Chapter 2 Criteria used in establishing guideline values

Relevant information on the pollutants was carefully considered during the process of establishing guideline values. Ideally, guideline values should represent concentrations of chemical compounds in air that would not pose any hazard to the human population. Realistic assessment of human health hazards, however, necessitates a distinction between absolute safety and acceptable risk. To produce a guideline with a high probability of offering absolute safety, one would need a detailed knowledge of dose–response relationships in individuals in relation to all sources of exposure, the types of toxic effect elicited by specific pollutants or their mixtures, the existence or nonexistence of "thresholds" for specified toxic effects, the significance of interactions, and the variation in sensitivity and exposure levels within the human population. Such comprehensive and conclusive data on environmental contaminants are generally unavailable. Very often the relevant data are scarce and the quantitative relationships uncertain. Scientific judgement and consensus therefore play an important role in establishing guidance that can be used to indicate acceptable levels of population exposure. Value judgements are needed and the use of subjective terms such as "adverse effects" and "sufficient evidence" is unavoidable.

Although it may be accepted that a certain risk can be tolerated, the risks to individuals within a population may not be equally distributed: there may be subpopulations that are at considerably increased risk. Therefore, groups at special risk in the general population must be taken specifically into account in the risk management process. Even if knowledge about groups with specific sensitivity is available, unknown factors may exist that change the risk in an unpredictable manner. During the preparation of this second edition of the guidelines, attention has been paid to defining specific sensitive subgroups in the population.

Information common to carcinogens and noncarcinogens

Sources, levels and routes of exposure

Available data are provided on the current levels of human exposure to pollutants from all sources, including the air. Special attention is given to atmospheric concentrations in urban and unpolluted rural areas and in the indoor environment. Where appropriate, concentrations in the workplace are also indicated for comparison with environmental levels. To provide information on the contribution from air in relation to all other sources, data on uptake by inhalation, ingestion from water and food, and dermal contact are given where relevant. For most chemicals, however, data on total human exposure are incomplete.

Toxicokinetics

Available data on toxicokinetics (absorption, distribution, metabolism and excretion) of air pollutants in humans and experimental animals are provided for comparison between test species and humans and for interspecies and intraspecies extrapolation, especially to assess the magnitude of body burden from long-term, low-level exposures and to characterize better the mode of toxic action. Data concerning the distribution of a compound in the body are important in determining the molecular or tissue dose to target organs. It has been appreciated that high-to-low-dose and interspecies extrapolations are more easily carried out using equivalent tissue doses. Metabolites are mentioned, particularly if they are known or believed to exert a greater toxic potential than the parent compound. Additional data of interest in

determining the fate of a compound in a living organism include the rate of excretion and the biological half-life. These toxicokinetic parameters should be compared between test species and humans for derivation of interspecies factors where this is possible.

Terminology

The following terms and definitions are taken largely from Environmental Health Criteria No. 170, 1994 *(1)*.

Adverse effect Change in morphology, physiology, growth, development or life span of an organism which results in impairment of functional capacity or impairment of capacity to compensate for additional stress or increase in susceptibility to the harmful effects of other environmental influences.

Benchmark dose (BMD) The lower confidence limit of the dose calculated to be associated with a given incidence (e.g. 5% or 10% incidence) of effect estimated from all toxicity data on that effect within that study *(2)*.

Critical effect(s) The adverse effect(s) judged to be most appropriate for the health risk evaluation.

Lowest-observed-adverse-effect level (LOAEL) Lowest concentration or amount of a substance, found by experiment or observation, which causes an adverse alteration of morphology, functional capacity, growth, development or life span of the target organism distinguishable from normal (control) organisms of the same species and strain under the same defined conditions of exposure.

No-observed-adverse-effect level (NOAEL) Greatest concentration or amount of a substance, found by experiment or observation, which causes no detectable adverse alteration of morphology, functional capacity, growth, development or life span of the target organism under defined conditions of exposure. Alterations of morphology, functional capacity, growth, development or life span of the target may be detected which are judged not to be adverse.

Toxicodynamics The process of interaction of chemical substances with target sites and the subsequent reactions leading to adverse effects.

Toxicokinetics The process of the uptake of potentially toxic substances by the body, the biotransformation they undergo, the distribution of the substances and their metabolites in the tissues, and the elimination of the substances and their metabolites from the body. Both the amounts and the concentrations of the substances and their metabolites are studied. The term has essentially the same meaning as pharmacokinetics, but the latter term should be restricted to the study of pharmaceutical substances.

Uncertainty factor (UF) A product of several single factors by which the NOAEL or LOAEL of the critical effect is divided to derive a tolerable intake. These factors account for adequacy of the pivotal study, interspecies extrapolation, inter-individual variability in humans, adequacy of the overall database, and nature of toxicity. The choice of UF should be based on the available scientific evidence.

