Australian Government

National Health and Medical Research Council

Ambient Air Quality Standards Setting

An Approach to Health-Based Hazard Assessment

National Health and Medical Research Council
and
Environmental Health Committee (enHealth)

Ambient Air Quality Standards Setting:
An Approach to Health-Based Hazard Assessment

Prepared by:

Professor Brian Priestly
Director, Australian Centre for Human Health Risk Assessment
Department of Epidemiology & Preventive Medicine, Monash University

Dr Roger Drew
Principal Toxicologist & Director, Toxikos Pty Ltd

Professor John McNeil
Head, Department of Epidemiology & Preventive Medicine, Monash University; Leader, Preventive Medicine Unit

Professor Michael Abramson
Deputy Head, Department of Epidemiology & Preventive Medicine, Monash University; Leader Clinical Epidemiology Unit

Associate Professor Andrew Forbes
Leader Biostatistics Unit,
Department of Epidemiology & Preventive Medicine, Monash University

ACKNOWLEDGEMENT

The work documented in this Report was funded by a contract with the National Health and Medical Research Council, and the Department of Health and Ageing.

The authors also acknowledge, with thanks, the assistance of Dr Andrew Harford, Dr Jennifer Ong, Mr John Frangos and Dr Martine Dennekamp with literature searches and preparation of the manuscript.

Thanks are also extended to the Working Committee for Air Standards Health Advice:

Mr Jim Dodds (Chair), Environmental Health Director, Department of Health Western Australia

Dr Andrew Langley, Director, Centre for Disease Control, Central Zone Public Health Network

Mr Chris Bell, Director, Water and National Program, EPA Victoria

Ms Deborah Willcocks, Team Leader, Rapid Risk Assessment, NICNAS

Dr Shannon Rutherford, Scientific Adviser on Environmental Health, Queensland Health

Dr Monika Nitschke, Scientific Officer, Department of Health, South Australia

Ms Jenni Paradowski, Director, Office of Health Protection, Department of Health and Ageing.

CONTENTS

EXECUTIVE SUMMARY AND RECOMMENDATIONS

This report reflects consensus between the health and environment sectors on a common, health based approach to the hazard assessment component of setting ambient air quality standards. It will enable health professionals to develop information in ways that effectively support an evidence-based decision-making process within the environment sector.

This work is the outcome of a national Workshop conducted in Canberra in July 2006, with fifty participants from the health and environment sectors collaborating to discuss a health-based methodology for setting ambient air quality standards. The workshop was hosted by the National Health and Medical Research Council and the Department of Health and Ageing. The discussion paper prepared for the workshop by consultants from Monash University forms the basis of this report, with subsequent input from submissions generated in targeted consultations and discussions at the workshop.

The report will inform the Environment Protection and Heritage Council Standards Setting Working Group and the National Environment Protection Council (NEPC) in NEPC's processes of reviewing the National Environment Protection Measure (NEPM) for Ambient Air Quality, due to be finalised in 2008. It has been approved by the Chief Executive Officer of the NHMRC, and by the NHMRC Council, the Environmental Health Committee (enHealth) and the Australian Health Protection Committee.

Issues discussed

Issues addressed at the Workshop, and in this Paper, include:

- what constitutes a critical adverse health effect associated with air pollution that would be suitable for standard setting;
- how to identify the studies that will be most useful for setting health-based air quality standards;
- where, in the standard-setting process, to consider factors that are indirectly related to health, such as impacts on quality of life;
- how epidemiological data should be used in standard setting, including:
  - the appropriate use of meta-analysis to improve the estimate of the association of health effects with air pollution; and
  - the relative weights to be given to data derived from studies based on Australian and overseas populations and localities;
- use of weight-of-evidence (WoE) analysis to determine the most appropriate studies and data for use in the standard-setting process. This includes analysis of potential cause-effect relationships, using the viewpoints established by Bradford Hill;
- the utility of biomarkers in identifying health effects associated with air pollutants;

- the interpretation of studies that suggest cancer as a potential adverse health effect associated with air pollutants, including criteria for judging the relevance, in standard setting, of carcinogenic outcomes in experimental animal models;
- the use and interpretation of experimental data from controlled exposures in humans and/or animals, including specific advice on appropriate methodology for dose-response modelling (DRM) and the use of DRM data in air quality standard setting;
- time scaling of data, so that studies deriving health effect dose-response relationships over varying exposure times may be adapted to setting averaging periods for air quality standards;
- how the health impacts on sensitive subgroups may affect the standard-setting process; and
- the sequential steps that could be used to develop air quality standards based on data from experimental and epidemiological studies.

Recommendations

Consideration of these issues led to the development of the following set of recommendations:

1. The guidance for environmental health risk assessment published by enHealth Council (2004) represents a suitable framework for the risk assessment component of setting air quality standards.
2. Air quality standards should take account of clinically relevant changes in physiological functions (particularly involving the sensory, respiratory and cardiovascular systems, even if the changes are relatively subtle) as well as more serious adverse health effects such as mortality, cancer and other delayed systemic toxicity. Standards should be based on robust dose-response modelling data, if available.
3. The Bradford Hill viewpoints are valid as a means of assessing whether specific air pollutants have a causal relationship with adverse health effects.
4. A weight-of-evidence analysis should be used to assess both qualitative and quantitative relationships between air pollutants and potential adverse health effects for those exposed to ambient air.
5. Where the key health effects are direct and of a relatively mild nature, experimental studies with humans (controlled chamber exposures) are likely to provide data of specific relevance to standard setting. However, where the health effects are severe or of a delayed onset, and experimental exposures of humans is unethical, well conducted epidemiological studies, supported by valid meta-analyses and/or animal experimental studies, are more likely to be useful.
6. Meta-analysis should be conducted only when original studies are of similar design, use comparable populations, and measure exposure and outcome in similar ways. Appropriate statistical methods should be used to assess meta-analyses, including a test for heterogeneity. Pooled results should not be relied upon for standard setting in the presence of significant heterogeneity between the studies.

7. Epidemiological data derived from overseas studies may be generalised for use in standard setting, provided the potential confounders (such as population demographics, meteorological differences and other such variables) are taken into account.

8. The possibility of exposure of sensitive sub-populations within the general population should be considered when setting an air quality standard. Where possible, chemical-specific adjustment factors should be developed. Where this is not possible, then the default composite uncertainty factor of 10 for human variability should be adopted, and this should be divided into kinetic and dynamic components, each being 3.16.

9. The default composite uncertainty factor of 10 for human variability adequately accounts for sensitive persons within the general population dose-response distribution. Only if the data demonstrate the presence of discrete genetic polymorphic sensitive populations would the default uncertainty factor need to be adjusted upwards. Different adjustment values should be used only when supported by relevant data.

10. Where potential carcinogenesis based on findings in animal studies is the critical health effect used for setting an air quality standard, the relevance to humans of the mode of action of the carcinogenic response should be assessed using criteria such as those recently published by the International Programme on Chemical Safety and the US Environment Protection Agency.

1 INTRODUCTION

There is a substantial body of evidence showing that air pollution in urban areas affects human health (AIRNET WG 4 2004; Brunekreef & Holgate 2002; Cohen *et al*. 2005). Given that air pollution has the potential to affect everyone in the community, and that individuals cannot readily control the extent to which they may be exposed to air-borne pollutants, there is a reliance on governments to ensure that appropriate levels of public health protection are enacted through air quality standards.

This Report represents consensus between the health and environment sectors on a common approach to the hazard assessment component of setting air quality standards. It will enable the health sector to develop information in ways that effectively support an evidence-based decision-making process within the environment sector.

1.1 SCOPE AND OBJECTIVES

The technical advice included in this document should assist jurisdictions to

- develop air quality standards or guidelines for air pollutants not covered by current National Environment Protection Measures (NEPMs), or
- revise existing standards or guidelines.

The difference between an air quality guideline or a standard in the context of this document relates to the way the numerical output from the process is used in risk management[1].

The advice provides a common platform that can be used by jurisdictions, industry, authorities or other stakeholders to establish numerical benchmarks to assist in the management of short-term or long-term air quality issues at the local, regional or national level. Specifically, the document provides overall direction and advice, relating to hazard assessment, on the choice and use of data to establish air guidelines or standards based on public health considerations. In establishing such standards, the detailed deliberations for each substance will be different and, in the context of health considerations, driven primarily by the extent and quality of data available.

As discussed by Maynard *et al*. (2003) in a review of research needs, there are many areas of uncertainty in standard setting, the resolution of which may require data. This document does not provide in-depth background information, nor detailed discussion of the many areas of debate associated with air standard setting, nor of the numerous forks in the process where a decision needs to be made, nor details of technical nuances for all situations that may arise when setting a standard.

[1] Although the regulatory difference between a guideline and a standard is recognised (see glossary), and in some regulatory or enforcement contexts the terms ‘guideline ‘ and ‘standard’ may be used interchangeably, the term ‘standard’ has been used preferentially in this document. Nevertheless, the document aims to provide guidance on methodologies which allow quantitative estimates of exposure that are protective of human health for a range of exposure scenarios, including those covered by standards.

The setting of health-based standards for air-borne chemicals is a multidisciplinary process. It involves, or should involve, epidemiologists, toxicologists, chemists, air dispersion modellers, and other disciplines as required. The key point is that these experts need to be proficient and active in their respective fields of expertise. Consequently not all people involved in health-based standard setting, who may refer to this report, will necessarily be able to apply advice that is outside their field of speciality.

The principles and advice in this document are pertinent to establishing air quality standards intended for national adoption, and management of large air sheds or of local specific pollutant sources. They are applicable to area-wide pollutants arising from diffuse sources, or to substances that have long-range transport, as well as those emitted from defined point sources that may only have potential for relatively local health impacts.

The advice provided relates only to potential health impact. It does not include risk management or other issues that may influence the final standard (eg available resources, air quality policy issues, capacity of air sheds, aspects of odour and well-being, ability to measure the pollutant, economic impacts). However, both 'health' and 'non-health' aspects should be carefully described in the documentation that supports and justifies the standard.

Air quality standards relate to public health—that is, to populations. The intention is to protect the vast majority of an exposed population from serious health effects. However, while sensitive sub-populations are considered in the standard-setting process, an air quality standard does not necessarily protect all of the people, all of the time, from all possible health effects; nor does it provide information on how a particular individual may perceive or respond to a pollutant present in air at levels that are at or below the level specified by the standard. Consequently, some people may occasionally sense the presence of a pollutant in the air or experience minor discomfort at levels that comply with the standard.

Standards should be established using the best information available at the time, and they should be amended in response to new information regarding the specific substance, or to advances in risk assessment.

1.2 RELATIONSHIP TO PREVIOUS GUIDANCE

The risk assessment paradigm and technical information provided by enHealth (2004) are recommended as the primary platform for setting health-based air quality standards in Australia, and the five-step process outlined by enHealth provides the basis for the framework set out in Figure 1.1.

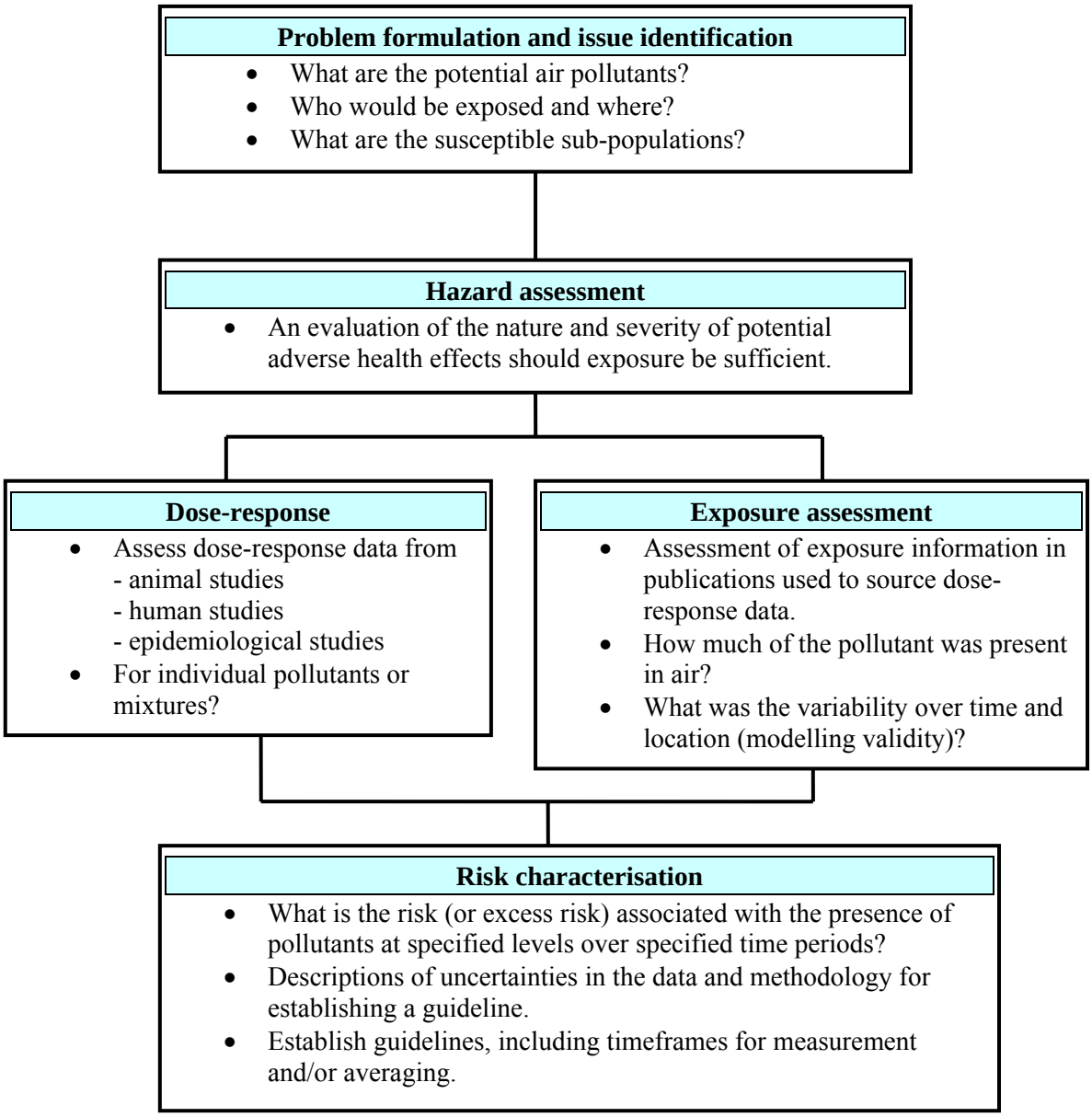

Figure 1.1: Risk assessment paradigm for air pollutants
Adapted from enHealth (2004).

The information provided by enHealth (2004) is complemented by other Australian documents:

- A recent Western Australia Health Department publication provides a concise overview of risk assessment, which consolidates and simplifies that given in the enHealth (2004) document (WA Department of Health 2006).
- The report of the Department of Environment and Heritage (2001), while relating primarily to assessment of indoor air quality, reviews national and international approaches to the assessment of air quality.
- Other guidance on the use of health risk assessment for the establishment of air quality standards is found in the outcomes of the Risk Assessment Task Force (RATF) of the National Environment Protection Council (RATF 2000).

As discussed above, standards should be established using the best information available at the time, and they should be amended in response to new information. Unthinking adherence to the guidance in enHealth (2004) may not, therefore, be appropriate in the face of new information on a substance and/or advances in risk assessment techniques. Any significant departure from the risk assessment advice in enHealth (2004) should be justified in the documentation supporting the numerical value in the standard.

1.3 SPECIAL ISSUES IN HEALTH RISK ASSESSMENT OF AIR POLLUTANTS

Appendix 2 of the enHealth guidance document (2004) notes the characteristics of air pollutants that dictate a process of health risk assessment different to that for other exposure routes:

- There is generally little individual choice over the quality of the ambient air that people breathe, or the extent to which they are exposed through normal day-to-day living.
- Air dispersion modelling[2] may be required to characterise point and diffuse sources of air pollution.
- Ambient air exposures are highly dependent on meteorological factors, which must be factored into any air dispersion modelling.
- For at least some air pollutants, human dose-response data are available from controlled exposure chamber studies.
- Irritation is often a critical determinant of an adverse health effect of an air-borne pollutant, and it can be associated with relatively short exposures to concentration 'spikes'. Irritant effects can occur when there is little or no systemic absorption of the chemical.
- Non-irritant or systemic toxicity may be a problem for some types of pollutants. For such substances, longer-term exposures are more likely to be important for standard

[2] Air dispersion modelling is frequently an inherent component of air guideline setting, especially for the management of local population impacts around point-source pollution emission. This topic is not addressed here, as comprehensive advice and technical requirements needed by Australian jurisdictions are available from the States.

setting than short-term exposures. Carcinogenesis is a prime example of such an effect. A substance that has both short- and long-term health effects may require more than one air quality standard to be assigned to it.

