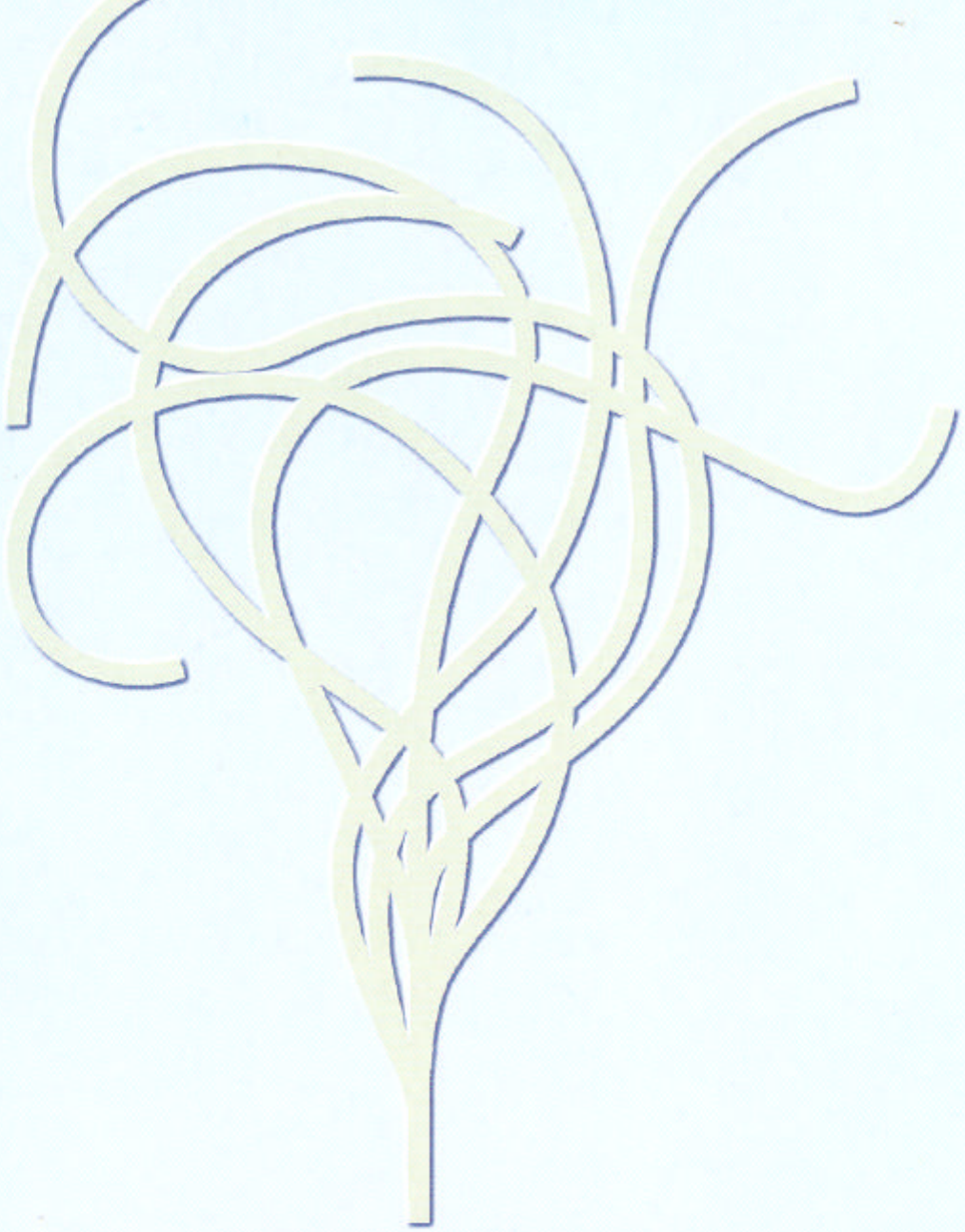

EVALUATION AND USE OF EPIDEMIOLOGICAL EVIDENCE FOR ENVIRONMENTAL HEALTH RISK ASSESSMENT

WORLD HEALTH ORGANIZATION
EUROPE

EVALUATION AND USE OF EPIDEMIOLOGICAL EVIDENCE FOR ENVIRONMENTAL HEALTH RISK ASSESSMENT

Guideline Document

World Health Organization
Regional Office for Europe

Scherfigsvej 8
DK-2100 Copenhagen Ø
Denmark

2000

EUR/00/5020369 E68940

EUROPEAN HEALTH21 TARGET 10

A HEALTHY AND SAFE PHYSICAL ENVIRONMENT

By the year 2015, people in the Region should live in a safer physical environment, with exposure to contaminants hazardous to health at levels not exceeding internationally agreed standards

(Adopted by the WHO Regional Committee for Europe at its forty-eighth session, Copenhagen, September 1998)

EUROPEAN HEALTH21 TARGET 19

RESEARCH AND KNOWLEDGE FOR HEALTH

By the year 2005, all Member States should have health research, information and communication systems that better support the acquisition, effective utilization, and dissemination of knowledge to support health for all

(Adopted by the WHO Regional Committee for Europe at its forty-eighth session, Copenhagen, September 1998)

ABSTRACT

Environmental health risk assessment is increasingly being used in the development of environmental health policies, public health decision-making, the establishment of environmental regulations and research planning. The credibility of risk assessment depends, to a large extent, on the strength of the scientific evidence on which it is based. It is, therefore, imperative that the processes and methods used to evaluate the evidence and estimate health risks are clear and explicit, and based on valid epidemiological theory and practice.

Evaluation and use of epidemiological evidence for environmental health risk assessment is a guideline document. The primary target audiences of the document are expert review groups that WHO (or other organizations) might convene in the future to evaluate epidemiological evidence on the health effects of environmental factors.

This Guideline Document identifies a set of processes and general approaches to assess available epidemiological information in a clear, consistent and explicit manner. The Guideline Document should also help in the evaluation of epidemiological studies with respect to their ability to support risk assessment and, consequently, risk management. Conducting expert reviews according to such explicit guidelines would make health risk assessment, and subsequent risk management and risk communication processes, more readily understood and likely to be accepted by policy-makers and the public. It would also make the conclusions reached by reviews more readily acceptable as a basis for future WHO guidelines and other recommendations, and would provide a more rational basis for setting priorities for future research.

Keywords

ENVIRONMENTAL HEALTH
RISK ASSESSMENT
EPIDEMIOLOGIC MEASUREMENTS
GUIDELINES
INTERNATIONAL COOPERATION

This document was text processed in Health Documentation Services
WHO Regional Office for Europe, Copenhagen

CONTENTS

Page

1. Introduction

Environmental health risk assessment contributes increasingly to policy development, public health decision-making, the establishment of environmental regulations and research planning. It also often plays an important role in cost-benefit analysis and risk communication. Its credibility depends, to a large extent, on the strength of the scientific evidence on which it is based. Epidemiology, toxicology, clinical medicine, and environmental exposure assessment all contribute information for risk assessment.

However, epidemiological studies play a unique role in the assessment of the health risk of environmental factors. Unlike laboratory experiments, epidemiology provides evidence based on studies of human populations under real-world conditions. It largely avoids the extrapolations across species and levels of exposure that are required for the use of data from animal experiments, and which contribute large uncertainties. In addition, epidemiology has often contributed to the recognition of new hazards, thereby stimulating new research and identifying new areas for public health action. Epidemiology's contribution to health risk assessment has been widely discussed (see e.g. NRC 1983, Gordis 1988, Federal Focus Inc. 1996, Hertz-Picciotto 1995, Samet & Burke 1998). However, epidemiological studies that report associations between measures of the health of populations and the presence of hazardous factors in the environment are frequently difficult to interpret (Neutra & Trichopoulos 1993, Taubes 1995). Therefore, a careful evaluation of all existing epidemiological evidence is necessary as part of the risk assessment process.

In order to provide authoritative assessments of environmental epidemiology research, public health and regulatory agencies may rely on expert review groups to evaluate the evidence, draw conclusions on the existence of hazard to health, and estimate the magnitude of associated health risks. These expert reviews may then be used to support actions that are difficult and expensive. It is, therefore, imperative that the processes and methods used to evaluate the evidence and estimate health risks are clear and explicit, and based on valid epidemiological theory and practice.

To improve the methodology used by the expert groups reviewing the evidence, the WHO European Centre for Environment and Health, Bilthoven Division, in collaboration with the International Programme on Chemical Safety, initiated the project "Accepting epidemiological evidence for health impact assessment". The results of this project are summarized in this report.

This project has been supported by a special grant from the Swiss Agency for the Environment, Forests and Landscape, which is gratefully acknowledged.

2. Scope and purpose

The purpose of this project is to develop guidelines, which identify a set of processes and general approaches to assess available epidemiological information in a clear, consistent and explicit manner. The guidelines should also help in the evaluation of epidemiological studies with respect to their ability to support risk assessment and, consequently, risk management.

Conducting expert reviews according to such explicit guidelines would make health risk assessment, and subsequent risk management and risk communication processes, more readily understood and likely to be accepted by policy-makers and the public. From WHO's standpoint, it would also make the conclusions reached by reviews more readily acceptable as a basis for future WHO guidelines and other recommendations, and would provide a more rational basis for setting priorities for future research.

This project focuses only on approaches to the evaluation and use of epidemiological evidence for health risk assessment. However, this should not be interpreted as implying that only epidemiological studies are important. The Working Group, and WHO, appreciate that data from toxicological, clinical, and other areas of research often play vital roles in both the characterization of health hazards and the estimation of risks to health, and may, in the absence of suitable epidemiological data, provide the sole basis for such activities.

Public health action (e.g. the reduction of population exposure to a suspected hazard or even its elimination from the human environment) must often proceed even when the scientific evidence is insufficient. Most of the Working Group members agree that the Precautionary Principle should play a role in guiding public health action where there is uncertainty.

The project considers two distinct activities of health risk assessment: Health Hazard Characterization and Health Impact Assessment. They correspond to components of risk assessment defined by both the US National Research Council (NRC) and WHO (NRC 1983, WHO 1999).

- **Health Hazard Characterization** involves the identification of environmental hazards via the collection, evaluation, and interpretation of available evidence from epidemiology and other scientific disciplines concerning the association between an environmental factor and human health.

- **Health Impact Assessment** involves the quantification of the expected health burden due to an environmental exposure in a specific population.

Health Hazard Characterization comprises the *hazard identification*, and elements of the *dose-response assessment* stages of the NRC Risk Assessment paradigm, at least as they apply to epidemiological studies. Health Impact Assessment combines the *exposure assessment, dose-response assessment*, and *risk characterization* stages of the NRC and WHO risk assessment paradigm. Thus the two stages overlap and interlock.

This project focuses on the evaluation and use of epidemiological evidence on associations between environmental factors and health. This evidence is used to support the assessment of health impact of certain exposures. The Health Impact Assessment discussed in this report is, therefore, not synonymous with the broader concept of assessment of health impacts of "specified action on the health of a defined population". The latter is an emerging tool in the evidence-based public health policy making. It can be applied to a wide range of actions, policies or projects on various determinants of health, such as behavioral factors, socioeconomic issues or health care system reforms. Environmental Health Risk Assessment, with its well established methodology, is a significant contribution to the wider concept of Health Impact Assessment (Scott-Samuel 1998).

The scope and purpose of the project with respect to the audience and the issues that it would address have been further refined as follows:

- The *target audiences* for these guidelines are expert review groups that WHO (or other organizations) might convene in the future to evaluate epidemiological evidence on the health effects of environmental factors.

- The Working Group, convened to develop this project, saw its role as providing future review groups with general recommendations and principles for conducting such evaluations, rather than providing formulae, or lists of "approved methods".