Criteria for endpoints other than carcinogenicity

Criteria for selection of NOAEL/LOAEL

For those compounds reportedly without direct carcinogenic effects, determination of the highest concentration at which no adverse effects are observed, or the lowest concentration at which adverse effects are observed in humans, animals or plants is the first step in the derivation of the guideline value. This requires a thorough evaluation of available data on toxicity. The decision as to whether the LOAEL or the NOAEL should be used as a starting point for deriving a guideline value is mainly a matter of availability of data. If a series of data fixes both the LOAEL and the NOAEL, then either might be used. The gap between the lowest-observed-effect level and the no-observed-effect level is among the factors included in judgements concerning the appropriate uncertainty factor. Nevertheless, one needs to consider that in studies in experimental animals, the value of the NOAEL (or LOAEL) is an observed value that is dependent on the protocol and design of the study from which it was derived. There are several factors that influence the magnitude of the value observed, such as the species, sex, age, strain and developmental status of the animals studied; the group size; the sensitivity of the methods applied; and the selection of dose levels. Dose levels are frequently widely spaced, so that the observed NOAEL can be in some cases considerably less than the true no-adverse-effect level, and the observed LOAEL considerably higher than the true lowest-adverse-effect level *(1)*.

A single, free-standing no-observed-effect level that is not defined in reference to a lowest-observed-effect level or a LOAEL is not helpful. It is important to understand that, to be useful in setting guidelines, the NOAEL must be the highest level of exposure at which no adverse effects are detected. It is difficult to be sure that this has been identified unless the level of exposure at which adverse effects begin to appear has also been defined. Opinions on this subject differ, but the working consensus was that the level of exposure of concern in terms of human health is more easily related to the LOAEL, and this level was therefore used whenever possible. In the case of irritant and sensory effects on humans, it is desirable where possible to determine the no-observed-effect level. These effects are discussed in more detail below.

On the basis of the evidence concerning adverse effects, judgements about the uncertainty factors needed to minimize health risks were made. Averaging times were included in the specification of the guidelines, as the duration of exposure is often critical in determining toxicity. Criteria applied to each of these key factors are described below.

Criteria for selection of adverse effect

Definition of a distinction between adverse and non-adverse effects poses considerable difficulties. Any observable biological change might be considered an adverse effect under certain circumstances. An adverse effect has been defined as "any effect resulting in functional impairment and/or pathological lesions that may affect the performance of the whole organism or which contributes to a reduced ability to respond to an additional challenge" *(3)*. Even with such a definition, a significant degree of subjectivity and uncertainty remains. Ambient levels of major air pollutants frequently cause subtle effects that are typically detected only by sensitive methods. This makes it exceedingly difficult, if not impossible, to achieve a broad consensus as to which effects are adverse. To resolve this difficulty, it was agreed that the evidence should be ranked in three categories.

1. The first category comprises observations, even of potential health concern, that are single findings not verified by other groups. Because of the lack of verification by other investigators, such data could not readily be used as a basis for deriving a guideline value. They do, however, indicate the need for further research and may be considered in deriving an appropriate uncertainty factor based on the severity of the observed effects.

2. The second category is a lowest-observed-effect level (or no-observed-effect level) that is supported by other scientific information. When the results are in a direction that might result in pathological changes, there is a higher degree of health concern. Scientific judgement based on all available health information is used to determine how effects in this category can be used in determining the pollutant level that should be avoided so that excessive risk can be prevented.

3. The third category comprises levels of exposure at which there is clear evidence for substantial pathological changes; these findings have had a major influence on the derivation of the guidelines.

Benchmark approach

The benchmark dose (BMD) is the lower confidence limit of the dose that produces a given increase (e.g. 5% or 10%) in the level of an effect to which an uncertainty factor can be applied to develop a tolerable intake. It has a number of advantages over the NOAEL/LOAEL approach *(2)*. First, the BMD is derived on the basis of the entire dose–response curve for the critical, adverse effect rather than that from a single dose group as in the case of the NOAEL/LOAEL. Second, it can be calculated from data sets in which a NOAEL was not determined, eliminating the need for an additional uncertainty factor to be applied to the LOAEL. Third, definition of the BMD as a lower confidence limit accounts for the statistical power and quality of the data; that is, the confidence intervals around the dose–response curve for studies with small numbers of animals or of poor quality and thus lower statistical power would be wide, reflecting the greater uncertainty of the database. On the other hand, better studies would result in narrow confidence limits, and thus in higher BMDs.

Although there is no consensus on the incidence of effect to be used as basis for the BMD, it is generally agreed that the BMD should be comparable with a level of effect typically associated with the NOAEL or LOAEL. Allen et al. *(4, 5)* have estimated that a BMD calculated from the lower confidence limit at 5% is, on average, comparable to the NOAEL, whereas choosing a BMD at 10% is more representative of a LOAEL *(6)*. Choosing a BMD that is comparable to the NOAEL has two advantages: (*a*) it is within the experimental dose-range, eliminating the need to interpolate the dose–response curve to low levels; and (*b*) it justifies the application of similar uncertainty factors as are currently applied to the NOAEL for interspecies and intraspecies variation. It should be noted, however, that the main disadvantage of the benchmark approach is that it is not applicable for discrete toxicity data, such as histopathological or teratogenicity data.