1.4 OTHER SOURCES OF GUIDANCE

This section provides a selection of guidance documents and advice relating to air quality standards and the processes for establishing them, that have been developed by international agencies. The list is not exhaustive, and inclusion does not indicate endorsement by the NHMRC. When using advice from any guidance document, it is important that it be based on sound science, and that it is transparent and compatible with science policy and general risk assessment practice in Australia—that is, with the general advice provided by enHealth (2004).

Guidance documents from international agencies include:

- US EPA (2006c) *Review of the process for setting national ambient air quality standards* ;
- California EPA (1999) *Determination of acute reference exposure levels for airborne toxicants;*
- WHO (2000a) *Air quality guidelines for Europe (2nd edition);*
- WHO (2000b) *Quantification of the health effects of exposure to air pollution;*
- WHO (2000c) *Evaluation and use of epidemiological evidence for environmental health risk assessment;*
- WHO (2005) *Global update of air quality guidelines;*
- US EPA (2004) *Introduction to air toxics risk assessment;* and
- US EPA (2005b) *Particulate matter health risk assessment for selected urban areas*.

The US EPA maintains an extensive reference library on air toxics on its website http://www.epa.gov/ttn/fera/risk_atra_vol1.html

The AIRNET project was initiated to develop an overarching Europe-wide framework for air pollution and health research. AIRNET collects, interprets and disseminates data from individual (EU-funded) projects, in order to strengthen the science-policy interface and draw policy-relevant recommendations.

In 2002, four of its working groups published reports on:

- Interpretation of exposure findings (WG1);
- Interpretation of epidemiological findings (WG2);
- Interpretation of toxicology findings (WG3); and
- Risk in health impact assessment (WG4).

The three tiers of the US EPA-promulgated Acute Exposure Guideline Levels (AEGL) offer information pertinent to graded protection of the general population against acute exposures from air-borne toxic chemicals (US EPA 2006a). While principally developed

for emergency planning purposes, and therefore of only indirect relevance to the setting of ambient air quality standards in Australia, AEGLs do have some elements in common with processes discussed in the current document.

2 ISSUE AND HAZARD IDENTIFICATION

The first step in standard setting is to identify and categorise air pollutants that may give rise to significant adverse health effects.

2.1 NATURE OF HEALTH EFFECTS ASSOCIATED WITH AIR POLLUTION

The section considers how an adverse health effect is defined, and how to differentiate those effects that may be considered a serious threat to health and well-being, from those that reflect a minor health impact or inconvenience. Building on this discussion, Section 5.2 considers what constitutes a critical health effect—that is, one that is of sufficient significance to drive the risk assessment process leading to an air quality standard.

2.1.1 General principles for defining an adverse health effect

Broadly speaking, an adverse health effect is anything that represents a departure from health as defined by the World Health Organization (1948)[3]; that is, 'a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity'.

In public health practice, readily measurable health objectives need to be established. These are usually based on an absence or presence of physical disability, or a significant impact on quality of life. Clinical thresholds chosen for these measurements will influence the frequency with which adverse health effects are recorded.

Well-being is broadly described as a person's self assessment of his or her state of happiness, healthiness and prosperity. It relates to quality of life and the ability to enjoy life, and is affected by a wide range of social and economic factors.

The National Health Council of New Zealand (NHC 2004) cites the following determinants of health and well-being:

- social and cultural factors (eg social support, participation, access to cultural resources);
- economic factors (eg income levels, access to employment);
- environmental factors (eg land use, air quality);
- population-based services (eg health and disability services, leisure services);
- individual/behavioural factors (eg physical activity, smoking); and
- biological factors (eg biological age, illness, anxiety).

While the more subtle and subclinical effects of air pollution may be widespread, the most severe health effects are experienced by only a small proportion of the general population. This is clearly demonstrated in the 'health effects pyramid' in the AIRNET WG4 (2004) document, reproduced as Figure 2.1.

[3] Although the document containing this definition has been reviewed by the WHO since 1948, the definition has not been amended.

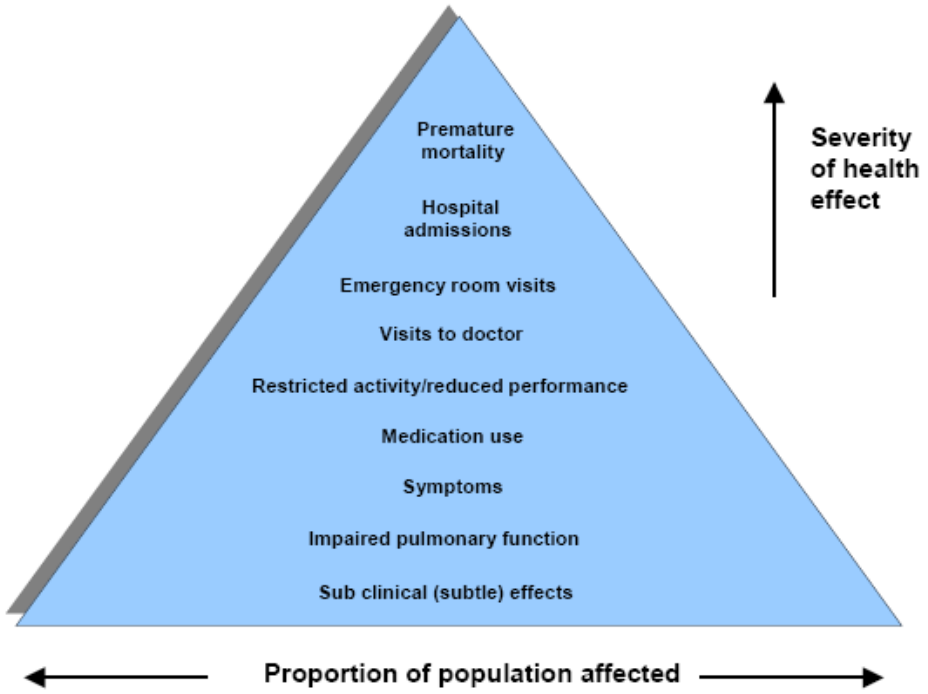

Figure 2.1: The air pollution health effects pyramid (AIRNET WG 4 2004)

Epidemiological studies usually focus on, or have the data to address, prevalence of only the most severe health effects—those at the top of the pyramid. Effects at the base of the pyramid involve many more people and therefore may have greater social, economic and individual impact. They are, however, more difficult to detect and define quantitatively, especially on a population basis, and they may not therefore be adequately addressed in the technical process of standard setting. There are a number of points in the technical process where this uncertainty can be addressed, most notably in the application of adjustment factors (Section 5) or in the risk management phase of standard setting.

The adverse effects of systemic toxicity (ie non-cancer effects) may lie along a continuum related to increasing exposure, which results in an imbalance in homeostatic physiologic functions (those that maintain the body's internal environment). Compensatory mechanisms may protect individuals at low levels of exposure (US EPA 2004). This relationship is expressed graphically in Figure 2.2.

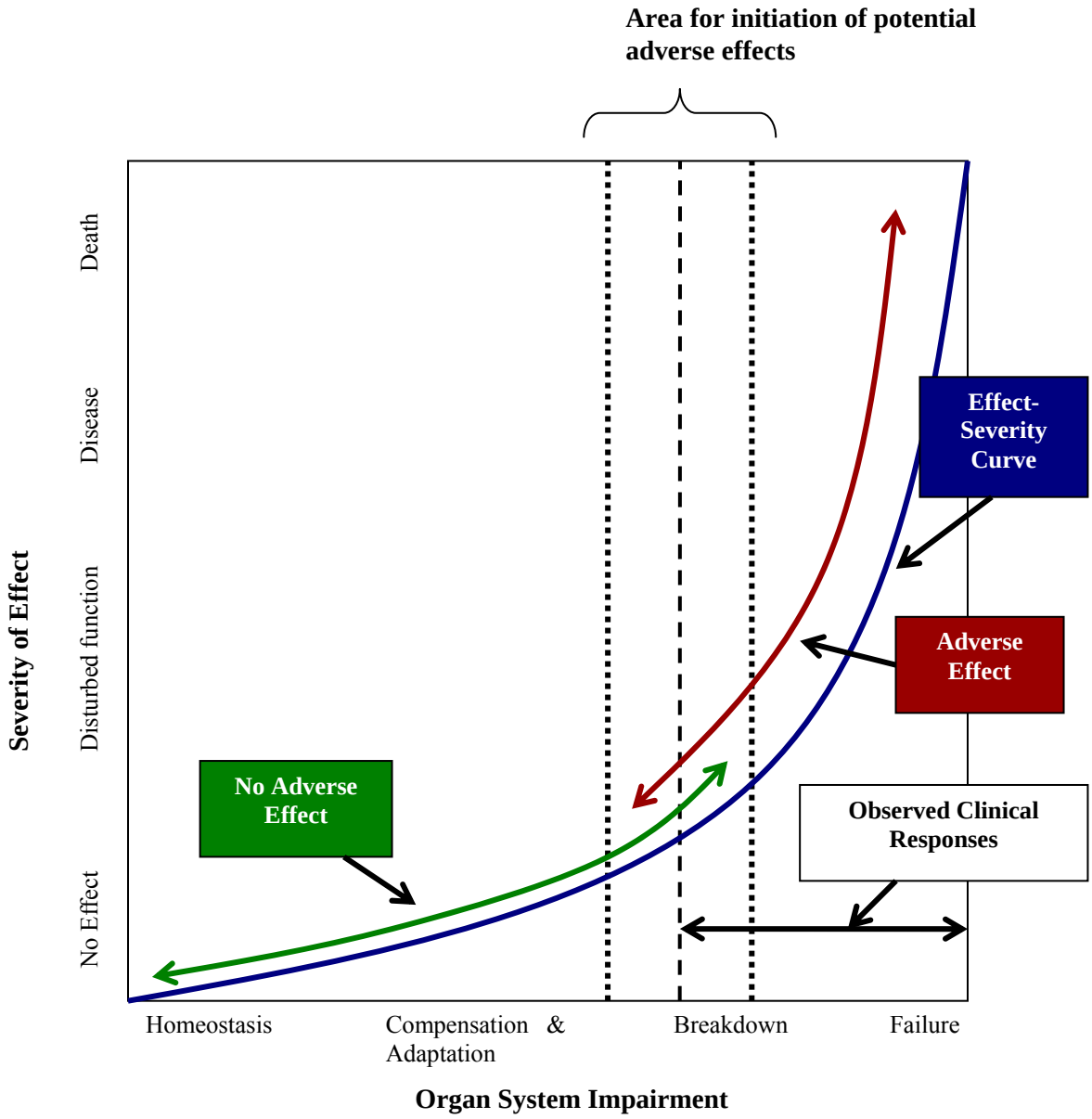

Figure 2.2: Schematic representation of the continuum of adverse and no-adverse health effects and the potential overlap
Adapted from US EPA (2004)

2.1.2 What constitutes an 'adverse health effect'?

This section relates specifically to health endpoints that may be identified by epidemiological and/or toxicological research.

Adverse health effects may be:

- *direct*, eg mesothelioma, lung-cancer, exacerbation of asthma or chronic obstructive pulmonary disease (typically measured objectively using pulmonary function tests); or
- *indirect*, eg reduced quality of life, increased hospitalisation or hospital emergency presentations, reduced exercise tolerance, increased medication use.

In standard setting, the advantage of using directly induced adverse health effects is that they can be most readily related to exposure.

Newer and more sensitive surrogate measures (eg the presence of inflammatory cells in sputum or bronchial washings) are being introduced and may, in time, enable exposure-effect studies to be undertaken with less discomfort to participants. However the relationship between the dose response for these effects and the dose response for clinical effects (eg bronchospasm) needs to be unambiguously established before these novel surrogate measures can be used quantitatively in standard setting (see also the discussion on biomarkers, Section 2.1.3). Non-specific airways hyper-responsiveness is falling out of favour as a physiological marker of asthma, and asthma research is focussing on more novel markers of airway inflammation such as the fraction of exhaled nitric oxide (fENO) and measurements of cytokines and inflammatory mediators in exhaled breath condensate (Horvath *et al*. 2005). There is a long way to go before this information can be defensibly used in standard setting.

Useful criteria on what constitutes an adverse respiratory health effect of air pollution are provided in the guidelines developed by the American Thoracic Society (ATS 1985; ATS Board of Directors 2000). This issue has not yet been addressed by the Thoracic Society of Australia and New Zealand (TSANZ) or other relevant local health authorities. An expert committee of the American Thoracic Society concluded that the following should be considered when assessing adverse health impacts of air pollutants:

- physiological impact;
- clinical symptoms;
- clinical outcomes;
- mortality;
- population health versus individual risk;
- other potential adverse respiratory effects;
- non-respiratory effects; and
- quality of life.

Each of these is discussed below.

Physiological impact

This is defined as loss of lung function attributable to air pollution exposure, including reversible loss in combination with the presence of symptoms, or any detectable level of permanent lung function loss.

Transient declines in forced vital capacity (FVC) have long been documented following exercise in high concentrations of pollutants (eg ozone) (McDonnell *et al*. 1983). However, a small transient loss of lung function by itself should not be considered adverse. The US EPA (1989) has offered an arbitrary (and apparently unvalidated) classification of changes in the forced expiratory volume in 1 second (FEV_1). There is now agreement between the American Thoracic Society and the European Respiratory Society that changes in FEV_1 of >12% and >200mL are likely to be clinically important (Pellegrino *et al*. 2005), and it is likely that this definition will be adopted in the future in Australia (although the TSANZ has favoured a cut-off of 15% to indicate significant short-term changes). Smaller reductions in FEV_1 could still be of public health importance, but reductions of < 8% would fall within the error of measurement and so are unlikely to be detected or considered significant (Pellegrino *et al*. 2005). However the combination of a reversible loss of lung function with symptoms such as wheezing, as may be seen during an asthma attack, is clinically important and should be considered an adverse effect.

Air pollution may limit lung growth during childhood and adolescence (Gauderman *et al*. 2000) and may also increase the age-related decline in lung function (Tashkin *et al*. 1994). Decline in FEV_1 is associated with risk of developing chronic obstructive pulmonary disease (COPD) (Fletcher & Peto 1977), and with mortality from all causes as well as from respiratory disease (Stavem *et al*. 2005).

Clinical symptoms

These are the symptoms related to air pollution that are associated with change in clinical status or diminished quality of life. Air pollution can provoke new symptoms in people who are otherwise asymptomatic and apparently healthy, and can trigger symptoms in people with asthma, COPD or cardiac disease.

Minor symptoms include infrequent eye irritation, sneezing or cough.

However, wheezing in young adults with asthma is associated with reduced quality of life, and more frequent wheezing, particularly in association with sleep disturbance or symptoms on waking, is associated with severe asthma. Symptom frequency is one component of the classification of asthma severity developed by the National Asthma Council of Australia (NACA 2002). A progression of asthma from mild to moderate, or from moderate to severe, associated with air pollution should be considered an adverse effect.

Clinical outcomes

Air pollution may lead to a range of measurable clinical outcomes, including:

- population level effects such as:
  - emergency department visits for asthma (Johnston *et al*. 2002) and COPD;

- hospital admissions for acute bronchitis, pneumonia (Barnett *et al*. 2005), asthma and COPD (Simpson *et al*. 2005a; Voigt *et al*. 1998);
- primary care presentations (Australian data are extremely limited); and

- individual exacerbations of asthma or COPD including requirements for reliever and/or preventer medication.

Such increased usage of health services and medications incurs additional costs to both individuals and the community. It is reasonable to conclude that any clinically detectable effect of air pollution should be considered adverse.

Mortality

Any increase in mortality related to air pollution is of serious concern; however mortality can be a relatively insensitive endpoint for identifying health effects that, as recent epidemiological studies show, primarily result in increased morbidity.