- The Working Group focused on the evaluation of evidence in the context of large-scale public health issues, as opposed to "local emergencies." The Working Group acknowledged that epidemiological studies of local environmental exposures (e.g. clusters of childhood leukaemia in the vicinity of nuclear power plants) might provide evidence about large-scale public health concerns, but felt that the evaluation of such "outbreaks" presents a unique set of problems, which warrant attention in their own right.

- The Working Group agreed that health impact assessments are conducted for a range of purposes and under a variety of conditions, and, therefore, the purpose of the health impact assessment will, and should, determine its scope, form and content.

- The principles described apply, in the first place, to chemical pollutants. In reviews concerning some other exposures, adjustments may be proposed. Should such deviations from the principles be applied, a clear justification must be given.

3. Process

The WHO Working Group of experts in epidemiology, public health, and environmental policy was assembled at the end of 1998 (see Annex 1). The experts were selected on the basis of their:

- experience in the scientific review of epidemiological evidence for the governmental bodies, WHO or other public and private sector organizations;

- involvement in risk assessment of environmental factors;

- involvement in communication related to health risk with general public or decision makers;

- representation of wide range of countries within the European Region of the WHO and the United States.

The experts were asked to prepare working papers presenting their views and proposals concerning the process of review of epidemiological evidence, as well as criteria for its acceptance and use in assessment of health risk of environmental factors. The papers were distributed to all Working Group members and provided a basis for the discussion at the meeting of the Working Group, convened in Il Ciocco, Italy, from 31 May to 2 June 1999. Dr Robert Maynard chaired the meeting, and Dr Aaron Cohen acted as its Rapporteur. The working papers are not included in this report, although the individual authors may choose to publish them elsewhere.

After a half-day plenary discussion to establish the exact scope of the meeting and the methods of working, two sub-Working Groups were formed: one to consider Health Hazard Characterization, and another to consider Health Impact Assessment. A third group met initially to discuss issues related to the broader social and public policy contexts in which environmental health risk assessment is used. Their views were ultimately incorporated into the chapters of the two main sub-Working Groups.

After an iterative process of subgroup discussion and plenary meetings, the sub-Working Groups summarized their discussions and drafted recommendations on their respective topics, which were further refined following the plenary meeting of the Working Group. These discussions are summarized in two papers prepared by the two working groups and are presented in Annex 3. Based on these materials the rapporteur of the meeting, assisted by the chairmen and rapporteurs of the subgroups, and the secretariat, prepared a draft of the meeting report. Prior to its finalization, that draft was presented to all members of the Working Group to ensure that it correctly reflected the Working Group's consensus on the recommendations and the rationale for them.

The draft report was discussed at a special WHO Symposium, organized at the joint conference of the International Society for Environmental Epidemiology and the International Society of Exposure Analysis in Athens, in September 1999 and made available, through the World Wide Web, for review. The comments that were received (see the list of their authors in Annex 2) were used in the preparation of the final draft of this Guideline Document in November 1999. Revisions based on these comments have focused on improving the depth and clarity of presentation of the Recommendations and Conclusions, rather than on additional detailed discussion of methodological issues of risk assessment presented in Annex 3. The draft was reviewed by the chairmen and rapporteurs of the Working Group (and subgroups) and accepted, with small editorial changes, as the Guideline Document in January 2000.

4. Conclusions and recommendations

This chapter lists the major recommendations made by the Working Group for the evaluation and use of environmental epidemiology studies for health risk assessment. They constitute the core of the *Evaluation and use of epidemiological evidence for environmental health risk assessment.* This Guideline Document comprises a set of general recommendations, followed by specific recommendations for evaluations of epidemiological research for Health Hazard Characterization, and use of epidemiological data for Health Impact Assessment.

4.1 General recommendations

1. Expert review groups should adopt **a systematic and explicit approach** to the assessment of epidemiological evidence for health risk assessment. The Working Group acknowledged that various expert review groups had, in the past, used a variety of methods and standards to assess epidemiological evidence, and that these were often inadequately described.

2. Expert review groups, and the agencies that sponsor them, should strive for **better communication** with stakeholders (e.g. citizens, private interests, and government agencies) regarding the process of evaluating, and drawing conclusions from, epidemiological evidence. The need for an evaluation of the epidemiological evidence often reflects the existence of divergent views among stakeholders about the true extent of the risk. When

expert review groups make explicit, and explain in clear terms, the methods they use to conduct their evaluations and reach their conclusions, they reduce the potential for those conclusions to be misunderstood and mistrusted by stakeholders.

3. To **improve the applicability of epidemiological research to Health Risk Assessment**, future epidemiological studies should seek where possible to provide results in a way that enhances the Health Risk Assessment at the interface of epidemiology, other fields of research, and policy-making. In particular, the study reports should describe as precisely as possible the exposure characteristics, shape of the exposure-response function, as well as a distinguish between the acute and chronic effects of exposure.

4. **The WHO secretariat of the future reviews should assess the feasibility of implementation of these recommendations** and the increased time and effort that will be needed and modify the Guideline Document as necessary. WHO should also attempt to assess whether their use leads to increased acceptance by stakeholders of the evaluations of environmental epidemiology research produced by expert review groups.

5. It has also been noted that the proposed more rigorous and thorough approach to the review of the evidence and its use in health impact assessment may require increased effort and resources. The transparency of the methods should, however, lead to a wider acceptance and applicability of the reviews, and may reduce the need for duplication of effort and facilitate updating.

6. The Working Group did not propose a “scale” or “rating” of the evidence, with respect to a “level of proof” required to support risk management decisions. Although Health Hazard Characterization precedes, and is often viewed as a prerequisite for, Health Impact Assessments, (because it provides the scientific justification for them, and provides data for the calculation of risk estimates), the existence of a specific level of scientific evidence required to justify either a Health Impact Assessment, or subsequent action, is controversial. For example, expert judgement that the available evidence is consistent with a causal relationship between exposure and health effect is considered a necessary condition for action by some, but not all (see Rothman 1988). Discussion of the level of evidence on hazard needed to conduct Health Impact Assessment is summarized at the end section A3.1.

4.2 Recommendations for the evaluation of epidemiological evidence for Health Hazard Characterization

The Working Group recommended five general guidelines for five aspects of the evaluation of epidemiological research:

1. **Development of a protocol for the review**. Expert assessments of epidemiological evidence for Health Hazard Characterization should be conducted systematically according to an explicit protocol, defined in advance. The objectives of a systematic review are transparency, avoidance of bias, validity, replicability, and comprehensiveness. A systematic approach provides an efficient way of updating the evidence base as new studies emerge, and will facilitate research planning. A protocol for the systematic review ensures that the expert group has a common understanding of its task, and will adhere to the systematic approach recommended by WHO. It is expected that revisions of the

protocol may be needed as new aspects of the task emerge during the review. The essential components of the protocol will be:

- Specification of the question(s) to be addressed by the Health Hazard Characterization.

- Justification of the expertise represented in the Health Hazard Characterization expert group. The criteria for selection should be based on having the appropriate mix of scientific expertise and experience. Within these criteria, WHO will also consider the need for geographical representation.

- Specification of the methods to be used for identification of relevant studies, assessment of evidence of the individual studies and interpretation of the entire body of available evidence (see below).

2. **Identification of relevant studies**. The assessment should be based on comprehensive identification of all relevant studies. A comprehensive bibliographic search would include the following:

   - involvement of qualified searchers (e.g. librarians, trained investigators);
   - definition of an explicit search strategy including identification of key words;
   - an effort to include all available studies;
   - searching of bibliographic databases;
   - inclusion of non-English reports.

Optional methods, which might be considered by the expert group include hand searching of journals, and inclusion of abstracts and unpublished data (including writing to authors of published data).

3. **Systematic assessment of the validity of epidemiological studies**. As Hill (1965) emphasized, this assessment should aim at answering the question "Is there any other way of explaining the set of facts before us [study results], is there any other answer equally, or more, likely than cause and effect?" The evaluation should consider:

   - The evidence on strength of association, its temporality, biological plausibility, coherence, consistency and specificity.

   - Characteristics of exposure response-relationships. The demonstration of specific patterns of association can provide strong support for causal interpretations if pathophysiological models agree with them. In such cases, more complex, and hence less implausible, patterns of confounding or bias are required as counter-explanations. In addition, the information on exposure-response relationships in particular study populations is an important component in Health Impact Assessments of other populations (see Section 4.3).

   - Alternative explanations for the observed associations. They fall into three categories: chance, bias (information, selection, analytic), and confounding.

   - Results of any sensitivity analysis. In such analysis the outcome variable(s) are examined with respect to (1) changes in expression of exposure variables, (2) addition of other plausible explanatory variables, and/or (3) introduction or removal of confounding variables.

4. **Conduct of systematic overviews of evidence from multiple studies: the use of meta-analysis**. Although meta-analysis is widely viewed as simply a method for statistically combining the results of multiple studies, it can contribute more to hazard characterization when viewed as a quantitative review of the literature, a "study of studies". Conducted in this way, a meta-analysis looks for consistent patterns among, and sources of discrepancies between, studies (Greenland 1987, Rothman & Greenland 1998). Expert groups should consider the following questions when conducting meta-analyses.

    - How will heterogeneity among studies be assessed?
    - Will summary effect estimates be calculated, and by which methods?

The Working Group recommends that expert review groups consider the following issues when designing and conducting quantitative reviews (meta-analyses) of epidemiological literature or assessing their findings:

- *Protocol*. Each meta-analysis must have its own protocol, perhaps "nested" within the overall protocol for the health hazard characterization. The protocol should include a clear statement of the objectives of the review, and the methods to be employed.

- *Inclusion criteria.* It is desirable for a meta-analysis to be inclusive rather than exclusive. Sensitivity to various inclusion criteria can then be examined.

- *Use of quality scores*. Reducing the features of a set of epidemiological studies to a single measure of "quality" is not recommended because these features may affect the results of the studies in different directions and to varying degrees. It is preferable to assess the characteristics of the primary studies individually.

- *Chance*. In meta-analysis, the results are usually weighted by the statistical precision (in general, by the amount of information) of each primary study. Adjustment for the amount of information can be achieved through either inverse-variance weighting or random effects models.

- *Publication bias*. The results of certain kinds of primary studies are more likely to be published than of the others. Publication bias can be detected, minimized or corrected. Its impact can, and should, also be assessed by sensitivity analysis.

- *Assessment of overall heterogeneity*. Systematic, quantitative assessment of heterogeneity may contribute significantly to the identification of both methodological and "natural" sources of variability of epidemiological effect estimates, including the identification of susceptible subgroups and exposure conditions.

- *Meta-analytic methods that may be used to compare studies*. E.g. stratified analysis or meta-regression.

- *Sensitivity analyses*. Such analyses might, for example, examine the sensitivity of summary estimates to reasonable alternatives with regard to the inclusion and exclusion of particular studies. One can also evaluate the sensitivity to alternative approaches to the extraction of results from published reports.

- *Methods to obtain summary estimates from different studies (aggregative meta-analysis).* Though quantitative summary estimates are not essential for health hazard characterization, they will be a particularly useful input to the Health Impact Assessment (see Section 4.3).

5. **Drawing conclusions from epidemiological evidence**. After the epidemiological evidence has been evaluated and appropriately summarized, as discussed above, expert judgement as to whether the observed associations are most consistent with a causal explanation or some alternative is required. This judgement should draw upon all the available epidemiological evidence, as well as on evidence from toxicology, clinical medicine, and other disciplines, as appropriate. The method of choice is critical scientific thinking: there are no formulas or checklists that will suffice, although, as noted above, Hill's attributes can provide useful guidance and focus. It is critical, however, that expert review groups make explicit the process of scientific reasoning that led to a judgement concerning causality. This explanation should include explanations of:

    - how expert reviewers weighted particular features of the epidemiological studies (e.g. assessments of bias, confounding, exposure-response) in reaching their judgement;
    - how expert reviewers used guidelines such as Hill's attributes;
    - how non-epidemiological sources of evidence figured in their interpretation of the epidemiological evidence, and how that evidence contributed to their overall judgement.

