (namely, k_h, where h ranges from 1 to 5) exhaust the possibilities. Then 135 risk estimates can be generated by combining the 27 D_i's with the 5 k_h's. Letting 'V' represent the convolution function and 'r_i' represent the $_i$th risk number:

$$V\,[I\,(d_1^1, e_1^1, d_2^1), k_1] = V\,(D_1, k_1) = r_1$$

$$V\,[I\,(d_1^2, e_1^1, d_2^1), k_1] = V\,(D_2, k_1) = r_2$$

$$\vdots \qquad\qquad \vdots$$

$$V\,[I\,(d_1^3, e_1^3, d_2^3), k_5] = V\,(D_{27}, k_5) = r_{135}$$

Or, alternatively:

$$R\,(d_1^1, e_1^1, d_2^1, k_1) = R\,(w_1) = r_1$$

$$R\,(d_1^2, e_1^1, d_2^1, k_1) = R\,(w_2) = r_2$$

$$\vdots \qquad\qquad \vdots$$

$$R\,(d_1^3, e_1^3, d_2^3, k_5) = R\,(w_{135}) = r_{135}$$

where $w_1 = (d_1^1, e_1^1, d_2^1, k_1)$, etc.

Now, with no loss of generality the subscripts can be rearranged on the risk estimates so that they are ordered from smallest to largest:

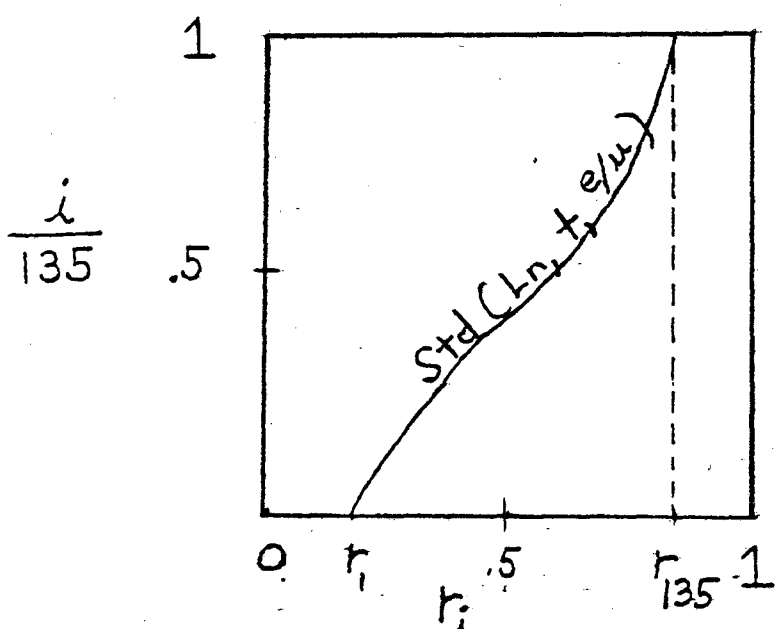
$$0 \leq r_1 \leq r_2 \leq \cdots \leq r_{135} \leq 1$$

Then, the 135 ordered pairs

$$(r_1, \frac{1}{135}), \quad (r_2, \frac{2}{135}), \ldots, \quad (r_{134}, \frac{134}{135}), \quad (r_{135}, 1)$$

can be plotted and a relatively smooth curve can be drawn through the plotted points; this is done for a hypothetical example in Figure 2-9.

Figure 2-9. Relative Frequency Plot



Even without its being given any further interpretation, the type of curve shown in Figure 2-9 is worthy of the attention of decision-makers and other interested parties:

(a) It gives a measure of the degree of consensus among the experts on the risk implications for the standard under consideration of the information in the criteria document; a curve that stands up relatively straight indicates that the data base reviewed in the criteria document

is strong enough in its implications for the standard under consideration to have lead to a high degree of consensus; a curve that stretches from corner to corner indicates a low degree of consensus for that standard.

(b) Let Std (L, 1 hr, 1 hr/yr) be a formal way of representing a standard where "L" represents the concentration level, the averaging time is one hour, and the expected number of exceedances per year is one. Then, a series of curves for different standard levels L_n serve to indicate the range of standard levels for which the risk is fuzzy; that is, if n is low enough the whole curve is far to the left, which means the risk is low no matter what combination of judgments is used; if n is large enough the whole curve is far to the right, which means the risk is high no matter what combination of judgments is used. A risk greater than 0.5 would be hard to interpret as providing a margin of safety, so if $r_1 > 0.5$ for a standard, the standard level can be regarded to be above the fuzzy risk range.

(c) The type of curve depicted in Figure 2-9 incorporates the secondary uncertainty about how to best represent the primary uncertainty on the air quality side as well; it simultaneously exhibits the consequences of the two secondary uncertainties. Also, in cases where geographical variation in the P_c distribution affects the resulting calculated risk significantly, a lower-bound P_c distribution and an upper-bound P_c distribution can be used to generate an upper-bound curve and a lower-bound curve for the same standard; the resulting risk ribbon indicates clearly that the risk

varies geographically throughout the United States for the same standard.

(d) Curves or ribbons of the type depicted in Figure 2-9 can be plotted for different combinations of health effect categories, for different definitions of the health effect threshold concentration, and for different numbers of exceedances of the threshold concentration.

Despite its usefulness even without being given further interpretation, the type of curve shown in Figure 2-9 needs to be given further interpretation and can be . It needs to be given further interpretation so that distinctions of degree can be made within the fuzzy risk range. It can be given further interpretation because:

(1) it has the formal properties of a cumulative probability distribution;

(2) it can be interpreted as a cumulative probability distribution on the threshold risk, although it may not be a cumulative probability distribution on the true threshold risk since there may be no such thing.

(3) it has a uniqueness that has importance for public regulatory decision-making.

Each of these will now be discussed in turn:

(1) It can be checked that the Kolmogorov Axioms are satisfied for the probability space $(\Omega, \mathcal{F}, P)$, where the sample space $\Omega = (w_1, w_2, \ldots, w_{135})$, the σ-field $\mathcal{F} = 2^{\Omega}$ (all of the subsets of Ω), and the probability measure, P, of a subset of Ω is the number of elements in the subset divided by 135. (21)

(2) The 135 r_i's are risk estimates, but they may not be estimates of some "true risk" in the same sense that an estimate of the length of a stick is an estimate of the actual length of the stick. There may be no such thing as "the true risk" associated with a standard in the fuzzy risk range. If there were a most rational P_c distribution and a most rational threshold distribution then the risk value obtained by convoluting the two most rational distributions could be considered the true risk; then the r_i's could be said to be estimates of the true risk. But since whether such most rational distributions exist is shrouded in philosophical obscurity, it cannot be assumed they do.

Unlike the P_c distribution, the threshold distribution does not fit the relative frequency mold, so the alternative available on the air quality side of defining the true P_c distribution, P_c^t, to be the limiting frequency distribution of the maximum pollutant concentration over the relevant period as the number of periods increases without limit when the given standard is just met, is not available on the threshold side. Hence, the alternative possibility of considering the convolution of the true P_c distribution with the true threshold distribution to be the true risk is not available.

Since the question of whether there exists a true risk for the r_i's to estimate is open, and is likely to remain open for the foreseeable future, the further interpretation of the curve in Figure 2-9 must be broken down into the two possible cases. The interpretations given for each case must be such that for practical purposes it will make no difference which case proves to be the actual case in the philosophical millennium.

Case 1: For a given body of evidence relevant to the value of a quantity, there exists (conceptually) a probability distribution which is the most rational representation of the uncertainty about the value of the quantity.

Let Std (L_i, 1 hr, 1 hr/yr) be a series of alternative standards; consider Std (L_n, 1 hr, 1 hr/yr). Let D^R be the most rational distribution representing the evidence on the health effect threshold; let $P_c^R(L_n)$ be the most rational distribution representing the evidence on the maximum pollutant concentration in the relevant period if Std (L_n, 1 hr, 1 hr/yr) is just attained. Call $R^T(L_n)$, the convolution of D^R with $P_c^R(L_n)$, the True Threshold Risk associated with Std (L_n, 1 hr, 1hr/yr). Let $| R^T(L_n) - r_i(L_n) | = G_i(L_n)$. Then the $r_i(L_n)$ which gives the minimum $G_i(L_n)$ is the Truest Risk Estimate (TRE) for Std (L_n, 1 hr, 1 hr/yr).

We have from (1) that the Kolmogorov Axioms are satisfied for a probability space which has $\Omega = (w_1, w_2, \ldots w_{135})$ as its sample space. If each $w_i = (d_i, e_i, d_j, k_h)$ is considered to be equally likely to give the TRE, then a curve of the type depicted in Figure 2-9 is a cumulative probability distribution on the TRE for the given standard level. Assuming arbitrariness has been minimized in the selection of the experts and each k_h is equally likely to give the best approximation of $P_c^t(L_n)$, from the public point of view each w_i is equally likely to give the TRE.

These circumstances suggest a public probability interpretation of the type of curve depicted in Figure 2-9. "Public probability" for two reasons: First, the particular experts contributing judgments were selected and encoded for the express purpose of assessing the risk associated with alternative standards by the public regulatory agency that has the responsibility of setting

the standards. Second, it is the public regulatory agency that would make the decision to regard each w_i as being equally likely to give the TRE.

This interpretation of probability is like, but not exactly the same as, both the subjective interpretation and the classical interpretation of probability. The classical interpretation is of a relative frequency type, but relative frequency in the sample space, not relative frequency in a random sample. Games-of-chance examples in textbooks often for the sake of simplicity use the classical interpretation. The 135 w's have the symmetric appearance of games-of-chance examples, but, just as in such games, someone has to decide to regard the elements of the sample space to be equally likely. In this case that someone is not a private individual, but rather a public regulatory agency; hence the interpretation is not a subjective interpretation.

This particular interpretation is the weak interpretation of public probability. The label for the vertical axis of the curve in Figure 2-9 under the weak interpretation would read "Estimated Public Probability that the Truest Risk Estimate is less than or equal to the given risk value". There is also a strong interpretation of public probability. The label for the vertical axis of the curve in Figure 2-9 under the strong interpretation would read "Estimated Public Probability", period. Under the strong interpretation a feature of the weak interpretation is dropped. The probability distribution is no longer on a point of the sample space.* The probability distribution is on threshold risk, not on the TRE. The public agency decision is to give each w_i equal weight in a cumulative probability distribution on threshold risk; that is, the decision is to regard the whole curve as a probability distribution on threshold risk. On the true threshold risk, $R^T(L_n)$? That depends on whether $R^T(L_n)$ exists.

---

*An endpoint adjustment is made in the way the curve is plotted; the points are $(r_i, \frac{i}{136})$, rather than $(r_i, \frac{i}{135})$, under the strong interpretation. This adjustment follows the recommendation of Gumbel (22).

Case 2: For a given body of evidence relevant to the value of a quantity, there does not exist (even conceptually) a probability distribution which is the most rational representation of the uncertainty about the value of the quantity.

In this case, $R^T(L_n)$, the true threshold risk, does not exist. So, under the strong interpretation of public probability the curve is a public probability distribution on threshold risk, but not on the true threshold risk since no such thing exists. The weak interpretation of public probability is not available in this case, because the TRE is not well-defined; none of the 135 r_i's can be nearest the true threshold risk if there is no true threshold risk.

Since the weak interpretation of public probability is not available in one of the two possible cases and the strong interpretation of public probability is available in both cases, it is best to adopt the strong interpretation. The curve can be thought of as a probability distribution on the true threshold risk if there is a true threshold risk--just on threshold risk otherwise. Threshold risk definitely exists, whether there is a true threshold risk or not.

The practical import of the probability distribution is the same in either case. The distribution indicates the public probability that threshold risk is less than or equal to various values. Thus, distinctions of degree can be made in a meaningful way within the fuzzy risk range. Each alternative standard has a public probability distribution (actually, a ribbon) associated with it and these distributions (ribbons) can be compared. The public has a meaningful way of comparing the alternative standard selected to the alternative not selected.

In generating public probability curves it is useful to recognize their classical structure in dealing with a third type of secondary uncertainty that can arise, disagreement among experts in their judgments on the number of threshold-independent categories. For example, suppose that half of the experts judge there to be three threshold-independent categories (only one of which requires a probability-of-existence judgment) and half judge there to be two in a situation similar in every other respect to the situation under consideration. Assuming there are two independent categories has given $3^3 \times 5 = 135$ risk estimates. Assuming there are three independent categories gives $3^4 \times 5 = 405$ estimates. Plotting 135 + 405 = 540 estimates would give undue weight to the judgments of those experts who judge there to be three categories; the resulting curve would not be a probability distribution on the TRE under the weak interpretation if the true risk exists. So, under the strong interpretation making no assumptions about the existence of a true risk, the correct approach should give the two groups of experts equal representation in the sample space; this is accomplished by counting three times each of the 135 r_i's calculated while working under the assumption that there are two independent categories. Then, $3 \times 135 + 405 = 405 + 405 = 810$ risk estimates are plotted.

(3) The public probability distribution for a given standard in a given location would include the judgments of every member of "the set of best-informed experts." Since the membership of this set could in principle be specified by having the public agency make (in some cases arbitrary) decisions on who is and who is not a member, the fact that it is unwise for the Agency to do so does not affect the conceptual existence of the public probability distribution. Another reason the label on the ribbon diagrams (see section 4.4) is "Estimated Public Probability" rather than just "Public Probability", is that on the air quality side the public probability distribution for a given standard and location

is the limit of successive approximations which can be made by increasing without limit the number of shape-factor values used (see table 4-3).

Obviously, the probability distribution which is used to estimate the public probability would be different if the subset of experts who contributed probability judgments were different. So, why should any one distribution be considered to give the estimated public probability? For the same sort of reason any one jury's decision should be considered the decision in a court case (assuming there was due process), even though another group of twelve people might have decided differently. In pragmatic decision-making, where subjective judgment inevitably plays a vital role and resource constraints are inevitably a factor, arbitrariness should be minimized, but it can't be avoided. The method suggested here attempts to identify and isolate those junctures at which there is inevitably some arbitrariness, so that how to proceed at such junctures in general can be carefully considered, and so that how the Agency does proceed in particular cases can be scrutinized by the public.

The jury analogy cannot be pressed too far, of course. The regulatory agency decision-makers, not the health experts,make the decision on what ribbon to accept.* The fact that in selecting a standard they are also accepting an associated risk ribbon separates the normative judgment of what risk ribbon to accept from the assessment of risk itself. These two types of judgment are best kept clearly separated.

Can the arbitrariness which is unavoidable in selecting a particular subset of health experts to make the judgments which help give rise to risk ribbons be avoided by returning to simpler times when risks, not risk ribbons, were accepted? An assumption of this rhetorical question is false. In

---

*The topic of risk acceptance is not addressed in this report; see Rowe (23).

accepting risk one is always accepting a risk ribbon, since a particular risk number is a special case of a ribbon — the special case in which one judgment is accepted at each point where there is a secondary uncertainty. Even if these judgments are made by highly qualified experts, there is more, not less, arbitrariness in ignoring the secondary uncertainty problem in assessing risks.

3.0 General Description of the Method

The previous section discusssed the underlying principles involved in assessing the risk to the most sensitive members of the population from a given air pollutant in an area just meeting a specified level of air quality. This section describes the method in more detail giving the basic tools for estimating the risks. The mathematical framework is presented showing how the uncertainties concerning the health effects thresholds and the maximum pollutant concentrations observed in a given time period are combined quantitatively to determine the risk. Next, the methods for obtaining the probability functions for maximum concentrations are discussed.

In the following subsections only the actual working equations used in the calculations are shown. Their derivations are given in Appendices B and C. An understanding of the underlying mathematics is not necessary to an understanding of the method or its application.

3.1 Mathematical Description of the Method

In the preceeding section it was pointed out that the risk of exceeding a true health effects threshold in a given period of time is determined by the uncertainty in the location of the health effects threshold (or in the

case of multiple thresholds, the uncertainty in the location of the lowest threshold) and the uncertainty in the maximum oxidant concentration over the given time period. It can be shown, by application of the theory of probability, that the following equation gives the relationship between the risks and the above uncertainties:

$$R = 1 - \int_0^\infty P_C(C) p_T(C) dC \qquad (1)$$

where:

R = Probability (Risk) that a true health effect threshold, or the lowest of a multiple number of thresholds, is exceeded one or more times in a given time period (e.g. one year, five years, etc.).

$P_C(C)$ = Probability that the highest observed time-averaged (e.g. one hour, two hour, etc.) pollutant concentration does not exceed the concentration C in the given time period.

$P_T(C)$ = The probability density function for the health effect threshold or in the case of multiple health effects the function for the lowest effect (the composite density function).

The derivation of Equation (1) is given in Appendix B.

Equation (1) also holds for a more general case in which R is defined as the risk of m or more exceedances of a threshold in a given time period, where m may have the integral values 1, 2, 3, etc. In this case P_C is redefined as the probability that the mth highest time averaged pollutant concentration does not exceed the concentration C in the given time period.

In other words, P_C is the cumulative distribution of the mth highest time averaged pollutant concentration in the time period of interest.

It was pointed out in the last section that specifying the national ambient air quality standard for a pollutant limits the range of $P_C(C)$ functions which will satisfy the air quality requirements of the standard. If this limitation can be expressed quantitatively, and the health effect threshold density functions have been determined, then for any given specification of the standard, a range of risks associated with that specification can be calculated from Eq. (1).

In practice, it is only convenient to use Eq. (1) directly to calculate risk when a single health effect with an existence probability of one is involved. When the risk that lowest of n health effects will be exceeded is to be calculated and when there is uncertainty as to whether one or more of the effects actually occur in the sensitive population at any attainable pollutant concentration, the following expanded version of Eq. (1) is used. (See Appendix B):

$$R = 1 - \int_0^\infty P_C(C) p_T^\circ \, dC - (1-e_1)(1-e_2)\cdots(1-e_n) \qquad (2)$$

where:

p_T° = The probability density function for the location of the lowest of n thresholds over the possible range of concentrations of the pollutant.

e_i = The probability that the ith health effect actually occurs in the most sensitive population in the possible range of concentrations of the pollutant.

The function p_T° is calculated from:

$$p_T^\circ = Q\left[\frac{e_1 p_1^\circ}{Q_1} + \frac{e_2 p_2^\circ}{Q_2} + \cdots + \frac{e_n p_n^\circ}{Q_n}\right] \quad (3)$$

where

p_i° = The probability density function for the threshold of the ith health effect assuming its $e_i = 1$. (That is, $\int_0^\infty p_i^\circ \, dC = 1$).

In practice p_i° is obtained by asking the subject health expert in an encoding session to first give his best judgment of the value of e_i and then encoding him as to the location of the threshold assuming that the effect actually occurs in the sensitive population at an attainable pollutant concentration. As was discussed in Section 2.2.7, the encoding procedure gives the cumulative distributions for the health effects thresholds. These functions are differentiated by numerical methods to obtain the p_i°.

The terms Q_i and Q in Eq. (3) are calculated from:

$$Q_i = 1 - e_i \int_0^C p_i^\circ \, dC \quad (4)$$

$$Q = Q_1 Q_2 Q_3 \cdots Q_n \quad (5)$$

The Q_i are seen to be functions of the cumulative distribution of the probability density functions p_i°.

Thus, given the $P_C(C)$ functions corresponding to different levels of the standard, the probability density functions p_i° for n independent health effects and their corresponding e_i, Eqs. (2) through (5) can be used to calculate the range of risks associated with alternate specifications of the ambient air quality standard. The risks can be calculated for the individual health effects and for composites of two or more of the effects in any combination. As will be discussed in later sections, calculating risks for individual health effects and various combinations can be of value where the effects differ significantly in their seriousness.

The calculations involved in Eqs. (2) through (5) can be most conveniently carried out with a computer. The function p_T° can also be calculated by differentiating its cumulative distribution function. The cumulative distribution is a function of the existence probabilities e_i and the cumulative distribution functions of the p_i°. (See Appendix B.) Less computational labor is involved with this method if there is no interest in knowing the density functions p_i°.

3.2 Obtaining the $P_C(C)$ Distributions

As indicated in earlier sections the $P_C(C)$ function is the cumulative distribution function for the highest time-averaged pollutant concentration for a specified period of time. If the risk is calculated for m or more exceedances, it is the cumulative distribution for the mth highest concentration. To simplify the following discussion it will be assumed that the concentration averaging time is one hour.

The $P_C(C)$ function for a pollutant in the air over a given region is a measure of the air quality for that region with respect to the pollutant. Specifying a National Ambient Air Quality Standard places a limitation on the range of $P_C(C)$ functions corresponding to air quality just meeting the standard. For example, if the ambient air quality standard specified an expected (average) value of maximum hourly average concentration for a one-year period was not to exceed a given level, this would immediately locate the mean value of the distribution of maximum values and thus define the concentration region in which the preponderance of maximum values must occur. However, depending upon the area and the control methods used to meet the standard, the distribution of maximum values about the mean could be relatively narrow or spread out. It is expected, however, that there would be practical limitations on the degree of spread of the distribution. Therefore, specifying the expected maximum concentration limits the $P_C(C)$ distributions just satisfying the standard.