The 1952 'killer fog' in London (Logan 1953) clearly established that death is a critical health effect of air pollution. Time series analyses of mortality data from four Australian cities (Simpson *et al*. 1997; Simpson *et al*. 2000; Simpson *et al*. 2005b) have demonstrated consistent effects of visibility-reducing particles on all-cause mortality, and of ozone on respiratory mortality. Methodological issues related to meta-analysis of such studies are discussed in Section 3.2.3.

In any association between exposure and mortality identified in short-term time series studies, the *extent* of life-shortening needs to be considered.

The American Thoracic Society position paper canvassed the possible phenomenon of 'harvesting'—the idea that there is a pool of frail elderly individuals whose death is advanced by a few days following an air pollution event. Careful statistical examination of this concern by Schwartz *et al*. (2000) suggests that harvesting on this short time scale probably does not occur to any significant extent. It does appear, however, that deaths may be advanced by a few months or more.

The WHO (2003) concluded that life expectancy was decreased by at least a few months as a result of long-term exposure to high particulate matter (PM) concentrations. It would appear that 'cumulative exposure of 1-2 months was actually more harmful (in terms of the daily number of deaths associated with them) than shorter exposures of similar magnitude' (Brunekreef & Hoek 2000).

A meta-analysis of data from four US cities also provided evidence that the association between PM_{10} and mortality was greater after longer time exposures (10 days to 2 months) than at time scales of a few days (Dominici *et al*. 2003). Only one Australian study has examined harvesting: Morgan *et al* (2003) used statistical methods similar to those used by Dominici *et al* to analyse mortality data from Sydney between 1994 and 2000. They concluded that 'the effects of particulate air pollution on mortality are larger at longer time scales (greater than one month) than at short time scales (1 to 4 days)'.

Population health versus individual risk

When it can be assumed that the relationships between risk factor and disease are causal, the American Thoracic Society considers that a shift in risk factor distribution, and hence

the risk profile of the exposed population, is adverse, even in the absence of the immediate occurrence of overt illness.

As levels of and exposure to air pollution rise, not only do more people experience adverse effects, which are potentially more severe, but in addition, individuals who were previously non-susceptible may be made more susceptible to other factors affecting respiratory function (eg allergens or viruses), and individuals with sub-clinical disease can be made symptomatic. For example, the responsiveness of sensitised asthmatics to an air-borne allergen (house dust mite) is enhanced by immediate prior exposure to NO_2 (Tunnicliffe *et al*. 1994).

Other potential adverse respiratory effects

Other potential health effects of air pollution include acute respiratory infections, chronic respiratory diseases, the degree of non-specific airways responsiveness, and lung cancer. Such diseases, however, have many other causes (eg allergens, cigarette smoking, occupational exposures). If study design and statistical analysis allow proper adjustment for confounders, derived exposure effect estimates may contribute to the process of setting air quality standards.

Non-respiratory effects

Adverse non-respiratory effects of air pollution are also documented. Changes in heart rate and rhythm have been documented following exposure to ozone (Rich *et al*. 2006), fine particles and other vehicle exhaust pollutants (Dockery *et al*. 2005; Holguin *et al*. 2003). However heart rate variability is at best a surrogate endpoint and there is no agreement as to what constitutes a clinically important change. Changes have also been observed in blood coagulability, viscosity and markers of systemic inflammation (Donaldson *et al*. 2001); but while these subclinical effects probably increase the risk of myocardial infarction and sudden cardiac death, further research is required.

Research on air toxics has also considered reproductive toxicity and developmental effects, reduced performance on neurobehavioural testing, and neuropsychological disease. These endpoints have not been used traditionally in standard setting. Individual papers reporting such effects in humans should be carefully evaluated in the context of comparable data from animal studies. Deliberations relating to cancer and sensory effects in standard setting are discussed in Sections 4 and 5.6.

Quality of life

Air pollution may have measurable negative effects on health-related quality of life, whether for people with chronic respiratory disease or the general population.

An increasing body of research on health outcomes has highlighted the importance of health-related quality of life. Quality of life includes the physical domain (eg symptoms, exercise capacity), the psychological domain and the socio-economic domain (social function). To measure quality of life, validated instruments that are sensitive to the impact of environmental factors have been developed. These include generic instruments such as the Short Form 36 (SF36) (Ware *et al*. 1994) and disease-specific instruments such as the Asthma Quality of Life Questionnaire (Marks *et al*. 1992), the Chronic Respiratory Questionnaire (Guyatt *et al*. 1987) and the St Georges Respiratory Questionnaire (Jones

1993). The minimum clinically important difference has been defined for each of these instruments (eg 4 points on the St Georges Respiratory Questionnaire). Where the exposure continues over a considerable duration, this concept can be extended to estimate Quality Adjusted Life Years (QALYs).

The issue of whether the more socio-economic aspects of quality of life (eg loss of amenity, negative impacts of odours) should be classified as adverse health effects is more controversial. While these matters should be considered in a standard-setting process, it is debatable whether the appropriate place is within the scientifically rigorous steps of risk assessment outlined in this guidance document, or during the consultative processes that accompany risk management.

Section 7 considers how quality of life issues might be included in standard setting.

2.1.3 Biomarkers and biomonitoring

The term 'biomonitoring' can be used to describe the measurement of biomarkers of exposure, biomarkers of effect, or biomarkers of susceptibility.

There are over 60 chemicals commonly monitored in the occupational environment. The reader is referred to the American Conference of Governmental Industrial Hygienists (ACGIH 2005) for a list of biological exposure indices (BEI). In the public health arena, there are only a few examples of biomarkers of environmental exposure being used in a regulatory setting, with blood lead level being the most well characterised (Sakai, 2000). Carboxyhaemoglobin is a specific biomarker of carbon monoxide exposure and is closely related to the mechanisms of carbon monoxide toxicity; its use as a biomarker is complicated by endogenous production and multiple environmental sources such as combustion sources, vehicle exhausts and tobacco smoking (WHO 2000a). Biomarkers for exposure to the other criteria pollutants are less well developed. There is considerable potential for the further development of biomarkers for those air toxics already having occupational health and safety standards such as BEIs.

Generally speaking, a validated biomarker of exposure can be readily adaptable to standard setting. Sensitive biomarkers of effect and susceptibility or surrogate endpoints are being developed as research tools, but there is an urgent need for their significance and relevance to adverse health effects to be investigated and understood before they can be used in standard setting. An additional problem is the difficulty faced in communicating the significance of any particular biomarker result to the public (Paustenbach & Galbraith 2006).

There is recent guidance on the use of biomarkers. Biomarkers of exposure may have a place in standard setting provided they are validated and meet the following criteria (WHO 2001):

- Sample collection and storage procedures are appropriate.
- Analytical methodology has been standardised and validated, and is specific for analytes of relevance.
- The biomarker is specific to the pollutant, or to a metabolite of the pollutant or a specific family of pollutants.

- The biomarker is able to be categorised as measuring exposure, as opposed to a measure of biological effect or susceptibility of an individual.
- The relationship between the presence of the biomarker and the relevant exposure period is well understood (ie there is knowledge of the persistence of the biomarker).
- The dose-response between the exposure level and presence of the biomarker is well characterised.
- There is a well characterised dose-response relationship between the biomarker and disease.

The European Centre for Ecotoxicology and Toxicology of Chemicals document *Guidance for the Interpretation of Biomonitoring Data* (ECETOC 2005) reviews the purpose and uses of biomonitoring data, offers guidance on interpretation, and suggests a framework for placing biomarker data into context.

Biomonitoring

Biomonitoring of the exposure to complex mixtures such as polluted ambient air, diesel exhaust or tobacco smoke is a particular challenge since these exposures have many constituents, and hence potential biomarkers, in common and many people are exposed to more than one of these mixtures. It is therefore problematic to demonstrate the specificity of some biomarkers for the pollutant of interest.

Nonetheless, there are significant biomonitoring programs underway. The National Health and Nutrition Examination Survey conducted by the US Centers for Disease Control and Prevention has now issued three reports (CDC 2005). The CDC's *National Report on Human Exposures to Environmental Chemicals,* and the Environment Protection Agency's *National Human Exposure Assessment Survey* (US EPA 2005d*)* were recently evaluated by the National Research Council (2006).

The integration of such biomonitoring programs with classical epidemiological and experimental toxicological studies should improve their utility in risk assessment (Angerer *et al* 2006).

Biomarkers of respiratory effect

The American Thoracic Society concluded that biomarkers of respiratory effects have not yet been validated sufficiently to enable a response to be used to define an adverse health effect from air pollution (ATS Board of Directors 2000). This current document takes a similar position on biomarkers of inflammatory effects. The field is, nevertheless, advancing rapidly, and biomarker changes related to air pollution can be detected at levels below those that produce clinically detectable disease (although not all such changes indicate potential disease or injury). Non-invasive measures of response will be increasingly used in clinical trials and epidemiological studies. Already the difference in fENO between asthmatic and healthy subjects, the response to anti-inflammatory therapy and its application to asthma management have been described (Kharitonov *et al*. 1994; Sandrini *et al*. 2003; Smith *et al*. 2005).

2.2 ASSOCIATION OR CAUSATION?

> '*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?' (Hill 1965)*

While it is not possible to prove absolutely that any exposure causes (or does not cause) a given disease, strong probability provides a compelling basis for action. The strength of the association between exposure and incidence ranges across a continuum, and 'proof' rests on an informed scientific interpretation of observations.

This interpretation is guided by the 'viewpoints' set out in 1965 by the British medical statistician Sir Austin Bradford Hill in his seminal paper on environmental causes of disease (Hill 1965). Bradford Hill's approach has been recently reappraised by Lucas and McMichael (2005) in the light of 40 years of epidemiological research.

Based on these and other publications, the following considerations are recommended for assessing potential cause-effect relationships:

(1) *Strength of association*: Strong associations are more likely to represent a causal relationship than weak associations. In epidemiological studies, particularly those undertaken in advanced industrialised nations where exposures are unlikely to be extreme, it is relatively unusual to identify very strong associations. In the case of weak associations, the difficulty of separating a true causal effect from the 'statistical noise' induced by imprecise information, uncontrolled biases and various forms of confounding often proves an insurmountable problem.

(2) *Consistency*: Causality is more likely if an observation has been made repeatedly in different settings, using perhaps different populations and study designs. The factors that may confound a relationship, however, may be the same in all observational studies.

(3) *Strength of study design*: Evidence from 'true experiments' is most compelling. For example, randomised controlled trials of exposure to gaseous pollutants have been performed with human volunteers. Often, however, such experiments are not feasible or ethical. It is then necessary to rely on weaker observational designs including (in descending order of preference): cohort studies, case-control studies, cross-sectional studies (surveys) and ecological studies.

(4) *Dose-response*: The data from observational studies can often be stratified according to the level of exposure. When the health effect appears greater amongst those with the higher levels of exposure (ie there is an apparent dose-response gradient), this may be a pointer to causality. In some settings, however, the higher the exposure, the higher are the levels of other confounders. For example, in early studies of the health effects of air pollution, it was noted that sulfur dioxide levels typically varied in accordance with particle levels, since both were partly derived from burning coal.

(5) *Temporality*: Exposure to the environmental cause must precede the development of disease. The correct temporal sequence can only be reliably established by cohort studies and randomised controlled trials.

(6) *Specificity*: One exposure should give rise to only one outcome. Whilst this requirement is satisfied for many infectious agents, it rarely applies to other environmental exposures (Lucas & McMichael 2005). This suggests that either a specific exposure, a specific outcome, or an outcome that only occurs under environmental conditions where genetic susceptibility is important, would be sufficient for the relationship to be considered causal.

(7) *Biological plausibility*: Arguably this is the most important consideration in assessing causation. When an observational study provides information that is in keeping with expectations from animal or *in vitro* research, the acceptability of claims of a causal association are considerably greater. Such research may cover a broad range, including animal toxicology and human volunteer studies.

(8) *Coherence*: Temporal patterns of exposure must fit with the observed pattern of disease. The hypothesis that fewer childhood infections are causing the rising prevalence of asthma (the popular 'hygiene hypothesis') is an example of such an association (Lucas & McMichael 2005).

(9) *Analogy*: While this is probably the least important consideration, the case for causation is strengthened if there is similarity to a previously established relationship. For example, it is plausible that diesel particle pollution could cause lung cancer because it contains many of the same polycyclic aromatic hydrocarbons as cigarette smoke.

While not generally regarded as one of the classical Bradford Hill viewpoints, the issue of *confounding* is integral to assessing causation. A confounder is any factor associated with the exposure of interest (in this case, air pollution) that is itself a determinant of the outcome of interest (for example, COPD). A good example of a confounder would be cigarette smoking. In many study designs, unless reliable and valid smoking data are collected and allowed for in an analysis, it is not possible to disentangle the effects of air pollution on COPD. The method employed in time series studies avoids having to control for confounders that do not vary over time.

Epidemiological studies may be conducted in relatively 'clean' settings where a single exposure dominates, or in 'dirty' settings where multiple exposures co-exist. In the latter, it is often difficult to separate the contribution of each exposure. In general, evidence of causality is more commonly derived from the studies where the influence of the potentially causal factor can be isolated, either because the study population has been exposed specifically to that agent, or because the potential confounders have been well measured and their effects allowed for adequately in the analysis.

The assessment of cause-effect relationships according to the approach described above will be strengthened by the application of the results of sensitivity analysis (WHO 2000c). If the introduction, deletion or adjustment of key variables in the dataset, or the introduction of other plausible explanatory factors, results in a significant change in the outcome of the analysis, there may be grounds for re-assessing the viability of the proposed model.

Specific guidance—causal criteria

The nine Bradford Hill viewpoints for assessing causal associations remain valid and should be used to assist assignment of weight of evidence (WoE) to data sets used to set air quality standards. The possibility of confounding needs to be assessed.

Should air quality standards be set on the basis of a certain level of causal evidence, and how should such a level be determined?

Choice of substance for standard setting should be based on a wide range of health and social / political criteria. The level of causal evidence used to set the quantitative aspects of a standard (ie the numerical value and averaging time) will vary from substance to substance, but will be driven by the WoE analysis, the quality and quantity of information, and quantitative interrogation of the dose-response relationships. If data are not amenable to quantitative dose-response analysis, they are not suitable for standard setting. Some quantitative descriptions of 'points of departure' on dose-response curves have already been established for standard setting in Australia (eg $m\text{BMD}_{05}$, NOAEL, LOAEL).

The question of what level of causal evidence from epidemiological studies should be used can be paraphrased as, 'what odds ratio or relative risk should be taken as evidence of cause and effect?' This is an inappropriate question, as the recommended approach for determining causality is application of WoE analysis to a systematic review of the available database and evaluation of all the data according to the Bradford Hill viewpoints.

How should the existence of associations between health effects and pollutants be considered?

The recommended approach is WoE analysis of the complete data sets and evaluation according to the Bradford Hill philosophy. In some situations it may be prudent and/or pragmatic to adopt the evaluations of competent overseas authorities for some aspects of the standard-setting process.

3 SELECTION OF STUDIES IDENTIFYING HEALTH EFFECTS ASSOCIATED WITH AIR POLLUTANTS

The quality and relevance of the studies used as a basis for setting health-based air quality standards is critically important. Data should meet the weight-of-evidence criteria outlined in Section 3.4, and, most importantly, should be robust to enable dose-response modelling (see Section 5.5).

Experimental or epidemiological studies based on observations in humans are likely to be more relevant than animal studies. Animal studies may nevertheless provide qualitative and quantitative (dose-response) information to complement the human data. Only if suitable human data are not available should animal data be the key element for determining the relationship between exposure to an air pollutant and a resultant adverse health effect.

While animal studies are limited in their ability to measure some adverse health effects relevant to humans (see Section 3.3.2), they may be the most suitable for determining adverse health endpoints that:

- have a delayed response or long latency (eg cancer, birth defects); or
- are manifested as systemic toxic effects that are most readily seen by histological examination of autopsy tissues; or
- have not been looked for in epidemiological studies (bearing in mind the limited ability of epidemiology studies to identify some chemical hazards).

The default position is that effects observed in animal studies are relevant for humans unless there is cogent scientific argument or data to the contrary. The absence of human data does not necessarily preclude the ability to establish a standard.

3.1 DEALING WITH APPARENTLY CONFLICTING DATA

In some instances, available data for a dose-response relationship may differ by a wide margin. For example, Jones-Otazo *et al*. (2005) commented on the wide range of risk assessment outcomes for carcinogens that have been promulgated by six international agencies, based on the selection of different exposure paradigms and toxicological reference values. Lewandowski and Rhomberg (2005), reviewing the epidemiological and animal bioassay literature on cancer risk assessments for trichloroethylene, found the cancer potency estimates varied by over 20-fold.