Expert judgments concerning the causal nature of observed associations are often accompanied by qualifications as to the degree of uncertainty. When the product of a Health Hazard Characterization is presented as a conclusion regarding the existence (or non-existence) of a hazard, the degree of uncertainty is sometimes expressed on a qualitative ("weak, moderate, strong evidence for hazard") or on a quantitative scale. If a quantitative scale is devised it should be capable of being reproduced by other experts. In either case, the use of a particular scale, and the meaning of its levels, should be clearly explained. More generally, it may be useful in the future to standardize such scales in order to avoid problems of non-comparability among the reviews produced by different expert review groups.

4.3 Recommendations for the use of epidemiological data for Health Impact Assessment

The Working Group made recommendations with regard to the use of epidemiological data for the design and implementation, as well as the interpretation, of Health Impact Assessments:

1. **The design and implementation of Health Impact Assessments**. Health Impact Assessments, which aim to quantify the expected health burden in a specific population(s), should be conducted according to **explicit protocols** that:

    - ***Specify the purpose of the assessment***. The purpose(s) of the health impact assessment should be made clear because decisions concerning the choice of epidemiological and other data and quantitative methods will depend on the objectives of the assessment. Ideally policy-makers, scientists and also stakeholders should be involved in defining the scope of the assessment, since different parties may have different questions about, and perspectives of, the same environmental health issue.
    - ***Specify the method(s) used to quantify uncertainty***. It should be made explicit in each Health Impact Assessment what the uncertainties are likely to be and how the assessors will deal with them. The choice of data and methods by which to quantify uncertainty may be determined by the specific objectives of the impact assessment

(e.g. identification of a maximum or minimum potential impact). The quantification of the uncertainty contributed by epidemiological effect estimates should consider both their statistical variability (i.e. precision), and non-statistical variability resulting from sources of error (e.g. bias and confounding) in the epidemiological data.

- ***Specify the metric of exposure to the specified hazards and methods to identify its distribution in the population for which assessment is requested.*** A clear and explicit definition of the metric of exposure, i.e. the operationalization of the "cause" considered in the Health Impact Assessment, should be provided. Health Impact Assessment will require information on the distribution of exposure in the target population, which will ultimately need to be combined with information on the exposure-response function in order to conduct the assessment. Depending on the available evidence (e.g. from epidemiological studies), the metric may need to incorporate temporal (e.g. induction period or latency) and compositional (e.g. mixtures and surrogates for them) aspects of exposure. The impact assessment should describe, and, whenever possible, quantify, the uncertainty contributed by the exposure assessment.

  The magnitude of the estimated impact will depend strongly on the level and range of exposure used in the Health Impact Assessment. The choice of a reference level for the impact assessment may be particularly complicated, and may require the consideration of epidemiological and other data with regard to issues such as the existence of thresholds and natural background levels. If exposures in the target populations exceed, or are below, those that have been studied epidemiologically it will be necessary to determine whether effects should be extrapolated. Ultimately, these choices will depend on both expert judgement and the perspective and purpose(s) of the assessment, but the basis for those choices should be clearly explained.

- ***Define the appropriate health outcomes.*** A particular Health Impact Assessment might focus on one or several health effects. If there is evidence of an environmental hazard being associated with several health effects, then ideally the impact should be assessed separately for each one. In practice, several aspects of the Health Impact Assessment, mainly its purpose and objectives, the definition of exposure, and the availability of the necessary data will guide the selection of health outcome(s). Based on these considerations, assessors may decide not to include all conceivable outcomes. These decisions and their rationales should be made explicit.

- ***Specify methods for estimating the exposure-response relationship.*** The quantitative association between the exposure and the health effect(s) is an essential component for the calculation of the attributable number of cases, and information about the exposure-response function is potentially the key contribution of epidemiological studies to a Health Impact Assessment.

  Due to both uncertainties in epidemiological studies, and true variability in the association between exposure and health outcomes within and among human populations, the available body of epidemiological evidence may provide different exposure-response functions for the same general exposure-outcome relationship. Thus, for a given Health Impact Assessment, the process used to derive the exposure-response function(s) must be well defined. It should, at a minimum, include a

systematic review of the available epidemiological information to obtain information on exposure-response relationships for every selected health outcome. All studies with quantitative information on exposure or which allow linkage to such information should be considered as potentially providing information for the exposure-response evaluation. The hazard characterization process normally will provide an inventory of the relevant studies.

Epidemiological studies identified as potentially providing useful exposure-response information may need to undergo an additional selection process which considers: (1) the quality of exposure measurement; (2) whether the exposure metric is the same as that available for the target population of the impact assessment; and (3) whether or not the estimated measures of effect are generalizable to the target population due to the influence of effect modifiers, such as local socioeconomic factors, or the prevalence of susceptible subgroups.

Projecting exposure-response relationships beyond the range of exposure observed in the available epidemiological studies, may be necessary for a given health impact assessment. However, the validity of such extrapolations should not be simply assumed, but rather, arguments for, and the limitations and potential impacts of, extrapolations should be carefully addressed in the Health Impact Assessment, including allowance for additional uncertainty.

An expert group may decide that combining exposure-response information from different epidemiological studies, for example via meta-analysis of published results or pooling of original data, is the best approach for deriving an exposure-response relationship for a given impact assessment. These approaches can potentially provide not only an overall summary of an exposure-response function, but, also (and perhaps as or more importantly) a range of estimates corresponding to possible sources of heterogeneity in the target population. Care should be taken to present estimates of the statistical and other sources of uncertainty in any combined estimates of the dose-response function.

- ***Specify approach for obtaining measures of baseline frequency of health outcomes in the target population***. Estimating the impact of exposure requires information on the baseline occurrence (rate, prevalence) of outcome(s) in the target population. Combined with the estimates of relative effect most often provided by epidemiological studies, it yields an estimate of impact of exposure in absolute terms, e.g. number of cases of disease or deaths (see below). While exposure-response relationships may be derived from the international literature, the baseline disease occurrence should preferably be obtained from data regarding the target population of the assessment. If such data are unavailable, or inadequate, health frequency data from other populations may sometimes be used. In such cases, the potential limitations of such substitutions should be considered and thoroughly discussed in the Health Impact Assessment.

- ***Specify methods for estimating the number of attributable cases***. The estimation of the burden of disease or mortality expected in the target population requires three basic elements whose estimation is discussed above:

  1. the distribution of the exposure in the target population;
  2. estimates of the epidemiology-based exposure-effect function;
  3. epidemiology-based estimates baseline frequency of the health measure of interest.

Using this information, and under the assumption that exposure causes the health outcome, an epidemiology-based Health Impact Assessment estimates the population attributable proportion (of disease or death) due to exposure, a measure described in standard epidemiological texts (see e.g. Rothman & Greenland 1998). When applied to the target population, the population attributable proportion yields an estimate of the expected number of cases attributed to the exposure.

In practice, both the estimation and interpretation of the population attributable proportion, and its application to the target population, may involve a number of subtleties. These involve, for example, the choice of relative risk estimator when there is evidence of confounding (see e.g. Rockhill et al. 1998, Greenland & Robins 1988). The assumptions underlying the statistical methods used to estimate attributable proportions, or other measures of impact, and their implications for interpretation of those estimates, should be discussed.

The uncertainties in the data that contribute to the impact assessment, as well as any natural sources of heterogeneity in the effect of exposure, will often require the calculation of a range of estimates in order to describe fully the likely impact of exposure and to better reflect the uncertainty.

2. **Issues in the interpretation of Health Impact Assessments**. The results of the impact assessment require not only clear presentation, but also coherent interpretation, including explicit discussion of assumptions and limitations. Specific components of the overall uncertainty and their potential impact on the results ought to be addressed, as discussed above. Sensitivity analyses, in which the effects of key assumptions are explored quantitatively, may provide a better sense of the overall uncertainty of the estimates than purely qualitative discussion, and should be performed when appropriate.

   In general, the direct effect of the removal of a particular exposure may only rarely be estimated. Depending on the health outcome, the specificity of the exposure, and the time frame of exposure and effect, the benefit (or reversibility) may be realized either much later than predicted, or not to the full extent. In particular, removal of the environmental hazard may not prevent the occurrence of the estimated number of cases due to how competing risks may come into play if one contributing cause (the exposure) is removed or reduced.

Annex 1

WORKING GROUP MEMBERS

Ross Anderson [a)]	St. George's Hospital Medical School, London, United Kingdom
Ben Armstrong [a)]	London School of Hygiene and Tropical Medicine, London, United Kingdom
Larry Chambers [b)]	McMaster University, Hamilton, Canada
Aaron Cohen [c)]	Health Effects Institute, Cambridge, USA
Lynne Edwards	European Commission, Brussels, Belgium
David Gee	European Environment Agency, Copenhagen, Denmark
Paul Harrison	MRC Institute for Environment and Health, Leicester, United Kingdom
Karl-Heinz Jöckel	Institut für Medizinische Informatik, Biometrie und Epidemiologie, Essen, Germany
Pierre Jouannet	Service d'histologie-embryologie oriente en biologie de la reproduction, Hospital Cochin, Paris, France
Matti Kamppinen	University of Turku, Turku, Finland
Klea Katsouyanni	University of Athens Medical School, Athens, Greece
Manolis Kogevinas	Universitat Autónoma de Barcelona, Barcelona, Spain
Michal Krzyzanowski [e)]	WHO European Centre for Environment and Health, Bilthoven, Netherlands
Nino Künzli [b)]	Institute of Social and Preventive Medicine, University of Basel, Switzerland
Erik Lebret	National Institute of Public Health and the Environment (RIVM) Bilthoven, Netherlands
Rolaf van Leeuwen	WHO European Centre for Environment and Health, Bilthoven, Netherlands
Marco Martuzzi	WHO European Centre for Environment and Health, Rome, Italy
Robert Maynard [d)]	Department of Health, London, United Kingdom
Göran Pershagen [f)]	Karolinska Institute, Stockholm, Sweden
Charles Poole	University of North Carolina School of Public Health, Chapel Hill, USA
Gerhard Raabe	Mobil Business Resources Corporation, New Hope, USA
Simon Rozendaal [f)]	Elsevier, Amsterdam, Netherlands
Bernd Seifert	Institute for Water, Soil and Air Hygiene of the Federal Environmental Agency, Berlin, Germany
Jack Spengler	Harvard School of Public Health, Boston, USA
John Vandenberg	U.S. Environmental Protection Agency, Research Triangle Park, USA
Paolo Vineis	Turin University, Turin, Italy
Douglas Weed	National Cancer Institute, Bethesda, USA
Bogdan Wojtyniak	National Institute of Hygiene, Warsaw, Poland
Maged Younes	International Programme on Chemical Safety, World Health Organization, Geneva, Switzerland
Denis Zmirou	Faculté de Médecine-Pharmacie, La Tronche, France

a) Subgroup Chairman
b) Subgroup Rapporteur
c) Rapporteur of the Working Group
d) Chairman of the Working Group
e) Scientific Secretary of the Project
f) Not present at Working Group Meeting

Annex 2

REVIEWERS

Ursula Ackermann-Liebrich, Switzerland
Jamie Bartram, WHO/headquarters
Piero Bellini, Italy
Ruta Bruvere, Lithuania
Devra Davis, USA
Norbert Englert, Germany
Anna Hansell, Paul Aylin, Lars Jarup, Mike Joffe and Jennifer Mindell, United Kingdom
Irva Hertz-Picciotto, USA
A.B. Miller, Germany
Emilia M. Niciu, Romania
Philip C. Nasca, USA
Lucas Neas, USA
Jonathan M. Samet, USA
Edward Sanders, Switzerland
Hana Šlachtová, Czech Republic
Allan H. Smith, USA
Peter Straehl, Switzerland
Richard Wilson, USA

Annex 3

REPORTS FROM WORKING SUBGROUP DISCUSSIONS

Sections A3.1 and A3.2 below summarize the discussions of the sub-Working Groups on Health Hazard Characterization and Health Impact Assessment, respectively. They were prepared over the course of the Working Group meeting and subsequently edited for clarity of presentation. They are included to provide the reader with a sense of the discussions that produced the major recommendations presented in Section 4 of this report. They were not intended as comprehensive reviews of methods for epidemiological research or quantitative risk assessment. These topics are covered more comprehensively in references cited in the text and bibliographies of each paper.