In applying the risk assessment method it is necessary to determine the range of $P_C(C)$ functions just meeting each alternate specification of the standard. While, in principle, this should be possible for almost any

type of standard it is more readily done for standards with statistical forms than for standards with deterministic forms.

The P_C distribution function is related to the distribution of hourly average concentrations. However, the presence of correlation between hourly average concentrations observed in different hours and the dependence of concentrations on time of day or period of the year can strongly affect this relationship. Air pollutants commonly show this correlation and time dependence and these effects must, therefore, be taken into account in developing suitable P_C functions. The approach to taking these effects into account that will be discussed here makes use of the case in which independence of hours and no time dependence of hours are assumed. This case will be discussed first.

If no correlation or time dependence of hours exists then it can easily be shown (See Appendix C) that:

$$P_C = [1 - G(C)]^n \tag{6}$$

where P_C is the distribution of the highest concentration for n hours. (n = 8760 hrs. for one year or 43,800 hrs. for 5 years.) The function G(C) is defined by:

$$G(C) = Pr[C_{obs} > C] \tag{7}$$

That is, G(C) is the probability that an observed hourly average concentration is greater than C. It is an alternative way of expressing the cumulative distribution of hourly average concentrations depicted in Fig. 2-6(a) in Section 2.3. If the distribution function in Fig. 2-6(a) is labeled $\hat{F}(C)$ then:

$$F(C) = Pr\,[C_{obs} \leq C] \tag{8}$$

from which it follows that:

$$G(C) = 1 - F(C) \tag{9}$$

If the broader definition of P_C is used, the expression is more complex.

$$P_C^{(m)} = \sum_{\nu=0}^{m-1} \frac{n!}{\nu!(n-\nu)!} G^{\nu} (1 - G)^{n-\nu} \tag{10}$$

where $P_C^{(m)}$ is the distribution of the mth highest hourly average concentration for n hours. (See Appendix C for derivation.)

Thus, if the distribution function G(C) is known, the desired P_C function can be obtained from application of Eq. (6) or (10). Studies have found that the distributions of short-term time averaged concentrations of air pollutants can usually be represented by lognormal, Weibull, or gamma distribution functions (24). Of these the Weibull function provides a good

fit to photochemical oxidant air monitoring data and is convenient to use since its G(C) function can be stated explicitly.

$$G(C) = e^{-(C/\delta)^k} \tag{11}$$

The parameter δ is referred to as the scale factor. It is the concentration corresponding to G(C) = 0.368. It establishes the approximate position of the mid-concentration values of the distribution. The parameter k is called the shape factor. It tends to be a measure of the spread of the distribution. The larger k the more compact the distribution. If the values of k and δ have been determined for a given geographic area the corresponding P_C functions can then be obtained by use of Eqs. (6) and (11) (or Eqs. (10) and (11)) at the given level of air quality.

For the risk assessment it is necessary to connect alternative levels of the ambient air quality standard with the corresponding P_C function. This is easily done through the Weibull distribution, Eq. (11). The proposed form for the ozone standard is:

> C_{STD} ppm hourly average concentration with an expected number of exceedances per year less than or equal to E.

It is shown in Appendix C that for any region to which the Weibull function applies and just meets the standard:

$$G(C) = e^{-(\ln (n_E/E))\ (C/C_{STD})^k} \tag{12}$$

where

C_{STD} = level of ambient air standard.

E = expected number of exceedances in n_E hours.

The term n_E is customarily the number of hours in one year or 8760 hours. The expected exceedance rate would normally be one for an air standard. In this case Eq. (12) becomes:

$$G(C) = e^{-9.078\,(C/C_{STD})^k} \tag{13}$$

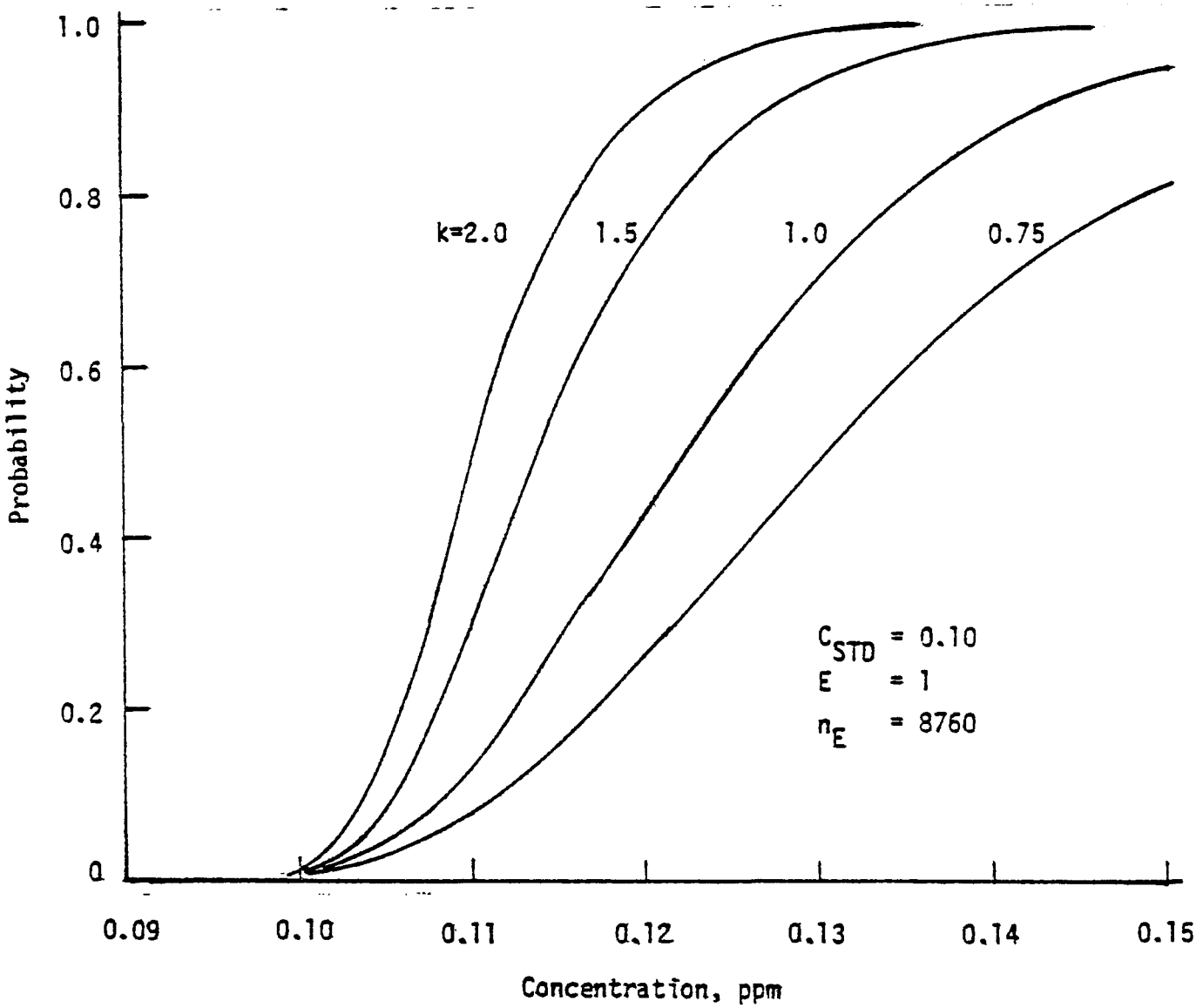
It should be pointed out that from the point of view of the risk assessment method developed in this report, the designation of the expected number of exceedances is a relatively arbitrary matter and could be set at any value that gives a convenient level for C_{STD} so long as the risk is the same (the value of m in $P_C^{(m)}$ of Eq. (11) has a more direct impact on health since it directly bears on the number of exceedances of a true threshold). For example, if it is decided that it is undesirable to have any exceedances of a true threshold over a given time period then the P_C used in the calculation is $P_C^{(1)}$ (see Eq. (10)) and once an acceptable level of risk is chosen any combination of E and C_{STD} values which yields this risk value in a given area give the same level of protection.

The combination C_{STD} and E determine the general location of P_C while k determines its spread. Figure 2-7 (Sec. 2.3) shows the P_C functions for a one-year period for a series of alternate levels of C_{STD} with E = 1 and for the Weibull shape factor k = 1. Figure 2-8 (Sec. 2.3) shows the effect

of changing the shape factor at C_{STD} = 0.1 ppm and E = 1. Figure 3-1 shows the effect of changing k for a P_C covering a 5-year period. It is seen from the three figures that changing C_{STD} displaces the P_C function over a wide range while having a relatively small effect on its shape. Changing k causes little actual displacement of the P_C for a one-year period but has a large effect on its shape. The changing k causes the P_C function in Figure 2-8 to pivot about the point (0.10, 0.368). The effect of k on the P_C for a five-year period (Fig. 3-1) is still largely in the shape of the function, but there also seems to be more displacement. This results from the fact that the pivot point for the distribution is now very close to the concentration axis (0.10, 0.007). In general, calculated risk values will be more sensitive to changing values C_{STD} than to changes in k over the usual ranges of these parameters.

From the preceeding discussion it is seen that the assumption of independence of hours and the use of the Weibull function to represent the distribution of hourly average concentrations readily yield P_C and $P_C^{(m)}$ functions. The Weibull can be used with little loss in accuracy even where other distributions such as the lognormal distribution provides a better fit to the concentration data. The primary concern in estimating the appropriate P_C function is to have it placed properly along the concentration axis and have the correct degree of spread. The parameters δ and k in the Weibull function provide wide flexibility in this regard. As shown above, the standard level essentially places the P_C function. The appropriate values of k can be obtained by fitting a Weibull distribution to hourly average concentration data obtained from air monitoring sites.

Figure 3-1. P_C Function for 5-Year Period for Different Values of Weibull Shape Factor, k



The range of applicable k values for standard-setting purposes should be determined by examining aerometric data in areas that are close to the concentration ranges of the alternate levels of the standard under consideration. Where this is not possible, the k values can be determined at existing levels of air quality and the results extrapolated to potential standard levels.

While using a Weibull distribution where a lognormal or gamma distribution might be more appropriate does not appear to lead to serious errors in the risk estimates, ignoring the possible dependence of hourly average concentrations can lead to significant error. An internal EPA study (25) showed that dependence of one hourly concentration on the value of another did not lead to serious errors if independence of hours was assumed. However, it also showed that dependence of concentrations on time of day or year can lead to P_c functions which were placed lower on the concentration axis than would be obtained assuming no time dependence. When this time dependence can be modeled it should be possible to generate the corresponding P_c tables of functions and express them mathematically. This was done in the EPA study for the daily maximum hourly average ozone concentration.

Another approach can be taken if the time dependence is such that the maximum concentration tends to occur only within some determinable period and the probability distribution of concentrations is approximately the same for all hours within the time period. In this case all hours outside the time period can be excluded and independence assumed for the hours within the period. A Weibull distribution then could be fit to the hours within the time period to determine the appropriate k values and Eqs. (6) and (12) used to calculate P_c. To the extent that the period under consideration is

also likely to contain the mth highest hourly average concentration, $P_C^{(m)}$ could be obtained using Eq. (10). The term n_E in Equation (12) would be set equal to the number of hours per calendar year of the time period. The term n in Equation (6) would be n_E times the number of years for which the risk is to be estimated. As will be discussed in Section 4, this procedure was applied to ozone.

4.0 Application of the Risk Assessment Method to Ozone

4.1 Introduction

The risk assessment method described in the previous sections has been developed during EPA's review of the Photochemical Oxidant NAAQS, which EPA proposes to rename the Ozone NAAQS. The initial application of the method has been to ozone. In order that the public be made aware of the method and some preliminary results of its application, a preliminary report was issued on January 6, 1978. Since that time two more health experts have contributed judgments, the question of the number of threshold-independent categories has been investigated further, and more suitable shape parameters for the P_c distributions have been chosen. All of these changes affect some or all of the risk estimates. The risk estimates are now presented in risk ribbon diagrams, as well as in tables as averages.

This section presents the final results of the ozone risk assessment. Subsection 4.2 presents the judgments of the health experts, including their subjective probability distributions. Subsection 4.3 presents the method used to derive the P_c probability distributions for ozone. Subsection 4.4 presents the results of the risk assessment.

4.2 The Judgments of Health Experts

Judgments were elicited from the nine health scientists listed in Table 4-1. There is a consensus that there are at least two threshold-independent health effect categories for ozone. The category of reduced resistance to bacteria infection was judged to be threshold independent of the remaining health effects of ozone. The evidence on this category of effect is mainly from animal toxicological studies, so the three toxicologists contributed judgments for that category.

TABLE 4-1

HEALTH EXPERTS PARTICIPATING IN THE ANALYSIS

Dr. David Bates	Clinical Investigator
Dr. Robert Carroll	Epidemiologist
Dr. Robert Chapman	Epidemiologist
Dr. Timothy Crocker	Toxicologist
Dr. Richard Erlich	Toxicologist
Dr. Bernard Goldstein	Toxicologist
Dr. Jack Hackney	Clinical Investigator
Dr. Steven Horvath	Clinical Investigator
Dr. Carl Shy	Epidemiologist

The Advisory Panel on Health Effects of Photochemical Oxidants, several of whose members supplied distributions for this assessment, has advised that the information in the revised criteria document indicates a one-hour averaging time represents a satisfactory estimate of the exposure duration which a primary ozone or oxidant NAAQS should protect against. Therefore, the health effect categories considered represent short-term effects only. Most of the evidence for the remaining short-term effects of ozone is from clinical and epidemiological studies. Hence, the six clinical investigators and epidemiologists contributed the judgments on these effects.

There is not a consensus on how many threshold-independent categories there are of the remaining short-term effects. Two experts felt that there are three additional threshold-independent categories. For one of these two experts the three additional categories are: reduction in pulmonary function; cough, chest discomfort, and irritation of mucous membranes of nose, throat, and trachea; and aggravation of asthma, emphysema, and chronic bronchitis. The other agrees, except that he feels emphysema and chronic bronchitis belong in the same category with cough, chest discomfort, etc. One expert judges that there are two additional threshold-independent categories; he would group the effects similar to the way the first expert who judges there to be four categories does, except he groups reduction in pulmonary function and aggravation of asthma, emphysema, and chronic bronchitis into the same category. Three experts feel that all of the remaining health effects are threshold interdependent. In summation, it is

uncertain whether there are two, three, or four threshold-independent categories for ozone.

Judgments were elicited from three experts for each of the following four categories: (1) reduction in pulmonary function; (2) cough, chest discomfort, and irritation of mucous membrances of nose, throat, and trachea; (3) reduced resistance to bacterial infection; and (4) aggravation of asthma, emphysema, and chronic bronchitis. The secondary uncertainty about how many of these categories are threshold independent is taken into account in the risk estimates. Of the nine sets of judgments elicited for categories (1), (2), and (4), two each were contributed by Drs. Bates, Chapman, and Shy, one each by Drs. Carroll, Hackney, and Horvath.

In this initial application of the method several subjective probability distributions were elicited from Drs. Shy and Goldstein before the final scheme for matching special fields of expertise to categories was determined. Several of those distributions did not fit the scheme arrived at later, and thus have not been used. It is contrary to the spirit of the method, of course, to encode more distributions than are intended to be used, and then select the ones to include in the assessment. When those distributions which do not fit the scheme arrived at later were encoded, they were intended to be used. In future applications the scheme will have been determined before judgments are elicited.

The definition of conditions of exposure will improve with time. They were not as precise as described in section 2.2.7 in two respects for the application to ozone. First, although the experts were asked to make their judgments for the ambient air of an average

United States city, to take additive effects from other pollutants in the air into account and to not double-count for other NAAQS pollutants, ambient levels to assume for other NAAQS pollutants were not specified. Second, the way in which adaptivity should be incorporated into their judgments was not specified precisely.

Although care was taken to emphasize that their best judgment, not the easiest-to-rationalize judgment, was desired, the health experts were encouraged to verbalize their thoughts about how they came to their probability judgments. These comments were noted and are used to sketch a very rough picture of the reasoning behind each expert's distribution.

(1) Reduction in Pulmonary Function

Three clinical investigators contributed subjective probability distributions for health effect category 1, reduction in pulmonary function. The amount of decrement in pulmonary function considered a health effect(as measured by percentage reduction in forced expiratory volume (FEV) response)was described by Expert B as 2-3% above the noise level, which will be different for different groups, and by Expert A as just above noise level. The percentage changes in mind when judgments were made were roughly 5-10% for Expert A, 15% for Expert B, and 10% for expert C. By consensus, asthma and infrequent exposure to high ozone levels were two characteristics of the most sensitive group. Expert A further characterized the most sensitive group as exercising, particularly those exercising at high altitudes. Expert B further characterized the most sensitive group for whom data was available to him, as a working population of asthmatic subjects. Expert C further characterized the most sensitive group as young (children) or elderly.

None of the three experts felt that one occurrence of the effect is serious; one occurrence is reversible, even for the most susceptible group. However, Expert A expressed concern about any impairment of functioning, and Expert C stressed that the seriousness goes up rapidly with the frequency.

The three subjective probability distributions elicited for reduction in pulmonary function are given in Figure 4-1. Although at most concentrations experts B and C differ some in their assessment of the probability that the health effect threshold is less than the given concentration, the trend of their judgments is close compared to those of expert A. The three medians (0.5 probability) are approximately 0.075 ppm for expert A, 0.175 ppm for expert B, and 0.18 ppm for expert C. Experts B and C assign a probability in the 0.03-0.05 range to the proposition that the threshold is less than 0.1 ppm; expert A assigns a probability of about 0.97 to the 0.18 ppm concentration which is roughly the median for both B and C.

In making his judgments, Expert A mentioned the results of DeLucia and Adams (1977) on the effects of ozone on exercising individuals. Experts B and C did not give the epidemiological studies reviewed in the criteria document and the von Nieding, et. al. (1976) clinical study very much weight. In light of the questions that have been raised, they will be skeptical about the von Nieding results until they are replicated. Expert B mentioned that he has noticed hints of an effect at about 0.2 ppm in his investigations of asthmatic subjects; taking variation in susceptibility into account, he estimated a median of 0.175 ppm. Expert C estimated that it would take about a two-hour exposure to 0.37 ppm to cause a health effect in normal people, about a one-hour exposure to 0.37 ppm

INTERVIEWER _____ SUBJECT _____ DATE _____

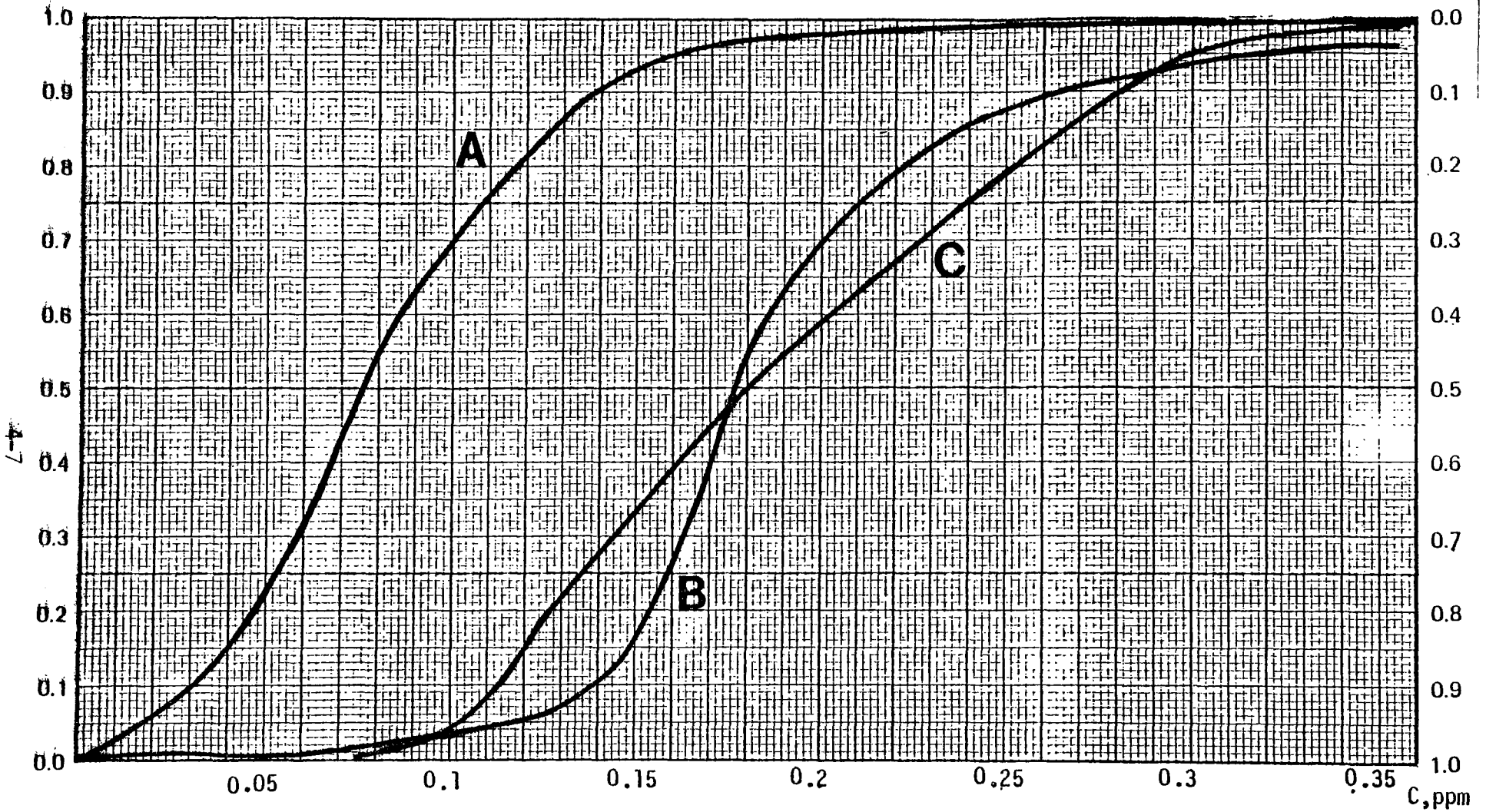


Figure 4-1. Reduction in Pulmonary Function (Experts A,B,&C)

VARIABLE _____

COMMENT _____

_____

to cause a health effect in very sensitive asthmatics such as those for whom the threshold is defined in this assessment.