The following steps are suggested to assess the validity and usefulness of individual studies as a means of addressing such conflicts:

(1) *Evaluate studies for internal validity*: the adequacy of study design and the extent to which it has validly measured what it intends or purports to measure.

(2) *Evaluate studies for external validity*: can the results be validly generalised, extrapolated or transferred to other settings?

(3) *Evaluate corroboration, contradiction and plausibility:* the Bradford Hill viewpoints may be useful here (see Section 2.2).

(4) *Make a choice*: select the study or studies that best represent the endpoint of most relevance for setting an air quality standard.

Table 3.1 provides further guidance for resolving data conflicts.

Table 3.1: Summary of criteria to be used in selecting human (experimental and epidemiological) and animal studies
Adapted from Lewandoski and Rhomberg (2005)

Human/epidemiological studies	Animal studies
Was the study an analytic study (cohort or case-control) or an ecologic (time series) study? Did the study address potential confounders or the effects of chance and bias? Was the statistical power of the study sufficient to detect effects if they occurred? Was the characterisation of exposure adequate? Could exposure have been misclassified? Was the duration of exposure long enough to assess potential carcinogenicity? Might biases have been introduced through incomplete follow up? Was the study sufficiently rigorous to support its purported outcomes? Were relevant outcomes measured?	Were an appropriate number and spread of doses used? Was the maximum tolerated dose reached or exceeded? Was the validity and/or purity of the test material specified? Were there potentially compromising diseases or pathogens within the test animal groups? Were the toxic endpoints well characterised? Were the dosing regimens well characterised in terms of stability in the dosing material and/or measurement of consumption? Was pathological assessment undertaken and, if so, was there a transparent method of grading the observed changes? Were the record-keeping, animal husbandry and clinical observation adequately rigorous, and did the study conform to Good Laboratory Practice or some other form of quality assurance?

3.2 HUMAN STUDIES

3.2.1 Experimental studies

Experimental studies of environmental air pollutants in humans are typically conducted in controlled environmental chambers, with the ability to add known concentrations of the agent(s) in question. In contrast to older style versions, modern chambers allow for adequate mixing and on-line monitoring of gas concentrations, and provide the capacity to vary pressure, temperature and humidity. Susceptible subjects can be studied under close clinical supervision and a range of parameters can be measured, including lung function and other indicators of human health effects. Importantly, mixtures of gases can be studied.

The quality of scientific information provided by adequately conducted chamber studies is generally higher than that from other forms of human studies. Such studies have some strengths but also severe limitations. Controlled clinical experiments of defined population

sectors may be valuable in refining acute dose-response analyses and developing population-specific adjustment factors. However, given the ethical constraints that limit the extent to which humans can be deliberately exposed to toxic chemicals, experimental chamber studies using humans are suitable for assessing only relatively mild, reversible and early-onset effects; this necessarily excludes direct study of the induction of all chronic diseases that generally have major impact on the affected individuals.

Furthermore, the study of reversible effects that require prolonged (eg seasonal) exposure of subjects is not practical. Perhaps the major limitation of clinical chamber studies is that only a relatively small number of individuals can be studied, and the statistical limitations this imposes make it difficult to study exposures that produce small or imprecisely measured effects. Another limitation is that, while some mixtures of pollutants can be approximated in the gas mix delivered into the chamber, the true impact of ambient exposure to many complex mixtures can not be achieved in such studies. A further weakness is that in order to decrease the heterogeneity of responses and increase internal validity, very homogeneous subject groups are often selected for study, raising questions about the ability to generalise findings to other segments of the population not represented by these samples (McDonnell 1993).

Specific guidance—use of chamber studies

Human studies involving controlled exposures, typically in an exposure chamber, may provide relevant and high quality hazard assessment information. Advantages of these studies include the following, although these advantages can also have a negative aspect, as discussed above:

- availability of data from single exposure at a precisely known concentration, so there is no confounding of the effect of one exposure with the effects of another (a common problem in epidemiology);
- ability to obtain precise information concerning dose-response;
- ability to measure earliest evidence of adverse effects (eg with pulmonary function testing);
- ability to study responses of specific sub-populations, including those likely to be most sensitive to the exposure; and
- ability to study adverse health effects during a full-range of human activities (eg during exercise, during varying levels of temperature and humidity).

3.2.2 Epidemiological studies

Since the early observations of increased cardiopulmonary mortality and morbidity during major air pollution episodes in the 1950s, epidemiological investigations have become more sophisticated, developing statistical modelling that enables daily mortality and morbidity counts to be linked to changing levels of air pollution on the same or recent days. This development led to recognition that significant health effects are associated with air pollution even at the lower end of exposure. Since the early 1990s, epidemiology's

ability to study long-term (cohort) and short-term health outcomes (time-series, panel studies) has provided a significant scientific tool in standard setting.

Positives attributes of epidemiology include the ability to:

- study large populations and extrapolate the results to the general population;
- carry out studies under realistic exposure scenarios that are relevant to the population in question;
- quantify mortality and morbidity;
- study a wide range of significant diseases in the general population and in sensitive sub-populations;
- study the health effects of air pollution without the ethical constraints placed on human experimental studies; and
- study the effect of exposures on health outcomes, such as quality of life, for socio-economically deprived populations with pre-existing poorer health status.

As mentioned, epidemiological studies focusing on environmental factors have been used extensively in setting air quality standards, and their principal advantage derives from their direct relevance to human health. They do, however, have limitations, including:

- misclassification of exposure and outcomes, leading to imprecise risk estimates;
- in large populations, substantial effects in small numbers of sensitive individuals may be swamped by a lack of effect in the majority; and
- confounding by closely linked co-exposures.

Epidemiology has its strengths in uncovering effects at the population level, but lacks the ability to study underlying biological disease processes. Other areas of scientific expertise are necessary to support epidemiological findings by teasing out biological pathways using human and animal experimental studies.

The use of epidemiological studies in standard setting is considered in Section 5.7.

3.2.3 Role of meta-analysis

Meta-analysis aims to provide a more precise effect estimate by pooling results across studies.

Where results are relatively homogeneous across studies despite differing study locations, study designs, data collection or analysis, then meta-analysis affords the potential for greater generalisability of results.

Where results are heterogeneous, meta-analysis allows this heterogeneity to be explored, though it does not necessarily afford potential for greater generalisability.

3.2.4 Methodological requirements for meta-analysis

Selection of studies for meta-analysis

Studies to be included in a meta-analysis must meet stringent convergence criteria.

Differences between studies that could influence results should be minimised, otherwise heterogeneity of results across studies may be due to study characteristics rather than truly differing effects of pollutants. Individual studies should, so far as possible, use the same study design, same data collection methods and same statistical model-building protocols and should report the same effect measures.

Differences in air pollutant effect estimates across cities may also be due to population demographic differences, and to the type and magnitude of ambient pollutant exposure levels, which may be due in part to differences in traffic congestion and topography.

The fixed effect model assumes that there is a single common effect of air pollution upon mortality or other outcomes. Random effects models explicitly allow for differences between studies and may be more appropriate if there is heterogeneity. The confidence intervals surrounding pooled effect estimates are typically wider for a random effects model than a fixed effect model.

General criteria for selecting individual time series studies

Time series studies included in a meta-analysis should use:

- pollutant measurements that are of adequate quality and reliability, and reflect the actual exposure in the population studied;
- valid and reliable health outcome measurements;
- statistical analyses (eg generalised additive models[4]) that control for long-term time trends, seasonality, day of week effects and other known confounders (eg temperature, humidity); and
- pollutant effect estimates that are robust to different statistical model specifications.

Effect estimates can also depend on the lag period chosen between pollutant level exposure and health outcome. For example, one could assess hospital admissions associated with pollutant level on the same day (lag 0), the day before (lag 1), two days before (lag 2) and so on. The use of a common lag period across studies in a multi-city analysis reduces criticism of *post-hoc* selection of lag, but also might introduce heterogeneity if lagged pollutant effects do truly vary across cities.

[4] Generalised additive modelling (GAM) must include recommended improvements to default convergence criteria of the estimation algorithms in software. Dominici *et al* (2002) noted that default criteria for declaring a convergent solution of the iterative estimation algorithms were too liberal (eg as in the GAM module of the statistical package SPlus version 3.4 or earlier). Simulations indicated that estimates using default criteria were biased upwards towards *overstatement* of effect of pollutants and with confidence intervals that were too narrow. Many analyses worldwide have been re-done with this more stringent convergence checking.

Specific guidance—meta-analysis

In what circumstances should risk estimates derived from meta-analysis / multi-city-analysis (analyses) be recommended for further use in risk characterisation, and under what circumstances should city-specific risk estimates be utilised? If meta-analysis / multi-city-analysis is considered appropriate, what methodological approaches for meta-analysis / multi-city-analysis should be used, and what criteria should be used for accepting or rejecting particular meta-analysis / multi-city analyses?

Risk estimates derived from meta-analyses of observational studies are useful only if the individual studies are of uniformly high quality. Otherwise the results of a meta-analysis of observational studies are often misleading.

The generalisability of epidemiological data relating to air-quality and health is dependent on the similarity of exposure. Where the epidemiological data are of high quality and the nature of the exposure is similar, then it is reasonable to extrapolate the results from one site to another.

Use of overseas and local meta-analyses

Australian exposures to ambient pollutants are typically at lower levels than in the USA or European cities, with a different mix of pollutants. Australians may, however, spend more time outdoors and therefore be exposed to ambient pollutants for longer periods.

The principal meta-analyses of overseas time series studies are:

- the *National Mortality, Morbidity and Air Pollution Study* of 90 cities in the USA (Samet *et al*. 2000);
- *Air Pollution and Health: a European Approach* (APHEA2) of 29 cities in Europe (Katsouyanni *et al*. 2002); and
- a meta-analysis of eight Canadian studies (Burnett *et al*. 2000).

Locally, Simpson *et al*. (2005b) followed the APHEA2 protocol in Brisbane, Melbourne, Perth and Sydney, with data from 1996-1999. Advantages of this study included a common protocol for all four cities, common exposure and outcome measures, and common statistical analysis methods, including specification of a common exposure lag (average levels over day 0 and day 1). The exposed population comprised 54% of the total Australian population. The study found no significant heterogeneity of mortality results across cities, but some for differences for hospital admissions. However, with only four cities, power to detect heterogeneity may have been low.

Studies in USA and Europe have found significant heterogeneity between cities. In Australia, despite similar geographical distances between study locations, there is less variation between cities in air pollution levels, air pollution composition, geographic / topographic characteristics, and population demographic and health profiles. Consequently these factors are less likely to be reasons for differences in results between cities, and it may be appropriate to use combined estimates from Australian studies.

Specific guidance—use of overseas data

For Australian standard setting, under what circumstances would it be acceptable, if at all, to use risk estimates from an overseas meta-analysis?

> Only meta-analyses of high quality observational studies would be acceptable for Australian standard setting. Such meta-analyses are relatively uncommon. More reliable data is likely to be derived from a qualitative review of high quality 'pivotal' studies.

For Australian standard setting, in what circumstances, would it be acceptable, if at all, to use Australian city-specific hazard assessment data?

> Standard setting typically involves consideration of multiple sources of data. High quality Australian data should be included amongst these data. However Australian data that are not of high quality should not take precedence over well conducted studies from elsewhere (See Table 3.1 regarding selection of studies).

Where Australian meta-analysis / multi-city-analysis data are available, what requirements need to be satisfied in order to recommend the use, in risk assessment, of the pooled risk estimates instead of the city-specific risk estimates for risk assessment?

> Data from a specific city might be used to establish city-specific risk estimates if they are of sufficiently high quality. As stated above, local data of limited quality should rarely take precedence over well conducted studies from elsewhere. Extrapolation of data from elsewhere requires that the nature of the exposure and characteristic of the population are similar. Pooled estimates are only reliable if there is homogeneity among the individual studies, and the exposure and effect measurements are valid.

How to read and critically appraise meta-analyses

In recent years, meta-analysis has been most commonly applied to systematic reviews of randomised controlled trials. Guidelines for the conduct of systematic reviews and meta-analysis published by the Cochrane Collaboration (Higgins & Green 2005) are directly relevant to short-term randomised controlled trials of gaseous pollutants in human volunteers. The reader of such systematic reviews is advised by the Evidence Based Medicine Working Group (Oxman *et al*. 1994) to answer the following series of questions:

Are the results valid?

1. Does the meta-analysis address a focussed clinical question?
2. Were the criteria used to select studies for inclusion appropriate?
3. Is it unlikely that important relevant studies were missed?
4. Was the validity of the included studies appraised?
5. Were assessments of studies reproducible?
6. Were the results similar from study to study?

What are the results?

1. Was a quantitative synthesis (meta-analysis) undertaken?

2. How were individual studies weighted?
3. What outcome measures were reported?
4. How precise were the results? Were 95% confidence intervals reported?

Will the results help in caring for patients?

1. Can the results be applied to your patients?
2. Were all clinically important outcomes considered?
3. Are the benefits worth the harms and costs?[5]

While the equivalent questions for meta-analyses of observational epidemiological studies have not been finalised, most of the above questions are directly applicable. To be valid, a meta-analysis needs to address a focussed public health or policy question. Appropriate data sets need to be included. It is particularly difficult for the reader to be confident that important relevant studies were not missed. Many of the original studies of the health effects of air pollution are published in government reports or conference proceedings rather than the peer reviewed literature, and thus cannot be identified by reviewers searching Medline and other bibliographic databases.

Separate meta-analyses should be performed for different study designs (time series analyses, case-control studies, case-crossover studies, cohort studies, randomised controlled trials), as even the summary measures of association (odds ratios, relative risks etc) across these study designs are different and cannot be meaningfully combined.

Time series analyses tend to be rated low in their potential to establish causation, as they are often considered by epidemiologists to be ecological designs. However their findings are unlikely to be seriously confounded by individual factors such as age, gender and socio-economic status, which do not vary on a day-to-day basis. A formal statistical test of heterogeneity is important in meta-analysis. If significant heterogeneity is demonstrated, further investigation (eg sensitivity analysis) should be undertaken to see if the source of the heterogeneity can be identified. As with the results of randomised controlled trials, some measure of precision (eg a 95% confidence interval) is more informative than a bland 'p' value.

In applying the findings of meta-analysis to standard setting, it is important that the populations studied are comparable to those for whom the standard is intended. For example, the prevalence of asthma in Australia is amongst the highest in the world and certainly much higher than in North America or Europe (Burney *et al*. 1996; The International Study of Asthma and Allergies in Childhood Steering Committee 1998). Thus, at least for pollutants to which people with asthma are more susceptible, it may be appropriate to give greater weight to the findings of local meta-analyses (Simpson *et al*. 2005b) than those conducted overseas.

Some further guidance on the interpretation of meta-analyses in environmental health risk assessment has been provided by the WHO (2000b). A WHO working group chaired by Professor Ross Anderson made the following recommendations:

- The meta-analysis should have a protocol that specifies the objectives and methods.

[5] It would be highly unusual for air pollution to have any beneficial effects; however its control certainly entails costs. This lies more in the province of risk management than health risk assessment.

- Inclusion criteria for studies should be broad rather than narrow.
- Characteristics of primary studies should be assessed qualitatively rather than using a global quality score.
- Meta-analysis can be performed by inverse variance weighting or random effects models.
- The effects of publication bias should be assessed by sensitivity analysis (graphical techniques such as the funnel plot may also be appropriate).
- Overall heterogeneity should be assessed as this may identify susceptible groups and exposure conditions.
- Meta-analytical methods that may be used to compare studies include stratified analysis and meta-regression.
- Sensitivity analysis should be performed to assess the robustness of summary estimates to the inclusion and exclusion of particular studies.
- Quantitative summary estimates provide useful input to health impact assessment.

3.3 ANIMAL STUDIES

Studies in animals should not be the primary basis for setting air quality standards where there are studies in humans (chamber exposures or epidemiological studies) that clearly define relevant concentration-effect relationships for the critical endpoint. They can, however, provide information on biological mode(s) of action and biological plausibility; for example, it is anticipated that animal toxicity data on particulates with different characterising parameters and potential modes of action will inform on the need for refinement of particulate standards in Australia[6].