Criteria for selection of uncertainty factors

In previous evaluations by WHO, uncertainty factors (sometimes called safety factors) have been applied to derive guidelines from evidence that conforms to accepted criteria for adverse effects on health *(7–9)*. Traditionally, the uncertainty (safety) factor has been used to allow for uncertainties in extrapolation from animals to humans and from a small group of individuals to a large population, including possibly undetected effects on particularly sensitive members of the population. In addition, uncertainty factors also account for possible

synergistic effects of multiple exposures, the seriousness of the observed effects and the adequacy of existing data *(1)*. It is important to understand that the application of such factors does not indicate that it is known that humans are more sensitive than animal species but, rather, that the sensitivity of humans relative to that of other species is usually unknown. It is possible that humans are less sensitive than animals to some chemicals.

In this second edition of the air quality guidelines, the terms "safety factor" and "protection factor" have been replaced by the term "uncertainty factor". It is felt that this better explains the derivation and implications of such factors. Of course, such a factor is designed to provide an adequate level of protection and an adequate margin of safety, because these factors are applied in the derivation of guidelines for the protection of human health. They are not applied in the derivation of ecological guidelines because these already include a kind of uncertainty factor with regard to the variety of species covered.

A wide range of uncertainty factors are used in this second edition, based on scientific judgements concerning the interplay of various effects. The decision process for developing uncertainty factors has been complex, involving the transformation of mainly non-quantitative information into a single number expressing the judgement of a group of scientists.

Some of the factors taken into account in deciding the margin of protection can be grouped under the heading of scientific uncertainty. Uncertainty occurs because of limitations in the extent or quality of the database. One can confidently set a lower margin of protection (that is, use a smaller uncertainty factor) when a large number of high-quality, mutually supportive scientific experiments in different laboratories using different approaches clearly demonstrate the dose–response, including a lowest-observed-effect level and a no-observed-effect level. In reality, difficulties inherent in studying air pollutants, and the failure to perform much-needed and very specific research, means that often a large uncertainty factor has to be applied.

Where an uncertainty factor was adopted in the derivation of air quality guidelines, the reasoning behind the choice of this factor is given in the scientific background information. As previously mentioned, exceeding a guideline value with an incorporated uncertainty factor does not necessarily mean that adverse effects will result. Nevertheless, the risk to public health will increase, particularly in situations where the most sensitive population group is exposed to several pollutants simultaneously.

Individuals and groups within a population show marked differences in sensitivity to given pollutants. Individuals with pre-existing lung disease, for instance, may be at higher risk from exposure to air pollutants than healthy people. Differences in response can be due to factors other than pre-existing health, including age, sex, level of exercise taken and other unknown factors. Thus, the population must be considered heterogeneous in respect of response to air pollutants. This perhaps wide distribution of sensitivity combined with a distribution of exposure makes the establishment of population-based thresholds of effect very difficult. This problem is taken up in the section on particulate matter. Existing information tends not to allow adequate assessment of the proportion of the population that is likely to show an enhanced response. Nevertheless, an estimate of even a few percent of the total population entails a large number of people at increased risk.

Deriving a guideline from studies of effects on laboratory animals in the absence of human studies generally requires the application of an increased uncertainty factor, because humans

may be more susceptible than laboratory animal species. Negative data from human studies will tend to reduce the magnitude of this uncertainty factor. Also of importance are the nature and reversibility of the reported effect. Deriving a guideline from data that show that a given level of exposure produces only slight alterations in physiological parameters requires a smaller uncertainty factor than when data showing a clearly adverse effect are used. Scientific judgement about uncertainty factors should also take into account the biochemical toxicology of pollutants, including the types of metabolite formed, the variability in metabolism or response in humans suggesting the existence of hypersusceptible groups, and the likelihood that the compound or its metabolites will accumulate in the body.

It is obvious, therefore, that diverse factors must be taken into account in proposing a margin of protection. The uncertainty factor cannot be assigned by a simple mathematical formula; it requires experience, wisdom and judgement.

Feasibility of adopting a standard approach

In preparing this second edition of the guidelines, the feasibility of developing a standard methodology for setting guidelines was discussed. It was agreed that Environmental Health Criteria No. 170 *(1)* was a valuable source of information. On the other hand, it was recognized that large variation in the data available for different compounds made the use of a standard approach impossible. Much of the difficulty concerns the adequacy of the database, and this has played a large part in controlling the methods of assessment adopted. This is illustrated in Table 1.

It will be seen that when the database is strong (that is, when a good deal is known about the human toxicology of the compound) it is suggested that expert judgement can be used to set a guideline. In such circumstances the level of uncertainty is low. If, on the other hand, the database is weak, then a larger level of uncertainty will exist and there is much to be said for using a standardized approach, probably involving the application of a substantial uncertainty factor. The dangers of replacing expert judgement and the application of common sense with advanced, complex and sometimes not intuitively obvious statistical methods for deriving guidelines was discussed. It was agreed that a cautious approach should be adopted.