(2) Cough, Chest Discomfort, and Irritation of Mucous Membranes of Nose, Throat, and Trachea

One clinical investigator (Expert A) and two epidemiologists (Experts B and C) contributed subjective probability distributions for health effect category 2. A health effect for this category was defined to be a coughing spell, sore throat, etc. sufficient to cause discomfort. By consensus of all three experts, the most susceptible group is exercising children. Expert C added that healthy children are probably the most sensitive since their nervous system would be the most acute. By consensus, category 2 is the least serious of the four categories of health effects; one occurrence is reversible.

The three subjective probability distributions elicited for category 2 are given in Figure 4-2. The three medians are about 0.13 ppm (Expert C), 0.15 ppm (Expert B), and 0.18 ppm (Expert A). Expert A assigns a probability of about 0.97 to the threshold being in the 0.1 ppm to 0.25 ppm concentration range; Expert B assigns a probability of about 0.95 to the threshold being in the 0.09 ppm to 0.3 ppm concentration range; and Expert C assigns a probability of only about 0.90 to the threshold being in the 0.05 ppm to 0.4 ppm concentration range. Hence, the three experts not only differ in the relative weight they put on various relevant studies, but also in the absolute weight they put on the body of evidence that is available.

INTERVIEWER _____ SUBJECT _____ DATE _____

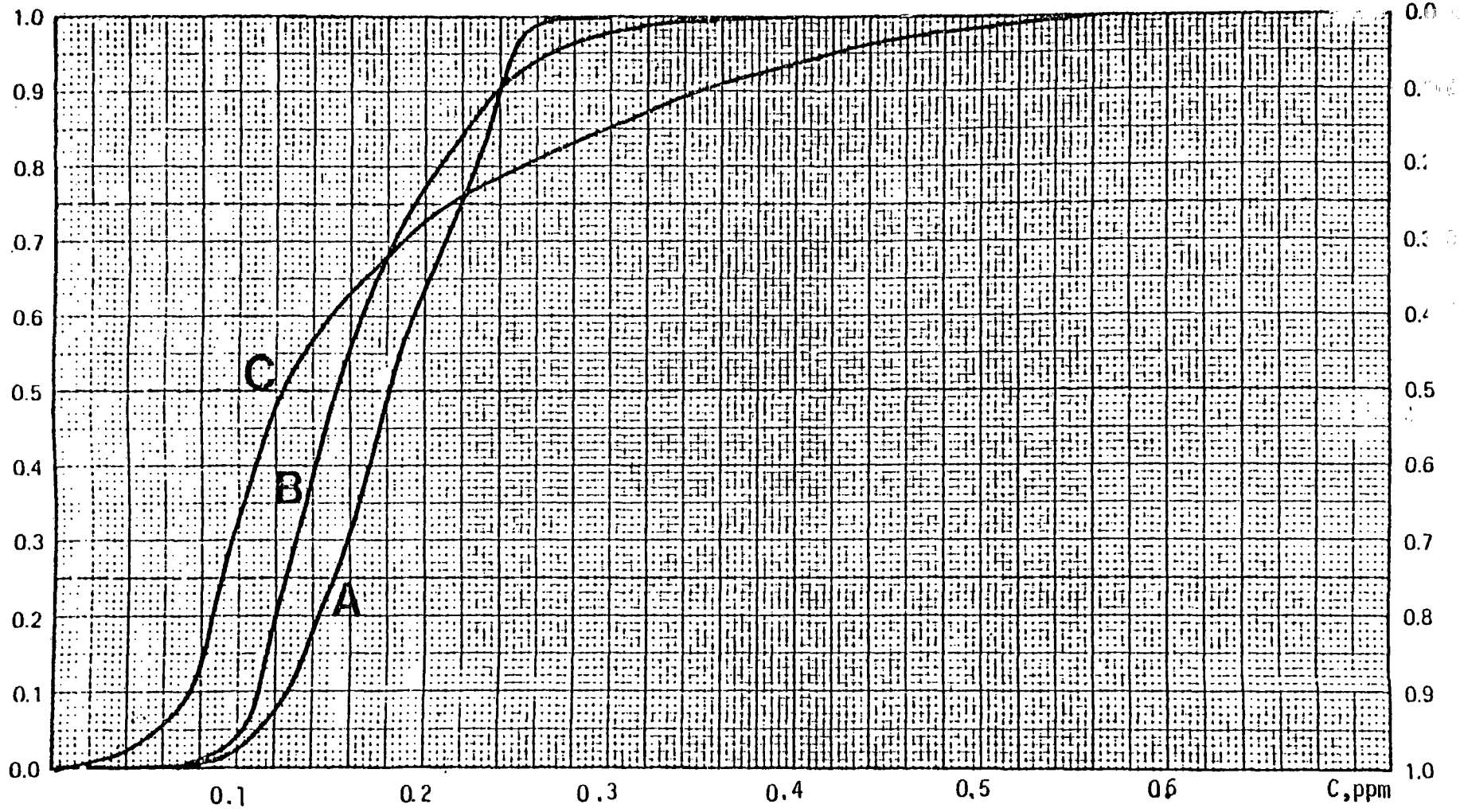


Figure 4-2. Cough, Chest Discomfort, and Irritation of Mucous Membranes of Nose, Throat, and Trachea (Experts A,B,&C)

VARIABLE _____

COMMENT _____

Different relative weight was put on the experimental findings of Bates and Hazucha (1973), Hammer's study of student nurses (1974), and the results of several Japanese epidemiological studies (Shimizu, 1975; Makino and Mizoguchi, 1975; Japan Environment Agency, 1976). Possible additive effects had to be factored into the weighting of these studies because the Hammer study and most of the Japanese results were for oxidants, whereas the Bates and Hazucha findings were for ozone. Also, the subjects of the Hammer and the Bates and Hazucha studies were young adults, whereas the subjects of the Japanese studies were school children.

All three experts considered the Japanese studies significant; they also found their implications hard to assess since total oxidants were measured, the relative contribution of other pollutants is not clear, and in some cases group dynamics may have been a factor. In making their judgments, Expert C, an epidemiologist, gave the Japanese studies the most weight and Expert A, a clinical investigator, gave them the least.

(3) Reduced Resistance to Bacterial Infection

The subjective probability distribution of three toxicologists was used for category 3, reduced resistance to bacterial infection. Most of the scientific basis for the health effect is a set of toxicological studies on animals. Experts A and C define the health effect for this category to be an increased incidence of bacterial infections in humans. Expert B defined the health effect to be an increased incidence of bacterial infections in humans or an increase in the severity of already occurring infections.

By consensus, the most susceptible group for category 3 is young children. Young children do not have fully developed lungs and immunological protection. Expert A characterized the most susceptible group more finely as: (a) prematurely born, since the lung development of prematurely born children lags behind that of the normal child until the age of about five; (b) not asthmatic; hence, airways are in good condition, and they will breathe deeply; (c) exercising vigorously.

Reduced resistance to bacterial infection itself was not described as being serious, but concern was expressed about: (a) increased severity of bacterial infections; (b) the obvious consequence to some of the people who have their resistance to bacteria infections reduced, namely, a bacterial infection they would not have had otherwise; and (c) the possibility of an increase in the risk of other, even more serious, health consequences due to the existence of the resulting bacterial infection. Expert B felt that the very old would belong to the group that is most susceptible in this sense.

Expert A's subjective probability of the health effect existing in humans was the lowest, namely 0.3. He believes that concentrations of ozone high enough to injure macrophages (which are a vital part of the body's defense mechanism against bacterial infection) may never reach the alveoli (air cells of the lung) of humans in commonly occurring ambient situations. First, he observed that there are anatomic differences between human and rodent lungs such that the exposure required to achieve the same dose to the alveoli is larger for humans. Second, he conjectured that irritation to the upper respiratory tract might be great enough to make most young children stop exercising before the dose necessary to cause the effect ever reached the lower lung. In support of this

conjecture, he referred to "a piece of work done at Riverside, California, in the 1960's" in which mice stopped running after a certain amount of exposure to ozone.

Because of the anatomical difference in the lungs of rodents and humans, and because of the difference between the conditions of actual human exposure in the ambient atmosphere and the conditions of exposure in the Gardner, et. al., infectivity model, Expert A felt it would take much higher concentrations than 0.1 ppm to cause the effect in humans, if it indeed caused the effect at all.

Expert B expressed a very different point of view on the probability of existence of the effect in humans. His subjective probability of existence in humans is 0.95. He expressed confidence that the lower lung of humans would receive the dose necessary to cause the effect if the exposure concentrations is high enough. This confidence was partially based on the results of the mathematical modeling approach to estimating ozone uptake in the deep lung done by Miller (Ph.D. Thesis). Some of the differences between Experts A and B on this question may be due to the fact that Expert B explicitly introduced an increase in severity of an already occurring bacterial infection into his definition of the health effect, whereas Expert A did not.

As can be seen in Figure 4-3, Expert B also has a very different view of the probability that the threshold is below various concentrations. His median is 0.11 ppm. He feels that the threshold for very susceptible humans is most likely about the same concentration as the concentration (0.1 ppm) which has been found to cause the effect in rodents.

INTERVIEWER _____ SUBJECT _____ DATE _____

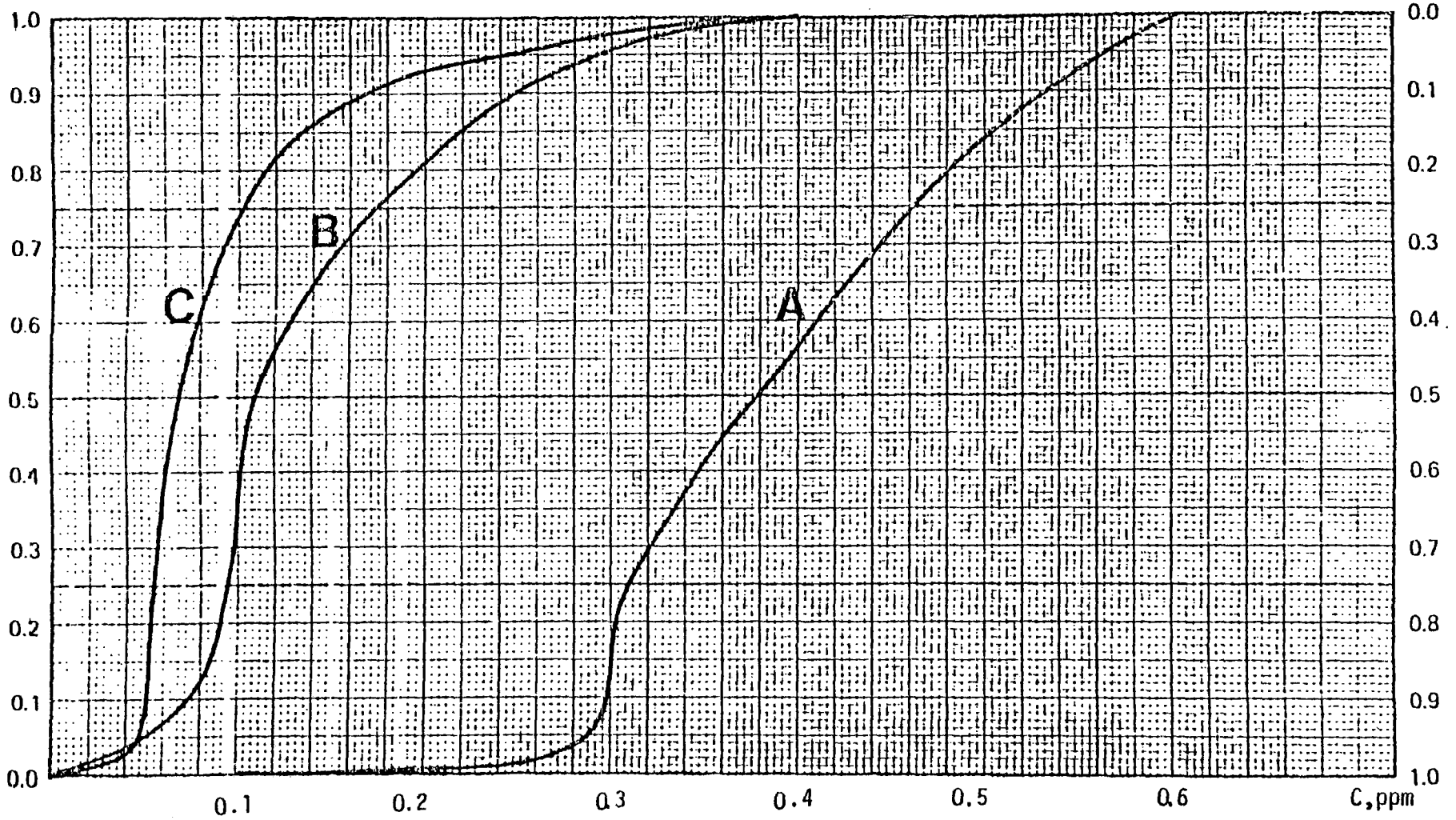


Figure 4-3. Reduced Resistance to Bacterial Infection (Experts A,B,&C)

VARIABLE _____

COMMENT _____

Expert C's subjective probability that ozone contributes to the effect in humans is 0.5. He feels that if the effect does occur in humans, the threshold for its occurrence in very susceptible individuals is most likely to be even less than the 0.1 ppm concentration which apparently will cause the effect in rodents. His median is about 0.07 ppm. He thinks there is a significant possibility that the threshold is as low as about 0.04 ppm. But, he acknowledges the possibility that the threshold might be as high as 0.35 ppm, either because humans are different than rodents or because, despite their apparent validity, the results at the 0.1 ppm level are misleading for rodents themselves.

(4) Aggravation of Asthma, Emphysema, and Chronic Bronchitis

Three epidemiologists have contributed distributions for health effect category 4, aggravation of asthma, emphysema, and chronic bronchitis. For expert B aggravation of asthma is threshold independent of aggravation of emphysema and chronic bronchitis. He groups aggravation of emphysema and chronic bronchitis with the effects of category (2). Expert C feels the evidence for this category is sparse, and that what little evidence there is applies mainly to aggravation of asthma.

A health effect for this category was defined to be an aggravation of one of the three lung diseases. It was observed that such aggravation not only increases discomfort, but also affects the individual's functioning, which is already restricted, and can have more serious irreversible consequences. The effect was described as serious. Expert C remarked that

very mild asthmatic attacks which are very easily triggered in sensitive asthmatics are clearly distinguishable from the severe attacks which are serious. He made his judgments on the latter.

Expert A judged the most sensitive group to be individuals who have asthma or emphysema. Experts B and C judged asthmatics to be the most sensitive group.

In figure 4-4 it can be seen that all three experts think it very unlikely this effect occurs below 0.06 ppm; expert A mentioned the negative result at that concentration of Rokaw and Massey (1962). Expert A mentioned studies by Molley, et. al. (1959), Remmers and Balcham (1965), and Schoettlin and Landau (1961) in coming to his probability judgments; he also mentioned that there are problems in interpreting them, especially the study by Schoettlin and Landau. Expert C also mentioned that the Schoettlin and Landau study was a basis for his judgments, despite his reservations about the study, because in his opinion, there is very little evidence other than that study to base judgments on.

The median for experts A and B is about 0.14 ppm; this may reflect the fact that Remmers and Balcham found a beneficial effect of air filtration in studies of four exercising patients at an ozone concentration of 0.13. Expert C's median is much higher at 0.25 ppm. Experts A and C remarked that there is a great deal of uncertainty about where the threshold is for this effect, and their spread-out distributions reflect this view. Their probability judgments that the threshold is less than 0.30 ppm are only 0.80 and 0.72, respectively. Expert B, whose distribution is similar to Expert A's up to the median, is almost sure the threshold is less than 0.30 ppm.

INTERVIEWER _____ SUBJECT _____ DATE _____

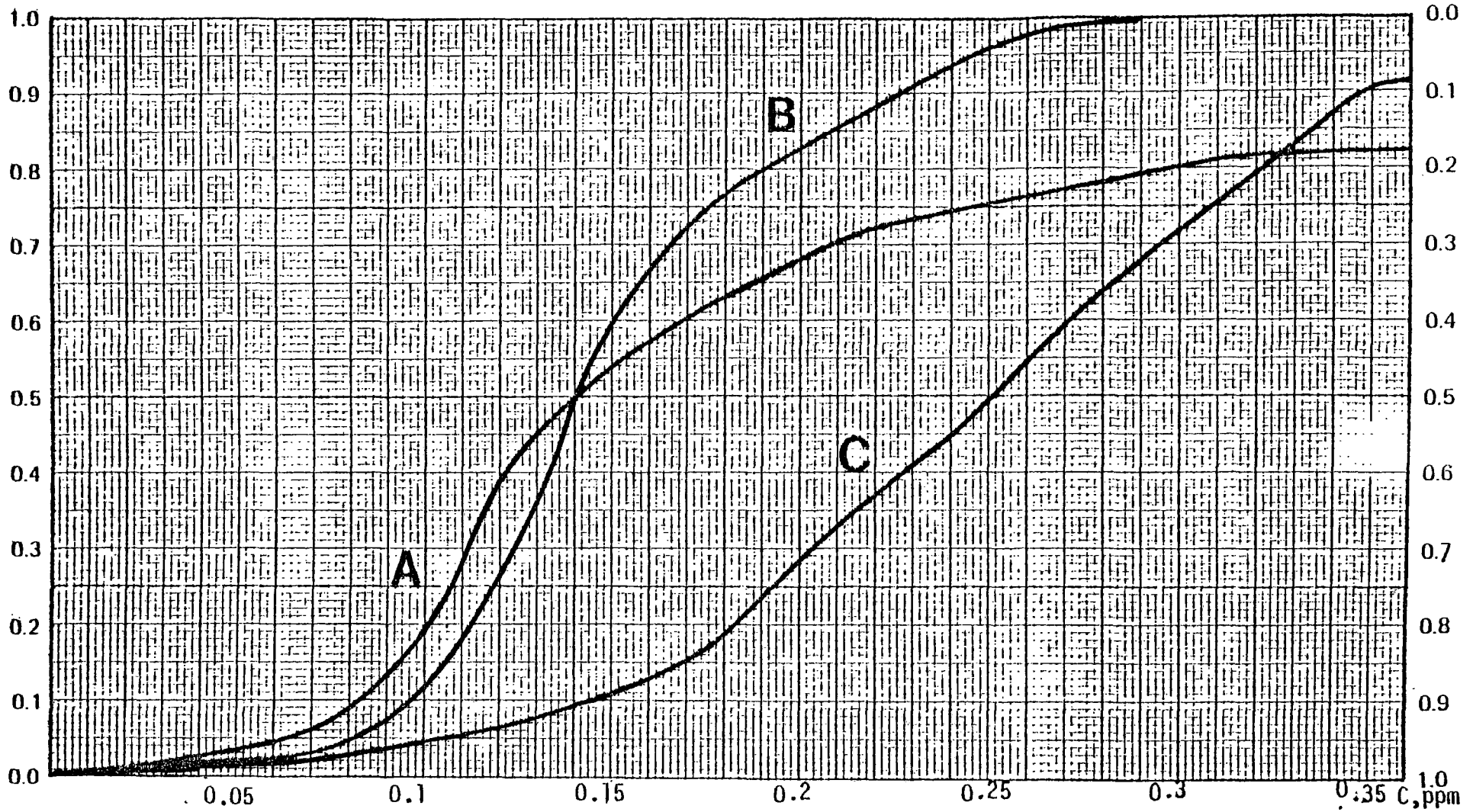


Figure 4-4. Aggravation of Asthma, Emphysema, and Chronic Bronchitis (Experts A,B,&C)

VARIABLE _____

COMMENT _____

4.3 Determination of P_C Functions for Ozone

As suggested in Sections 2 and 3 estimating the appropriate P_C function involves: 1. determining the mathematical function or functions which best describe the distribution of time-averaged ambient concentrations of the pollutant; 2. given the distribution of time-averaged concentrations, the standard, and the nature of the correlation between concentrations in neighboring time periods and dependence upon time of day and year, deriving suitable P_C functions; and 3. estimating the range of values of parameters appearing in the P_C function.