A3.1 Evaluation of epidemiological evidence for Health Hazard Characterization

Introduction

The World Health Organization requires a reliable, transparent and broadly acceptable approach to the identification of potential environmental hazards and to the assembling, evaluation and interpretation of available evidence concerning the causality of associations between a potential hazard and health. We call this process Health Hazard Characterization.

The process of Health Hazard Characterization is enhanced by epidemiological evidence. The purpose of this section is to justify and describe an approach to the evaluation of epidemiological evidence. We recommend that expert groups convened by WHO and regulatory bodies to inform policies concerning environment and health follow this approach.

Systematic review of epidemiological evidence for Health Hazard Characterization

In the view of the Working Group, Health Hazard Characterization should include a systematic review of all of the relevant evidence, and the process and methods of that review should be clearly documented. Adoption of this approach would help to ensure that the conclusions of the review are transparent, unbiased, replicable, and valid. Such a systematic approach would also provide a foundation for the continued monitoring of additional evidence as it emerges.

Objectives of a systematic review

The following objectives should be achieved by the review:

- *Transparency*. This means that what is done is clear to the expert group, other scientists, policy-makers and the public.

- *Avoidance of bias.* A systematic approach will help to ensure that each step of the Health Hazard Characterization has been carefully considered to prevent the introduction of bias into the process of review.

- *Validity*. Validity refers to the degree to which the conclusions of the review are likely to be the correct ones. Validity will be increased if the Health Hazard Characterization is systematically conducted so that inferences can be drawn from its findings, especially generalizations extending beyond the studies used in the Health Hazard Characterization.

- *Replicability*. A systematic and clearly described method of Health Hazard Characterization will allow replication and a basis for comparison with reviews of future evidence as it becomes available.

- *Cover all relevant issues*. A systematically conducted Health Hazard Characterization will help ensure that all relevant issues are considered.

- *Improve efficiency in updating evidence.* A systematically conducted Health Hazard Characterization will provide an efficient way of updating the evidence base as new studies emerge.

- *Research planning.* Identification of gaps and uncertainties in data and methods to be addressed in with additional research is a valuable by-product of the review process.

Increased expectation that evidence in the health sciences is systematically reviewed

Most health sciences journals now insist on meta-analyses, overviews, and reviews being systematically conducted and methods of conducting such systematic reviews have been published. (Badgett et al. 1997, Bero & Jadad 1997, Blair et al. 1995, Chalmers & Lau 1993, Dickersin & Berlin 1992, Greenland 1987, Moolgavkar 1995, Ohlsson 1994, Thacker 1988, Wong & Raabe 1996). There are several ways of performing such reviews, depending on the aims. Some examples of approaches to systematically reviewing evidence in the health sciences are outlined here.

The monographs prepared by the International Agency for Research on Cancer (IARC) which evaluate the carcinogenicity of specific substances comprise narrative summaries of individual studies (and any existing meta-analyses), prepared according to specified guidelines by ad hoc working groups. IARC provides specific guidelines to ad-hoc working groups who prepare the monographs (see, for example the introductory text of IARC Monographs, (IARC 1999: http://www.iarc.fr/, accessed 1 March 2000)). Before convening, the IARC staff provides working groups with the primary studies and background papers are prepared. The working group then assesses the evidence at the time of the meeting. No explicit criteria are used for selecting the participants in the working groups. Although extensive searches of the literature are conducted, and only the "informative" studies are summarized or included in the evaluation, there are no explicit criteria for selection of the primary studies. The summaries prepared by the working groups offer brief comments on the quality of individual studies, but formal meta-analyses are rarely conducted. The epidemiological evidence on the association between a risk factor and cancer is then summarized according to discrete categories (sufficient, limited, inadequate, no evidence). The distinctions among categories are chiefly based on (a) reproducibility of the evidence; (b) validity (absence of bias and confounding); (c) role of chance.

The systematic reviews of randomized clinical trials (RCT) prepared by the Cochrane Collaboration review groups begin with the selection of the primary studies according to strict criteria, e.g. only randomized controlled trials that provide a description of the randomization method are selected (Cochrane 1999. http://www.update-software.com/ccweb/cochrane/hbook.htm, accessed 1 March 2000). One or two individuals who have undergone training in, and agreed to comply with, the rules of the Cochrane Collaboration usually perform reviews of randomized clinical trials on a particular therapeutic intervention. They assess the quality of the primary studies according to predefined criteria, and a score is assigned and used to exclude some studies. Summary effect estimates are then calculated by combining the results of those RCTs meeting the criteria. The outcome of the Cochrane meta-analyses is a meta-relative risk estimate with a confidence interval (i.e. a quantitative assessment) accompanied by a commentary by the authors of the review. This commentary deals with relevant aspects of the evidence that are not provided by the summary estimates of effect.

Protocol for Health Hazard Characterization

The first step in ensuring a systematic approach in conducting a Health Hazard Characterization is to adopt the general protocol for Health Hazard Characterization proposed by this Working Group. The protocol will help ensure that an expert group has a common understanding of its task and will adhere to the systematic approach recommended by WHO.

A protocol approach requires the following decisions:

1. Agreement on the question(s) to be dealt with in the Health Hazard Characterization.
2. Justification of the expertise represented on the Health Hazard Characterization expert group.
3. Specification of the methods to be used in the Health Hazard Characterization.
    - identification of relevant studies (including reviews)
    - assessment of quality of the studies
    - interpretation of the evidence.

Although these elements should be agreed upon before the Health Hazard Characterization proceeds, it is accepted that mid-Health Hazard Characterization revisions of the protocol may be necessary as information not appreciated at the outset of the Health Hazard Characterization emerges.

The success of the Health Hazard Characterization process will depend on various factors including the adequate and timely preparation of material, the composition of the Expert Group, and the availability of input from scientists (or the general public) outside the Expert Group. The whole process has to be efficient and benefit the conclusions of the Health Hazard Characterization.

The composition of the Expert Group is critical. The criteria for selection should be based on having the appropriate mix of scientific expertise and experience. Within these criteria, WHO also needs to consider the need for geographical representation.

Components of reviews of the epidemiological evidence in a Health Hazard Characterization

The view of this WHO Working Group is that there are three fundamental steps in the systematic review of the epidemiological evidence in a Health Hazard Characterization.

1. Comprehensive identification of all relevant studies
2. Systematic assessment of the quality of the available studies
3. Interpretation and conclusions from the body of epidemiological and other evidence.

The assessment of studies quality can be divided to the review of each of the individual studies and to the joint analysis of all identified the studies. The recommended approach at each of these steps is described in the following sections.

Comprehensive identification of all relevant studies

A comprehensive bibliographic search would include the following:

- involvement of qualified searchers (e.g. librarians, trained investigators);
- definition of an explicit search strategy including identification of key words;
- an effort to include all available studies;
- searching of bibliographic databases;
- inclusion of non-English reports.

Optional methods, to be specified by the review protocol, would include hand searching of journals, and inclusion of abstracts and unpublished data (including writing to authors of published data).

After identifying an initial list of members of the Expert Group, the preparation for the first Expert Group meeting, should include the results of the initial identification of all relevant studies together with a summary of the quality of the studies. The group will need to be consulted about the search strategy and criteria for quality. All potentially relevant studies should be identified at the first stage of the Expert Group's work. However, depending on the types of exposure evaluated, whole categories of studies may be excluded in the second step, on the basis quality criteria (see below). In the evaluation of a specific pesticide, for example, studies of routine statistics on cancer risk in agricultural workers might be excluded because assessment of exposure to a specific pesticide in such studies is problematic.

Systematic assessment of the quality of primary studies

WHO expert review groups should consider each of the following questions when assessing a particular study for use in a Health Hazard Characterization:

- Is the study question clear?
- Was the exposure assessed using valid and reliable measures?
- Was the health outcome(s) assessed using valid and reliable measures?
- Was the study design appropriate?
- Did the analysis of the data take into consideration chance, confounding, and bias (information, selection, and analytic)?
- Were the conclusions consistent with the results of the data analysis?

As Hill (1965) emphasized, the *fundamental question* in assessing epidemiological research for hazard identification (i.e. assessment of causality) is this: "Is there any other way of explaining the set of facts before us, is there any other answer equally, or more, likely than cause and effect?" This concept is appropriate in every assessment of the quality of epidemiological studies whether individual or collective. Hill offered several attributes (which he pointedly did not call "criteria") to bear in mind while contrasting causality with alternative explanations. Hill's attributes are helpful in the assessment of confounding and bias. One such attribute is the strength of association. If the estimated relative risk is 2.0, for example, a single confounder cannot fully explain the departure of this estimate from the null value of 1.0 unless that confounder at least doubles the risk of the disease and the confounder is at least twice as common among exposed persons as among unexposed persons.

Another of Hill's attributes is specificity of cause. Bias from exposure misclassification may be suspected as an explanation for a positive association between a specific exposure of interest and a health effect. However, the same bias may also be suspected with regard to other exposures examined in the same study. If so, then if disease is only associated with the exposure of interest, misclassification would be an implausible explanation.

Yet another of Hill's attributes is temporality: hypothetical causes must precede their hypothetical effects. Thus, an important feature of exposure and disease classification is to establish their proper sequence in time, with exposure occurring before disease. Failure to establish this temporal sequence makes the study difficult (if not impossible) to interpret with respect to a causal association.

It is important to note that Hill confined his attention to the context of "positive" studies, i.e. studies that appear to suggest the presence of a cause-and-effect relationship. In a comprehensive, weight-of-evidence approach, so-called "negative" studies, i.e. studies that appear not to show an association between exposure and disease, must be considered as well. For these studies, one must consider alternatives to the absence of cause-and-effect.

For either purpose, the alternative explanations may be divided into five general categories: chance, confounding, information bias, selection bias (including publication bias) and analytic bias. In a review aiming at the health hazard characterization, these issues should be considered first in the evaluation of individual studies and then in the evaluation of an epidemiological literature as whole. Short discussion on these issues is presented in the following paragraphs.

- *Chance.* In the context of estimating a measure of effect (for instance, some form of relative risk), the width of a confidence interval should be used to gauge the precision of the estimate. For instance, two studies might each produce relative risk estimates of 1.0, suggesting that exposure has no effect. However, the confidence interval around that estimate might in one case be 0.5 to 2, and in the other be 0.25 to 4.0. The confidence interval conveys clearly the relative imprecision of the latter study, allowing one to infer, for example (by comparison of the upper bounds), that the more precise result is consistent with only a doubling of disease incidence or risk.

  When exposures are either measured on an ordinal scale, or are measured on a continuous scale and then divided into multiple ordered categories for data analysis, the most important information concerns the change in the effect measure, say the relative risk, over the ordered categories. Therefore, it is the statistical precision of the estimate of the slope or trend in the relative risk that should be assessed, e.g. via the width of a confidence interval, rather than the precision of the category-specific relative risks (Rothman & Greenland 1998).

  Epidemiologists have increasingly eschewed statistical significance testing for the interpretation of study results in favour of interval estimation. However, in a significance testing framework, if the result is strongly statistically significant, chance does not usually need to be considered. If the result is not statistically significant, the power of the statistical test must be taken into account.