Table 1. Size and completeness of database in relation to assessment method

Examples	Completeness/ size of database	Uncertainties	Feasibility of expert judgement	Need for standardized approach
Nitrogen dioxide, ozone, lead	+++	+	+++	+
Manganese, nickel	++	++	++	++
Volatile organic compounds	+	+++	+	+++

Criteria for selection of averaging times

The development of toxicity is a complex function of the interaction between concentration of a pollutant and duration of exposure. A chemical may cause acute, damaging effects after peak exposure for a short period and irreversible or incapacitating effects after prolonged exposure to lower concentrations. Our knowledge is usually insufficient to define accurately the relationship between effects on the one hand and concentration and time on the other. Expert judgement must be applied, therefore, based on the weight of the evidence available *(10)*.

Generally, when short-term exposures lead to adverse effects, short-term averaging times are recommended. The use of a long-term average under such conditions would be misleading, since the typical pattern of repeated peak exposures is lost during the averaging process and the risk manager would have difficulties in deciding on effective strategies. In other cases, knowledge of the exposure–response relationship may be sufficient to allow recommendation of a long averaging period. This is frequently the case for chemicals that accumulate in the body and thereby produce adverse effects. In such cases, the integral of concentration over a long period can have more impact than the pattern of peak exposure.

It should be noted that the specified averaging times are based on effects on health. Therefore, if the guidelines are used as a basis for regulation, the regulator needs to select the most appropriately and practically defined standards in relation to the guidelines, without necessarily adopting the guidelines directly. It was appreciated that monitoring techniques for some pollutants would not allow reporting of data in terms of the averaging times recommended in the guidelines. Under such circumstances, a compromise between the averaging time specified in the guidelines and that obtainable in practice has to be reached in setting an air quality standard.

A similar situation occurs for effects on vegetation. Plants are generally damaged by short-term exposures to high concentration as well as by long-term exposures to low concentration. Therefore, both short- and long-term guidelines to protect plants are proposed.

Criteria for consideration of sensory effects

Some of the substances selected for evaluation have malodorous properties at concentrations far below those at which toxic effects occur. Although odour annoyance cannot be regarded as an adverse health effect in a strict sense, it does affect the quality of life. Therefore, odour threshold levels have been indicated where relevant and used as a basis for separate guideline values.

For practical purposes, the following characteristics and respective levels were considered in the evaluation of sensory effects:

- intensity, where the *detection threshold level* is defined as the lower limit of the perceived intensity range (by convention the lowest concentration that can be detected in 50% of the cases in which it is present);

- quality, where the *recognition threshold level* is defined as the lowest concentration at which the sensory effect, such as odour, can be recognized correctly in 50% of the cases; and

- acceptability and annoyance, where the *nuisance threshold level* is defined as the concentration at which not more than a small proportion of the population (less than 5%) experiences annoyance for a small part of the time (less than 2%); since annoyance will be influenced by a number of psychological and socioeconomic factors, a nuisance threshold level cannot be defined on the basis of concentration alone.

During revision of the guidelines, the problems of irritation (for example, of the skin) and headache were also considered as possible problems of annoyance. It was agreed that headache should be regarded as a health endpoint and not merely as a matter of annoyance.

Criteria for carcinogenic endpoint

Cancer risk assessment is basically a two-step procedure, involving a qualitative assessment of how likely it is that an agent is a human carcinogen, and a quantitative assessment of the cancer risk that is likely to occur at given levels and duration of exposure (*11)*.

Qualitative assessment of carcinogenicity

The decision to consider a substance as a carcinogen is based on the qualitative evaluation of all available information on carcinogenicity, ensuring that the association is unlikely to be due to chance alone. Here the classification criteria of the International Agency for Research on Cancer (IARC) have been applied (Box 1). In dealing with carcinogens, a "general rule" and exceptions from this were defined. The "general rule" states that for compounds in IARC Groups 1 and 2A (proven human carcinogens, and carcinogens with at least limited evidence of human carcinogenicity), guideline values are derived with the use of quantitative risk assessment with low-dose risk extrapolation. For compounds in Groups 2B (inadequate evidence in humans but sufficient evidence in animals), 3 (unclassifiable as to carcinogenicity in humans) and 4 (noncarcinogenic), guideline values are derived with the use of a threshold (uncertainty factor) method. For compounds in Group 2B, this may incorporate a separate factor for the possibility of a carcinogenic effect in humans.

In case of sufficient scientific evidence, one may be justified in deviating from the "general rule". First, a compound classified in Group 1 or 2A may be assessed with the use of the uncertainty factor methodology, provided that there is strong evidence that it is not genotoxic as judged from a battery of short-term test systems for gene mutation, DNA damage, etc. In such cases it can be established with certainty that an increase in exposure to the compound is associated with an increase in cancer incidence only above a certain level of exposure. It was considered that this required a level of understanding of the mechanisms of action not presently available for the compounds classified as Group 1 or 2A on the current list. Second, a compound in Group 2B may be assessed with the use of quantitative risk assessment methods instead of the uncertainty factor approach. This may be considered appropriate where the mechanism of carcinogenesis in animals is likely to be a non-threshold phenomenon as indicated, for example, by the genotoxic activity of the compound in different short-term test systems.