A study performed under contract for EPA involving 14 sites scattered around the United States and involving 22 site-years of data showed that the Weibull Distribution (Eq. (11), Section 3.2) provides an excellent fit to hourly ozone concentrations(17). In only two cases did a lognormal distribution give a superior fit.

Ozone hourly average concentrations exhibit strong correlation and strong dependence on time of day and time of year. The day of the week also has some effect. As indicated in Section 3, the correlation between neighboring concentrations does not appear to have an important effect on the P_C function; however, the time dependence does.

The method of dealing with the time dependence discussed in Section 3 was to find a time period in which the time dependence was relatively constant and which was highly likely to contain the maximum hourly average concentration for the time period. If such a time period existed and a distribution could be fit to the time-averaged concentrations within this period, then the appropriate P_C function could be derived by assuming

complete independence of the time-averaged concentrations. For ozone this period is during the midday hours of the warm months of the year.

In a continuation of the above-mentioned study (17) it was found that Weibull distributions could be fit to data obtained between 11 AM and 6 PM both from May through September and July through August. It was further shown that the maximum ozone concentration had a significantly greater chance of occurring during the longer period. Therefore, this time period, which contains 1071 hours, was used in the derivation of the P_C function.

The form of the National Ambient Air Quality Standard proposed for ozone is: C_{STD} ppm hourly average concentration with an expected number of exceedances per year less than or equal to one. For a region whose air quality just meets this standard, the Weibull distribution of hourly averages for the hours 11 AM to 6 PM, May through September, would be according to Eq. (12) in Section 3.3:

$$G(C) = e^{-(\ln 1071)(C/C_{STD})^k} \tag{12'}$$

And the P_C function for a period of n_y years would be from Eq. (6) in Section 3.3.

$$P_C = (1-e^{-(\ln 1071)(C/C_{STD})^k})^{1071\, n_y} \tag{6'}$$

By substituting alternate levels, C_{STD}, of the standard into Eq. (6') the P_C function for n_y years needed for the risk assessment can be obtained. However, before this can be done it is necessary to estimate the range of values of the parameter k (Weibull shape factor) over different regions in

the U.S. Table 4-2 shows measured k values for the 1071-hour time period as well as other time periods. The range for the values based on 1071 hours is 1.31 to 2.04.

For the risk assessment best estimates were made of the lower-bound and upper-bound values of k. This was done by probability encoding two researchers involved in the development of the Weibull distributions and the P_C functions. The information base was the data in Table 4-2, plus the data developed during the Weibull distribution studies. The data shown in Table 4-2 are, by and large, for geographic regions above the range of alternative ozone standards considered. The Weibull studies suggest that the k factors at the standard levels would be somewhat higher than those shown in Table 4-2. This factor was taken into account in the encoding. The median values for the location of the lower- and upper-bound shape factors were 1.36 and 2.54 respectively. The distributions obtained in the encoding sessions are shown in Table 4-3. Since the range of k values varies somewhat with the standard level, the range, strictly speaking, should be estimated for each alternative level of the standard. In the case of ozone the difference is not likely to be large enough to seriously affect the risk estimates.

Given the above range of values for the Weibull shape factors, Eq (6') can be used to calculate the P_C functions. Figure 4-5 shows the P_C functions used in succeeding sections to calculate risk estimates for a standard level of 0.10 ppm and an expected exceedance rate of once per year. The time period is five years. That is, the functions in Figure 4-5 can be used to calculate estimated risks of exceeding an ozone health effect threshold one or more times in five years when the standard level is at 0.10 ppm. The functions were calculated

Table 4-2 Change in Weibull Distribution Shape Factors With Changing Time Segment in Which Hourly Average Concentrations are Collected[a]

Site	Year	Shape Factor, k		
		Full Year[b]	May-Sept.[c] 11 AM-6PM	July-Aug.[d] 11 AM-6PM
Kansas City, Kansas	1975	1.24	--	4.34[e]
Des Moines, Iowa	1975	1.92	2.04	2.40
Louisville, Kentucky	1974	0.80	1.66	2.02
Memphis, Tennessee	1974	1.32	2.28	2.34
Mamaroneck, New York	1975	0.80	1.31	1.57
Racine, Wisconsin	1974	1.49	--	1.99[e]

a Reference 17.

b 8760 hours.

c 1071 hours

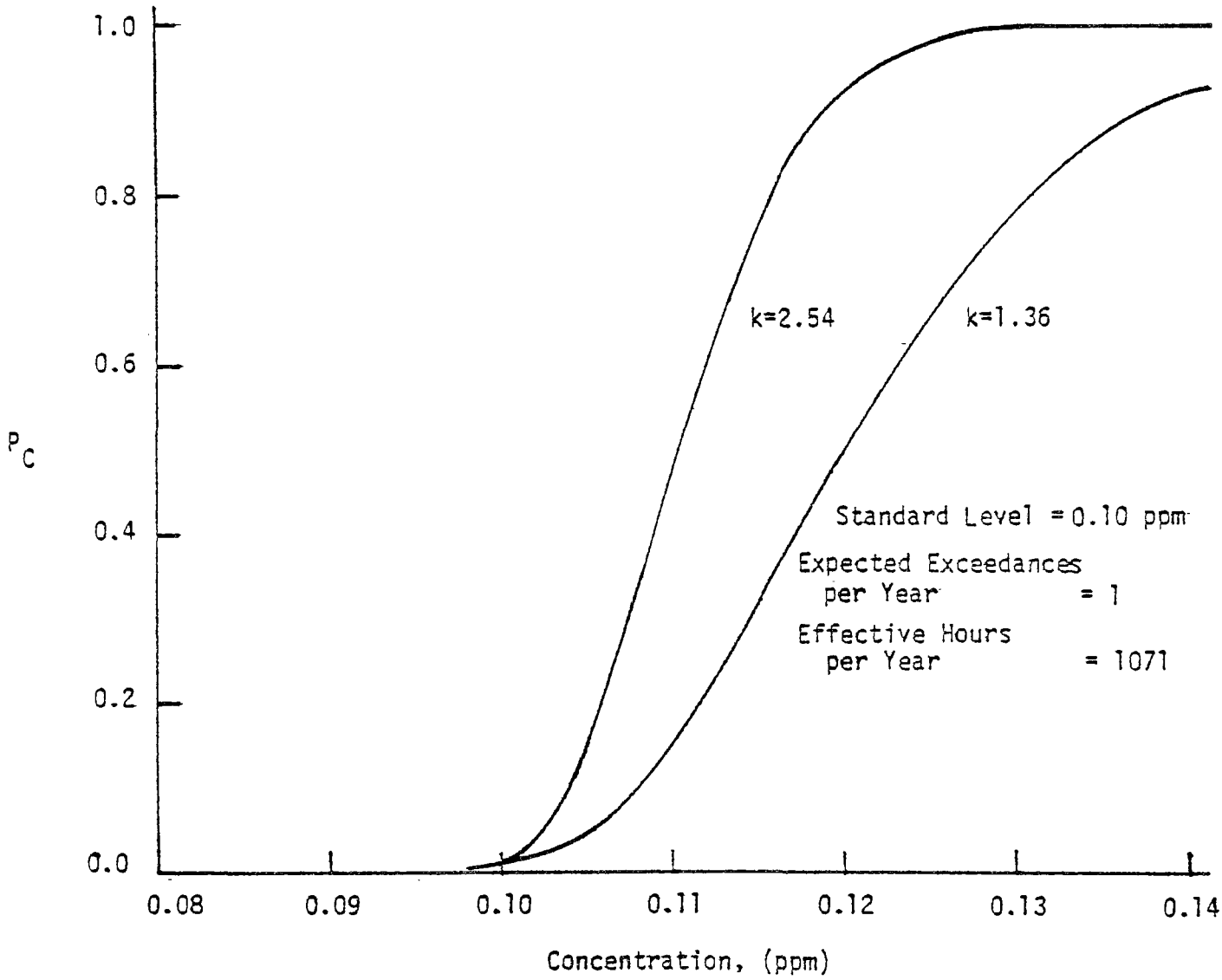
d 434 hours.

e Poor Weibull fit.

Table 4-3 Subjective Probability Distributions for Upper- and Lower-Bound Weibull Shape Factors (k) for Distributions of Hourly Average Ozone Concentrations

Probability that k is below specified value	Lower-Bound k	Upper-Bound k
0.10	1.15	2.41
0.30	1.30	2.50
0.50	1.36	2.54
0.60	1.41	2.66
0.90	1.47	2.98

Figure 4-5 P_C Function for Estimating Risks of Exceeding an Ozone Health Effects Threshold One or More Times in Five Years



using the median lower-bound and median upper-bound shape factors. The two curves in the figure therefore bound the large majority of P_c functions (for the above parameters) that would be encountered in different areas of the U.S.

Because the standard level enters into Eq (6') as a divisor of the concentration, Figure 4-5 can also be used to indicate the P_c function for other standard levels. This is done by relabeling the concentration axis. For example, if the function is desired for a new level, C_{STD}, each number on the concentration axis is multiplied by $C_{STD}/0.10$.

Estimates were also made of the risk of exceeding an ozone health effect threshold five or more times in five years. The corresponding $P_c^{(5)}$ functions for the median upper- and lower-bound shape factors are shown in Figure 4-6. It is seen that these functions are much closer together than those in Fig. 4-5. As a result, the range of risk estimates for five or more exceedances of a health effects threshold in a five-year period will be much smaller than for one or more exceedances in the same period.

In dealing with risks of multiple exceedances of a health effect threshold it should be noted that a risk of five or more exceedances in five years is not the same as the risk of one or more exceedances in one year since the respective $P_c^{(m)}$ functions are different. This point is illustrated in Fig.4-7 which shows a series of $P_c^{(m)}$ functions where m varies from one to five years. It is seen that the "one or more in one" risk will be greater than the "five or more in five" risk.

It is tempting to think of the risk of five or more exceedances of a threshold in five years as equivalent to the risk of an average of one or more exceedances of a threshold per year. This is not the case. An average of one or more exceedance per year would correspond to the limiting $P_c^{(m)}$ for m years when m increases without limit. For the situation depicted in Figure 4-7, this limiting $P_c^{(m)}$ is a step function which is zero for concentrations below 0.10 ppm (the standard level) and one for concentrations of 0.10 ppm and above. Depending upon the health effect threshold with which the $P_c^{(m)}$ is correlated, the

Figure 4-6 P_C Functions for Estimating Risks of Exceeding an Ozone Health Effects Threshold Five or More Times in Five Years

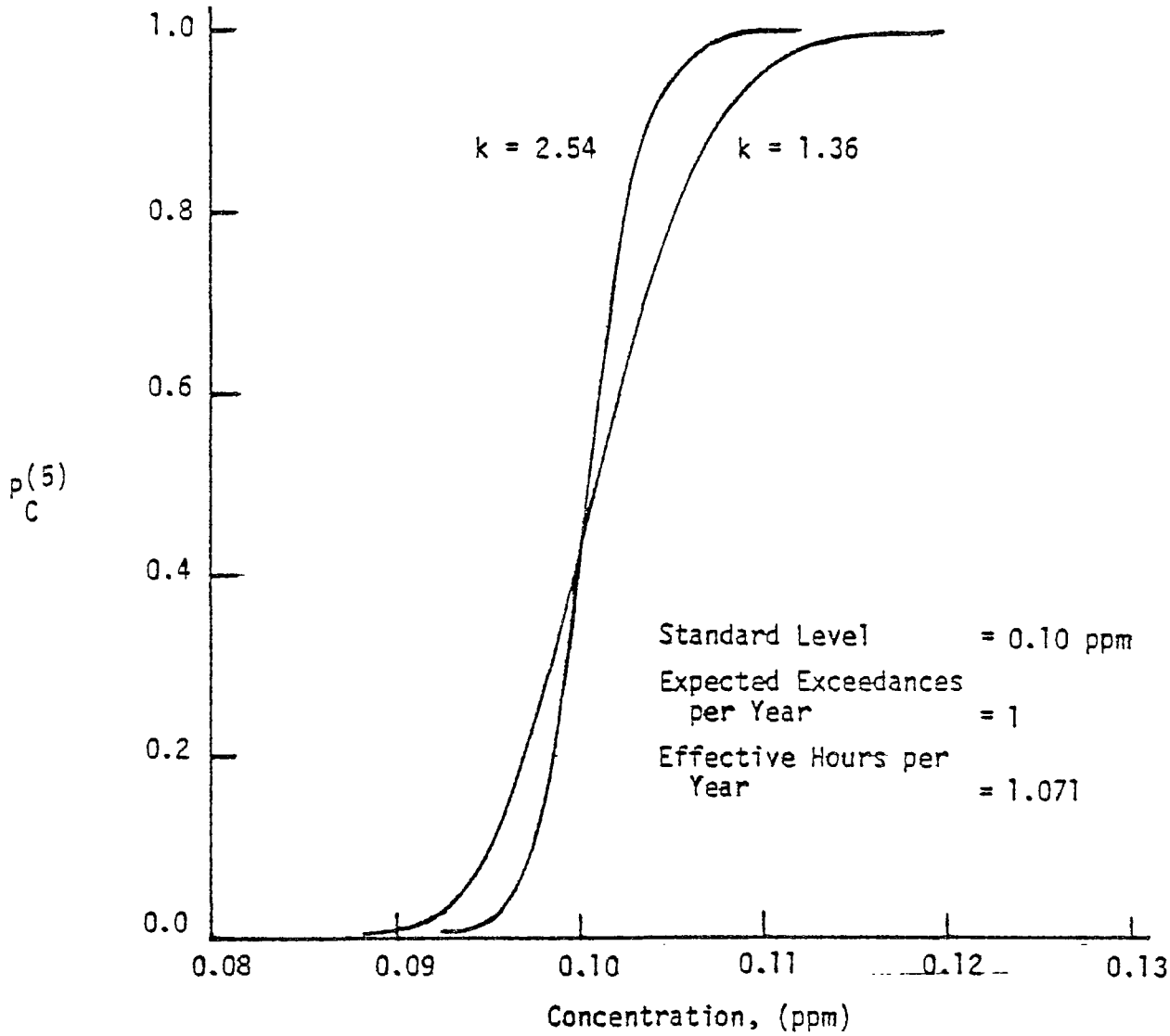
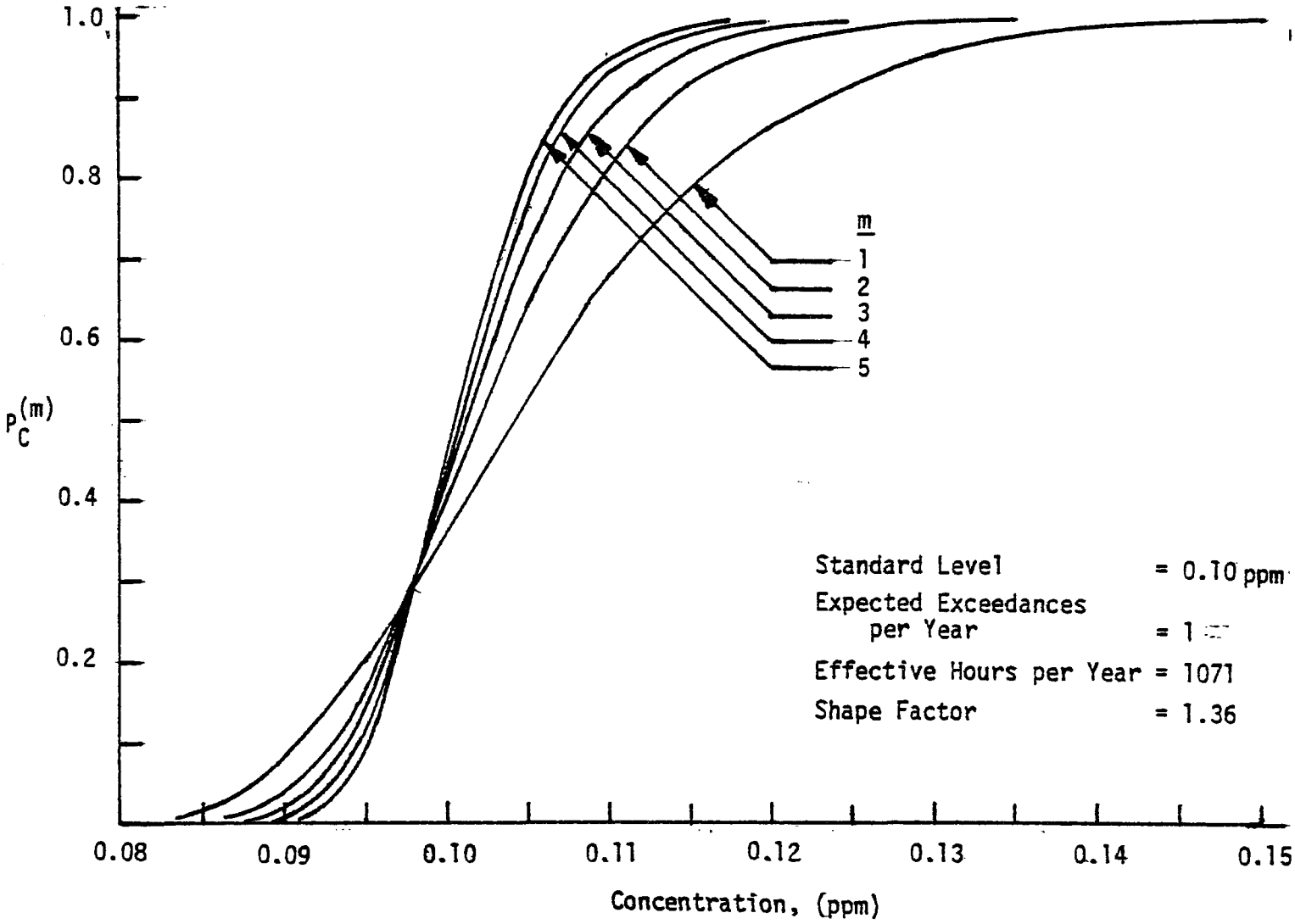


Figure 4-7 P_C Functions for Estimating Risks of Exceeding an Ozone Health Effects Threshold "m" or More Times in "m" Years



average risk can be significantly different from the risks for low values of m. In general, it would give a lower risk than that obtained using the $P_c^{(m)}$ for lower values of m.

4.4 Risk Tables and Risk Ribbons

This section presents the quantitative results of the ozone risk assessment. Table 4-4 presents average threshold risk estimates for several alternative standards. Table 4-4(a) presents estimates of the risk that the health effect threshold will be exceeded one or more times in five years if the given standard is just met. Table 4-4(b) presents estimates of the risk associated with the given standard that the health effect threshold will be exceeded five or more times in five years if the given standard is just met. The average risk estimates are obtained by averaging the risk estimates which result when the single most likely value of the shape parameter for the P_c distribution is used.

Figures 4-8 through 4-17 are a series of risk ribbons. Section 2.4 explains how these ribbons are derived. The plots are ribbons rather than simply curves because, as is explained in section 2.3, the risk varies over the United States. The lower-bound and upper-bound curve for each alternative standard are obtained by estimating the extremes for the shape parameter of the P_c distribution, as is explained in section 4.3.

Figure 4-8 presents the ribbons which are obtained for the two most serious health effect categories when the health effect threshold for the most sensitive group is defined to be the health effect threshold for the least sensitive member of the most sensitive 1 percent of the most sensitive group; figure 4-9 similarly for the least sensitive member of the most sensitive 5 percent of the most sensitive group; figure 4-10 similarly for the least sensitive member of the most sensitive 10 percent of the most sensitive group. The risk ribbons shift toward lower risk values as the definition changes from the 1 percent definition to the 5 percent definition to the 10 percent definition.