- *Confounding.* Important confounders are the known, strong risk factors (causative or preventive) for the disease that might be strongly associated with the exposure and that are not consequences of the health effect or the exposure. For example, cigarette smoking would be an important potential confounder for a study of any environmental exposure and lung cancer. Confounding may be upward (i.e. toward spuriously high estimates of effect) or downward (i.e. toward spuriously low estimates), depending on the directions of the associations between the confounder and the exposure and between the confounder and the health effect.

- *Information bias.* Bias can be produced by poor quality of information on the exposure, the health effect, or potential confounders. The direction of the bias depends on whether the quality of the information differs between groups being compared. If such errors vary approximately at random, and do not differ in frequency between compared groups (i.e. are non-differential), the bias is usually in the direction of underestimating the magnitude of any true association, though there are exceptions to this general rule. Differential information bias may spuriously elevate or reduce observed associations depending on the relative degree of bias between the groups being compared (Rothman & Greenland 1998).

- *Selection bias.* Bias can be introduced by the method of selection of people for studies, by incomplete participation, and by missing information for some study subjects. As with information bias, the direction and magnitude of selection bias depends on particulars of the absolute and relative frequencies with which different kinds of people are included or excluded.

- *Analytic bias.* Biases can be produced by the manner in which epidemiological data are analysed. Categorization of continuously measured exposures, or the assumption of a linear dose-response relationship when the true relationship is non-linear are important examples.

Epidemiological studies of environmental agents will be particularly useful for health hazard characterization if they provide estimates of exposure response-relationships (i.e. which levels of

exposure might be expected to affect human health, and the degree of harm expected at various exposure levels). The demonstration of specific patterns of association over ordered categories of exposure, such as monotonic increases in the relative risk, can provide strong support for causal interpretations if they cohere with pathophysiologic models, and because more complex, and hence implausible, patterns of confounding or bias are required as counter-explanations.

The information that existing epidemiological studies provide about exposure-response relationships in particular study populations are important components in Health Impact Assessments of other populations.

The credibility of a study is enhanced if its results are confirmed in a sensitivity analysis. In such analysis the outcome variable(s) are examined with respect to (1) changes in expression of exposure variables, (2) addition of other plausible explanatory variables, and/or (3) introduction or removal of controlling variables. Inclusion of some form of sensitivity analysis is becoming the norm in published epidemiological studies. As part of the overall evaluation of epidemiological evidence, sensitivity analysis of primary studies as well as any differences among studies should be explicitly identified.

Sometimes information from one study may be used to adjust the results of another study. For example, one study may have measured exposure in two ways, of which one is more accurate than another. Comparisons between the two methods may be used to adjust the results of a study in which only the inferior method was used (Greenland 1987).

Conduct of systematic overviews of evidence from multiple studies: the use of meta-analysis

WHO expert review groups should consider each of the following questions when designing and conducting systematic overviews of epidemiological studies:

What is the question(s) that the review proposed to answer?

- Is there a comprehensive strategy for searching the literature?
- How will the quality of the individual studies, and their findings, be assessed?
- How will the reliability of the reviewers' assessments of the quality of each study be evaluated?
- How will the results of individual studies be summarized?
- How will heterogeneity among studies be assessed?
- Will summary effect estimates be calculated, and by which methods?

Meta-analytic techniques may be profitably used to summarize the available epidemiological studies. Although meta-analysis is often viewed simply as the statistical combination of results across studies, i.e. as focusing on the last bullet point alone, meta-analysis has also been described as an approach to the quantitative review of the literature (Greenland 1987), a "study of studies", which provides a quantitative assessment of the extent to which bias might account for observed results, and of the patterns, and sources of heterogeneity. It is this latter approach, the critical quantitative review, which potentially affords the most insight for hazard characterization. The use of meta-analysis (and data pooling, discussed below) in risk assessment has been discussed recently by Samet & Burke (1998).

Meta-analyses are usually conducted on the published results of studies, which are often highly summarized. Alternatively, and less often, the data on individual subjects in several studies are included in a pooled analysis. If results from two or more studies are to be aggregated, a decision needs to be made whether to base the aggregation on the published results or whether to obtain the individual subject data from the original investigators. Published results have the disadvantage that they are often already highly summarized, sometimes in ways that make them difficult to combine (e.g. the use of different category boundaries for categorising continuously measured exposures). Individual subject data have the disadvantage of being difficult and time-consuming to obtain; often they are available for only a small and perhaps non-representative subset of all the studies that have been done. When there are important

analytic questions about key studies, however, re-analysis of the individual subject data can be highly informative and thus worth the extra time and expense. This would be a matter for the discretion of the WHO expert group.

The Working Group recommends that expert review groups consider the following issues when designing and conducting quantitative reviews (meta-analyses) of epidemiological literatures:

- *Protocol.* Each meta-analysis must have its own protocol, perhaps "nested" within the overall protocol for the health hazard characterization. The protocol should include a clear statement of the objectives of the review, and the methods to be employed.

- *Inclusion criteria.* It is advisable for a meta-analysis to be inclusive rather than exclusive. The criteria that might be used to exclude studies from a meta-analysis that aggregated effect estimates across studies (an aggregative meta-analysis) can then be used to test hypotheses relating to these criteria in a meta-analysis that focuses on specific study characteristics (a comparative meta-analysis).

- *Overall quality scores* Reducing the features of a set of epidemiological studies to a single measure of "quality" is not recommended because these features may affect the results of the studies in different directions and to varying degrees. It is preferable to assess the characteristics of the primary studies individually.

- *Chance.* In meta-analysis, the results are usually weighted by the statistical precision (in general, by the amount of information) in each primary study. Technically, in its simplest form, the weights are inversely proportional to a statistical measure known as the "variance" of the study's estimate of effect. Studies with more information (e.g. studies that are "larger") produce estimates with narrower confidence intervals and lower variances. Thus, inverse-variance weighting assigns more weight to the studies based on more information.

- *Publication bias.* The results of certain kinds of primary studies are more likely to be published than of the others. Publication bias can be:
  - minimized – by doing a comprehensive literature search (e.g. including, if possible, unpublished results);
  - detected – by funnel graphs, tests (e.g. Dickersin & Berlin, 1992);
  - corrected – by statistical models in which missing data are imputed;
  - assessed by sensitivity analysis – by determining how many studies of what characteristics and with what results would have to be missing to give the literature a substantially different appearance from the one it currently has.

- *Assessment of overall heterogeneity.* The three characteristics of primary studies, which are examined in heterogeneity analyses are: definition of populations, exposure characteristics and exposure contrasts, and, research methods.
  - *Definition of populations*: Similar populations are needed if results obtained in one population are to be used to predict effects in another. However, when studies of dissimilar populations yield similar measures of effect then causal interpretation is strengthened. Heterogeneity of the association may, however, indicate an existence of population(s) with different sensitivity to the exposure and this possibility must be a subject of careful analysis.
  - *Exposure characteristics and exposure contrasts*: Did the studies measure the same exposures in the same way using the same metrics and did they compare risk between or among similar levels of exposure? The less similar the studies are in these regards, the less advisable it is to combine their results.

- *Research methods*: Confounding control, selection bias (e.g. characteristic for cohort vs. case-control studies), information bias (e.g. blinded vs. not-blinded interviews) and data analysis methods (e.g. cut-point choices for exposure categories) in each primary study must be assessed before deciding to combine the studies in a summary analysis.

In general, standard statistical tests of heterogeneity, which do not account for specific study characteristics (such as those listed above) that could produce heterogeneous results, are insensitive, and have low statistical power to detect heterogeneity. For this reason, stratified analysis and visual, including graphical, inspection of stratum-specific results are also valuable tools. Some recommend that heterogeneity tests be performed at less stringent alpha-levels (e.g. 0.1 or 0.2), as well.

- *Meta-analytic methods that may be used to compare studies*
  - Stratified analysis: this examines one characteristic at a time. It is less formal and more useful with a small number of studies.
  - Meta-regression: here, the dependent variable is the estimate of effect, and the independent variables are study characteristics (see above). This technique is more formal, can examine more than one characteristic at a time if there are a sufficient number of studies.

- *Sensitivity analysis of meta-analysis.* The results of meta-analytic summaries should be subjected to sensitivity analyses to test their robustness to alternative data specifications and analytic approaches, in the same way that one would apply sensitivity analysis in the evaluation of the results of a single study. In the meta-analytic context, such analyses might, for example, examine the sensitivity of the results to reasonable alternatives with regard to the inclusion and exclusion of particular studies. One can also evaluate the sensitivity to alternative approaches to the extraction of results from published reports; the latter task often requires the exercise of professional judgement on the part of the analyst.

- *Summary estimates from different studies (aggregative meta-analysis).* If health effect, exposure, and methods (for example, the choice of exposure metric) are similar and there is no appreciable evidence of publication bias, overall heterogeneity, or specific study characteristics associated with results, then results from more than one study may be combined to form summary estimates. Conversely, the existence of any one of these considerations may be a contraindication to aggregation.

Conclusions from the body of epidemiological and other evidence

After the epidemiological evidence has been evaluated and appropriately summarized, as discussed above, expert judgement as to whether the observed associations are most consistent with a causal explanation or some alternative is required. A process of scientific reasoning must be followed to arrive at this judgement: a process that draws on all the available epidemiological evidence, as well as evidence from toxicology, clinical medicine, and other disciplines, as appropriate. The method of choice is critical thinking: there are no formulas or checklists that will suffice, although, as noted above, Hill's attributes can provide useful guidance and focus.

It is critical, however, that expert review groups make explicit the process of scientific reasoning that led to a judgement concerning causality. This explanation should include explanations of:

- how expert reviewers weighted particular features of the epidemiological studies (e.g. assessments of bias, confounding, exposure-response) in reaching their judgement;
- how expert reviewers used guidelines such as Hill's attributes;
- how non-epidemiological sources of evidence figured in their interpretation of the epidemiological evidence, and how that evidence contributed to their overall judgement.

Expert judgements concerning the causal nature of the observed association will need to be accompanied by statements regarding the degree of uncertainty that the expert reviewers attach to it. When the product of a Health Hazard Characterization is presented as a conclusion about the existence (or not) of a hazard, the degree of uncertainty could be expressed on a qualitative ("weak, moderate, strong evidence for hazard") or on a quantitative scale. If a quantitative scale is devised it should be calculable, and capable of being reproduced by other experts. In either case, the use of a particular scale, and the meaning of each level of it, should be clearly explained. There may be a need to standardize such scales in order to avoid problems of non-comparability among the reviews produced by different expert review groups.

Role of the Precautionary Principle

The Precautionary Principle, and its application, was succinctly described by Horton (1998).

> We must act on facts, and on the most accurate interpretation of them, using the best scientific information. That does not mean that we must sit back until we have 100% evidence about everything. Where the state of the health of the people is at stake, the risks can be so high and the costs of corrective action so great, that prevention is better than cure. We must analyse the possible benefits and costs of action and inaction. Where there are significant risks of damage to the public health, we should be prepared to take action to diminish those risks, even when the scientific knowledge is not conclusive, if the balance of likely costs and benefits justifies it.

The Precautionary Principle, as Horton describes it, provides a guide to action under (perhaps considerable) uncertainty: the condition under which most, if not all, public health decision-making on environmental issues takes place. It assumes that the available scientific evidence, the "facts", has been objectively assessed, and that the uncertainty of that assessment has been made explicit.

The Working Group discussed the role that the Precautionary Principle should play in Health Hazard Characterization. Although most Working Group members agreed that the Precautionary Principle should play a role in public health decision-making, there was no agreement on what role, if any, the Precautionary Principle should play in the evaluation of epidemiological evidence, per se. One proposal called for defining

> ... a minimum level of acceptable epidemiological evidence, which would act as a point of reference -rather than a criterion- for subsequent public health action or inaction. The extent to which external or extra-scientific factors affect decision-making (in either direction) [would] depend upon the particulars of the situation. (D. Weed, Epidemiological Evidence and the Precautionary Principle. Unpublished working paper prepared for the WG Meeting).