Quantitative assessment of carcinogenic potency

The aim of quantitative risk assessment is to use information available from very specific study situations to predict the risk to the general population posed by exposure to ambient levels of carcinogens. In general, therefore, quantitative risk assessment includes the extrapolation of risk from relatively high dose levels (characteristic of animal experiments or occupational exposures), where cancer responses can be measured, to relatively low dose

Box 1. Classification criteria of the International Agency for Research on Cancer

Group 1 – the agent (mixture) is carcinogenic to humans.
The exposure circumstance entails exposures that are carcinogenic to humans.
This category is used when there is *sufficient evidence* of carcinogenicity in humans. Exceptionally, an agent (mixture) may be placed in this category when evidence in humans is less than sufficient but there is *sufficient evidence* of carcinogenicity in experimental animals and strong evidence in exposed humans that the agent (mixture) acts through a relevant mechanism of carcinogenicity.

Group 2
This category includes agents, mixtures and exposure circumstances for which, at one extreme, the degree of evidence of carcinogenicity in humans is almost sufficient, as well as those for which, at the other extreme, there are no human data but for which there is evidence of carcinogenicity in experimental animals. Agents, mixtures and exposure circumstances are assigned to either group 2A (probably carcinogenic to humans) or group 2B (possibly carcinogenic to humans) on the basis of epidemiological and experimental evidence of carcinogenicity and other relevant data.

Group 2A – the agent (mixture) is probably carcinogenic to humans.
The exposure circumstance entails exposures that are probably carcinogenic to humans.
This category is used when there is *limited evidence* of carcinogenicity in humans and sufficient evidence of carcinogenicity in experimental animals. In some cases, an agent (mixture) may be classified in this category when there is inadequate evidence of carcinogenicity in humans and *sufficient evidence* of carcinogenicity in experimental animals and strong evidence that the carcinogenesis is mediated by a mechanism that also operates in humans. Exceptionally, an agent, mixture or exposure circumstance may be classified in this category solely on the basis of limited evidence of carcinogenicity in humans.

Group 2B – the agent (mixture) is possibly carcinogenic to humans.
The exposure circumstance entails exposures that are possibly carcinogenic to humans.
This category is used for agents, mixtures and exposure circumstances for which there is *limited evidence* of carcinogenicity in humans and less than *sufficient evidence of* carcinogenicity in experimental animals. It may also be used when there is *inadequate evidence* of carcinogenicity in humans but there is *sufficient evidence* of carcinogenicity in experimental animals. In some instances, an agent, mixture or exposure circumstance for which there is *inadequate evidence* of carcinogenicity in humans but *limited evidence* of carcinogenicity in experimental animals together with supporting evidence from other relevant data may be placed in this group.

Group 3 – The agent (mixture or exposure circumstance) is not classifiable as to its carcinogenicity to humans.
This category is used most commonly for agents, mixtures and exposure circumstances for which the evidence of carcinogenicity is inadequate in humans and inadequate or limited in experimental animals. Exceptionally, agents (mixtures) for which the evidence of carcinogenicity is inadequate in humans but sufficient in experimental animals may be placed in this category when there is strong evidence that the mechanism of carcinogenicity in experimental animals does not operate in humans. Agents, mixtures and exposure circumstances that do not fall into any other group are also placed in this category.

Group 4 – The agent (mixture) is probably not carcinogenic to humans.
This category is used for agents or mixtures for which there is *evidence suggesting lack of carcinogenicity* in humans and in experimental animals. In some instances, agents or mixtures for which there is *inadequate*
evidence of carcinogenicity in humans but *evidence suggesting lack of carcinogenicity* in experimental animals, consistently and strongly supported by a broad range of other relevant data, may be classified in this group.

Source: IARC *(12)*.

levels, which are of concern in environmental protection and where such risks are too small to be measured directly, either by animal studies or by epidemiological studies.

The choice of the extrapolation model depends on the current understanding of the mechanisms of carcinogenesis *(13)*, and *no* single mathematical procedure can be regarded as

fully appropriate for low-dose extrapolation. Methods based on a linear, non-threshold assumption have been used at the national and international level more frequently than models that assume a safe or virtually safe threshold.

In these guidelines, the risk associated with lifetime exposure to a certain concentration of a carcinogen in the air has been estimated by linear extrapolation and the carcinogenic potency expressed as the incremental unit risk estimate. The incremental unit risk estimate for an air pollutant is defined as "the additional lifetime cancer risk occurring in a hypothetical population in which all individuals are exposed continuously from birth throughout their lifetimes to a concentration of 1 μg/m^3 of the agent in the air they breathe" *(14)*.