Table 4-4(a). Risk that Health Effect Threshold Will be Exceeded 1 or More Times in 5 Years for Alternate Standard Levels*

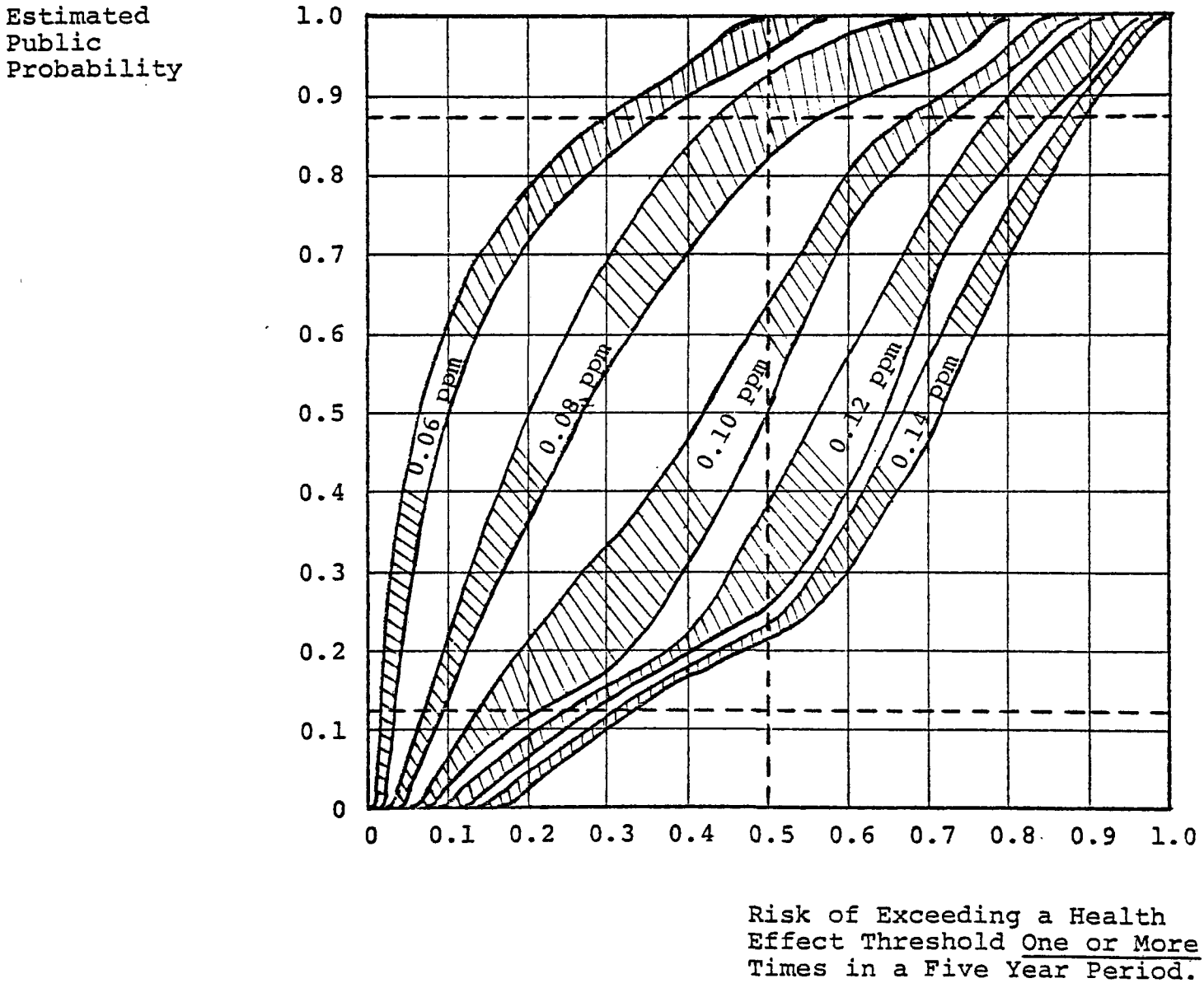
Hourly average standard level (1 expected exceedance per year)	(1)Reduction in pulmonary function	(2)Chest discomfort and irritation of the respiratory tract	(3)Reduced resistance to bacterial infection (animal studies)	(4)Aggravation of asthma emphysema, and chronic bronchitis	Risk of exceeding 1 or more of the thresholds for the individual categories
0.06 ppm	0.14-0.16	0.03-0.05	0.03-0.12	0.03-0.04	0.25-0.30
0.08 ppm	0.22-0.26	0.09-0.14	0.17-0.20	0.10-0.15	0.42-0.50
0.10 ppm	0.31-0.36	0.21-0.27	0.24-0.26	0.22-0.29	0.60-0.67
0.12 ppm	0.41-0.47	0.34-0.42	0.28-0.29	0.36-0.41	0.74-0.79
0.14 ppm	0.52-0.60	0.47-0.56	0.31-0.32	0.45-0.50	0.83-0.87

Table 4-4(b). Risk that Health Effect Threshold Will be Exceeded 5 or More Times in 5 Years for Alternate Standard Levels*

Hourly average standard level (1 expected exceedance per year)	(1)Reduction in pulmonary function	(2)Chest discomfort and irritation of the respiratory tract	(3)Reduced resistance to bacterial infection (animal studies)	(4)Aggravation of asthma emphysema, and chronic bronchitis	Risk of exceeding 1 or more of the thresholds for the individual categories
0.06 ppm	0.11	0.03	0.08	0.02	0.21
0.08 ppm	0.19-0.20	0.06	0.15	0.06	0.36
0.10 ppm	0.27	0.15	0.21	0.16	0.52
0.12 ppm	0.35	0.27	0.26	0.28-0.29	0.67
0.14 ppm	0.44	0.39	0.29	0.40	0.78-0.80

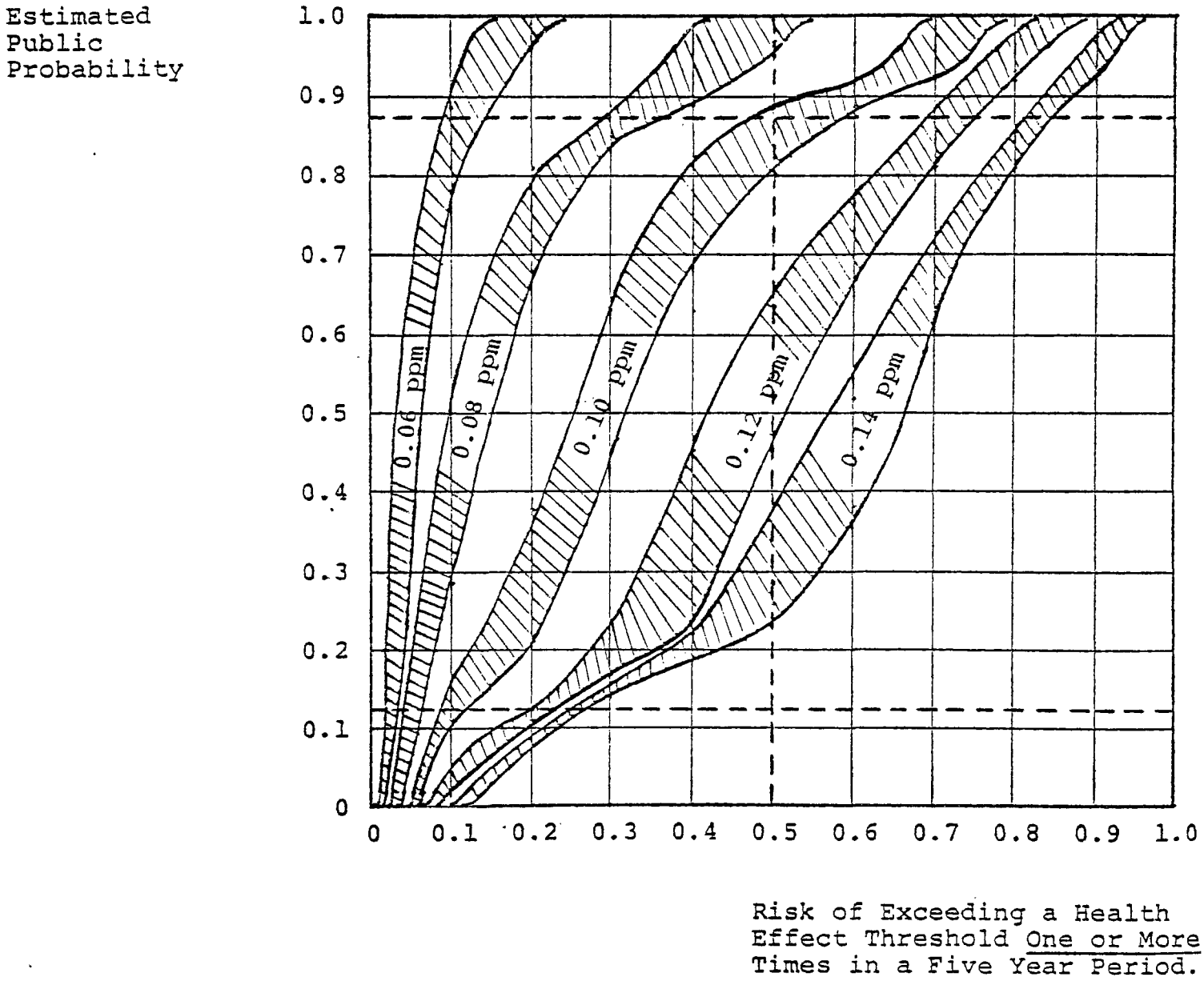
*Risk values expressed as range to reflect estimated variation in risk throughout U.S.

Figure 4-8. RISK OF EXCEEDING THE THRESHOLD* OF AT LEAST ONE OF THE FOLLOWING HEALTH EFFECTS:
- o REDUCED RESISTANCE TO BACTERIAL INFECTION
- o AGGRAVATION OF ASTHMA, EMPHYSEMA, AND CHRONIC BRONCHITIS



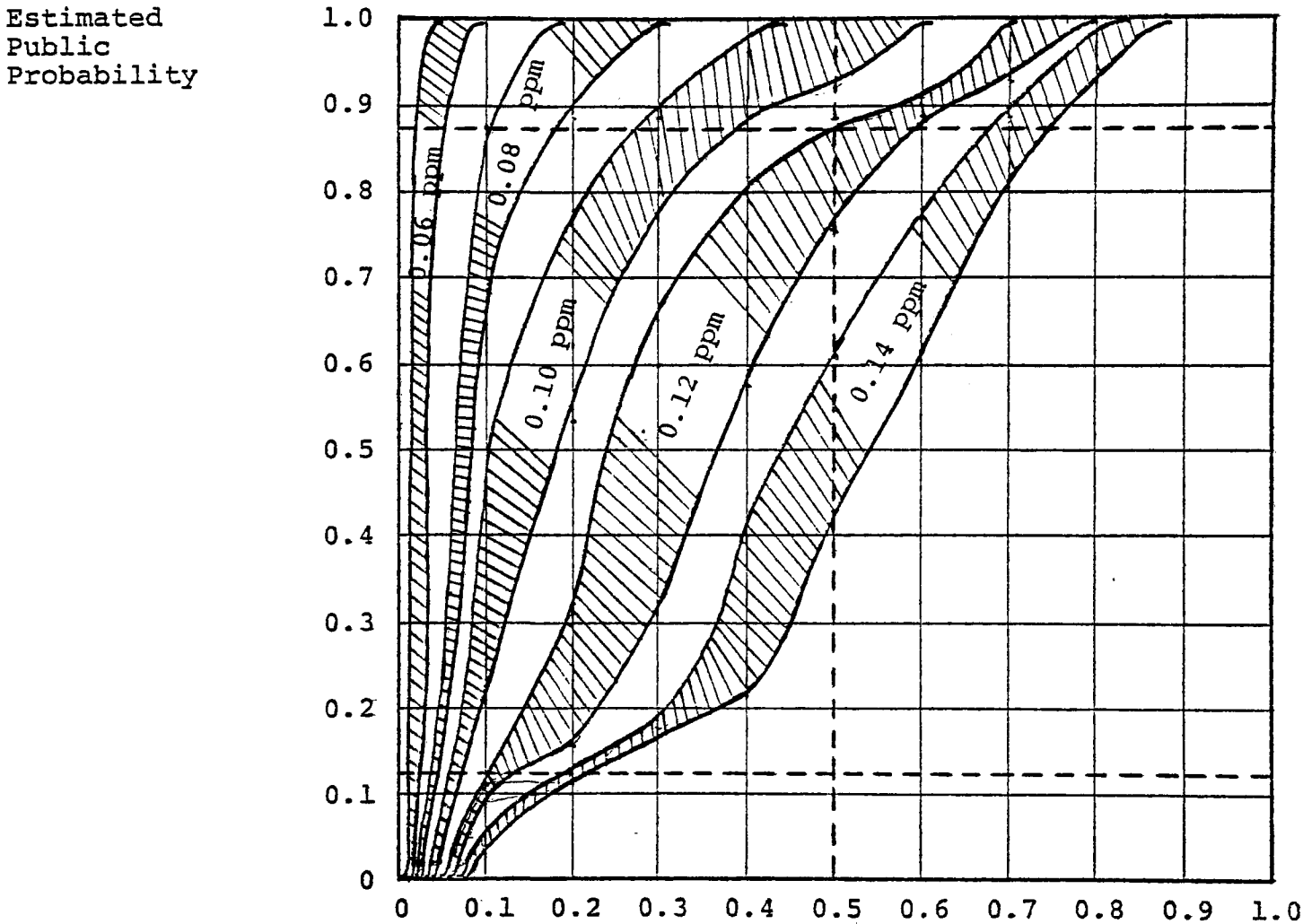
* HEALTH EFFECT THRESHOLD: Concentration at which 1% of group most sensitive to the effect would suffer the effect if the whole group were exposed under specified conditions.

Figure 4-9. RISK OF EXCEEDING THE THRESHOLD* OF AT LEAST ONE OF THE FOLLOWING HEALTH EFFECTS:
- o REDUCED RESISTANCE TO BACTERIAL INFECTION
- o AGGRAVATION OF ASTHMA, EMPHYSEMA, AND CHRONIC BRONCHITIS



* HEALTH EFFECT THRESHOLD: Concentration at which 5% of group most sensitive to the effect would suffer the effect if the whole group were exposed under specified conditions.

Figure 4-10. RISK OF EXCEEDING THE THRESHOLD* OF AT LEAST ONE OF THE FOLLOWING HEALTH EFFECTS:
- o REDUCED RESISTANCE TO BACTERIAL INFECTION
- o AGGRAVATION OF ASTHMA, EMPHYSEMA, AND CHRONIC BRONCHITIS



* HEALTH EFFECT THRESHOLD: Concentration at which 10% of group most sensitive to the effect would suffer the effect if the whole group were exposed under specified conditions.

Figure 4-11 presents the ribbons for the risk of exceeding the health effect threshold for the two most serious categories five or more times in a five year period when the threshold is given the 1 percent definition. Figure 4-12 similarly for the 10 percent definition. The ribbons generally collapse to curves when the risk values are rounded to two decimal places for five or more exceedances because the shape of the P_c distribution for the fifth high hourly average ozone concentration is not as sensitive to changes in its shape parameter k (see section 4.3).

Figure 4-13 presents the risk ribbons for all four health effect categories when the health effect threshold is given the 1 percent definition. These are the only risk ribbons presented which must deal with the fact that there is secondary uncertainty about the number of threshold-independent health effect categories for ozone. See section 2.4 for an explanation of how this type of uncertainty is handled. The weights used were

3 - two independent categories

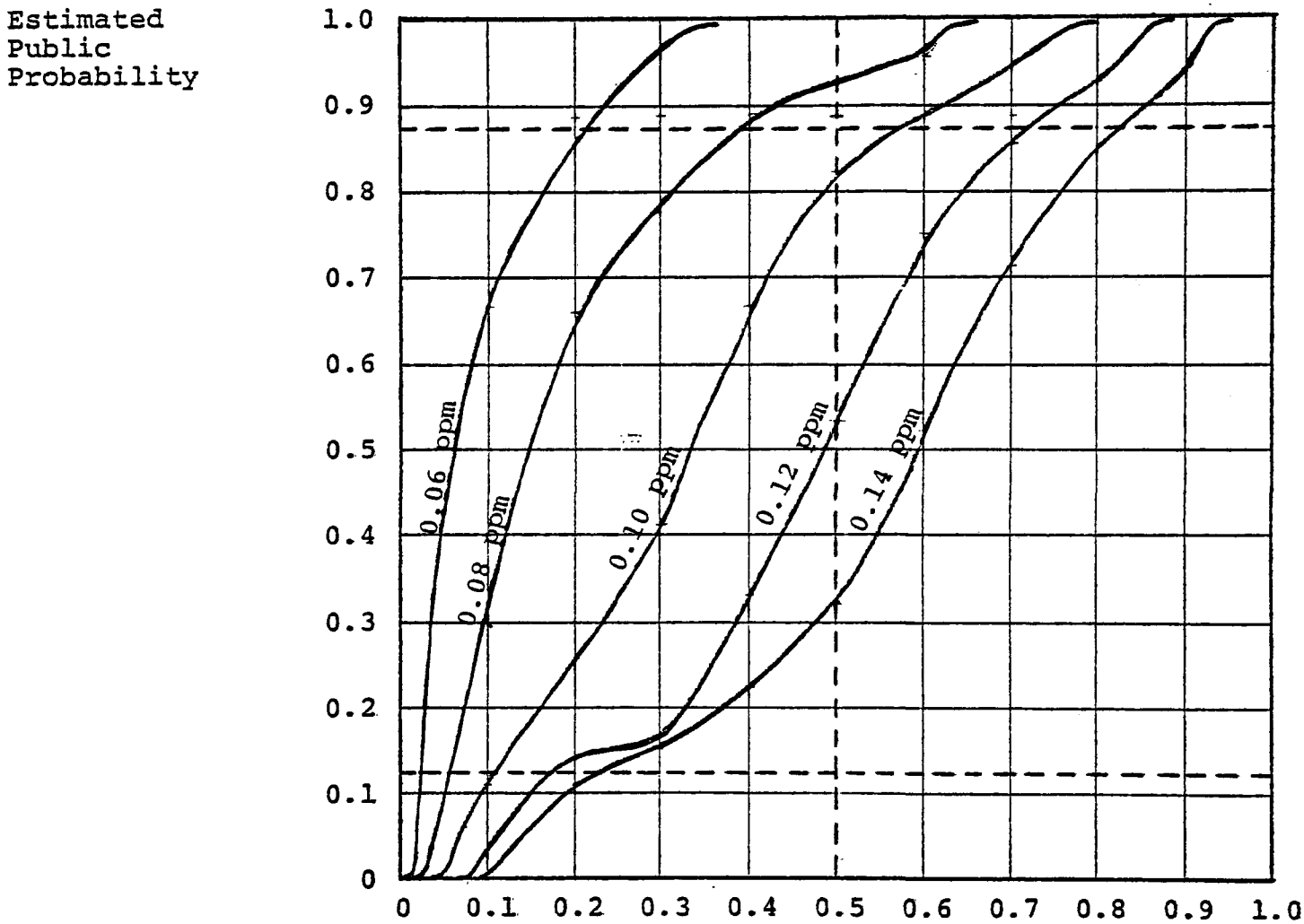
1 - three independent categories

2 - four independent categories

in view of the judgments related in section 4.2.