Other Working Group members felt that because regulators need to depend on an objective evaluation of the existing evidence, it is important for expert review groups to maintain a clear demarcation between their reviews and public health decision-making. Still others stressed the importance of making clear the remaining uncertainties, and their implications. They pointed out that a corollary of this view is that Health Hazard Characterizations should describe the additional research needed to resolve the uncertainties.

A3.2 Evaluation of epidemiological evidence for Health Impact Assessment

Introduction

Health Impact Assessment in the meaning specific for this project, aims at quantitative expression of the expected number of people with a health effect that could be attributed to a specific exposure situation. There are a number of common reasons for Health Impact Assessment, but the reason may be quite different in nature, depending on the goals and scale of the Health Impact Assessment. *Inter alia*, Health Impact Assessment may play a role in one of the following:

- Environmental Impact Assessment required by law to evaluate future impacts during or after execution of large infrastructure projects;
- Evaluation of different policy scenarios, e.g. smog alert policies, or assessment of new environmental quality directives;
- Providing an estimate of expected morbidity for local evaluation of suspected disease clusters;
- Analysis of a local pollution situation;
- Projects to calculate the external (monetary) costs of environmental pollution or the benefits of risk management proposals or actions.

This chapter outlines the process of quantifying the health impact of an environmental hazard. Particular consideration is given to the contribution and role of epidemiological studies and the evidence they produce for Health Impact Assessment. Ideally, the output of the Health Impact Assessment process will be the number of cases or events attributable to some hazard in a target population. Whenever available, epidemiological studies will play a central role in estimating attributable cases. However even in the absence of epidemiological data, the epidemiological expertise is essential in the interdisciplinary process of Health Impact Assessment.

The major steps that play a role in such quantification processes are the following:

1. Specify the purpose and framework of the impact assessment;
2. Specify the method(s) used to quantify uncertainty;
3. Specify the measure(s) of exposure;
4. Specify the range of exposure to be considered;
5. Derive the population exposure distribution;
6. Specify the time window between exposure and effect;
7. Select appropriate health outcome(s);
8. Estimate the exposure-response relationship in the population of interest;
9. Derive population baseline frequency measures for the relevant health outcomes;
10. Calculate the number of attributable cases.

The first two steps emphasize general conceptual considerations. Steps three to five relate to exposure, whereas step six also refers to health outcome. Steps seven, eight and nine address health aspects. The last step describes the final quantification process. These steps will be outlined in more detail below. Particular focus will be given to the role and contribution of epidemiological studies to the process of health impact assessment.

The above steps describe an iterative process rather than a “cook book” with a fixed sequence of steps. Some steps may be less important in one Health Impact Assessment but be a crucial point of discussion in other Health Impact Assessments.

“Causality” – A requirement for Health Impact Assessment?

In conducting studies to quantify the health impact of environmental hazards, it is assumed that the scientific evidence for “causality” is strong enough or the possible consequences of a neglected hazard are sufficiently large to justify the assessment of the *potential* health impact. The process of Health Hazard Characterization is described in section A3.1. It must be kept in mind that in the epidemiology based Health Impact Assessment, not only well specified biologically plausible, chemically or physically characterized agents, e.g. benzo-a-pyrene, but also *indicators of exposure*, e.g. cigarettes may be considered a “cause” and, thus, the subject of Health Impact Assessment. In fact, epidemiology based Health Impact Assessment may make use of a special strength of environmental epidemiology. The epidemiological approach allows the assessment of the effects of physically or chemically unknown, or ill defined and complex, causes (e.g. “traffic”, a water system, etc), using indicators (or proxy measures) of

exposures. These indicators may be useful measures of contributing or sufficient causes (Rothman & Greenland 1998) which cannot be directly measured. In the framework of Health Impact Assessment, the definition of "causes" must be made explicit.

The preference for using epidemiological data in the Health Impact Assessment relates to the advantage of epidemiological measures to approximate the required information under "true life condition" of human beings. However, the assessment of causality should not only rely on epidemiology but should include all scientific information available. In fact, there may be environmental hazards, which have been identified primarily with other scientific approaches, as discussed in the earlier sections.

It is assumed in this chapter that the valid epidemiological studies have already been selected, as they are essential in establishing causality. Health Impact Assessment itself will make use of epidemiological studies. However, it might well be that not all available and valid studies may be useful in the quantification of the health impact. As an example, studies, which may be important for the health hazard characterization, may have used a health outcome measure not applicable to Health Impact Assessment. Consequently, Health Impact Assessment may be based on a subset of studies. On the other hand, some epidemiological studies may contribute to Health Impact Assessment but not to the process of hazard characterization. For example, due to weaknesses in the design or analysis, an epidemiological study may be considered of little value for etiologic inference or may not even address the association of some environmental hazard with health. However, this study may provide a valid estimate of the baseline frequency of a health condition in the population, which may in itself be an important information for Health Impact Assessment (see the following steps). In conclusion, the set of epidemiological studies contributing to Health Impact Assessment may be different from the studies considered in the process of Health Hazard Characterization.

Ten recommendations for the Health Impact Assessment

The following section summarizes actions related to each of the ten steps listed in the introduction.

Specify the purpose and framework of the impact assessment

Health Impact Assessment is an interdisciplinary task. The purpose of the health impact assessment should be made clear as several decisions related to the steps explained below will be influenced by the general framework. Furthermore, the questions from interested parties, e.g. policy-makers, may have to be rearranged to be able to conduct a Health Impact Assessment. As an example, it is particularly important to decide whether an economic valuation will be attached to the Health Impact Assessment. In such scenario, it is important to estimate non-overlapping, distinct quantities for which economists have some monetary value at hand. In another situation, Health Impact Assessment may be required for one specific aspect of an environmental problem, for example of a potential crash close to the airport, although other aspects of the airport may be of greater health relevance, e.g. noise in the neighbourhoods of the airport. Ideally policy-makers, scientists and also stakeholders should be involved in defining the scope of the assessment, since different parties may have different questions about and perspectives of the same environment and health problem (Staatsen et al. 1994). Ethical considerations may also influence the scope of Health Impact Assessment. If relevant questions come up during the iterative Health Impact Assessment process that have not been addressed initially, the project may need adaptation.

Specify the method to deal with uncertainty

Health Impact Assessment has inherent uncertainties and requires a set of assumptions. It should be made explicit in each Health Impact Assessment what the uncertainties are and how the assessors deal with them.

First, it may be that Health Impact Assessment is required to provide the minimum or the maximum number of cases attributable to some hazard. This requirement would influence each step of quantification where uncertainties come into play and assumptions that have to be made. Choices may be influenced by

the consequences of providing a wrong estimate. For example, it may be preferable to adopt a quantification process resulting in the minimum ("at least") number of cases that may be expected to be attributable to a complex exposure. On the other hand, policy-makers may require estimating the "worst case scenario" (according to the precautionary approach) for the health impact of a new substance not yet on the market or for the contamination causing public outcry. For example, in the early phase of an assessment of potential effects of elevated concentrations of dioxins in dairy products distributed to a large population, it may be informative for the public discussion and decision-making process to assume "worst case scenario" in a Health Impact Assessment. It might also be useful to consider a broader range of potential health effects (e.g. reproductive or immune system effects besides the elevated cancer risk), even if there are inadequate data to quantify these.

Second, to make the extent of uncertainty explicit, the impact may be quantified as a range, or distribution of possible values, rather than as a single point estimate only (Morgan & Henrion 1990, Rai et al. 1996). Some epidemiology-based uncertainties may be quantified whereas others may only be expressed qualitatively. The underlying assumptions as well as the implications of providing ranges or distributions versus single point estimates should be discussed early, including how to communicate results expressed in this form.

There are a variety of methods available to estimate the interval for parameter describing the magnitude of the impact. For example, uncertainties at each methodological step may be weighted and combined in sophisticated calculations (US EPA 1996). The level of sophistication adopted to derive a range of uncertainty for the impact must be weighted against the inherent limitations of Health Impact Assessment. Thus, a pragmatic approach may be adopted rather than the complicated calculation of "pseudo-accurate" ranges.

Epidemiological studies may be one source to provide a quantitative range of uncertainty, as results in epidemiological studies usually include estimates of the uncertainty in the precision of the effect estimates. Upper and lower boundaries of confidence intervals or similar statistical measures of variability may be used in the process of Health Impact Assessment to delimit a range of possible health impact. Impact assessors, however, should consider the different concept and meanings of the statistical measures of variability in epidemiological studies and any range of uncertainty given for the Health Impact Assessment. If the health effects are certain (i.e. established causality), the lower limit of the Health Impact Assessment may be greater than zero, i.e. at least *some* detrimental effect of the exposure can be assumed. For example, an environmental factor may well be accepted as a hazard or cause for health effects even if some single epidemiological studies may have shown statistically non-significant results. The use of a confidence interval from a single study to provide a range of impact may, thus, include "no impact" which inherently conflicts with the underlying assumption that the Health Impact Assessment is being conducted for an established causal agent.

On the other hand, confidence intervals reflect only statistical uncertainty, not uncertainty in the many assumptions required to determine the statistical model. Thus uncertainty ranges based on confidence intervals should generally be interpreted as minimum estimates of uncertainty – true uncertainty will be greater.

In some situations, Health Impact Assessment may be required for environmental factors which are suspected to cause some health effects but where the evidence is not clear. If the health impact is less than certain, the lower limit of the Health Impact Assessment may be zero, but not lower than zero (DOH 1999), unless there is plausible evidence for a preventive effect of the suggested hazard on the specified health effect.

Specify exposure

The definition of the measure of exposure, i.e. the operationalization of the "cause" considered in the Health Impact Assessment, has to be made explicit. This requires interdisciplinary iterative discussions as "exposure" plays an important role in two steps. First, the Health Impact Assessment requires knowledge

about the population exposure distribution, or, in the simplest case of dichotomous classification of exposure, the proportion of "exposed". Ideally, this exposure information might be available. In other cases, exposure may have to be estimated from available data, or obtained via an exposure assessment in the framework of the Health Impact Assessment. Second, if epidemiological evidence is used for Health Impact Assessment, epidemiological exposure response functions will be needed, that provide the quantitative association between exposure and health outcome.

The definition of exposure in the two steps should be coherent. Health Impact Assessment needs the population exposure distribution for the exposure indicator or biomarker used in the epidemiological or experimental studies from which the exposure-response function was derived. The compatibility between exposure data used in epidemiological studies and the data on the population exposure distribution may be poor. For example, if in a Health Impact Assessment of noise on sleep disturbance in the vicinity of an airport it is intended to use the exposure-response function describing the number of sleep interruptions as a function of the average noise level, dB(A), above some no-effect level, one would ideally need the noise level distribution in the population, measured in dB(A), for the night time, rather than the number of take-offs from the airport.

If "exposure" represents a mixture, this must be made clear and the selection of the most reasonable indicator(s) of the mixture has to be discussed. It should be carefully considered whether it is warranted to assess several indicators of the same or of strongly correlated sources of exposure, and if, and how, to combine the estimated impacts. For example, each of single pollutants which stand for the same mixture or exposure (e.g. traffic related air pollutants such as TSP, PM10, PM2.5, NOx, etc.) may result in the some health outcomes in the same segments of the population (e.g. increased incidence of respiratory disease in schoolchildren). The number of cases attributed to each of the indicator pollutants in Health Impact Assessment should not be summed since each separate estimate may represent the same affected group. (Seethaler 1999, Sommer et al. 1999).

Finally, attention should be paid to the time dimension of exposure, including averaging times and duration.

All these exposure-related issues may be a further source of uncertainty and, therefore, should be considered at the introductory steps of the Health Impact Assessment.