The methodology, interpretation and extrapolation of inhalational toxicology studies in animals are discussed in the comprehensive monograph edited by Salem and Katz (2006).

An understanding the relative advantages and limitations inherent in the design of animal studies informs their use in standard setting to complement relevant studies in humans.

3.3.1 Advantages

The major advantage of animal studies is the ability to control exposures. Studies are usually designed such that the highest exposure results in measurable adverse effects. While ethical constraints limit the extent to which humans can be deliberately exposed to toxic chemicals, it is possible to design ethically acceptable animal studies that allow valid adverse effect data to be collected.

In addition, adverse health effects can be monitored through a combination of in-life observation and measurements (blood chemistry, haematology, urinalysis, lung function etc) and necropsy examination of tissues. The latter measurements are particularly

[6] For example, whether an urban $PM_{10/2.5}$ standard should be different from a $PM_{10/2.5}$ standard for rural or sparsely populated areas.

important to identify target organs for pathological damage (eg liver, brain) and are of critical importance if cancer is a suspected endpoint (see Section 4).

3.3.2 Limitations

The most obvious limitation in using animal studies is the uncertainty in extrapolating the findings to humans. This is generally managed by the use of uncertainty factors (or 'safety factors'). The default values for uncertainty factors take into account potential inter- and intra-species variations in sensitivity. The use of such uncertainty factors generally adds a degree of conservatism to the extrapolated data. Uncertainty may relate to:

- the route via which the chemical enters the animal's system: Studies are often designed to expose animals (usually rodents) to controlled concentrations in exposure chambers for significant periods of time (including lifetime exposures). Where whole-body exposures occur, there may be some uncertainty about the relative extent to which systemic absorption occurs via inhalation versus ingestion, particularly for particulates, where feed contamination and ingestion via grooming may confound the analysis. Studies may be designed with nose-only exposure, using suitable air delivery systems, but the degree to which the animals must be restrained generally limits the duration and frequency of the exposures;
- differences in respiration rate, tidal volumes, anatomy and oronasal breathing patterns; and
- the extent to which deposition on the surfaces of the delivery and exposure equipment may reduce the actual exposure concentrations. These same problems apply to chamber exposure studies with humans, and they may be overcome by appropriate air sampling and monitoring.

Unless the study includes very careful observation, it may be difficult to assess sensory deficits and adverse effects on behaviour. Headache, nausea and mucous membrane irritation (eye and respiratory passages) are likely to be of particular significance in humans exposed to noxious air-borne pollutants, but they are difficult, if not impossible, to assess in animal studies.

Animal studies usually employ very high exposures where even the lowest level exceeds concentrations likely to be encountered in ambient air. Various mathematical approaches have been used to extrapolate dose response downward from the experimental range. Care must be taken to assess whether effects observed at high doses in animals are relevant to lower exposures. There are many instances in the literature where specific mechanisms of toxicity only become operational at high doses. A breakpoint in the dose-response curve is usually a clue to the possibility that the response seen in animals may be concentration-dependent.

3.4 Weight-of-Evidence Criteria

Where a range of evidence is available to support a standard-setting process, it is important to be clear about the relative weights to be given to the different data sources, and how this weighting is achieved.

Weed (2005) noted that there is no clear or universally accepted definition of what constitutes a weight-of-evidence (WoE) approach, and pointed out that WoE can be taken to include processes where the approach is:

- *metaphorical*: where it is implied that a collection of data or studies has been taken into consideration, but there is no clear indication of what type of weighting has been applied, or whether there has been any objective analysis;
- *methodological*: where systematic interpretative analyses have been applied to a complete or nearly complete data set, such as systematic narrative reviews, meta-analysis or some other careful review of the quality criteria of the studies included; and
- *theoretical*: where the term is simply applied as a conceptual framework.

The methodological approach, while it is likely to differ for different pollutants due to the varying quality and quantity of data available, should yield the most defensible outcomes in any standard-setting process. Most meta-analyses should fall into this category, provided they satisfy the caveats outlined in Section 3.2.3.

Weed (2005) also noted that, of 276 papers published from 1994 to 2004 that purported to apply a WoE analysis, about half could be classified as simply applying a metaphorical meaning to the WoE concept, with no elaboration of how the concept was applied.

Any study that professes to include WoE to underscore its outcomes should therefore be carefully examined for the real meaning of the claim.

The extent to which 'expert judgement' has been incorporated into a claimed WoE approach should also be considered. While Weed (2005) acknowledged that judgement is 'a kind of intellectual glue cementing together the evidence and methods', it can be a relatively vague concept that may hide potential biases and preconceived ideas. It is important to be clear who is responsible for such judgement, and the basis on which it has been formed.

General guidance on how to carry out a systematic review of scientific literature has been summarised in the context of preparing clinical practice guidelines (NHMRC 2000) and some of these principles will be applicable to reviewing environmental health data from a WoE perspective. Intuitive WoE rankings have been commonly applied in the clinical trial literature, with randomised trials given the highest weight in any hierarchy, and case reports or expert opinion appearing the lowest. However, application of a parallel hierarchy to the types of studies used to support the setting of air quality standards in Australia is likely to be more difficult. It may even be counter-productive to try to reduce the features of a dataset to a single measure or 'quality score', since individual studies may influence the overall quality of the dataset in different ways and to varying degrees.

It is preferable to assess the characteristics of primary studies on an individual basis (WHO 2000c). WHO has developed guidelines for the assessment and use of epidemiological studies in health impact assessment for air pollution, and these provide useful guidance on how such evaluations should be conducted (WHO, 2000c). These guidelines have been used in evaluating the literature as part of the recent review of the WHO air quality guidelines (WHO 2006). The US EPA provide similar guidance in the criteria documents prepared as part of the development of air quality standards in the US (US EPA, 2004a [PM], 2004b [ozone]). These documents, together with the NHMRC document, provide

useful references to assist in the evaluation of epidemiological literature for use in the setting of air quality standards in Australia.

Application of the WoE approach, and its place in integrating data from experimental and epidemiological studies, is discussed in Section 5.1.

Specific Guidance—Weight of Evidence

Which general principles, criteria and definitions should be used for a WoE methodology?

> The following is a useful working definition of WoE for use in setting Australian air quality standards:
>
> WoE consists of considerations in assessing the interpretation of published information about epidemiology and toxicity of chemicals. It involves evaluation of quality of testing methods, size and power of study design, consistency of results across studies, and biological plausibility of exposure-response relationships and statistical associations.

Should the WoE conclusion be based on (1) criteria scoring, (2) an explicit panel decision, (3) other defined scientifically valid process, or (4) some combination of (1), (2) and (3)?

> Wherever possible, WoE analysis should be objective and systematic. Criteria scoring and other analysis tools will aid this objective, but there are no 'off-the-shelf' scoring systems that could be applied. Probably the most expedient way to develop a generic tool for Australia would be to assess a few data sets with quantitative WoE analysis in mind. Panel decisions, backed up by an appropriate consultative mechanism (eg peer review), are a means of bringing a consensus view to determining the WoE to be applied to a particular data set, but it may be a lengthy process and consensus may be difficult to reach. The preparation of high quality technical drafts produced by a technically competent person or small team and released to stakeholders for comment within a defined timeframe is the favoured model. Points of technical reconciliation could be put to a committee if needed. Ultimately, a case-by-case consideration of the available data is needed, using a combination of expert advice and rational analysis (option 4).

The role and composition of technical or expert panels (as described above), which would deal primarily with the hazard identification, study selection, WoE assessment and DRM components, will be determined on a case-by-case basis. Such panels should include a range of scientific disciplines appropriate to the standard under consideration. The specific processes of consultation with stakeholders are likely to be detailed and/or mandated by legislation, and may occur at many stages of the overall standard-setting process. It is imperative, however, that they include an adequate mechanism for peer review. Dialogue between risk assessors, risk managers, analytical chemists and other stakeholders will be most effective if it occurs at times when it can inform the relevant stage of the standard-setting process. For example, technical input into what air pollutants can be measured and over what time periods may be critical to selecting those studies that will carry most weight in a WoE analysis.

3.5 SELECTING STUDIES IN THE AUSTRALIAN CONTEXT

The type of study likely to be of most relevance to a standard-setting process will depend on the nature of the pivotal health effects, as well as the proposed use of the standard (eg whether it aims to manage ambient air exposures or exposures associated with point sources of toxic air pollutants).

Studies that may form the basis for standard-setting could include:

- well controlled experimental studies in humans (normal subjects and those with disease pre-disposing them to increased susceptibility) with single or mixed chemical exposures (chamber studies), noting that such studies would only be ethically acceptable where a relatively mild adverse effect is monitored (eg mucous membrane irritation, mild respiratory impairment);
- well conducted individual epidemiological studies, acknowledging the caveats outlined in Section 3.2.2, which suggest that such studies would be likely to have limited power in detecting subtle adverse effects;
- animal studies where the primary adverse effect is systemic toxicity that may not be readily detected in experimental human or epidemiological studies (eg cancer, reproductive or developmental toxicity, frank neurotoxicity);
- meta-analyses of observational epidemiological studies, where some type of grading of exposure categories has been analysed, and the potential impact of publication bias can be addressed.

The selection of studies to be used in a standard-setting process will also involve making judgements about:

- the WoE of the available data;
- the severity of the effects; and
- the latency period between exposure and onset of effect.

The usefulness of studies for standards setting is assessed on an individual basis and will depend on the nature of the exposure, the adverse health effect and the availability of data. For example, human experimental studies, by virtue of the available controls over exposure and measurement of clinical outcomes, would be seen as very useful when assessing the health effects of mild irritants. However, such studies would be expected to be less useful, or not useful at all, where the effect is chronic or of delayed onset, or where the consequences are dire (eg cancer, birth defects). In these circumstances, the results of animal experimentation will take on greater prominence. The usefulness of meta-analyses will be enhanced where the meta-analysis is limited to include only well conducted and locally relevant cohort studies rather than time-series analysis.

4 ASSESSMENT OF POTENTIAL CARCINOGENS

While epidemiological studies may have sufficient power to identify air pollutants capable of increasing cancer risk in humans, it is generally difficult to establish dose-response relationships.

This is not always the case; for example, studies on occupational exposures to benzene (NICNAS 2001), asbestos and vinyl chloride monomer have enabled dose-response relationships to be sufficiently well defined for them to be used for setting health-based standards relevant to air-borne exposures.

For most air-borne, potentially carcinogenic chemicals in a mixture, however, quantitative risk assessment will rely on extrapolation of data for single substances from controlled animal studies. The duration of such studies should be as long as possible. Studies that use lifetime exposure by the inhalational route are likely to yield the most relevant data for defining a dose-response relationship. Inhalational studies of a shorter duration may provide concentration-effect relationships suitable for quantitative risk assessment, but such studies would need adjustment for less-than-lifetime exposure (Kimmel 2002).

For many chemicals of interest, the only lifetime exposure studies available may be those using the oral route. Such studies, when combined with an appropriate battery of *in vitro* and *in vivo* studies of genotoxicity, can be useful for making a qualitative judgement on whether a chemical has carcinogenic potential. However, these studies generally use doses that approach the maximum tolerated dose, which may increase cancer incidence by non-genotoxic mechanisms, and their relevance for quantitative risk assessment at much lower doses may be questionable. Extrapolation from oral cancer studies for setting inhalation standards involves many uncertainties and is not recommended unless tools such as physiologically based pharmacokinetic (PBPK) dose-response modelling and quantitative analysis of target tissue dose with effect can be applied.

The WoE assessment of whether an air pollutant represents a carcinogenic risk for humans should include an assessment of the relevance to humans of the proposed carcinogenic mode of action. A suitable model for this process is provided in a risk assessment project conducted by the International Life Sciences Institute as described by Meek *et al*. (2003), which expanded on earlier advice from the US EPA and the IPCS on the use of mode-of-action information.

Evaluation of the human relevance of animal carcinogenesis data involves four steps:

1. Establish whether there is sufficient weight of evidence to establish a mode of action for the observed carcinogenic response in animals.
2. Establish whether the key events in the animal mode of action are plausible in humans.
3. Establish whether the key events in the animal mode of action are plausible in humans, taking into account kinetic and dynamic factors.
4. Provide a statement of confidence about the analysis and implications for human risk assessment.

Assessment of the mode of action includes consideration of:

- evidence for a genotoxic or non-genotoxic mechanism;
- dose-response relationships;
- temporal association of key events leading to cancer;
- strength, consistency and specificity of the relationships between the key events;
- biological plausibility and coherence of key events;
- consideration of alternative mode(s) of action; and
- uncertainties and inconsistencies in the data and gaps in knowledge.

Application of this framework in relation to six model carcinogens resulted in the development of a decision-tree of a type depicted in Figure 4.1.

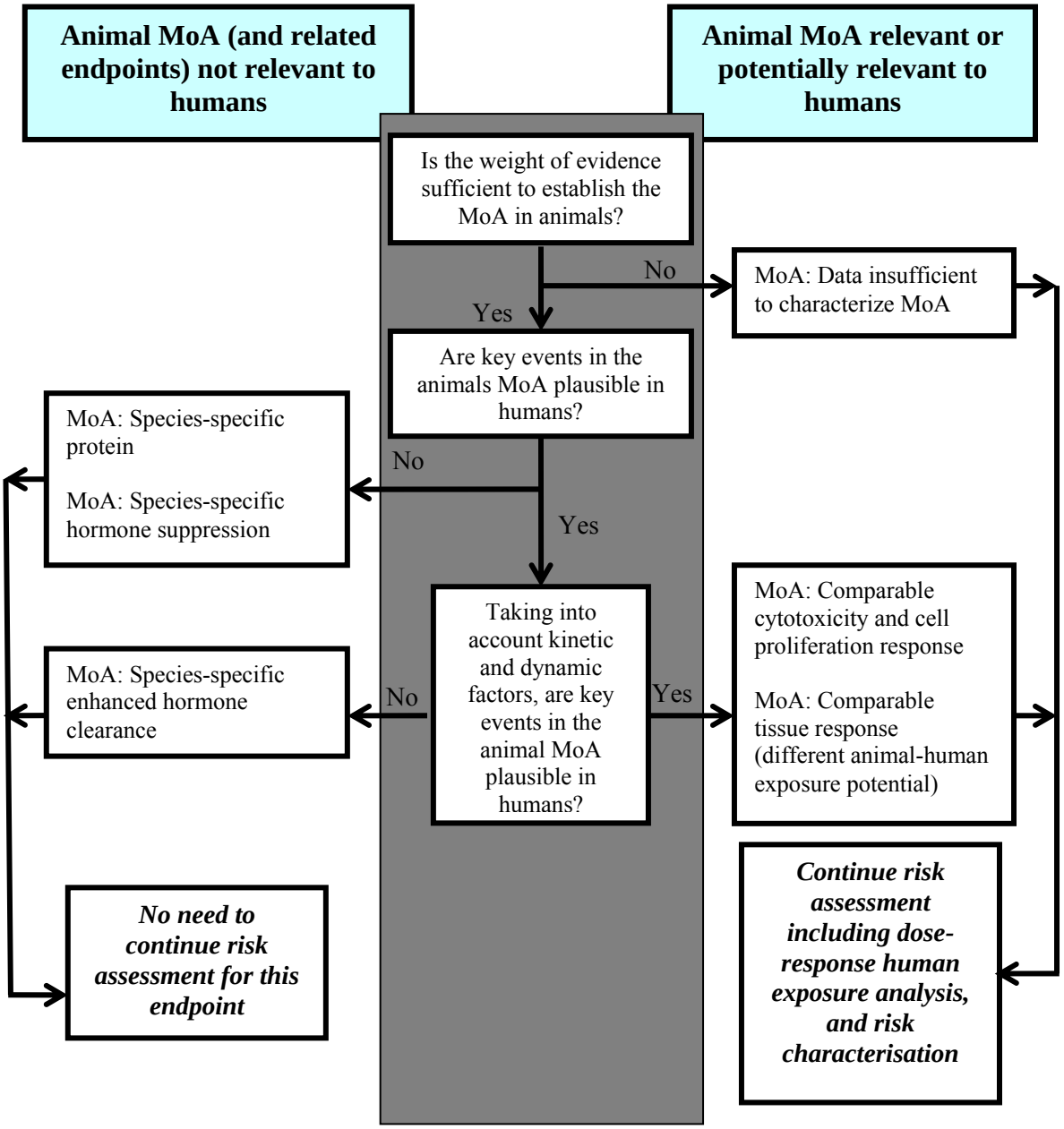

Figure 4.1: Schematic representation of the major steps for determining relevance of animal carcinogenic studies to humans

MoA = Mode of Action.