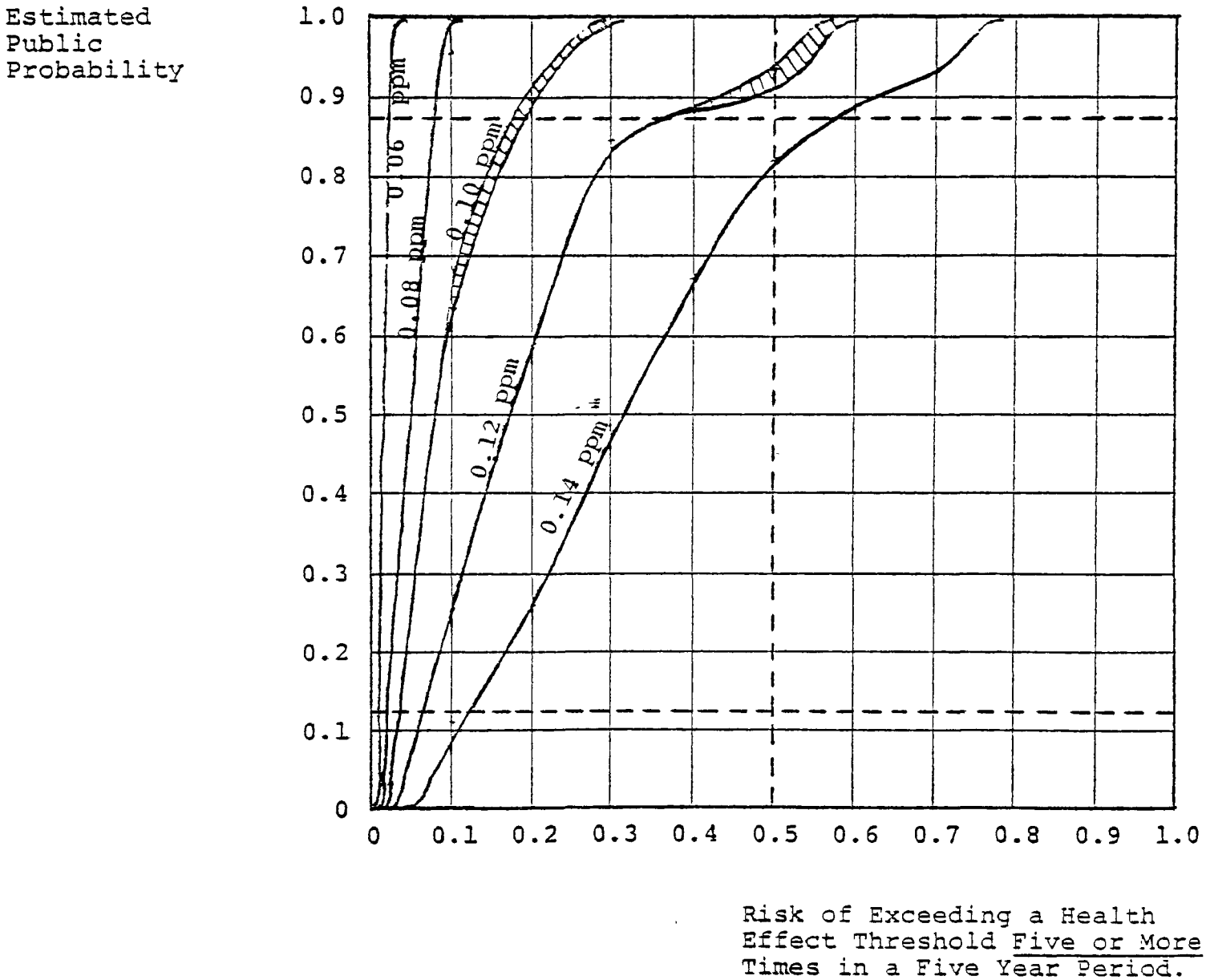
Figures 4-14 through 4-17 present the risk ribbons for the individual categories. They are not as smooth because there are only 15 (= 3 x 5) points being plotted, except for category 3. Category 3 requires a probability-of-existence judgment so 45 (= 3^2 x 5) points are plotted for it.

Figure 4-11. RISK OF EXCEEDING THE THRESHOLD* OF AT LEAST ONE OF THE FOLLOWING HEALTH EFFECTS:
- o REDUCED RESISTANCE TO BACTERIAL INFECTION
- o AGGRAVATION OF ASTHMA, EMPHYSEMA, AND CHRONIC BRONCHITIS



* HEALTH EFFECT THRESHOLD: Concentration at which 1% of group most sensitive to the effect would suffer the effect if the whole group were exposed under specified conditions.

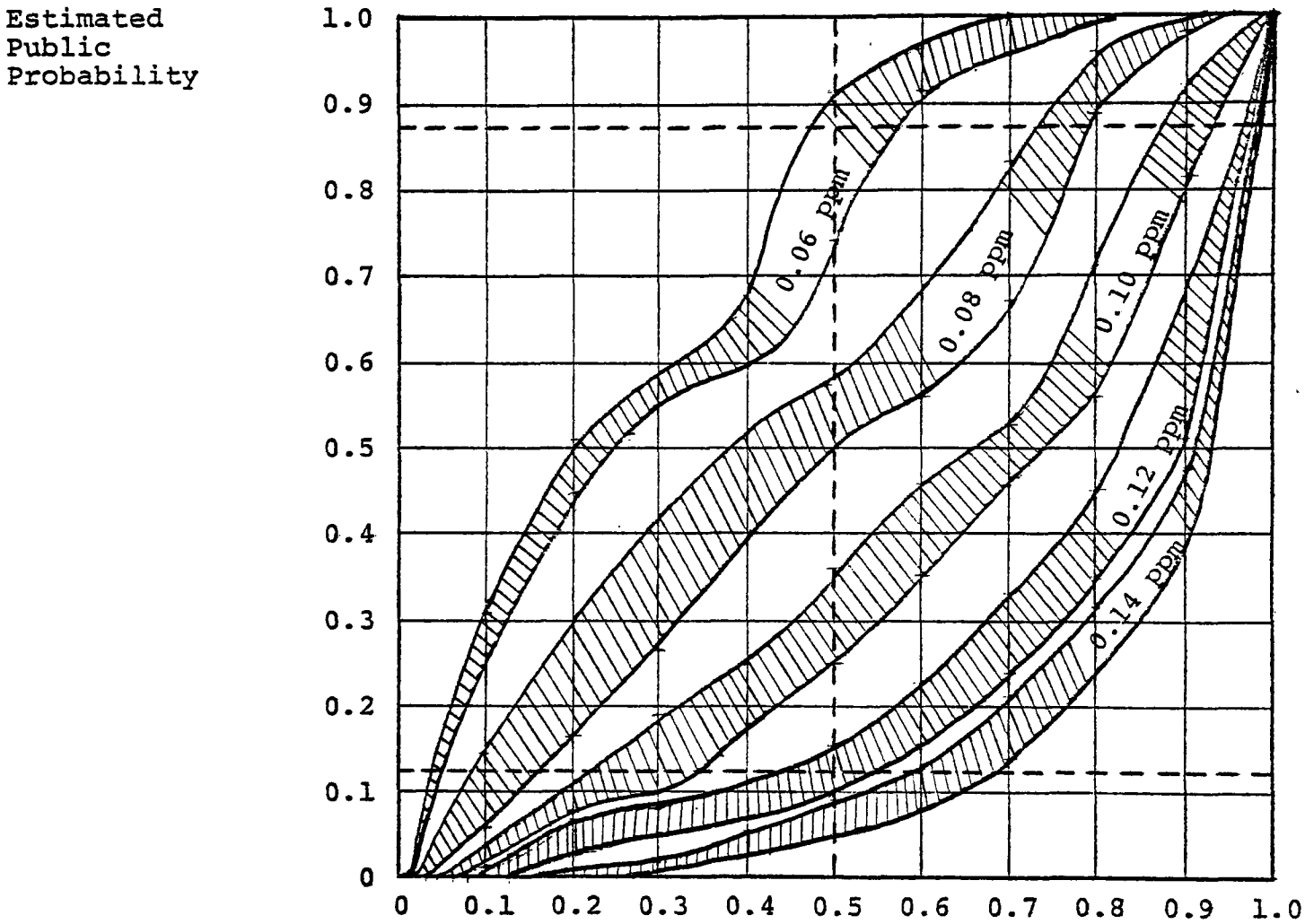
Figure 4-12. RISK OF EXCEEDING THE THRESHOLD* OF AT LEAST ONE OF THE FOLLOWING HEALTH EFFECTS:
- o REDUCED RESISTANCE TO BACTERIAL INFECTION
- o AGGRAVATION OF ASTHMA, EMPHYSEMA, AND CHRONIC BRONCHITIS



* HEALTH EFFECT THRESHOLD: Concentration at which 10% of group most sensitive to the effect would suffer the effect if the whole group were exposed under specified conditions.

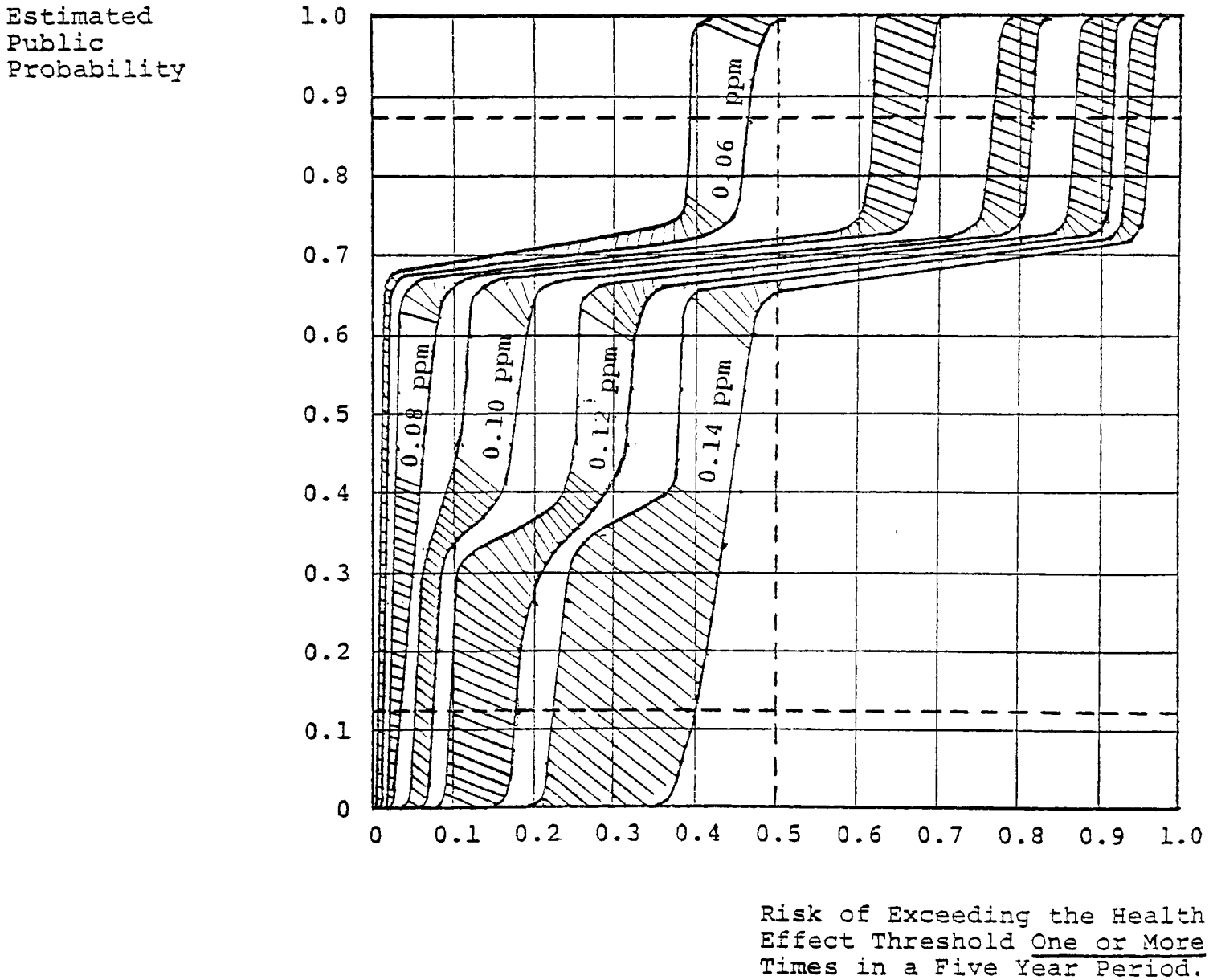
Figure 4-13. RISK OF EXCEEDING THE THRESHOLD* OF AT LEAST ONE OF THE FOLLOWING HEALTH EFFECTS:

- o REDUCTION IN PULMONARY FUNCTION
- o COUGH, CHEST DISCOMFORT, AND IRRITATION OF THE NOSE, THROAT, AND TRACHEA
- o REDUCED RESISTANCE TO BACTERIAL INFECTION
- o AGGRAVATION OF ASTHMA, EMPHYSEMA, AND CHRONIC BRONCHITIS



* HEALTH EFFECT THRESHOLD: Concentration at which 1% of group most sensitive to the effect would suffer the effect if the whole group were exposed under specified conditions.

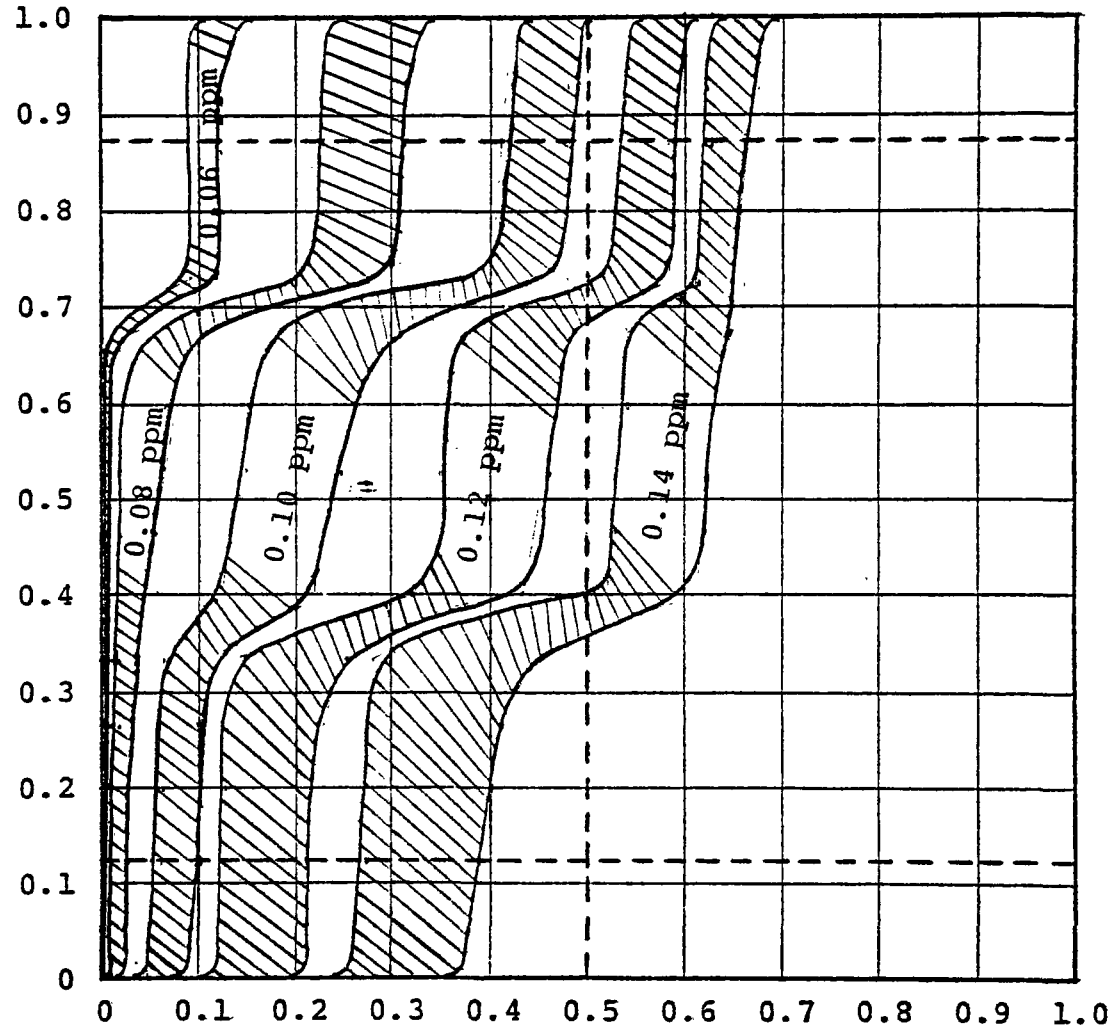
Figure 4-14. RISK OF EXCEEDING THE THRESHOLD* OF THE FOLLOWING HEALTH EFFECT:
- o REDUCTION IN PULMONARY FUNCTION



* HEALTH EFFECT THRESHOLD: Concentration at which 1% of group most sensitive to the effect would suffer the effect if the whole group were exposed under specified conditions.

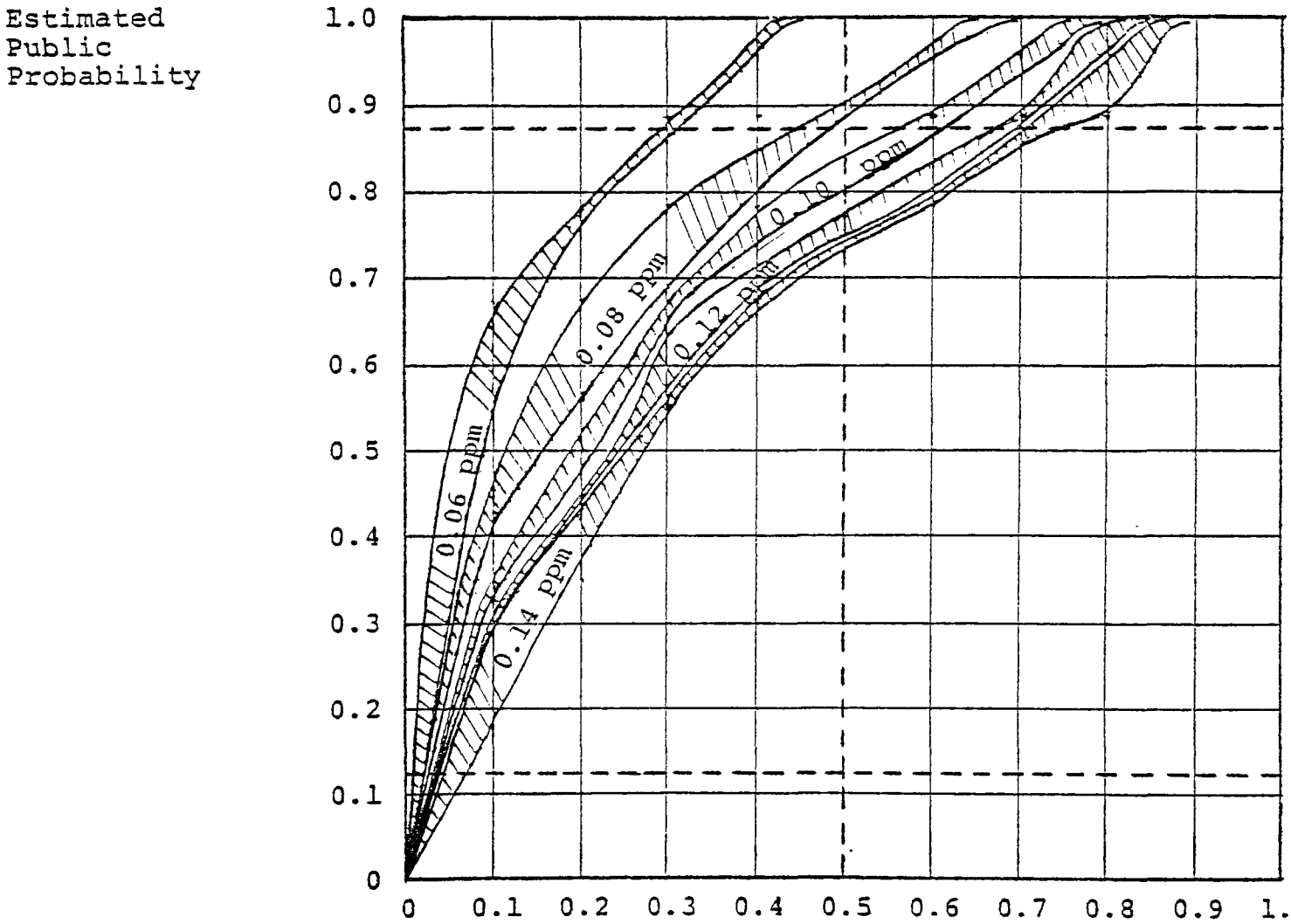
Figure 4-15. RISK OF EXCEEDING THE THRESHOLD* OF THE FOLLOWING HEALTH EFFECT:

- COUGH, CHEST DISCOMFORT, AND IRRITATION OF MUCOUS MEMBRANES OF NOSE, THROAT, AND TRACHEA



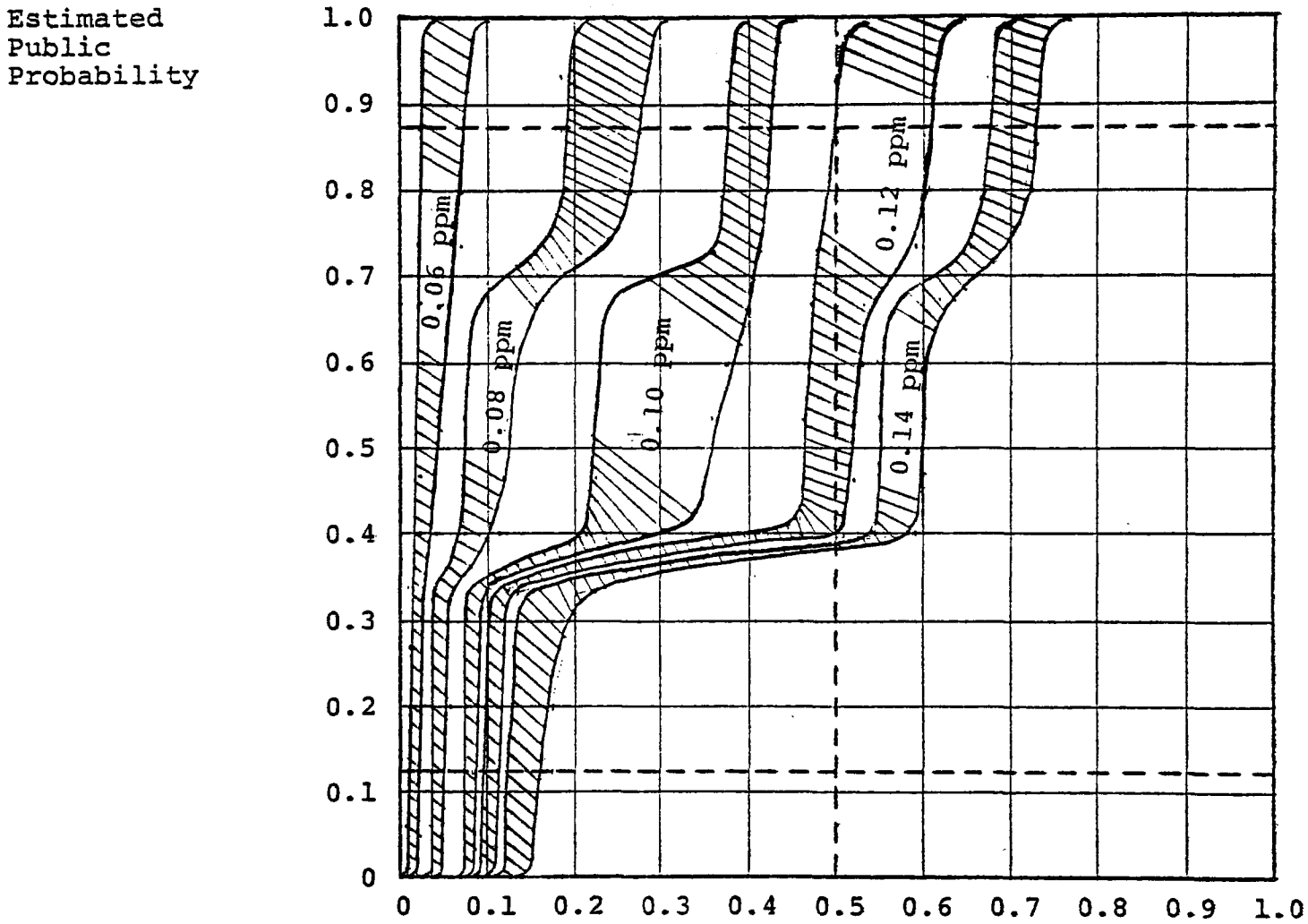
* HEALTH EFFECT THRESHOLD: Concentration at which 1% of group most sensitive to the effect would suffer the effect if the whole group were exposed under specified conditions.