The results of calculations expressed in unit risk estimates provide the opportunity to compare the carcinogenic potency of different compounds and can help to set priorities in pollution control, taking into account current levels of exposure. By using unit risk estimates, any reference to the "acceptability" of risk is avoided. The decision on the acceptability of a risk should be made by national authorities within the framework of risk management. To support authorities in the decision-making process, the guideline sections for carcinogenic pollutants provide the concentrations in air associated with an excess cancer risk of 1 in a population of 10 000, 1 in 100 000 or 1 in 1 000 000, respectively, calculated from the unit risk.

For those substances for which appropriate human studies are available, the method known as the "average relative risk model" has been used, and is therefore described in more detail below.

Several methods have been used to estimate the incremental risks based on data from animal studies. Two general approaches have been proposed. A strictly linearized estimate has generally been used by the US Environmental Protection Agency (EPA) *(14)*. Nonlinear relations have been proposed for use when the data derived from animal studies indicate a nonlinear dose–response relationship or when there is evidence that the capacity to metabolize the polluting chemical to a carcinogenic form is of limited capacity.

Quantitative assessment of carcinogenicity based on human data

Quantitative assessment using the average relative risk model comprises four steps: (*a*) selection of studies; (*b*) standardized description of study results in terms of relative risk, exposure level and duration of exposure; (*c*) extrapolation towards zero dose; and (*d*) application to a general (hypothetical) population.

First, a reliable human study must be identified, where the exposure of the study population can be estimated with acceptable confidence and the excess cancer incidence is statistically significant. If several studies exist, the best representative study should be selected or several risk estimates evaluated.

Once a study is identified, the relative risk as a measure of response is calculated. It is important to note that the 95% confidence limits around the central estimate of the relative risk can be wide and should be specifically stated and evaluated. The relative risk is then used to calculate the excess lifetime cancer risk expressed as unit risk (UR) associated with a lifetime exposure to 1 μg/m^3, as follows:

$$UR = \frac{P_0(RR - 1)}{X}$$

where: P_0 = background lifetime risk; this is taken from age/cause-specific death or incidence rates found in national vital statistics tables using the life table methodology, or it is available from a matched control population

RR = relative risk, being the ratio between the observed (O) and expected (E) number of cancer cases in the exposed population; the relative risk is sometimes expressed as the standardized mortality ratio $SMR = (O/E) \times 100$

X = lifetime average exposure (standardized lifetime exposure for the study population on a lifetime continuous exposure basis); in the case of occupational studies, X represents a conversion from the occupational 8-hour, 240-day exposure over a specific number of working years and can be calculated as X = 8-hour TWA $\times$ 8/24 $\times$ 240/365 $\times$ (average exposure duration [in years])/(life expectancy [70 years]), where TWA is the time-weighted average ($\mu g/m^3$).

It should be noted that the unit lifetime risk depends on P_0 (background lifetime risk), which is determined from national age-specific cancer incidence or mortality rates. Since these rates are also determined by exposures other than the one of interest and may vary from country to country, it follows that the UR may also vary from one country to another.

Necessary assumptions for average relative risk method

Before any attempt is made to assess the risk in the general population, numerous assumptions are needed at each phase of the risk assessment process to fill in various gaps in the underlying scientific database. As a first step in any given risk assessment, therefore, an attempt should be made to identify the major assumptions that have to be made, indicating their probable consequences. These assumptions are as follows.

1. *The response (measured as relative risk) is some function of cumulative dose or exposure.*

2. *There is no threshold dose for carcinogens.*
Many stages in the basic mechanism of carcinogenesis are not yet known or are only partly understood. Taking available scientific findings into consideration, however, several scientific bodies *(8, 15–17)* have concluded that there is no scientific basis for assuming a threshold or no-effect level for chemical carcinogens. This view is based on the fact that most agents that cause cancer also cause irreversible damage to deoxyribonucleic acid (DNA). The assumption applies for all non-threshold models.

3. *The linear extrapolation of the dose–response curve towards zero gives an upper-bound conservative estimate of the true risk function if the unknown (true) dose–response curve has a sigmoidal shape.*

The scientific justification for the use of a linear non-threshold extrapolation model stems from several sources: the similarity between carcinogenesis and mutagenesis as processes that both have DNA as target molecules; the strong evidence of the linearity of dose–response relationships for mutagenesis; the evidence for the linearity of the DNA binding of chemical carcinogens in the liver and skin; the evidence for the linearity in the dose–response relationship in the initiation stage of the mouse 2-stage tumorigenesis model; and the rough consistency with the linearity of the dose–response relationships for several epidemiological studies. This assumption applies for all linear models.

4. *There is constancy of the relative risk in the specific study situation.* In a strict sense, constancy of the relative risk means that the background age/cause-specific rate at any time is increased by a constant factor. The advantage of the average relative risk method is that this needs to be true only for the average.

Advantages of the method

The average relative risk method was selected in preference to many other more sophisticated extrapolation models because it has several advantages, the main one being that it seems to be appropriate for a fairly large class of different carcinogens, as well as for different human studies. This is possible because averaging doses, that is, averaging done over concentration and duration of exposure, gives a reasonable measure of exposure when dose rates are not constant in time. This may be illustrated by the fact that the use of more sophisticated models *(14, 18, 19)* results in risk estimates very similar to those obtained by the average relative risk method.