Figure adapted from Meek et al. (2003)

5 INTEGRATION OF ANIMAL AND HUMAN DATA

5.1 GENERAL CONSIDERATIONS

The WoE approach recommended in Section 3.4 requires that the complete data base for the chemical of interest—epidemiological (general population and occupational), controlled clinical studies and animal toxicology data—be systematically evaluated, along the lines set out by Bradford-Hill, and integrated. Summary tables, figures and weighting / classification schemes will help to make this process objective and transparent, while visual presentation of the data in this way allows an impartial assessment of apparently outlying studies. The US EPA staff paper on the review of the air quality standard for particulate matter (US EPA 2005c) provides an example of this approach, using the same effect metric (excess risk of effect per 25 $\mu g/m^3$ incremental increase in $PM_{2.5}$) to compare a variety of epidemiological time series studies that investigate different health endpoints.

Dose-response modelling (DRM), discussed in Section 5.5, is a powerful tool in WoE evaluations, and should be performed in some form on as many data sets as possible—it should not be restricted to the chosen pivotal study, nor to a specific type of data set (as has been done in some past standard-setting exercises). A screening step using DRM methods will help to identify uncertainties in the method, early in the standard-setting process. Given the demands that DRM places on resources and time, it may be necessary to develop screening steps that use pared down DRM methods—maybe simply graphing of dose-response data sets on the same piece of paper.

WoE evaluation can be an onerous task for data-rich substances. Reviews by overseas agencies might therefore be used to assist in hazard identification and location of pivotal studies, meta-analyses and PBPK dose-response modelling. The US EPA generates substantial documentation for many of the air pollutants that are of interest in Australia. Other potential sources include WHO Concise International Chemical Assessment Documents and Environmental Health Criteria reviews, National Industrial Chemicals Notification and Assessment Scheme (NICNAS) Priority Existing Chemical reviews, Dutch RIVM and TNO reports, the Agency for Toxic Substances and Disease Registry, Californian Reference Exposure Level documentation, and Canadian Priority Substances Reviews (note that European and WHO documentation supporting air standards is not as comprehensive as that generated by the US EPA). In using reviews by overseas agencies in the standard-setting process:

- Science policy and regulatory differences in areas such as definitions of adverse effects and quantification / classification of severity of effect must be taken into account;
- The currency of the information and completeness of information retrieval must be considered, and a literature search undertaken to discover information not included in the reviews;
- The original key articles should be obtained and evaluated for suitability. The notable exception would be if an agency has obtained proprietary or other data not obtainable

in Australia, in which case the interpretation and analysis by the overseas agency would have to be relied on; and

- The extent to which alternative approaches have been advanced or canvassed should be considered.

The deliberations, and even the conclusions and recommendations of overseas agencies might be adopted for standard setting in Australia when data/study choices, description and manipulation are clearly presented, such that an independent reviewer can reach a conclusion regarding the reasonableness and applicability of the information for Australia.

While guideline setting should, if possible, be based on health effects information from general public exposures[7], it may be necessary to place greater weighting on quantitative information from other sources (eg controlled clinical studies, animal studies) if the study data are not amenable to quantitative interrogation. More weighting is placed on animal studies that use the inhalational route of exposure compared to the oral route. If there is a dearth of these studies, consideration may be given to extrapolating from ingestion data, but this should be avoided if possible.

Data from epidemiological studies, controlled human studies, and animal studies are complementary, and none should be evaluated in isolation from others for the purposes of hazard identification, pivotal study selection or quantitative analysis.

Quantitative integration of human and animal data is best achieved by applying several methodologies for DRM assessment and standard development (eg *m*BMD, BMD, NOAEL, linear extrapolation) to different data sets in order to assess the influence of methodology and study selection on standard outcome. Such exercises, which can be of a screening nature (as discussed above), will assist in data and method selection for a detailed analysis. The different data (animal and human) should also be used for 'reality' checking. Part of the checks and balances in deriving a standard is an erudite presentation and comparison of the logic underpinning standards established by other agencies.

Quantitative accounting of sensitive or vulnerable sub-populations requires documented consideration of both human and animal data. The animal data will inform regarding matters such as potential overall kinetic handling of a substance in mammals, metabolic pathways, likelihood of reactive metabolites being formed, and the relationship of these processes to production of adverse health effects. Kinetic data are needed to generate and assess PBPK models used to estimate human internal or tissue dose metrics. The decision on whether to develop population-specific adjustment factors or use the default IPCS (2005) process should be based on both the population description within human studies and animal toxicokinetic information.

[7] General population studies are preferred primarily because they should inherently include sensitive sub-populations and real-life exposures. This, however, is balanced against compromises in exposure estimation, and if sensitive sub-populations are not characterised in the study then its usefulness may be limited. A negative response in a large population study may mask an effect in a sensitive sub-population. Furthermore, as discussed in Section 3.2.2, well conducted epidemiological studies amenable to DRM are relatively rare.

5.2 SELECTING ADVERSE HEALTH EFFECT ENDPOINTS FOR STANDARD SETTING

5.2.1 Principles for defining a 'critical health effect' from exposure to ambient air

Where a range of adverse effects is identified for a particular agent, a critical health effect needs to be identified as the basis for standards setting. This is usually based on:

- weight of evidence;
- severity;
- dose-response in the anticipated exposure range; and
- knowledge of mechanisms and structure/activity relationships.

Other factors that may be considered include:

- dread;
- reversibility;
- response to treatment; and
- interactions with other agents that may be present.

The nature, severity, pattern and mechanisms of the other health effects will also have some bearing on the standards setting process.

5.2.2 What is a 'critical health effect'?

A 'critical health effect' is defined as one that is of such significance, either in terms of severity or concern, or the number of people likely to be affected, that it drives the risk assessment process leading to the air quality standard. A qualifying factor is that the data defining the dose-response relationships must be sufficiently robust to support the DRM component of the standard-setting process.

Mortality is a critical health effect of air pollution. There is considerable weight of evidence supporting a causal relationship with air pollution and, at least for particulate matter (as PM_{10}), the dose-response relationship is characterised. Estimates range from 0.2 to 1.0% increase in all-cause mortality per 10 $\mu g/m^3$ increase in PM_{10} (Dockery & Pope 1994; Simpson *et al*. 2005b). Understanding of basic mechanisms, however, is still incomplete and current investigations are focussing on the ultrafine fraction, diesel particles and metallic contaminants. There are interactions between particles and gaseous pollutants, although by itself SO_2 does not seem to be responsible for increasing all-cause mortality, and the link between increased all-cause mortality and particles is still observed in cities with relatively low ambient levels of SO_2 (Simpson *et al*. 2005b).

There is weight of evidence that some clinical outcomes, such as hospital admission rates, could also be considered a critical health effect. Dose-response relationships are not entirely consistent, with some unexplained city-to-city variation in the effects of fine particles, ozone and NO_2 (Simpson *et al*. 2005a); however, at least for the gaseous pollutants, physiological mechanisms are understood to some extent. These effects are

usually reversible and respond to treatment. There are often prominent interactions with infectious agents, allergens, cigarette smoking and other individual factors.

5.2.3 Presenting information about health effects

Reports prepared for standards setting related to ambient air need to draw on the lessons of risk communication in other areas. The health effects of air pollution need to be presented in plain language: lay audiences cannot be presumed to have a detailed understanding of, for example, pulmonary physiology or the statistical assumptions underlying time series analysis. Effects on quality of life, symptoms and clinical outcomes should be readily comprehensible. The results of meta-analysis are often presented graphically as forest plots, and the scale, odds ratios and relative risks, 95% confidence intervals and pooled estimates all need to be explained in plain language. Similarly, homogeneity and heterogeneity need to be explained (eg as the 'apples and oranges' problem of meta-analysis), so it can be understood why not too much importance should be attached to pooled effect estimates in the presence of significant heterogeneity (see also Section 3.2.4).

5.3 TRANSFERABILITY OF RESULTS

There is an ongoing debate on whether epidemiological studies done in overseas locations can be generalised to Australia. This debate is canvassed to some extent in Section 3.2.4, in relation to meta-analyses, with some discussion of the transferability of data from large US, European and Canadian studies.

It is difficult to make hard and fast rules as to whether such studies are directly applicable to conditions in Australian cities or non-urban areas, and the comparability of the conditions of the relevant studies needs to be considered. Large meteorological differences between regions might limit the transferability of data.

As a general rule, well conducted epidemiological studies that are able to control for confounders and have been able to estimate exposures offer data that could be useful for setting air quality standards in Australia. Where regional nuances cannot be taken into account, such studies should be used with caution. For example, epidemiological studies from China, South East Asia or the sub-continent would need to take account of regional differences such as cooking practices with unflued devices, nutritional status, spread of cigarette smoking and health service delivery.

5.4 ADDRESSING SENSITIVE SUB-POPULATIONS

Within the general population, there may be sub-groups who are potentially more susceptible than others to the effects of air-borne chemicals. Human variability (also called intra-species, or inter-individual variability) may arise through toxicokinetic or toxicodynamic variability (Dybing & Soderlund 1999), both of which may be due to acquired and/or inherent factors.

5.4.1 Epidemiological principles

A sensitive sub-population is one where:

- an adverse response to an air pollutant occurs at concentrations substantially lower than affect the majority of the population. The concept applies principally to irritant and allergenic compounds; or
- the consequences of exposure are more significant than in the majority of the population. For example children may be considered a sensitive population because any irreversible adverse effects may influence their health throughout their life. The elderly, especially those with specific co-morbid effects such as cardiac or respiratory failure, may also constitute a sensitive group because the secondary consequences (eg pneumonia, worsening cardiac failure) may be more serious than in the remainder of the population.

Sensitive sub-populations may be characterised by:

- clinical history—for example, asthma, cardiac failure, chronic bronchitis, cystic fibrosis;
- evidence of airways hyper-responsiveness—for example, using methacholine or more specific challenge tests;
- demographic factors—for example, age (elderly, young); and
- genetic factors—for example, cystic fibrosis.

The following broad principles should be taken into account in air quality standard setting to allow for sensitive populations:

- Irritant substances are likely to be of greatest concern to sensitive sub-populations.
- Acceptable levels of such agents should be established such that they afford substantial protection to sensitive sub-populations.
- The extent of targeted protection must take account of the feasibility of establishing and enforcing specific exposure limits, population-wide consequences (especially the numbers affected and severity of the consequences), and the availability of other management strategies for those affected (eg anti-asthma drug therapy).
- No-effect levels in sensitive sub-populations may be best established by controlled exposure studies, but other relevant information may be derived from epidemiological studies. While controlled exposure data for both normal and sensitive sub-populations will often not be available, results from epidemiological studies may be influenced by the presence of members of sensitive subgroups (eg asthmatics).
- Safety margins employed in standard setting should be sufficient to include NOAEL in most members of the most sensitive subgroups. Typically these will be individuals with moderate to severe asthma.

5.4.2 Applying the principles in risk assessment and standard setting

From a physiological standpoint, any person who has decreased functional reserve in an organ system is theoretically less able to cope with additional environmental stressors

(non-chemical or chemical[8]). Whether an individual will respond adversely to a inhaled stressor depends upon the balance between the extent of physiological compromise (in some cases this is proportional to disease severity) and the extent of exposure (usually the concentration of the stressor in air). In many situations, acquired susceptibility (eg through illness or old age) shifts a person towards the sensitive tail of the population dose-response curve for the pollutant. These individuals nonetheless remain part of the continuum of the overall population dose response, and they experience either similar effects to others but at lower exposures to pollutants, and/or more intense effects at equivalent exposures[9].

Genetic variability can make an important contribution to human variability, for example, in the form of polymorphic genes for metabolism or tissue repair from toxic insult. Although it has long been recognised that genetic polymorphism plays an important role in driving the variability in xenobiotic metabolism, and genetic polymorphisms have been used as biomarkers of potential effect (Scherer 2005), this awareness has typically not translated into quantitative use of the data in risk assessment or standard setting (Haber *et al*. 2002; US EPA 2002). This is likely due to gaps in our knowledge on, for example:

- the prevalence of polymorphism;
- a defined link between genetic polymorphism and an adverse effect;
- the extent of induction / inhibition through co-exposure with other substances, lifestyle or diet;
- the relative contribution of multiple enzyme systems;
- allelic frequencies for major ethnic groups;
- common occurrence of low-frequency alleles in a population;
- chemical-specific phenotype data; and
- differences within or between *in vitro* and *in vivo* kinetic data.

Genotyping individuals from a sample of DNA is becoming increasingly easy. It is already possible to genotype people for loci that are thought to control kinetic and dynamic susceptibility to chemicals. However, before this information can be used in chemical risk assessments, a much better understanding is needed of the genetics of susceptibility to xenobiotics, including prevalence, and the mode of action of toxicity for the pollutant being considered. A considerable amount of research is needed to fill these knowledge gaps and truly enable chemical-specific adjustment factors for human variability to be determined (Festing 2001).

Inherent variability factors are usually related to genetic variability where carriers of certain genetic information are rendered more susceptible to the effects of pollutant exposure. These individuals may experience the same effects as other members of the

[8] In relation to air as the exposure medium and the respiratory system as the primary target organ, non-chemical stressors include infections, loss of reserve capacity due to normal aging, non-anthropomorphic allergens, physical injury and, for some asthmatics, exercise in cold dry air. Chemical stressors include products of combustion (eg smoking, vehicle exhausts, gas cookers, environmental fires and wood heaters), emissions from domestic products, cooking, and industrial activity.

[9] The effects may be largely concentration-dependent, or be a function of concentration and exposure time (see Section 5.6).

population but at lower exposure levels, or their reaction may be idiosyncratic in nature and not predictable from the type of effects observed in the wider population. In either case, genetic polymorphic groups are not necessarily captured within the continuum of the overall population dose-response curve, but from a distinct population with its own dose-response characteristics. The dose-response distributions for both populations may, however, overlap such that the least susceptible individuals in the sensitive polymorphic group experience effects at exposures affecting the most sensitive persons in the general population group. Often it is difficult to distinguish two distinct populations, and those rendered sensitive to the effects of pollution by genetic traits constitute the sensitive tail of the overall dose-response distribution.

Thus in the context of setting standards, the dose-response information for both genetic polymorphic groups and the wider population is central for addressing variability in human responses to air-borne pollutants.

The International Programme on Chemical Safety has produced a guidance document for the development of chemical-specific adjustment factors (CSAF) for interspecies and human variability (IPCS 2005). Advice for data-driven CSAFs in the IPCS document utilises decision trees that incorporate default positions for development of adjustment factors for protecting public health. The IPCS guidance is recommended for adoption in Australia for setting health-based chemical standards.

In brief, the IPCS process involves analysis of the dose-response distribution(s) to give a point exposure estimate related to a percentile response within the distribution. The final percentile to be used would be a policy decision and could be influenced by aspects such as the severity of the effect, the robustness of the data, the nature of the distribution, and risk management considerations. Examples of potentially suitable percentiles that might be provided to the risk manager are the 90^{th}, 95^{th} or 97.5^{th} percentile. For Australia the 95^{th} percentile would be appropriate and consistent with the risk level chosen as the point of departure in DRM for estimation of the modified benchmark dose (*m*BMD) (NHMRC 1999). The CSAF is calculated as the effect concentration (or exposure) at the population response percentile of interest (the 95^{th}) divided by the concentration estimate at the population mean response (ie the ratio of the 95^{th}:50^{th} percentiles). The CSAF is then applied to the 'point of departure' on the dose-response curve (ie the NOAEL, LOAEL, or *m*BMD) as normal. The procedure also allows for use of PBPK modelling. Where there are discrete subgroups of the population, the IPCS recommends that the CSAFs for different percentiles should be calculated based on data for the whole population, including the subgroup, and also for the subgroup separately. Both sets of results should be provided to the risk manager.

In the vast majority of cases, the IPCS recognises that the quantitative toxicokinetic or toxicodynamic data necessary to define a CSAF will not be available, and hazard characterisation will be necessary using the usual NOAEL / BMD / uncertainty factor approach. The default uncertainty factors (UF) for inter-human variability in kinetics (HK_{UF}) and dynamics (HD_{UF}) recommended below are the same as those recommended by IPCS (1994) and NHMRC (1999). Hence the IPCS (2005) guidance remains compatible with the current default procedures of the WHO and Australia.