Figure 4-16. RISK OF EXCEEDING THE THRESHOLD*
OF THE FOLLOWING HEALTH EFFECT:
o REDUCED RESISTANCE TO BACTERIAL INFECTION



* HEALTH EFFECT THRESHOLD: Concentration at which 1% of group most sensitive to the effect would suffer the effect if the whole group were exposed under specified conditions.

Figure 4-17. RISK OF EXCEEDING THE THRESHOLD* OF THE FOLLOWING HEALTH EFFECT:
- o AGGRAVATION OF ASTHMA, EMPHYSEMA, AND CHRONIC BRONCHITIS



* HEALTH EFFECT THRESHOLD: Concentration at which 1% of group most sensitive to the effect would suffer the effect if the whole group were exposed under specified conditions.

REFERENCES

1. Air Quality Criteria for Ozone and Other Photochemical Oxidants, U.S. EPA, Office of Research and Development, EPA-600/8-78-004, April, 1978.

2. National Academy of Sciences, Decision Making for Regulating Chemicals in the Environment, Washington, D.C., 1975.

3. National Academy of Sciences, Decision Making in the Environmental Protection Agency, Washington, D.C., 1977.

4. Clean Air Act, Section 109, 42 U.S.C.

5. William W. Lowrance, Of Acceptable Risk, William Kaufmann, Inc., Los Altos, California, 1976.

6. Senate Committee on Public Works, A Legislative History of the Clean Air Amendments of 1970, 1974.

7. Carl S. Spetzler and C.A.S. Stael von Holstein, "Probability Encoding in Decision Analysis", Management Science, Vol. 22, No. 3, November 1975.

8. Daniel J. Quinn and James E. Matheson, The Use of Judgmental Probability in Decision Making, prepared for U.S. EPA, May 1978.

9. Ian Hacking, The Emergence of Probability, Cambridge University Press, 1975.

10. Henry E. Kyburg, Jr., and Howard E. Smokler (eds.), Studies in Subjective Probability; John Wiley and Sons, Inc.; New York, 1964.

11. Robert L. Winkler and William L. Hays, Statistics, 2nd edition, Holt, Rinehart and Winston, New York, 1975.

12. Bernard Altshuler, "A Bayesian Approach to Assessing Population Risks from Environmental Carcinogens," Proceedings of SIMS Research Application Conference on Environmental Health. Alta, Utah; July 5-9, 1976; Revised November 2, 1976.

13. V. Hasselblad, W.C. Nelson, and G.R. Lowrimore, "Analysis of Health Effects Data: Some Results and Problems", in John W. Pratt (ed.), Statistical and Mathematical Aspects of Pollution Problems, Marcel Dekker, Inc., New York, 1974.

14. Glenn Shafer, A Mathematical Theory of Evidence, Princeton University Press, Princeton, New Jersey, 1976.

15. Rudolf Carnap, Logical Foundations of Probability, 2nd edition, The University of Chicago Press, 1962.

16. L. Jonathan Cohen, The Probable and the Provable, Oxford University Press, 1977.

17. "The Validity of the Weibull Distribution as a Model for the Analysis of Ambient Ozone Data." Draft Report to EPA by PEDCo Environmental, November 17, 1977.

18. Harold A. Linstone and Murray Turoff (ed's.), The Delphi Method, Addison-Wesley Publishing Co., Inc., Reading, Massachusetts, 1975

19. Steve Leung, Elliot Goldstein, and Norman Dalkey, "Human Health Damages from Mobile Source Air Pollution: A Delphi Study-Volume I," performed for California Air Resources Board, April, 1977.

20. Morris H. DeGroot, "Reaching a Consensus," Journal of the American Statistical Association, March 1974, Volume 69, Number 345.

21. Terrence L. Fine, Theories of Probability, Academic Press, Inc., New York, 1973.

22. E. J. Gumbel, Statistics of Extremes, Columbia University Press, New York, 1958.

23. William D. Rowe, An Anatomy of Risk, John Wiley and Sons, New York, 1977.

24. T. C. Curran and N. H. Frank, "Assessing the Validity of the Lognormal Model When Predicting Maximum Air Pollution Concentrations" Presented at 68th Annual Meeting of the Air Pollution Control Association, Boston, June 15-20, 1975.

25. Joel Horowitz and Sam Barakatz, "Statistical Analysis of the Maximum Concentration of an Air Pollutant: Effects of Autocorrelation and Nonstationarity", submitted for publication to Atmospheric Environment.

26. "Summary Statement From the EPA Advisory Panel on Health Effects of Photochemical Oxidants," Prepared for EPA under the supervision of the Institute for Environmental Studies of the University of North Carolina at Chapel Hill, January, 1978.

APPENDIX A

THE NEED FOR A RISK ASSESSMENT

How is it to be determined which alternative standards would provide an adequate margin of safety? Assume for the present that short term (one hour) exposure to pollutant X results in only one adverse human health effect, that the suitably defined threshold concentration for this effect is L, and that the maximum hourly concentration of X in a given period of time can be set at any concentration desired by adoption of a suitable control program. One approach would be to consider various margins of concentration, as measured in the direction of lower levels from some nonarbitrary point of reference, and try to determine the smallest margin that provides an adequate margin of safety. The point of reference could be either L_d, the lowest concentration for which the effect is judged to be demonstrated to a given (high) degree, or L_p, the concentration which is judged to be the most probable concentration of the threshold.

Even in the simplified context this approach will not suffice to put EPA in position to make the most meaningful judgment as to which alternative standards provide an adequate margin of safety. For consider Figure 1 where m_1 and m_2 are any two candidates for the "margin of safety," L_d, is the point of reference (the argument is similar for L_p), and L_1 and L_2 are the two potential levels of the standard corresponding to m_1 and m_2. Obviously, the greater the margin the more protection provided by the standard, so m_2 provides more protection than m_1. But without further information nothing can be said in absolute terms about the safety or degree of protection provided by either m_1 or m_2.

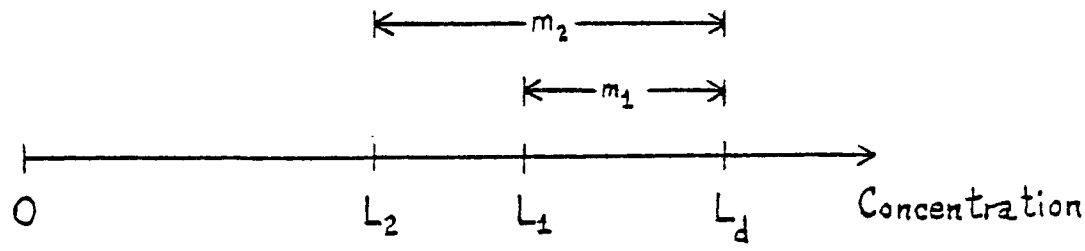
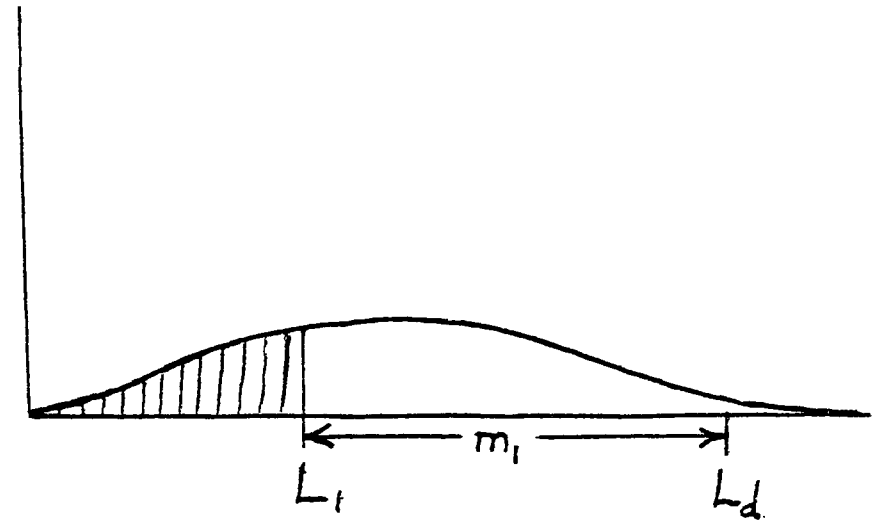


Figure A-1. "Margins of Safety"

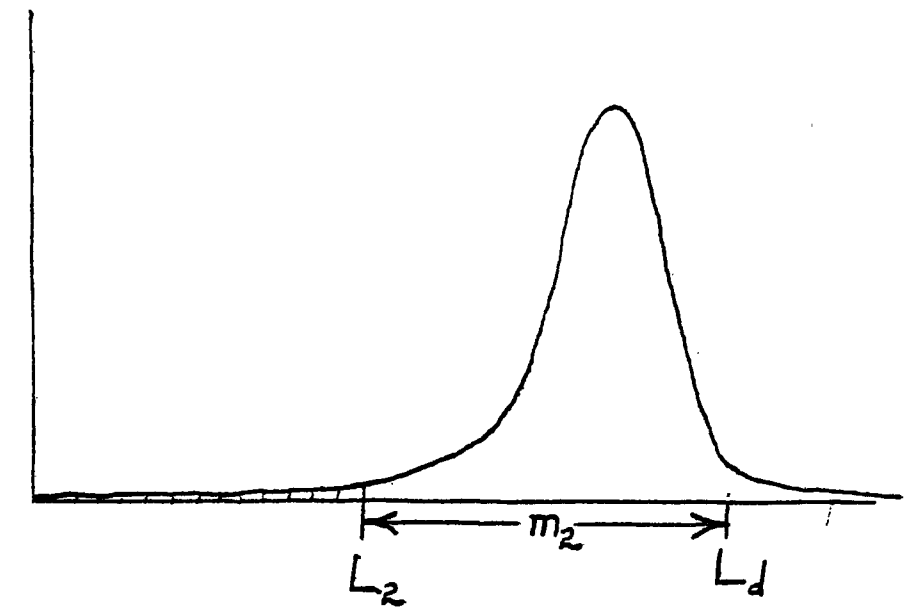
The approach just outlined falls short because it does not provide for an estimation of the risk associated with each possible standard. The risk associated with each possible standard is the risk to the most susceptible group in the general population of suffering adverse health effects when air quality just achieves that standard. In order to make a meaningful judgment on whether a possible standard provides an adequate margin of safety, a conception is needed of the risk associated with the possible standard. If the risk associated with the possible standard is deemed to be acceptable in view of the circumstances, then that standard is judged in a meaningful way to allow an adequate margin of safety.

To see clearly why the approach of comparing margins outlined above has inherent problems, consider the two hypothetical probability density distributions shown in Figure 2. If L_d is the lowest concentration at which human health effects are judged to be scientifically demonstrated to a certain (high) degree, the probability density distribution representing uncertainty about the health effect threshold concentration could be flat, as in (a), or relatively spiked, as in (b). That is, the criteria document could either have a sparse scientific data base which leaves a wide range of uncertainty



(a)

$m_1 > m_2$



(b)

Figure A-2. Larger Margin with Higher Risk

about the threshold concentration, or it could have a more complete data base which strongly suggests the threshold concentration lies within a relatively narrow range. The margin m_1 is greater than the margin m_2. But, if (a) is the case m_1 gives a standard level, L_1, that has more risk associated with it than the standard level m_2 gives, L_2, if (b) is the case.

The point made by the example does not assume that there would ever be agreement among experts as to the exact shape of a probability distribution only that in the one case they would agree that the range of uncertainty is relatively large and the probability distribution relatively flat, whereas in the other case they would agree that the range of uncertainty is relatively small and the probability distribution relatively spiked. The point reinforced by the example is that in setting a NAAQS one should do more than just identify a nonarbitrary point of reference and then choose the length of a margin to be called "the margin of safety." The Clean Air Act requires that the standard provide an adequate margin of safety, not identify some margin as "the margin of safety". Hence, setting a NAAQS is fundamentally a matter of choosing the least stringent standard which has an acceptable level of risk associated with it. Only then can any margin be meaningfully identified as an adequate margin of safety.

To be sure, one of the circumstances EPA decision-makers may want to consider in determining what alternative standard to set are the concentrations at which it is scientifically demonstrated to a given degree in the judgment of health experts that the pollutant in question contributes to various

adverse human health effects.* As mentioned above, the difference between the lowest such concentration and the standard level could be called "the margin of safety". Everything else being equal a decision-maker may prefer to accept more risk the larger "the margin of safety" in this sense. But if this were the case it would not affect the fact that the so-called "margin of safety" is not a measure of the margin of safety provided by the standard; the measure of the margin of safety provided by the standard is the risk associated with the standard.

The above arguments can be reformulated so that they show that the ratio m/T_d (or m/T_p), the ratio of "the margin of safety" to the demonstrated effects level, is not a measure of the margin of safety provided by a standard. However, this point can be made more succinctly by considering more realistic situations than the oversimplified one we have considered so far. Dropped first is the unrealistic assumption that the maximum hourly concentration of the pollutant in a given period of time can be set at any concentration desired by adoption of a suitable control program.

*The appropriate health experts to make these judgments may be the authors and reviewers of the criteria document

Let Std(L, t, e/u) be a formal way of representing a standard, where 'L' represents 'the concentration level,' 't' represents 'the averaging time,' 'e' represents 'the number of expected exceedenaces of the level,' and 'u' represents a unit of time. Now, the analysis of how to estimate the risk associated with a standard presented in the main body of the report shows that Std(L_1, 1 hr. 1 hr/yr) provides more protection, that is a greater margin of safety, than Std(L_1, 1 hr, 2 hrs/yr). Yet, the so-called "margin of safety", whether defined in terms of m or m/Td, is the same for the two standards. The same would be true for Std(L_1, 1 hr, 1 hr/yr) and Std(L_1, 1 hr, n hrs/yr.), no matter how large n. Yet, if n is large enough the degree of protection or safety provided by the two standards can differ substantially.

Compare Std(L_1, t, e/u) with Std(L_2, t, e/u), where $L_2 < L_1$. If either m or m/L_d were a good measure of the margin of safety provided by a standard, the difference in the degree of protection provided by the two standards would be the same in cases (a) and (b) of Figure 3. But in fact if the evidence indicates something like (a) is the case, there is a small difference in the degree of protection or risk associated with the two standards, whereas if the evidence indicates something like (b) is the case, there is a large difference.

When the simplifying assumption that the pollutant only contributes to one health effect is dropped there are additional problems with using m or m/L_d as a measure of the margin of safety provided by a standard. Let X and Y be two pollutants for which L_d is the same concentration; suppose there is a thin criteria document for X which indicates X only contributes to one health effect; suppose there is a thick criteria document for Y which indicates that there are h health effects which, by agreement of the experts, group into j

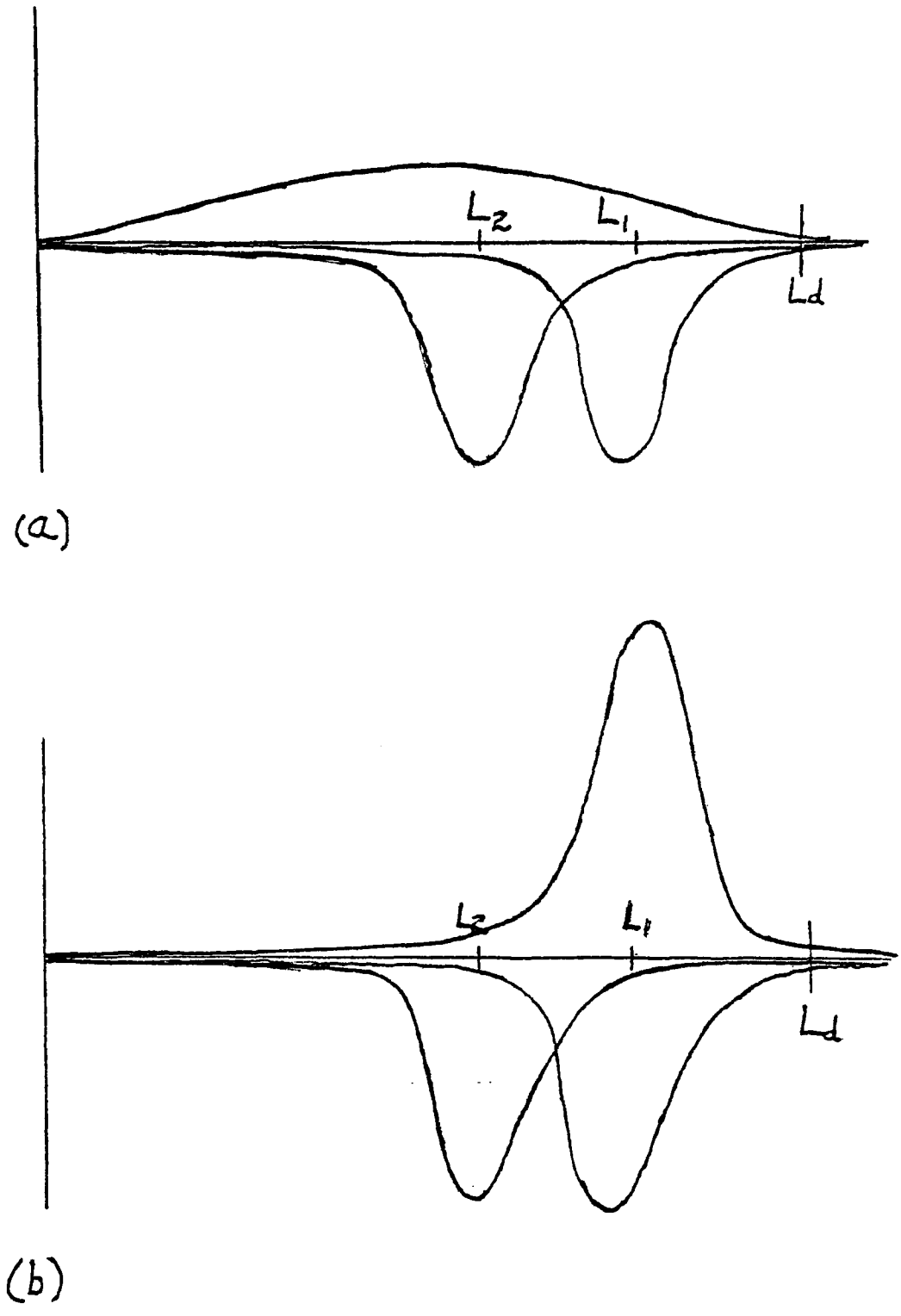


Figure A-3. Small Change in Risk vrs. Large Change in Risk for Same Change in Standard Level

independent categories; suppose also that the demonstrated effect levels for each of the j-i categories that don't have the lowest demonstrated effect level are at concentrations only slightly greater than L_d. Then, as the analysis in the main body of the report shows, the risks associated with Std(L, t, e/u) are not the same for pollutants X and Y; yet, they have the same so-called "margin of safety" as measured by m or m/L_d.