Another advantage of the method is that the carcinogenic potency can be calculated when estimates of the average level and duration of exposure are the only known parameters besides the relative risk. Furthermore, the method has the advantage of being simple to apply, allowing non-experts in the field of risk models to calculate a lifetime risk from exposure to the carcinogens.

Limitations of the method

As pointed out earlier, the average relative risk method is based on several assumptions that appear to be valid in a wide variety of situations. There are specific situations, however, in which the method cannot be recommended, mainly because the assumptions do not hold true.

The cumulative dose concept, for instance, is inappropriate when the mechanism of the carcinogen suggests that it cannot produce cancer throughout all stages of the cancer development process. Also, specific toxicokinetic properties, such as a higher excretion rate of a carcinogen at higher doses or a relatively lower production rate of carcinogenic metabolites at lower doses, may diminish the usefulness of the method in estimating cancer risk. Furthermore, supralinearity of the dose–response curve or irregular variations in the relative risk over time that cannot be eliminated would reduce the value of the model. Nevertheless, evidence concerning these limitations either does not exist or is still too preliminary to make the average relative risk method inappropriate for carcinogens evaluated here.

A factor of uncertainty, rather than of methodological limitation, is that data on past exposure are nearly always incomplete. Although it is generally assumed that in the majority of studies

the historical dose rate can be determined within an order of magnitude, there are possibly greater uncertainties, even of more than two orders of magnitude, in some studies. In the risk assessment process it is of crucial importance that this degree of uncertainty be clearly stated. This is often done simply by citing upper and lower limits of risk estimates. Duration of exposure and the age- and time-dependence of cancer caused by a particular substance are less uncertain parameters, although the mechanisms of relationship are not so well understood *(11)*.

Risk estimates from animal cancer bioassays

Animal bioassays of chemicals provide important information on the human risk of cancer from exposure to chemicals. These data enhance our confidence in assessing human cancer risks on the basis of epidemiological data.

There is little doubt of the importance of animal bioassay data in reaching an informed decision on the carcinogenic potential of a chemical. The collection and use of data such as those on saturation mechanisms, absorption, distribution and metabolic pathways, as well as on interaction with other chemicals, is important and should be continued. Regrettably, these data were not always available for the air pollutants evaluated during the update and revision of the guidelines. The process of evaluating guidelines and the impact of exposure to these chemicals on human health should continue and be revised as new information becomes available.

Several chemicals considered in this publication have been studied using animal cancer bioassays. The process is continuing and new information on the potential carcinogenicity of chemicals is rapidly appearing. Consequently, the status of chemicals is constantly being reassessed.

There is no clear consensus on appropriate methodology for the risk assessment of chemicals for which the critical effect may not have a threshold, such as genotoxic carcinogens and germ cell mutagens. A number of approaches based largely on characterization of dose–response have been adopted for assessment of such effects:

- quantitative extrapolation by mathematical modelling of the dose–response curve to estimate the risk at likely human intakes or exposures (low-dose risk extrapolation);
- relative ranking of potencies in the experimental range; and
- division of effect levels by an uncertainty factor.

Low-dose risk extrapolation has been accomplished by the use of mathematical models such as the Armitage-Doll multi-stage model. In more recently developed biological models, the different stages in the process of carcinogenesis have been incorporated and time to tumour has been taken into account *(20)*. In some cases, such as that of butadiene, uncertainty regarding the metabolism in humans and experimental animals precluded the choice of the appropriate animal model for low-dose risk extrapolation. In other cases where data permitted, attempts were made to incorporate the dose delivered to the target tissue into the dose–response analysis (physiologically based pharmacokinetic modelling).

During revision of the guidelines, other approaches to establishing guideline levels for carcinogens were considered. Such approaches involve the identification of a level of exposure at which the risk is known to be small and the application of uncertainty factors to

derive a level of exposure at which the risk is accepted as being exceedingly small or negligible. This approach has been adopted in the United Kingdom, for example. It was agreed that such an approach might be applicable on a national or smaller scale, but that it was unlikely to be generally applicable.

Interpretation of risk estimates

The risk estimates presented in this book should *not* be regarded as being equivalent to the true cancer risk. It should be noted that crude expression of risk in terms of excess incidence or numbers of cancers per unit of the population at doses or concentrations much less than those on which the estimates are based may be inappropriate, owing to the uncertainties of quantitative extrapolation over several orders of magnitude. Estimated risks are believed to represent only the plausible upper bounds, and may vary widely depending on the assumptions on which they are based.

The presented quantitative risk estimates can provide policy-makers with rough estimates of risk that may serve well as a basis for setting priorities, balancing risks and benefits, and establishing the degree of urgency of public health problems among subpopulations inadvertently exposed to carcinogens. A risk management approach for compounds for which the critical effect is considered not to have a threshold involves eliminating or reducing exposure as far as practically or technologically possible. Characterization of the dose–response, as indicated in the procedures described above, can be used in conjunction with this approach to assess the need to reduce exposure.