The composite default uncertainty factor used to account for human response variability is 10. This is split equally between kinetic and dynamic considerations, each being 3.16 (ie $10^{0.5}$). For air pollutants whose critical health effect is systemic, both default kinetic and dynamic uncertainty factors (ie the full composite factor of 10) should be applied as a default option unless data suggest they should be reduced or increased. For substances that have point-of-contact health effects only, HD_{UF} (ie 3.16) should be considered because the health effect is not reliant on the biological processes that HK_{UF} addresses[10] (ie absorption into the blood, distribution to the target tissue and internal tissue metabolism).

The following question needs to be considered when addressing sensitive populations:

> *Do the usual adjustment (ie uncertainty or safety) factors applied to experimental human and/or animal data to account for inter-individual human response variability adequately:*
>
> - *protect sensitive people at the tail of the general population dose-response curve, and*
> - *cater for the spread of adverse health responses that may occur within a distinct polymorphic sensitive sub-population?*

In considering this, it is crucial to remember that uncertainty factors in standard setting are applied to concentrations at the bottom end of the dose-response curve—that is, they are being applied to account for human variability in the lowest sensitive percentiles (say the 5th percentile and below) of the population; they are not accounting for the complete spread of responses across the entire population, from least to most sensitive. In many cases the point of departure for application of uncertainty factors is a concentration that does not cause an effect (ie the NOAEL). One would therefore intuitively anticipate the default safety factors to be protective of the majority of the population, including sensitive members.

The National Research Council of the US National Academy of Sciences, when preparing their advice on science and judgement in risk assessment, noted that the common predisposing factors for increased sensitivity to chemicals conferred only marginal increases in relative risk to those affected (less than doubling of susceptibility). Many of the other predisposing factors, recognised as conferring high relative risks, tended to be uncommon, so few individuals were affected (NRC 1994). Recognising that it may not be possible for a standard adequately to protect very sensitive individuals, the IPCS (1994) stated that although application of uncertainty factors seeks to provide protection for sensitive members of the population, 'idiosyncratic hyper-susceptibility (excessive reaction following exposure to a given dose of a substance compared with the large majority of those exposed to the same dose) in a few individuals would not be the basis for the derivation of the TI [tolerable intake].'

A number of researchers have evaluated human data on variability in the context of evaluating whether the default 10-fold inter-variability factor appropriately accounts for the variability between the average and sensitive human in response to chemicals. In general, data from all of those studies indicate that the default value of 10 for human

[10] There may be circumstances where metabolism in the contact tissue affects the likelihood of an adverse health impact. In these situations it may be prudent to consider a value of HK_{UF} >3.

variability is protective when starting from a median response, or by inference, from the NOAEL derived from an average group of humans. Although some of these analyses noted a range of variability greater than 10-fold, it is because these authors evaluated the total range of human variability, rather than considering that the uncertainty factor of 10 is applied to account for the degree of variability between the population average (ie 50^{th} percentile) and the sensitive human ($<5^{th}$ percentile) (Dourson *et al*. 2002; Haber *et al*. 2002). Burin and Sanders (1999) considered the default factor to be relatively robust, protecting over 99% of the population, including sensitive subgroups. Similarly, Whalan *et al*. (2006), in their review of inhalation risk assessment procedures conducted by the US EPA, quoted the conclusion of Renwick and Lazarus (1998) that a 10-fold factor would cover the vast majority of the population, more than 99%.

The US EPA (2002) review on derivation of the reference concentration (RfC) indicated that the 10-fold intra-species factor appears sufficient in most cases and that reduction of the factor from a default of 10 should be considered only if data representative of susceptible sub-populations are available. The US EPA noted that the 10-fold factor may be too small for influences that have a large impact on susceptibility, such as some genetic polymorphisms. The IPCS (1994) also considered that where susceptible sub-populations exist, the default factor may not adequately cover the additional variability and modification of the default should be considered or special management strategies adopted for the sensitive sub-group. However if the risk assessment is based on *in vivo* data from the sensitive group, then the composite factor can be reduced to less than 10.

The following recommendations are made in the light of the above information:

- The possibility of sensitive sub-populations within the general population should be considered when setting an air quality standard.
- Where possible, chemical-specific adjustment factors should be developed. Where this is not possible, then the default composite uncertainty factor of 10 for human variability should be adopted, and this should be divided into kinetic and dynamic components, each being 3.16.
- The default composite uncertainty factor of 10 for human variability adequately accounts for sensitive persons within the general population dose-response distribution.
- Only if the data demonstrate the presence of discrete genetic polymorphic sensitive populations would the default uncertainty factor need to be adjusted upward. Different adjustment values should be used only when supported by relevant data.
- Where an uncertainty factor other than the default factor has been used in setting a standard, the sub-population of interest needs to be identified and its dose-response sensitivity described as the basis for such decisions.

Children as special sub-populations

Much has been written on the theoretical reasons why children may be more susceptible than adults to air-borne pollutants (eg Ginsberg *et al*. 2002; US EPA 2006b; WHO 2005). The potential impact on child development is of concern for a number of reasons; however review of these is beyond the scope of this document.

If specific data on the effects of an air pollutant on children are available, they should be incorporated in the standard-setting process. In the absence of such data, then—as with other sensitive sub-groups—the question must be asked: Do the default adjustment factors applied to non-cancer end points (including non-genotoxic carcinogens) provide adequate protection for children?

According to Doursen *et al*. (2002), virtually all the available studies suggest that a high percentage of the population, including children, is protected by using a 10-fold uncertainty factor for human variability, or using a 3.16-fold factor each for toxicokinetic and toxicodynamic variability. Specific comparisons for newborns, infants, children, adults, and those with severe disease suggest that between 60% and 100% of the population is protected. Studies conducted in larger populations that include sensitive individuals suggest that the proportion of people protected is closer to 100%. Similarly, the US National Academy of Sciences has concluded that a 10-fold intra-species factor provides adequate protection for infants and children (Bruckner 2000).

The US EPA (2006b) draft framework for child risk assessment indicates that an age-dependent adjustment factor should be incorporated into the risk assessment and, by implication, into the standards. For chemicals with mutagenic modes of action where data concerning early life susceptibility are lacking, a 10-fold factor for children under 2 years and a 3-fold factor for children between 2 and 16 years should be used. No factor is needed for children over 16 or for non-mutagenic modes of action or other end points.

The following recommendations are made, in the light of this information:

- The impact of air pollution on children's health should be actively considered when setting air quality standards.
- Provided data warrant it, the standard should be adjusted to account for possible increased sensitivity of children.
- In the absence of information showing that children have increased sensitivity to a particular air pollutant, no adjustment for child exposure is needed since the default adjustments for human variability within the adult population adequately protect children as well.
- Chemicals with a mutagenic mode of action should be evaluated on a case-by-case basis.

Pollutant-mediated increased susceptibility

Some air pollutants have the potential to make susceptible individuals more responsive to another environmental agent. This is discussed briefly in Section 2.1.2. The extent to which this is factored into the numerical value of an air quality standard should be driven by data and considered on a case-by-case basis. The questions should be considered during such deliberations:

- Does the increased responsiveness occur at environmentally relevant concentrations?
- Are the concentrations at which pollutant A is able to induce demonstrable increased responsiveness to pollutant B, higher or lower than the exposure required to produce the critical effect of pollutant A?

- If a standard is established without specific numerical factoring of the potential for enhanced responsiveness, does the standard provide an acceptable margin of safety below the concentration known to have caused increased responsiveness?

Co-exposure of pollutants

While risk assessment methodologies are being developed for co-exposure to chemical pollutants (eg ATSDR 2001; HCN 2002; US EPA 2000), by and large these techniques are not formally incorporated into the standard-setting process, primarily because of the infinite number of concentration combinations that might occur over time in a mixture of substances in air. There has generally been an expectation that the margin of exposure between the air standard for a particular substance and the exposure required to cause the most critical health in a sensitive individual[11] caters for situations where co-exposure of pollutants might occur. While this may be appropriate for pollutants that have well defined thresholds and the standard allows a large demonstrable margin of exposure, it may not be so for substances that have low or non-identified thresholds[12]. Methods to address co-exposure and potential interactions between air-borne chemicals will need to be developed as-needed.

Before the standard for an individual substance is modified to account for possible interactions with other pollutants:

- There should be data indicating that an interaction (either toxicokinetic or toxicodynamic) between substances might occur.
- There should be a realistic (rather than merely hypothetical) probability that co-exposure of the general public to the chemicals could occur.
- The interaction should be likely to occur at environmentally relevant exposures, and not just at the high experimental doses often encountered in animal experiments.
- The health end points resulting from the interaction should be relevant for public health.
- The question should be asked, is the interaction antagonistic, additive or synergistic? In the absence of data to the contrary (and providing the above points are satisfied), the default position in developing a method to account for mixture exposure in standard setting is to assume the interaction is additive.

Differential vulnerability to pollutants

Some sectors of the population may be more likely than others to experience the effects of air pollution because of their socio-economic status. Typically, their exposure is the result

[11] A sensitive individual, in this context, is one who experiences a non-immune mediated adverse effect from an air pollutant at lower concentrations than the majority of the population. The term 'sensitive' does not imply that a person has become allergic to the substance, or that the toxicological mode of action involves the immune system.

[12] For example, the epidemiological information (time series studies) for particulates does not identify a clear threshold for associated mortality and acute hospitalisation. Care must be taken when describing lack of identified thresholds, as this lack of data does not equate with the absence of a threshold in either the general or susceptible population. Not being able to identify a threshold may mean the threshold is quite low and investigation techniques are not sensitive enough to see it.

of living near an industrial pollutant source because they cannot afford to live elsewhere. The health effects people in these situations experience from exposure to pollutants may be exacerbated by lower health status, poorer nutrition, less access to health support services, and lower educational levels, compared to the general population. People who fall into these groups may also be less likely to seek assistance.

As with pollutant-mediated increased susceptibility and co-exposure to pollutants, social disadvantage is not usually specifically taken into consideration when establishing an air standard. Many epidemiological studies, however, stratify the exposed population according to socio-economic status, or study these factors and/or attempt to control for them. Human studies should be carefully interrogated to determine if the dose-response data is for, or is inclusive of, vulnerable populations. The supporting documentation for the standard should specifically address this issue.

5.5 DOSE-RESPONSE CHARACTERISATION

Ideally, standard setting will be informed by a quantitative description of how the nature and intensity of the most sensitive adverse effect varies with exposure to the pollutant. This is best achieved through mathematical modelling of the dose-response relationship. Dose-response modelling (DRM):

- enables the prediction of responses at concentrations below those documented in animal and epidemiological studies[13];
- facilitates estimates of possible risk at exposures above a standard; and
- provides the basis for an objective comparison of risk and benefits, and objective assessment of the impact of alternative standards or management options.

DRM thus provides a platform upon which uncertainties can be rationally addressed—often through allocation of uncertainty factors (or safety factors) for specific areas of technical doubt, applied to a point of departure[14] on the exposure-response curve (see Section 5.4). Although the NOAEL and LOAEL are not necessarily determined via mathematical data modelling, they nonetheless constitute a semi-quantitative description (albeit crude) of the dose-response relationship at the low end of the curve[15]. From a risk

[13] Important for standard setting when animal or occupational data form the basis of the DRM.

[14] The point of departure on the exposure-response curve may be the NOAEL, LOAEL, a benchmark dose (BMD) for a specific incidence level of effect, or some predetermined 'acceptable risk' for the critical health effect being modelled. The point of departure may be subject to a variety of uncertainty or modifying factors, such as those described in Section 5.4.2, to establish the final numerical value of the standard. The BMD and "acceptable risk" points of departure are particularly amenable in situations where a data set, whether animal toxicity or epidemiological, does not include identification of a NOAEL. In these circumstances the exposure-response modelling may provide a change of effect per incremental increase in exposure. For example, in a population of 100,000 there may be "x" number of additional hospitalisations for asthma exacerbation per "y" $\mu g/m^3$ increase in PM_{10}. The setting of a standard for such a substance therefore depends on the policy choice for the number of hospitalisations acceptable to the community. Such policy decisions are often informed by sound knowledge of "background" concentrations of the pollutant, incidence of the effect being modelled and the incidence of susceptible individuals in the Australian population to which the standard will apply.

[15] It should be noted that the NOAEL and LOAEL are not properties of the chemical being investigated. The NOAEL and LOAEL are very sensitive to experimental design, particularly the population size, dose spacing

management perspective, on-going use of estimates from DRM can give an improved characterisation for decision-making (IPCS 2004).

DRM should therefore be undertaken if possible (see Figure 5.1), and in selecting studies as the basis for an air quality standard, the amenability of the data to DRM is a major consideration. However, where the necessary quantity and quality of data or the resources for DRM are not available, it may only be possible to identify an experimental NOAEL or LOAEL. Nevertheless, even for pollutants for which a threshold may exist, the use of the NOAEL or LOAEL should only be considered when DRM is not feasible.

The WHO International Programme on Chemical Safety draft document on modelling dose-response for risk assessment of chemicals (IPCS 2004) offers a useful reference source. It provides an overview of data selection and various mathematical models for DRM, and describes the benchmark dose (BMD) method in detail.

A modification of the BMD method has been developed in Australia for application to risk assessment of carcinogens in soil (NHMRC 1999). The modified benchmark dose (*m*BMD) method (NHMRC 1999) offers a potentially useful approach applicable to all end points and exposure media, and is consistent with WHO (IPCS 2004) recommendations[16]. The methodology includes guidance on how to apply uncertainty factors to the *m*BMD to derive a *guideline dose* for exposure to a carcinogen. NHMRC (1999) identified a CSIRO software program as an appropriate mathematical modelling tool; however this is not readily available and is under refinement. Software tools for DMR and determination of the BMD and *m*BMD are available from the US Environment Protection Agency (US EPA 2003).

Where data are not amenable to *m*BMD modelling, alternative mathematical approaches are necessary. Traditionally, DRM of genotoxic carcinogenic effects has been performed differently from DRM of health effects for which a threshold response is assumed. The method used has been linear multistage mathematical modelling or, more recently, low dose linear extrapolation from an appropriate departure point on the dose-response curve [17] (US EPA 2005a). The linear extrapolation techniques raise a number of technical and science policy issues and there is not consensus between jurisdictions or within the general scientific community regarding their validity. The *m*BMD was developed to address these problems; however, for reasons discussed below, it may not be practical for the *m*BMD to be estimated. In such instances, bearing in mind that standards are subject to periodic review, an interim approach may be to use linear dose-response models to establish an air standard. In Australia this has historically been restricted to adoption of cancer unit risk values that either the US EPA or the WHO has developed, often without adequate

---

in animal experiments and exposure stratifications in epidemiological studies. The NOAEL and LOAEL can vary from study to study and therefore have statistical bounds (Leisenring & Ryan 1992)

[16] The principal difference between the WHO (BMD) and the NHMRC (*m*BMD) methodology is that the WHO recommends the use of the 95% lower confidence limit of the BMD as a point of departure for risk assessment (eg reduction by the application of uncertainty factors), whereas the NHMRC *m*BMD methodology specifies the use of the best estimate (mean value of the 5% risk estimate BMD_{05}) as the point of departure for application of uncertainty factors.

[17] BMD departure points of 1%, 5%, or 10% incidence have been used as the position on the dose response from which linear extrapolation to zero incidence is undertaken. The slope of the resulting line is used to calculate the potency of a genotoxic carcinogen. The potency is expressed as a unit risk for inhalation exposure, ie incremental increased incidence of cancer per $\mu g/m^3$ $[\mu g/m^3]^{-1}$.

documentation[18]. As summarised in Figure 5.1, it is recommended that other DRM methods be explored before such information is adopted for setting air standards in Australia.

Although a research team may have gathered toxicological or epidemiological data suitable for DRM, the published information often includes only descriptive statistics. DRM of such data is usually difficult. While it may be theoretically possible for Australia to obtain the raw data, other authorities may have more success at securing overseas information. Even if data are amenable to DRM, the necessary technical expertise or other resources may not be available in Australia. In the absence of modelling resources and/or suitable data, it may be possible (though not ideal) to adopt the DRM conducted by a competent overseas authority (a regulatory agency or a well respected scientific group that has published in the peer reviewed literature). If this approach is used, the following criteria should be met to ensure the information being used is true and robust:

- The dose-response data modelled must be adequately described. For example, in the case of animal experiments, one should be able to confirm that the modelled data were sourced from a study conducted according to good scientific principles or Good Laboratory Practice.
- The raw data used as input parameters for modelling should be presented.
- Any data manipulation should be clearly described.
- A description and, ideally, the output files of the DRM should be provided.
- The model predictions (eg from PBPK) should be validated against experimental data, human biomonitoring data and/or similar compounds.
- There should be a discussion of the influence of alternative assumption choices on the output of the model. This could take the form of a sensitivity analysis.