Suppose Z were a pollutant just like Y except that its criteria document indicates there is evidence from animal studies that Z may contribute to a j + 1st human health effect category; suppose the experts agree that if the effect is an effect of Z in man as it is in animals, then its threshold concentration is most likely less than the threshold concentrations of the other j categories of health effects; suppose the expert judgments of the probability that Z contributes to the effect in man are neither close to 0 or 1.0. Then, clearly the risk associates with Std(L, t, e/u) is even greater for Z than for Y; yet m and m/L_d remain the same.

Illustration has been given of several ways an approach to setting NAAQS's, that does not involve risk assessment, can fail to put EPA in a position to make the most meaningful judgment as to which alternative standards provide an adequate margin of safety. In most real world cases several of the logical difficulties with alternate approaches illustrated by the above hypothetical situations will obtain. Therefore, risk assessment should be an integral part of setting NAAQS's.

APPENDIX B

DERIVATION OF BASIC EQUATIONS FOR ASSESSING HEALTH RISKS ASSOCIATED WITH ALTERNATIVE AIR QUALITY STANDARDS

1.0 Risk of Exceeding a True Health Effect Threshold or the Lowest of Several Thresholds

First determine probability P where:

P = Probability that no hourly average concentration exceeds the true health effect threshold or the lowest of several thresholds in a given period.

Let

$P_C(C)$ = Probability that no time averaged concentration exceeds the concentration C in the given period.

$P_T(C)$ = Probability density function expressing the uncertainty in the location of the true threshold or the lowest of several thresholds.

Note

$\int_a^b P_T(C)dC$ = Probability that a true threshold or the lowest of several thresholds is in the interval a,b.

It follows that:

$P_C(C)p_T(C)dC$ = Probability that the true threshold or the lowest of several thresholds is contained in the interval dC and this value is not exceeded by any hourly average concentration in the given period.

By integrating over all values of C the quantity P, defined above, is determined:

$$P = \int_0^\infty P_C(C)P_T(C)dC \qquad \text{(B-1}$$

If R is the probability or risk of exceeding a health effects threshold or the lowest of several thresholds one or more times in the given period, then:

$$R = 1 - P$$

$$R = 1 - \int_0^\infty P_C(C)p_T(C)dC \qquad \text{(B-2}$$

Note that if the function P_C is defined as the probability that no time averaged concentration exceeds the concentration C it must also be the probability that the highest time averaged concentration observed in the time period does not exceed C (Therefore, is $\leq$ C). In other words, P_C is the cumulative distribution of the highest time averaged concentration occuring in the given time period.

It is further noted that if it is desired to estimate the risk of exceeding a health effects threshold or the lowest of several threshold "m" or more times in a given time period, then P_C becomes the probability that a threshold will not be exceeded more than m-1 times in the given time period. In this case P_C is the probability distribution of the mth highest time averaged concentration.

2.0 Determination of Probability Density Function for the Lowest of Several Health Effect Threshold

Assume n different health effects and that a probability density function has been obtained for each. If the function $P_C(C)$ is known, Eq. (B-1) can then be used to evaluate the individual probabilities for each health effect.

$$P_i = \int_0^\infty P_C(C) P_{T_i}(C)\,dC \qquad \text{(B-3)}$$

Let:

P = The probability that no hourly average concentration exceed any of the n true thresholds in the given time period.

To evaluate this probability for all possible configurations it is assumed that each threshold in turn is the lowest threshold.

Differential elements of the following type can be formed:

$$P_C(C) P_{T_1}(C)\,dC \cdot \int_C^\infty P_{T_2}(x)\,dx \cdot \int_C^\infty P_{T_3}(x)\,dx \cdots \int_C^\infty P_{T_n}(x)\,dx$$

This term is the probability that: 1. no hourly average concentration exceeds the value C; 2. the threshold T_1 is in the interval dC; and 3. all the other thresholds are above this value.

If:

$$Q_i(C) = \int_C^\infty P_{T_i}(x)\,dx$$

And:

$$Q(C) = \prod_{i=1}^{n} Q_i(C)$$

the above expression simplifies to:

$$P_C \frac{Q}{Q_1} p_{T_1}\,dC$$

Integrating over the concentration range gives:

$$\int_0^\infty P_C \frac{Q}{Q_1} p_T \, dC$$

This integral is the probability that no hourly average concentration exceeds any of the n thresholds in the time period and T_1 is the lowest threshold. In order to enumerate all cases, it is necessary to form the sum of integrals of this type in which each threshold in turn is assumed to be the lowest.

Therefore:

$$P = \sum_{i=1}^{n} \int_0^\infty P \frac{Q}{Q_i} p_{T_i} \, dC$$

or:

$$P = \int_0^\infty P_C \left[Q \sum_{i=1}^{n} p_{T_i}/Q_i\right] dC$$

or:

$$P = \int_0^\infty P_C \, p_T \, dC$$

This equation is, as would be expected, identical to Eq. (B-1) except that p_T is now a composite probability density function such that:

$$p_T = Q \prod_{i=1}^{n} p_{T_i}/Q_i \qquad \text{(B-4)}$$

where:

$$Q_i(C) = \int_C^\infty p_{T_i} \, dC \qquad \text{(B-5)}$$

and

$$Q = \prod_{i=1}^{n} Q_i \qquad \text{(B-6)}$$

The quantity $p_T(C)dC$ is the probability that the lowest of the n thresholds is in the interval dC.

An alternate method of deriving the composite probability density function is to start with the composite cumulative distribution for the health effects threshold. Let:

$$D_T = Pr\ (T_{lowest} \leq C)$$

and:

$$D_{T_i} = Pr\ (T_i \leq C)$$

It follows from probability theory that:

$$1-D_T = \prod_{i=1}^{n} (1-D_{T_i}) \qquad \text{(B-7)}$$

The composite probability density function is by definition:

$$P_T = \frac{dD_T}{dC} \qquad \text{(B-8)}$$

Differentiating (B-7) yields:

$$\frac{dD_T}{dC} = \left(\prod_{i=n}^{n} (1-D_{T_i}) \right) \sum_{i=1}^{n} \frac{1}{1-D_{T_i}} \frac{dD_{T_i}}{dC} \qquad \text{(B-9)}$$

Since by definition:

$$1-D_{T_i} = \int_C^{\infty} p_{T_i}\, dC = Q_i \qquad \text{(B-10)}$$

$$Q = \prod_i (1-D_{T_i})$$

$$p_{T_i} = dD_{T_i}/dC \qquad \text{(B-11)}$$

Equation (B-9) is equivalent to Equation (B-4).

Equations (B-7) and (B-8) provide an alternate route to obtaining the composite density function, p_T, which can then be used in Equation (B-2) to estimate risk.

3.0 Inclusion of Uncertainty That One or More Health Effects Exist

For some health effects there may be uncertainty that the effect actually occurs in humans. It would be desirable to include this consideration when considering the uncertainty in the location of the threshold on the concentration axis. Given uncertainty only in the location of the threshold it has been shown that Eq. (B-2) gives the risk that the true threshold T_i is exceeded one or more times in a given period.

If there is uncertainty as to whether the ith effect occurs in humans, assign the probability e_i that the effect does exist. Then choose a value of C = u such that u is many times larger than any concentration likely to be encountered. In other words, u is many times beyond the concentration range of interest. In this case, if $p^{\circ}_{T_i}(C)$ is defined as the probability density function for the location of the i th effect if it does exist, a new overall probability density function can be written:

$$p_{T_i}(C) = e_i p^{\circ}_{T_i} + (i-e_i)\,\delta(C-u) \qquad i = 1\cdots\cdots, n \qquad \text{(B-12)}$$

where $\delta(C-u)$ is the Dirac delta function. It has the property:

$$\int_a^b \delta(C-u)\, dC \quad \begin{cases} = 1 & \text{if } a \leq u \leq b \\ = 0 & \text{if } a > u \text{ or } b < u \end{cases} \tag{B-13}$$

In other words, if the interval of integration includes C = u the value of the integral is unity. If it does not include C = u the integral is zero. The Dirac delta function itself may be considered as zero for all values of C except u. At u it has an infinite value. It is also assumed that $p^{\circ}_{T_i}(C)$ is essentially zero in the vicinity of C = u and above.

Equation (B-12) is based on the premise that saying an effect does not exist is mathematically equivalent to saying that the true threshold is at a very high concentration which is above any concentration likely to be encountered. Thus, if the probability e can be assigned to the certainty that it does exist, the fraction e of the total area under the probability density curve can be assigned to the concentration range in which the effect is thought to be located if it does occur. The rest of the area, (1-e), can be assigned to a range above any concentration likely to occur.

Note also, the use of the Dirac delta function is only a matter of convenience since it leads to a simple form of Eq. (B-12). Any function can be used in this outer region as long as it has the value zero in the region of interest and has the area (1-e). The Dirac function is simply a convenient form for including the desired property in the probability density function.

Substituting Eq. (B-12) into (B-3) gives:

$$P_i = e_i \int_0^{\infty} P_C\, p^{\circ}_{T_i}\, dC + (1-e_i) \int_0^{\infty} P_C\, \delta(C-u) dC$$

Since P_C will be essentially 1 at C = u, the above equation yields:

$$P_i = e_i \int_0^{\infty} P_C p^{\circ}_{T_i}\, dC + (1-e_i) \qquad i = 1, \cdots, n \tag{B-14}$$

To find the probability that no hourly average concentration exceeds any of n thresholds, it is necessary to derive the appropriate form of the composite threshold probability density function. Its general form is given by Eq. (B-4).

Substituting Eq. (B-12) into Eq. (B-4) and rearranging gives:

$$p_T = Q \sum_{i=1}^{n} \frac{e_i}{Q_i} p^{\circ}_{T_i} + \left[Q \sum_{i=1}^{n} \frac{1-e_i}{Q_i} \right] \delta(C-u) \qquad \text{(B-15}$$

The first term on the right side of the Eq. (B-15) is evaluated in the ambient concentration range of the pollutant well below the value C = u. In this region the functions Q_i have a simple interpretation. Substituting Eq. (B-12) into Eq. (B-5).

$$Q_i = e_i \int_C^{\infty} p^{\circ}_{T_i} dC + (1-e_i) \int_C^{\infty} \delta(C-u) dC \qquad \text{(B-16}$$

For all C < u

$$Q_i = e_i \int_C^{\infty} p^{\circ}_{T_i} dC + (1-e_i) \qquad \text{(B-17}$$

The second term on the right side of the Eq. (B-15) is evaluated in the vicinity of C = n, far above the ambient concentration range. In this region the behavior of the Q_i functions needs to be more carefully considered. The problem is the behavior of the second term on the right side of Eq. (B-16) as C passes through u.

Note that in the vicinity of C = u the first term on the right side of Eq. (B-16) is zero by definition.

Thus:

$$Q_i = (1-e_i) \int_C^{\infty} \delta(C-u)dC \qquad \text{for } C \approx u \tag{B-18}$$

Expanding the second term of Eq. (B-15) gives:

$$[\, Q_2Q_3\cdots Q_n(1-e_1) + Q_1Q_3\cdots Q_n(1-e_2) + \cdots + Q_1Q_2\cdots Q_{n-1}(1-e_n) \,]\, \delta(C-u)$$

which on substituting Eq. (B-18) yields:

$$\left\{ n\, \delta(C-u) \left[\int_C^{\infty} \delta(C-u)dC \right]^{n-1} \right\} \prod_i (1-e_i)$$

The portion of this term in large brackets can be shown to have the properties of a Dirac delta function (Eq. (b-13)) as follows:

Let:

$$\psi(C-u) = n\, \delta(C-u) \left[\int_C^{\infty} \delta(C-u)dC \right]^{n-1} \tag{B-19}$$

and

$$\phi(C) = \int_C^{\infty} \delta(C-u)dC \tag{B-20}$$

then

$$\frac{d\phi(C)}{dC} = -\delta(C-u) \tag{B-21}$$

Substituting into the bracketed terms:

$$\psi(C-u) = -n\phi^{n-1} \frac{d\phi}{dC}$$

or

$$\psi(C-u) = -\frac{d\phi^n}{dC} \tag{B-22}$$

To show that the derivative has the properties of the Dirac delta function, integrate between the limits a and b.

$$\int_a^b \psi(C-u)dC = -\int_a^b d\phi^n = \phi(a)^n - \phi(b)^n$$

$$= \left[\int_a^\infty \delta(C-u)dC\right]^n - \left[\int_b^\infty \delta(C-u)dC\right]^n \tag{B-23}$$

Using Eq. (B-13) to evaluate the quantities in brackets for different values of a and b it follows that:

$$\int_a^b \psi(C-u)dC \quad \begin{array}{l} = 1 \text{ if } a \leq u \leq b \\ = 0 \text{ if } a > u \text{ or } b < u \end{array} \tag{B-24}$$

where $b > a$

Therefore, $\psi(C-u)$ is also a Dirac delta function.

We can, therefore, write:

$$p_T = \left(Q\sum_i \frac{e_i}{Q_i}\ p^\circ_{T_i}\right) + \delta(C-u)\ \prod_i^n (1-e_i) \tag{B-25}$$

where:

$$Q_i = e_i \int_C^\infty p^\circ_{T_i} dC + (1-e_i)$$

or alternatively:

$$Q_i = 1-e_i \int_0^C p^\circ_{T_i} dC \tag{B-26}$$

and

$$Q = Q_1 Q_2 \cdots Q_n \tag{B-27}$$

Substituting Eq. (B-25) into Eq. (B-1) gives:

$$P = \int_0^\infty P_C p^\circ_T dC + \prod_i (1-e_i) \tag{B-28}$$

and therefore:

$$R = 1 - \int_0^\infty P_C p^\circ_T dC + \prod_i (1-e_i) \tag{B-29}$$

The Eqs. (B-25) through (B-29) are the basic working equations for estimating risk when a multiple number of health effects thresholds are involved, and where one or more of the health effects may not occur in humans.

The above results can be derived in an alternate manner starting with the cumulative distribution functions for the location of the health effects thresholds. Note that Eq. (B-7) relating the composite cumulative distribution to the individual distributions continues to hold, although now the composite and individual density functions are given by Eqs. (B-25) and (B-12). The definition of the cumulative distribution for the ith effect can be written:

$$D_{T_i} = e_i D^\circ_{T_i} + (1-e_i) \int_0^C \delta(C-u)dC \tag{B-30}$$

By differentiating Eq. (B-30) it can be shown to be the appropriate distribution function for the density function Eq. (B-12).

$$\frac{dD_{T_i}}{dC} = e_i \frac{dD^{\circ}_{T_i}}{dC} + (1-e_i) \frac{d}{dC} \int_0^{\infty} \delta(C-u)du = p_{T_i}$$

or

$$p_{T_i} = ep^{\circ}_{T_i} + (1-e_i)\, \delta(C-u)$$

which is Eq. (B-12).

Next, it can be shown that the use of Eq. (B-30) and Eq. (B-7) yields the correct form of the composite, p_T. Substituting Eq. (B-30) in Eq. (B-7) gives:

$$(1-D_T) = \prod_i [1-e_i D^{\circ}_{T_i} - (1-e_i) \int_0^{C} \delta(C-u)dC] \qquad \text{(B-3}$$

Note that for $c < u$:

$$(1-D_T) = \prod_i (1-e_i\, D^{\circ}_{T_i})$$

For $C \sim u$

$$D^{\circ}_{T_i} = 1$$

and, therefore:

$$(1-D_T) = \prod_i [\, 1-e_i - (1-e_i) \int_0^{C} \delta(C-u)dC \,]$$

For $C > u$

$$1-D_T = 0 \quad \text{or} \quad D_T = 1$$

It is, therefore, possible to write:

$$1-D_T = \prod_i (1-e_i D^\circ_{T_i}) - \left(\prod_i (1-e_i) \right) \int_0^C \delta(C-u)du \tag{B-32}$$

where Eq. (B-32) gives the same result as Eq. (B-31) over the whole range of the concentration, C.

If D° is defined by:

$$1-D^\circ_T = \prod_i (1-e_i D^\circ_{T_i}) \tag{B-33}$$

then Eq. (B-32) can be written as:

$$D_T = D^\circ_T + \left(\prod_i (1-e_i) \right) \int_0^C \delta(C-u)dC \tag{B-34}$$

which is the composite analog of Eq. (B-30).

The derivative of Eq. (B-34) is taken to obtain p_T. But first, note from the definition of Q_i in Eq. (B-26)

$$Q_i = 1-e_i D^\circ_{T_i} \tag{B-35}$$

and from Eq. (B-27)

$$Q = \prod_i Q_i$$

Taking the derivative of (B-34) then:

$$\frac{dD_T}{dC} = p_T = \frac{-dQ}{dC} + \prod_i (1-e_i) \frac{d}{dC} \int_o^C \delta(C-u)dC$$

or

$$p_T = \frac{-dQ}{dC} + \delta(C-u) \prod_i (1-e_i)$$

However:

$$\frac{dQ}{dC} = Q \sum \frac{1}{Q_i} \frac{dQ_i}{dC} = - Q \sum \frac{e_i p^\circ_{T_i}}{Q_i}$$

or

$$p_T = Q \sum \frac{e_i p^\circ_{T_i}}{Q_i} + \delta(C-u) \prod_i (1-e_i)$$

which is the desired composite density function and is identical to Eq. (B-25).

Note that starting with the $D^\circ_{T_i}$ and their corresponding e_i it is possible to first form the composite D°_T and then by:

$$\frac{dD^\circ_T}{dC} = p^\circ_T \qquad \text{(B-36)}$$

obtain the function p°_T for insertion into Eq. (B-29) to estimate the risk of exceeding the lowest of n thresholds.

APPENDIX C

DERIVATIONS RELATED TO THE P_C FUNCTION

1.0 Derivation of P_C Function Given Distribution of Time Averaged Concentrations and No Interactions or Time Dependence

The following basic relationship from probability theory may be used to relate the P_C function to the distribution function for the time averaged concentrations G(C) which is defined by Eq. (8) of Section 3.2.

$$P_\nu = \frac{n!}{\nu!(n-\nu)!} p^\nu (1-p)^{n-\nu} \qquad \text{(C-1)}$$

Where P_ν is the probability that a certain event, whose probability of occuring in a single trial p, will occur ν times in n independent trial. Here a trial is the period covered by one averaging time. For example, for an averaging time of one hour a trial is a single hour. Thus, a year will have 8760 trials. The event is the occurence of an observed concentration above some specified value C.

The collection of values $P_0, P_1, P_2, \cdots$ are referred to as the binomial distribution.

From the above definition of the event it follows that:

$$p = G(C) \qquad \text{(C-2)}$$

Where G(C) is distribution of the averaged concentrations defined by Eq. (8) of Section 3.2.

If there are to be no exceedances of the concentration C, then m =0 and it follows from Eqs. (C-1) and C-2) that:

$$P_0 = (1-G(C))^n \qquad \text{(C-3)}$$

which is Eq. (10) of Section 3.2.

If the P_C function is to be used to calculate the risk that the threshold will be exceeded m or more times, then it is labeled $P_C^{(m)}$ and corresponds to the probability that a threshold will be exceeded not more than m-1 times in the time period.

In this case:

$$P_C^{(m)} = \sum_{\nu=0}^{m-1} P_\nu \tag{C-4}$$

where P_ν is given by Eq. (C-1). If Eqs. (C-1) and (C-2) are substituted into Eq. (C-4):

$$P_C^{(m)} = \sum_{\nu=0}^{m-1} \frac{n!}{\nu!(n-\nu)!} G^\nu (1-G)^{n-\nu} \tag{C-5}$$

Eq. (C-3) is the special case of Eq. (C-5) corresponding to m = 1.

2.0 Distribution of Time Averaged Concentrations When Represented by Weibull Distribution Just Meeting an Air Quality Standard.

The National Ambient Air Quality Standard is assumed to have the following form:

> C_{STD} ppm hourly average concentration with an expected number of exceedances per year less than or equal to E.

The Weibull distribution is defined by Eq. (12) of Section 3.2:

$$G(C) = e^{-(C/\delta)^k} \tag{C-6}$$

G(C) can be interpreted as a relative frequency. As such, the quantity E/n_E (where n_E = 8760 hrs in a year) is the expected frequency of occurence of concentrations above the level $C = C_{STD}$.

Therefore:

$$\frac{E}{n_E} = e^{-\left(\frac{C_{STD}}{\delta}\right)^k} \tag{C-7}$$

Solving this expression for the parameter δ gives:

$$\delta = C_{STD}/(\ln(n_E/E))^{\frac{1}{k}} \tag{C-8}$$

Substituting this value in Eq. (C-6) yields:

$$G(C) = e^{-(\ln(n_E/E))\ (C/C_{STD})^k} \tag{C-9}$$

which is Eq. (13) of Section 3.2.