Specify the range of exposure to be considered

The magnitude of the impact of a health hazard strongly depends on the level and range of exposure for which the Health Impact Assessment is required to estimate attributable cases. Depending on the purpose of the Health Impact Assessment, delimitations of considered exposure range may apply to either lower or upper end of the exposure distribution, or to its both ends. At the lower end of the range, a "reference level" of exposure below which no impact will be considered may be part of the Health Impact Assessment assumptions. The issue of "reference levels" may have different aspects. These aspects may relate to epidemiology, toxicology or the policy-maker perspective. The following are typical situations where the "reference level" may be of conceptual importance.

- Exposure below some level is assumed to have no measurable effect (threshold; see also section on exposure-response relationship).
- Exposure below some level may be considered "natural" (e.g. natural background level of tropospheric ozone) and the respective impact is excluded from the Health Impact Assessment.
- Exposure may be due to anthropogenic pollution sources, but it may be impossible to attain concentrations below some level under the considered exposure reduction strategy.
- It may only be required to assess the impact of an environmental exposure above some defined level (e.g. above the air quality guideline level).

Depending on the perspective and the purpose of the Health Impact Assessment, these aspects may be weighted in different ways. It should be clearly stated whether any reference level will be adopted, and why.

The outcome of the Health Impact Assessment is also influenced by the range of exposure observed in the population to which the Health Impact Assessment is applied. It might be that a proportion of the population for which Health Impact Assessment has to be conducted, lives under exposure conditions which are higher than those observed in the epidemiological studies providing the exposure-response function. Thus, it has to be discussed whether extrapolation of the exposure response function up to these levels is adequate. It might be considered whether the impact should only be quantified up to a certain level, perhaps the maximum observed in the relevant epidemiological studies, ignoring possible additional impact beyond these levels.

To illustrate the issue, reference can be made to projects assessing the health impact of air pollution. For example, the number of cases of premature death, cardio-respiratory hospital admissions, incidence of chronic bronchitis in adults, acute bronchitis in children, restricted activity days, and asthma attacks which can be attributed to PM10 exposure above 7.5 $\mu g/m^3$ have been estimated in the trinational epidemiology based impact assessment (Seethaler 1999, Künzli et al. 1999). A similar US study did not quantify any impact for PM10 levels below 15 $\mu g/m^3$ and the study assumed a concentration of only 50 $\mu g/m^3$ for the population living in areas with PM10 annual mean levels higher than 50 $\mu g/m^3$. Obviously, these assumptions have a strong impact on the overall results.

Derive the population exposure distribution

To calculate the attributable number of cases, population exposure distribution estimate is required for Health Impact Assessment. The availability of such information may be a limiting factor in Health Impact Assessment. In some cases, the underlying epidemiological studies provide data regarding exposure distributions. In other situations, exposure data may be available from monitoring systems or other exposure assessment studies. In sophisticated Health Impact Assessment approaches, available environmental monitoring data may be used to spatially model pollution levels. Such pollution maps may be combined with demographic data on population density to estimate the exposure (Filliger et al. 1999). In other settings, it may be warranted to use a single estimate of the overall average exposure. This may be appropriate if concentration levels are rather homogeneous across large areas and affect everybody, e.g. ambient outdoor air pollutants such as PM2.5 or ozone. It may be more problematic for heterogeneous exposures. In a simplified Health Impact Assessment, an estimate of the proportion of "exposed" (considering all the others as non-exposed) may be useful.

As stated in further sections, it is important to consider the same definition of exposure as that used in the epidemiological studies from which the exposure-response functions have been derived. For example, if epidemiological studies present the risk function between ambient outdoor average concentrations of particulate matter and mortality (Katsouyanni et al. 1997), Health Impact Assessment requires the population exposure distribution for the ambient particle levels. It would not be appropriate to apply these risk functions to the distribution of personal PM10 exposure – if available – which reflects both indoor and outdoor pollution with particles. Such personal exposure data would only be compatible with exposure-response functions from studies in which personal exposure and health effects have been assessed.

Specify the time window between exposure and effect

When the available evidence suggests a certain temporal scale for an exposure-outcome association, then studies allowing for such latency period should be given higher priority. In any case, it should be made clear whether the assessed health impact relates to immediate or delayed effects of exposure as the interpretations of the results, e.g. by policy-makers or economist, have to take the time window into account.

Select appropriate health outcome(s)

As in case of exposure, the health outcome measures have to be defined. Depending on the goal of the Health Impact Assessment, focus might be on only one or on several health effects, ranging from disturbance to death. If there is evidence for an environmental hazard being associated with several health effects, the impact should be assessed separately for each health end point. However, the selected endpoints may be overlapping measures, reflecting similar aspects of health. For example, when morbidity and mortality are both considered, a single individual may be affected by both, disease and then death. As long as this possibility is acknowledged and described, Health Impact Assessment may justifiably include evaluation of all considered endpoints. (In effect, this is the explicit goal of "burden-of-disease" approaches like Disability-Adjusted-Life-Years (DALY), where time spend in ill health and premature mortality are combined in a composite index (Murray and Lopez 1996, de Hollander et al 1999)). In other cases of Health Impact Assessment, in particular where estimating a monetary cost is required, the impact assessors must clarify whether the health endpoints considered in the assessment may be overlapping entities and whether this may cause "double counting" of the same effects. For example, the tri-national air pollution impact assessment team (Seethaler 1999) decided to only quantify the short-term effects of air pollution on hospital admissions, but not on emergency room visits. This was justified by the observation that, in many health care systems, there is large overlap of the two measures of health care system usage: cardio-respiratory hospitalizations may often go through emergency rooms.

The selection of health outcome(s) will be guided by several aspects of the Health Impact Assessment, mainly the purpose and use of the Health Impact Assessment, the definition of "exposure", and the availability of the respective data. Therefore, for any of these aspects, assessors may decide not to include all measures of health in the Health Impact Assessment for which epidemiological evidence is available. These decisions and the underlying arguments should be made explicit.

Specify the exposure-response relationship

The quantitative association between the hazard and the outcome frequency is crucial information for calculation of the attributable number of cases. The exposure-response function is, in fact, the key contribution of epidemiological studies to Health Impact Assessment. The function may be reported as the slope of a regression line, as a relative risk measure for a given change in exposure or comparing "exposed" with "unexposed". Due to the many sources of uncertainty in observational science, different epidemiological studies may lead to different exposure-response functions. Thus, for the Health Impact Assessment, the process used to derive a exposure-response function (or functions) must be defined. The following issues have to be considered:

- Available epidemiological information should be systematically reviewed in order to obtain information on reliable exposure-response relationships for every selected health outcome. The hazard identification process normally will provide an inventory of the relevant studies that are considered of acceptable quality. All studies with quantitative information on exposure or which allow linkage to such information should be considered for the exposure-response evaluation.

- The process of combining studies for deriving an overall exposure-response relationship may be based on formal meta-analytic methods, pooled analyses, or on expert judgement (Blettner et al. 1999). Published meta-analyses may also be useful, provided they are based on studies that are considered to be eligible for Health Impact Assessment purposes. Measures of uncertainty around central point estimates should be derived and information on heterogeneity between studies (for example from published meta-analyses) should be considered.

- The studies selected during hazard identification may need to undergo an additional selection process and may have to be weighted for the purpose of evaluating the exposure-response relationship for Health Impact Assessment, on the basis of the following aspects:

  (a) The quality of exposure measurement needs to be considered.

(b) Studies based on the same exposure metric as that used in the population for which the impact assessment is required will have highest priority; studies based on different metric, but for which it is possible to convert results to the selected metric, will be given less weight.

(c) Studies will also be evaluated on the basis of whether or not the estimated risks might apply to the population for which the Health Impact Assessment is being conducted (i.e. generalization from one to another population). E.g. information on the possible presence of effect modifiers, such as local socioeconomic factors, or the importance of susceptible subgroups, such as asthmatics, that may drive the observed effects is valuable and should be taken into account.

- It is possible that the body of evidence will provide an estimated exposure-response relationship for a medium range of exposure levels, while Health Impact Assessment is required for a population mainly exposed to much lower or much higher levels. Projecting exposure-response relationships beyond the range of exposure observed in the underlying studies normally involves uncertain extrapolations. The arguments for and the limitations or potential impacts of extrapolations ought to be carefully addressed in the Health Impact Assessment. Knowledge of the biological mechanisms underlying the specified effect may support the decision to extrapolate. In any case, allowance for additional uncertainty should be made.

- The shape of the exposure-response function should be specifically evaluated in all available studies. Particularly, the possible existence of threshold levels ("no effect level") may be very important for the Health Impact Assessment.

Derive population baseline frequency measures for the health outcomes under consideration

In epidemiological studies, effects are most often reported as the relative change in risk rather than the absolute increase in number of subjects affected. Therefore, the Health Impact Assessment process to quantify the impact requires information on the baseline occurrence (rate, prevalence) of the selected outcome. With this information at hand, it is possible to calculate how many additional cases may be expected or may be attributed to some level of exposure. While exposure-response relationships may be derived from the international literature, the baseline disease occurrence should preferably be obtained from data regarding the population for which Health Impact Assessment is being made. If such local data are not available, health frequency data from other populations may sometimes be used. For example, if it is known that an environmental hazard increases the number of asthma attacks among asthmatics, quantification of the impact requires information about the number of asthmatics in the population as well as the average number of asthma attacks per asthma patient. Such information may be difficult to get for the target Health Impact Assessment population and data from other sources may be used (Künzli et al. 1999). Apart from situations with complete availability of population health measures (e.g. death statistics), baseline frequency data are estimates which are subject to errors and uncertainties.

Calculate the number of attributable cases

The epidemiology based Health Impact Assessment essentially relies on the attributable risk concept described in most epidemiology textbooks or specific examples of impact assessment (e.g. Doll & Peto 1981, Rothman & Greenland 1998). It consists in combining the major outputs from the steps described above, which are:

1. estimates of the epidemiology based exposure-effect function, i.e. the mathematical link function between the degree of exposure and the expected change in health state;

2. estimates of the epidemiology based baseline frequency of the health measure of interest;

3. the distribution of the exposure in the target population.

Based on the observed frequency of health outcomes (incidence, prevalence) and the observed actual level of exposure, the expected number of cases will be calculated for an assumed baseline (or "reference") level of exposure.

In theory, the *(population) attributable proportion* (AP) is the fraction of all cases attributed to a specified (dichotomous) exposure causing the health outcome:

$$AP = [p (RR - 1)] / [1 + p (RR - 1)]$$

where RR = relative risk for the health outcome due to the exposure, and
p = the proportion exposed in the population.

The important assumptions at this stage are that there is a causal relationship between the exposure and the health outcome, that the relative risk estimate applies to the all exposed group and that there is no confounding of the observed effect.

For example, if the whole population is exposed, p=1 and the (population) attributable proportion is equivalent to the *attributable fraction among the exposed (PA_e)* (Last & Abramson 1995). A general formula to derive AP may be applied in the situations with two or more exposure categories ($n \geq 2$) (Walter 1976, Krzyzanowski 1997):

$$AP = S \{ [RR(c) - 1] p(c) \} / S [RR(c) p(c)] \qquad (c=1,\ldots n)$$

Where:
RR(c) = relative risk for the health outcome in category c of exposure (by definition, c=1 is a reference category with RR(1) =1),
p(c) = proportion of the target population in category c of exposure, S p(c) = 1.

S denotes summation for c=1 to n.

Multiplying the observed frequency of health outcomes (incidence, prevalence) and the AP estimated according to the above formulas, the expected frequency attributable to the exposure can be calculated. Product of this frequency and the size of the population under the study gives the expected number of cases attributed to the exposure.

In practice, both the estimation and interpretation of the population attributable proportion, and its application to the target population, may involve a number of subtleties. They regard, for example, the choice of relative risk estimator or existence of confounding. Consideration of those factors would involve more complex statistical calculations than this basic description implies (see e.g. Rockhill et al. 1998, Greenland & Robins 1988). The assumptions made and their expected impact on the impact estimates should be described in detail.