Combined exposures

Exposure to combinations of air pollutants is inevitable. Data dealing with the effects of co-exposure to air pollutants are, however, very limited and it is not possible to recommend guidelines for such combinations. Of course, measures taken to control air pollution frequently lead to the reduction in concentrations of more than one pollutant. This is often achieved by controlling sources of pollutants rather than by focusing on individual pollutants.

Ecological effects

The importance of taking an integrated view of both health and ecological effects in air quality management was recognized from the beginning of the project. Ecological effects may have a significant indirect influence on human health and wellbeing. For example, most of the major urban air pollutants are known to have adverse effects at low levels on plants, including food crops. A consultation group was therefore convened to consider the ecological effects on terrestrial vegetation of sulfur dioxide, nitrogen dioxide, and ozone and other photochemical oxidants. These substances are important both because of the high anthropogenic amounts produced and because of their wide distribution. They deserve special attention because of significant adverse effects on ecological systems in concentrations far below those known to be harmful to humans.

The pollutants selected for consideration here form only part of the vast range of air pollutants that have ecological effects. The project timetable permitted only an evaluation of adverse effects on terrestrial plant life, although effects on animal and aquatic ecosystems are also of great concern in parts of Europe. Nevertheless, even this limited evaluation clearly indicates the importance attached to the ecological effects of such pollutants in the European Region.

References

1. *Assessing human health risks of chemicals: derivation of guidance values for health-based exposure limits.* Geneva, World Health Organization, 1994 (Environmental Health Criteria, No. 170).
2. CRUMP, K.S. A new method for determining allowable daily intakes. *Fundamental and applied toxicology*, **4**: 854–871 (1984).
3. US ENVIRONMENTAL PROTECTION AGENCY. Guidelines and methodology used in the preparation of health effect assessment chapters of the consent decree water quality criteria. *Federal register*, **45**: 79347–79357 (1980).
4. ALLEN, B.C. ET AL. Dose–response modeling for developmental toxicity. *Toxicologist*, **12**: 300 (1992).
5. ALLEN, B.C. ET AL. Comparison of quantitative dose response modeling approaches for evaluating fetal weight changes in segment II developmental toxicity studies. *Teratology*, **47**(5): 41 (1993).
6. FARLAND, W. & DOURSON, M. Noncancer health endpoints: approaches to quantitative risk assessment. *In*: Cothern, R., ed. *Comparative environmental risk assessment.* Boca Raton, FL, Lewis Publishers, 1992, pp. 87–106.
7. VETTORAZZI, G. *Handbook of international food regulatory toxicology. Vol. 1. Evaluations.* New York, SP Medical and Scientific Books, 1980.
8. *Guidelines for drinking-water quality. Vol. 1. Recommendations.* Geneva, World Health Organization, 1984.
9. *Air quality guidelines for Europe*. Copenhagen, WHO Regional Office for Europe, 1987 (WHO Regional Publications, European Series, No. 23).
10. *Air quality criteria and guides for urban air pollutants: report of a WHO Expert Committee*. Geneva, World Health Organization, 1972 (WHO Technical Report Series, No. 506).
11. PEAKALL, D.B. ET AL. Methods for quantitative estimation of risk from exposure to chemicals. *In*: Vouk, V.B. et al., ed. *Methods for estimating risk of chemical injury: human and non-human biota and ecosystems.* New York, John Wiley & Sons, 1985.
12. *Polychlorinated dibenzo*-para-*dioxins and polychlorinated dibenzofurans.* Lyon, International Agency for Research on Cancer, 1997 (IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Vol. 69).
13. ANDERSON, E.L. Quantitative approaches in use in the United States to assess cancer risk. *In*: Vouk, V.B. et al., ed. *Methods for estimating risk of chemical injury: human and non-human biota and ecosystems.* New York, John Wiley & Sons, 1985.
14. *Health assessment document for nickel*. Research Triangle Park, NC, US Environmental Protection Agency, 1985 (Final Report No. EPA-600/8-83-12F).
15. ANDERSON, E.L. ET AL. Quantitative approaches in use to assess cancer risk. *Risk analysis*, **3**: 277–295 (1983).
16. NATIONAL RESEARCH COUNCIL. *Drinking water and health.* Washington, DC, National Academy Press, 1977.
17. *Risk assessment and risk management of toxic substances. A report to the Secretary, Department of Health and Human Services.* Washington, DC, US Department of Health and Human Services, 1985.
18. *Health assessment document for chromium.* Washington, DC, US Environmental Protection Agency, 1984 (Final report EPA-600-8-83-014F).
19. *Health assessment document for inorganic arsenic.* Washington, DC, US Environmental Protection Agency, 1984 (Final report EPA-600-8-83-021F).

20. MOOLGAVKAR, S.H. ET AL. A stochastic two-stage model for cancer risk assessment. I. The hazard function and the probability of tumor. *Risk analysis,* **8**: 383–392 (1988).