Consideration of uncertainties in risk estimates and in exposure distribution results in an estimate of a range of impact estimates, and not in a single number. In more complex approaches, a probability distribution of impact is estimated, e.g. by Monte Carlo techniques (Covello & Merkhofer 1993).

If one purpose of Health Impact Assessment will be the estimation of the monetary cost of the health impact, we also need economic valuation tools of the estimated public health impact.

Discussion and weighting of the overall results

The results of the 10-step procedure outlined above should be interpreted coherently and limitations should be made explicit. Aspects of the overall uncertainty and its potential impact on the results ought to be addressed. A sensitivity analysis may provide a better evaluation of the overall uncertainty of the estimates.

In a Health Impact Assessment, the purpose of which is to estimate the attributed cases for an existing hazard, it should be made clear that the impact, i.e. the attributable number of cases, may not simply be interpreted as the "preventable fraction". In other words, it may not be by default true that removal of the environmental hazard will prevent the occurrence of the full number of cases obtained in the Health Impact Assessment. The quantitative reversibility function, i.e. the direct effect estimate of the removal of an exposure may only rarely be assessed. Depending on the health outcome, the specificity of the exposure and the time frame of exposure and effect, the benefit (or reversibility) may be harvested much later or not to the full extent predicted. This is due to the open question of how competing risks may come into play if one contributing cause is removed or reduced. This caveat, however, inherently relates to the concept of attributable cases which has been applied to impact quantification in public health for decades (Doll & Peto 1981, Rothman & Greenland 1998).

Improvement of epidemiological studies for Health Impact Assessment

Health Impact Assessment is an interdisciplinary process. Epidemiological data can often play a crucial role in it. To improve the applicability of epidemiological research to Health Impact Assessment, future epidemiological studies should seek where possible to provide results in a way that enhances the Health Impact Assessment at the interface of epidemiology, other fields of research, and policy-making. As stated in the above sections, the following aspects should be considered:

- The lack of comparability of the exposure data available for steps related to the exposure distribution and exposure-response in the procedure summarized in the earlier section often stems from a lack of interdisciplinary communication. Thus, from the Health Impact Assessment perspective, future epidemiological studies should ideally provide adequate and complete information of the exposure used, including definition, measurements, exposure distribution and ranges of observed exposure.

- To facilitate Health Impact Assessment, epidemiological studies ought to provide the characteristics of their exposure metrics and give extensive descriptive statistics of exposure. In cases where several indicators may be available to describe exposure, the quantitative relationship between the indicators should be characterized. However, to get population exposure distribution data, separate studies may be required for Health Impact Assessment.

- The distinction and overlap between long-term and short-term effects should be specifically addressed in future investigations as the definition of the time-window is important in the Health Impact Assessment.

- More emphasis should be placed in epidemiological research on the explicit assessment of no-effect thresholds of exposure. The better determined the shape of the exposure-response curve, the more reliable the Health Impact Assessment.

References

BADGETT, R.G. ET AL. Using systematic reviews in clinical education. *Annals of internal medicine*, **126**: 886–891 (1997).

BERO, L.A. & JADAD, A.R. How consumers and policymakers can use systematic reviews for decision making. *Annals of internal medicine*, **127**: 37–42 (1997).

BLAIR, A. ET AL. Guidelines for application of meta-analysis in environmental epidemiology. *Journal of regulatory toxicology and pharmacology*, **22**: 189–197 (1995).

BLETTNER, M. ET AL. Traditional reviews, meta-analyses and pooled analyses in epidemiology. *International journal of epidemiology*; **28**: 1–9 (1999).

CHALMERS T.C. & LAU, J. Meta-analytic stimulus for changes in clinical trials. *Statistical methods in medical research*, **2**: 161–72 (1993).

COVELLO, V.T. & MERKHOFER, M.W. *Risk assessment methods. Approaches for assessing health and environmental risks.* New York, Plenum Press, 1993.

DE HOLLANDER, A.E.M. ET AL. An aggregate public health indicator to represent the impact of multiple environmental exposures. *Epidemiology*, **10**: 606–617 (1999).

DICKERSIN, K. & BERLIN, J.A. Meta-analysis: state of the science. *Epidemiological reviews*, **14**: 154–176 (1992).

Department of Health. *Economic appraisal of the health effects of air pollution*. London, The Stationery Office, 1999.

DOLL, R. & PETO, R. *The causes of cancer. Quantitative estimates of avoidable risks of cancer in the United States today*. Oxford/New York, Oxford Medical Publications, 1981.

*Principles for evaluating epidemiologic data in regulatory risk assessmen*t. Washington, DC, Federal Focus, Inc. 1996.

FILLIGER, P. ET AL. PM10 Population exposure. *Health costs due to road traffic-related air pollution. An impact assessment project of Austria, France and Switzerland. Prepared for the Third WHO Ministerial Conference on Environment and Health, London, 16–18 June 1999.* Berne, Federal Department for Environment, Transport, Energy and Communications Bureau for Transport Studies, 1999

GORDIS, L. ED. *Epidemiology and health risk assessment.* New York, Oxford University Press, 1988.

GREENLAND, S. Quantitative methods in the review of epidemiologic literature. *Epidemiological reviews*, **9**: 1–30 (1987).

GREENLAND, S. & ROBINS, J. Conceptual problems in definition and interpretation of attributable fractions. *American journal of epidemiology*, **128**: 1185–1197 (1988).

HERTZ-PICCIOTTO, I. Epidemiology and quantitative risk assessment: a bridge from science to policy. *American journal of public health*, **85**: 484–491 (1995).

HILL, A.B. The environment and disease: association or causation. *Proceedings of the Royal Society of Medicine*, **58**: 295–300 (1965).

HORTON, R. The *new* new public health of risk and radical engagement. *Lancet*, **352**: 251–252 (1998).

KATSOUYANNI, K. ET AL. Short term effects of ambient sulphur dioxide and particulate matter on mortality in 12 European cities: results from time series data from the APHEA project. *British medical journal*, **314**: 1658–1663 (1997).

KRZYZANOWSKI, M. Methods for assessing the extent of exposure and effects of air pollution. *Occupational and environmental medicine*, **54**: 145–151 (1997).

KÜNZLI, N. ET AL. Technical report on epidemiology. *In: Health costs due to road traffic-related air pollution. An impact assessment project of Austria, France and Switzerland. Prepared for the Third WHO Ministerial Conference on Environment and Health, London, 16–18 June 1999.* Berne, Federal Department for Environment, Transport, Energy and Communications Bureau for Transport Studies, 1999.

LAST, J.M. & ABRAMSON, J.H. *A dictionary of epidemiology,* 3rd ed. New York, Oxford, Toronto, Oxford University Press, 1995.

MOOLGAVKAR, S.H. When and how to combine results from multiple epidemiological studies in risk assessment. *In:* Graham, J.D., ed. *The role of epidemiology in regulatory risk assessment.* Amsterdam, Elsevier Press, 1995.

MORGAN, M.G. & HENRION, M. *Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambridge, Cambridge University Press, 1990.

MURRAY, C.J.L. & LOPEZ, A.D., ED. *The global burden of disease; a comprehensive assessment of mortality and disability from disease, injury, and risk factors in 1990 and projected to 2020*. Cambridge, MA, Harvard University Press, 1996 (Global burden of disease and injury series, Vol. 1).

NATIONAL RESEARCH COUNCIL. *Risk assessment in the Federal Government: managing the process*. Washington, DC, National Academy Press, 1983.

NEUTRA, R.R. & TRICHOPOULOS, D. The place of epidemiology in environmental decisions: needed support for the development of risk assessment policy. *Environmental health perspectives supplements,* **101**(4): 67–69 (1993).

OHLSSON, A. Systematic reviews – theory and practice. *Scandinavian journal of clinical & laboratory investigation,* Suppl. No. 219: 25–32 (1994).

RAI, S.N. ET AL. A general framework for the analysis of uncertainty and variability in risk assessment. *Human and ecological risk assessment,* **2**(4): 972–989 (1996).

ROCKHILL, B. ET AL. Use and misuse of population attributable fractions. *American journal of public health*, **88**: 15–19 (1998).

ROTHMAN, K.J. ED. *Causal inference.* Chestnut Hill, MA, Epidemiology Resources Inc., 1988.

ROTHMAN, K.J. & GREENLAND, S. *Modern epidemiology,* 2nd ed. Philadelphia, Lippincott-Raven, 1998.

SAMET, J.M. & BURKE, T.A. Epidemiology and risk assessment. *In:* Brownson, R.C. & Petitti, D.B., ed. *Applied epidemiology: theory and practice.* New York, Oxford University Press, 1998, pp. 137–175.

SCOTT-SAMUEL, A. Health impact assessment – theory into practice. *Journal of epidemiology and community health,* **52**: 704–705 (1998).

SEETHALER, R. Synthesis report. *In: Health costs due to road traffic-related air pollution. An impact assessment project of Austria, France and Switzerland. Prepared for the Third WHO Ministerial Conference on Environment and Health, London, 16–18 June 1999.* Berne, Federal Department for Environment, Transport, Energy and Communications Bureau for Transport Studies, 1999 (Order No. 801.633e).

SOMMER, H. ET AL. Economic evaluation. Technical report on economy. *In: Health costs due to road traffic-related air pollution. An impact assessment project of Austria, France and Switzerland. Prepared for the Third WHO Ministerial Conference on Environment and Health, London, 16–18 June 1999.* Berne, Federal Department for Environment, Transport, Energy and Communications Bureau for Transport Studies, 1999.

STAATSEN, B.A.M. ET AL. *Health impact statement Schiphol Airport.* Executive summary [in English]. Bilthoven. RIVM, 1994 (Report 441520003).

TAUBES, G. Epidemiology faces its limits. *Science*, **269**: 164–169 (1995).

THACKER, S.B. Meta-analysis: a quantitative approach to research integration. *Journal of the American Medical Association*, **259**: 1685–1689 (1988).

The benefits and costs of the Clean Air Act. Washington, DC, US Environmental Protection Agency, 1996.

WALTER, S.D. The estimation and interpretation of attributable fraction in health research. *Biometrics*, **32**: 829–849 (1976).

WONG, O. & RAABE, G.K. Application of meta-analysis in reviewing occupational cohort studies. *Occupational and environmental medicine*, **53**: 793–800 (1996).

Principles for the assessment of risks to human health from exposure to chemicals. Geneva, World Health Organization, 1999 (Environmental Health Criteria 210).

Environmental health risk assessment is increasingly being used in the development of environmental health policies, public health decision-making, the establishment of environmental regulations and research planning. The credibility of risk assessment depends, to a large extent, on the strength of the scientific evidence on which it is based. It is, therefore, imperative that the processes and methods used to evaluate the evidence and estimate health risks are clear and explicit, and based on valid epidemiological theory and practice.

Evaluation and use of Epidemiological Evidence for Environmental Health Risk Assessment is a guideline document. The primary target audiences are expert review groups that WHO (or other organizations) might convene in the future to evaluate epidemiological evidence on the health effects of environmental factors.

This Guideline Document identifies a set of processes and general approaches to assess available epidemiological information in a clear, consistent and explicit manner. The Guideline Document should also help in the evaluation of epidemiological studies with respect to their ability to support risk assessment and, consequently, risk management. Conducting expert reviews according to such explicit guidelines would make health risk assessment, and subsequent risk management and risk communication processes, more readily understood and likely to be accepted by policy-makers and the public. It would also make the conclusions reached by reviews more readily acceptable as a basis for future WHO guidelines and other recommendations, and would provide a more rational basis for setting priorities for future research.

World Health Organization
Regional Office for Europe

Scherfigsvej 8
DK-2100 Copenhagen Ø
Denmark

Telephone +45 39 17 17 17
Telefax: +45 39 17 18 18
Telex: 12000 who dk

Electronic mail: postmaster@who.dk
World Wide Web Address: http://www.who.dk