The time and effort required to develop health-based standards within the risk assessment paradigm depend on the complexity of the data and the degree of transparency required. The level of scientific detail addressed by DRM and the level of documentation needed may vary depending on the motivations for producing the guideline and/or the expected impact of the standard and its enforcement. Not all guideline or standard-setting situations may warrant the cost and effort of comprehensive DRM.

Summary

The recommended DRM options are summarised in Figure 5.1.

Scientific judgement is required at all steps. Where more than one option is possible, it may be prudent to develop a standard by more than one of the pathways in Figure 5.1 to enable the influence of methodology on the derived air standard to be assessed. Each step in the process should be described and justified in the documentation for the standard.

Mathematical modelling of exposure-response relationships for air-borne chemicals and pollutants offers distinct advantages for the standard-setting process. Done properly, DRM

[18] While the unit risk approach generates an estimate of risk at low levels of exposure, in order for this to be used in standard setting, there needs to be consensus on the "target risk". Target risk values used in various countries for setting standards range from 1 in 10,000 to 1 in a million.

offers objectivity and the ability to test various management options. As with any modelling, the outcome of DRM is highly dependent upon the starting information. Many animal and epidemiological data sets are not amenable to DRM, primarily because the studies were designed for hazard identification and not for DRM. However, because more jurisdictions around the world are requiring information for DRM, better data sets are likely to become available in the future. This is a trend that Australia encourages in research into the health impact of air pollution in this country. Where DRM is not feasible, other approaches for characterising the exposure-response relationship or the adoption of overseas DRM may need to be considered.

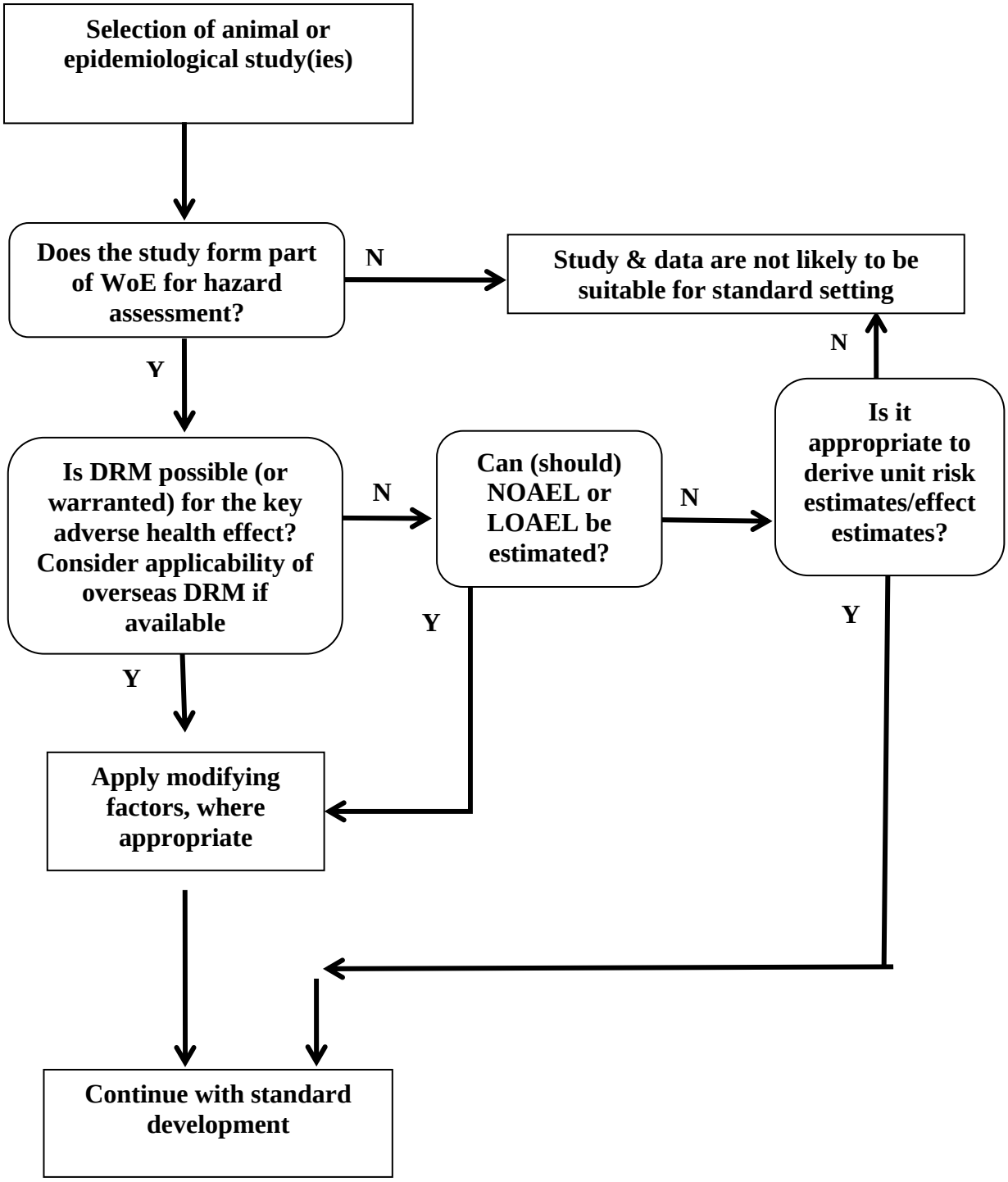

Figure 5.1: Schematic summary of dose-response modelling for setting air standards

DRM = dose-response modelling, see text
NOAEL = no observed adverse effect level
LOAEL = lowest observed adverse effect level
WoE = weight of evidence, see Section 3.4

5.6 TIME SCALING - AVERAGING TIMES FOR AIR STANDARDS

As a general rule, averaging times attached to health-based air standards should be consistent with or, where practical, shorter than the elicitation time for the effect of concern. This is an appropriate and conservative approach for public health protection. Thus:

- For compounds whose effects require chronic exposure, the averaging time should be a year.
- For pollutants whose effects are rapid in onset and/or readily reversed, the averaging time should be of the order of days, hours or less.

Ideally, the averaging time will be determined from the experimental data used to identify and measure the key adverse effect (Figure 6.1). However issues such as the practicality of measurement, the desire for uniformity[19], or the possibility of short-term spikes within the averaging period[20] may influence the final regulatory time frame assigned to an air standard.

For acute systemic adverse effects, toxicokinetic information on the active compound (either the parent molecule or metabolite) may inform selection of an appropriate averaging time. For example, where an adverse effect that is not dependent upon accumulation of toxicity is identified in a clinical or animal study, but is observed some time after steady state blood or body burden concentrations are achieved, the averaging time for a standard could be determined by adjusting the experimental observational period (often this is the same as the exposure period) according to the half-life of the active molecule. It takes approximately 5 half-lives to achieve steady state conditions for blood or body burden concentrations of the pollutant. This means that an air standard averaging time less than the equivalent of 5 half-lives will confer additional conservatism if the numerical value of the standard has been established on a NOAEL identified from the longer experimental observational period. This is discussed below in relation to application of Haber's rule for time scaling experimental data and averaging times for the standard.

These considerations for non-cumulative toxicity lead to two important considerations for setting standard averaging times[21] based on experimental data:

---

[19] Although considerations of measurement practicality and uniformity of averaging times are best incorporated into the guideline at the risk management stage, they can be addressed *inter alia* in the process described in this document, provided there is ongoing consultation, during the derivation process, between the risk assessor deriving the guideline and the risk manager(s). A transparent, defensible description of the logic for the averaging time is always required.

[20] Generally for guidelines set on acute, rapid-development health effects, the possibility of short-term concentration spikes within an averaging period is addressed by adjusting the numerical value for concentration in the guideline. This can be done using default peak-to-mean ratios as described in Katestone Scientific (1998), NSW DoEC (2005), and MfE (2002). The risk assessor is encouraged to consult a competent air dispersion modeller regarding appropriate peak-to-mean ratios. Alternatively, *continuous* measurement data from an appropriately sited monitoring station can provide specific data for peak-to-mean adjustments. The UK Department of Environment (UK DoE 2000) purposefully considered short-term concentration spikes within the averaging time when setting their 15 minute short-term air guideline for SO_2.

[21] These general rules do not apply in situations such as where the adverse effect is the result of additive toxic hits (eg accumulation of DNA damage from subsequent exposures), or where the air pollutant increases susceptibility to other disease-causing agents (eg by acting as an adjuvant).

- A NOAEL identified in an experimental exposure time frame *greater* than that required to achieve steady state will also be a NOAEL for both shorter and longer exposure times. The averaging time might therefore be determined as a multiple (less than 5) of the half-life, depending on the degree of concern for the key adverse health effect.
- A NOAEL identified in an experimental exposure time frame *less* than that required to achieve steady state will be a NOAEL for shorter exposure periods but not necessarily for longer times. As a rule of thumb, averaging times should not be set greater than the experimental exposure period identifying the critical health effect.

Adjustment of averaging times

More often than not, the averaging time for an air standard may be established on considerations other than the biological and kinetic mechanisms that give rise to the health effect. In such cases, it should be clearly stated that the averaging time does not reflect exposure times associated with health effects, but has been set on other grounds that are transparently articulated in the supporting documentation for the standard.

Notwithstanding the motivation for the averaging time ultimately assigned to an air quality standard, the experimental data should always be the starting point for setting averaging times. This is achieved by extrapolating exposure concentrations from the available data set(s) to the exposure duration specified in the nominated averaging time for the standard[22]. Many organisations[23] apply Haber's Law to perform the required math for the extrapolation. While there is uncertainty in this exercise, the procedure described below is considered, within the limits of current knowledge, to be appropriately conservative.

Haber's Law states that the product of the concentration (C) and the time of exposure (t) is equal to a constant level or severity of response (K) for a specific toxicological effect. Thus

$$C \times t = K \quad \text{.....Equation 1}$$

This assumes that concentration and time of exposure are equally important in producing the effect[24]. Equation 1 is equivalent to the area under the exposure curve; the area under the dose-response curve that measures the total delivered dose to a target tissue is an analogous concept. However, not all substances follow this simple relationship, and a more general exponential relationship of $C^n t = K$ (where 'n' is a chemical-specific parameter[25] that is also specific for a specified health end point) is required to describe the effects of concentration and exposure time for some toxicological endpoints (ten Berge *et al*. 1986). While it is uncertain whether, for a given substance, the value of 'n' can be extrapolated to other biologic effects (Krewski *et al*. 2004), the extrapolation is nevertheless often done.

[22] Extrapolation is applied to the 'point of departure' (PoD) concentrations identified by quantitative examination of dose-response relationships, ie ideally PoDs are determined from mathematical dose-response modelling. These PoDs may be the NOAEL, LOAEL, BMD or *m*BMD.

[23] For example the US National Advisory Committee in setting AEGLs (NRC 2001, Krewski et al. 2004), ASTDR (Miller et al. 2000), and California EPA (OEHHA 1999).

[24] Haber's Law is assumed to be operable in the often-applied linear extrapolation of effects observed with discontinuous experimental inhalation exposures in animals (eg 6 hr/d, 5d/wk) to an assumed continuous exposure of humans (24hr/d, 7d/wk) by dividing the exposure concentration by 5.6 (24/6 x 7/5).

[25] The value of n is determined from concentration versus response relationships for several different exposure times. Haber's general rule of $C^n t = K$ can be written as $\ln t = \ln K - n \ln C$. This simple linear form allows easy determination of the value of n.

Since the effect is assumed constant for a given concentration (C_1) and exposure time (t_1) product, the concentration (C_2) required to produce the same effect at exposure time t_2 (the nominated averaging time for the standard) can be easily calculated from:

$$C_1^n t_1 = K = C_2^n t_2 \quad \text{.....Equation 2}$$

C_2 is thus the numerical value (ppm or $\mu g/m^3$) adjusted for the nominated averaging time (t_2).

Not all substances for which an air standard value may need to be developed have experimentally derived values for the exponent 'n'; for 20 substances, ten Berge *et al*. (1986) empirically estimated 'n' to vary between 0.8 and 3.5. The Office of Environmental Health Hazard Assessment (OEHHA 1999) expanded the number of substances to 57 and the derived values for 'n' ranged from 0.8 to 4.6; the mean value rounded to 2, while the interquartile range (25 – 75%), where most 'n' values are found, was 1 – 2.2.

Many researchers have noted that, when extrapolating from a low experimental exposure time to a higher exposure (standard averaging) time, the use of a value of n>1 may not be conservative (Gaylor 2000; OEHHA 1999; ten Berge *et al*. 1986). Based on the relationship between concentration and duration to produce an equal incidence of mortality, it is suggested that conservative extrapolation[26] to shorter durations of exposure should be based on a value of n = 3 (Gaylor 2000). California EPA has a default averaging time of 1 hour for acute ambient air standards; they use a value of n = 2 for downward extrapolation from experimental exposures to the 1 hour averaging time and n = 1 when extrapolating upward to the 1 hour average.

In the absence of experimentally derived chemical-specific values for 'n' and within the time span constraints discussed below, it is recommended that[27]:

- for downward extrapolation from experimental exposures to a standard averaging time shorter than the experimental exposures, a value of n = 3 be used in the general exponential form of Haber's Law. This is considered suitably conservative for public health purposes; and
- for upward extrapolation from short experimental exposure times to a longer standard averaging time, a value of n = 1 be used.

Over what time frames is extrapolation appropriate?

It is inappropriate to extrapolate effect results from experimental exposures over a long time, especially for a substance with a long half life, to establish a short-term standard. It is recommended that extrapolation be limited to relatively small differences between the

[26] If a concentration C_1 is considered safe for an exposure duration of t_1, the 'safe' concentration C_2 for a shorter duration of t_2 is based on $(C_1^3 \times t_1) = (C_2^3 \times t_2)$, giving $C_2 = C_1 \times (t_1/t_2)^{1/3}$. For example, if the exposure time is reduced by a factor of 8, the concentration should only be increased by a factor of $(8)^{1/3} = 2$; whereas Haber's Law would allow an increase in the concentration by a factor of 8 if n = 1. Hence in downward extrapolation to shorter averaging times, it is more conservative to use a higher value of the exponent 'n'. The reverse is true for upward extrapolation to a longer averaging time. If the exposure time is increased from t_1 = 1hr to t_2 = 8hr, then C_2 becomes 0.125 of C_1 if n = 1, but 0.5 of C_1 if n = 3.

[27] These recommendations are made purely on the conservatism the math confers to the downward and upward extrapolations if n = 3 & 1 respectively.

experimental exposure and the desired averaging time[28]. As the gap between experimental exposure times and averaging time widens, the uncertainty associated with applicability of Haber's Law increases (Miller *et al*. 2000; Rozman 2000).

It is generally agreed that Haber's Law could not apply for an infinite time of exposure, or there would be no safe exposure limits for prolonged or repeated exposures (Witschi 1999). Further, the applicability of Haber's Law to some health endpoints involving site-of-contact effects is limited. In relation to air standards, this is especially pertinent for substances that elicit odour adverseness or sensory irritation, mainly because the effects are receptor-mediated and have very rapid onset at effective air concentrations, as discussed below.

Sensory irritation and standard averages

An ambient air standard is established to protect the general population (including sensitive sub-groups) against the critical health effect associated with the chemical in question. The critical health effect is usually the one that occurs with the lowest level of exposure; and often the sensitive effects identified in population-based studies are associated with sensory irritation and/or odour. It is debatable whether odour and mild sensory irritation can be considered as true health effects, but they can be readily be categorised as 'quality of life' influencers (see Section 7), and air guidelines have been established on the basis of sensory irritation or odour as the critical effects[29].

Odour and sensory irritation occur with very short-term exposures, and may contribute to a decreased sense of well-being. They are also associated with industrial emissions and polluted air; indeed, the perception of odour as being the 'Trojan Horse' for sinister toxic compounds in industrial emissions and/or that the air is polluted by industrial activity can, of itself, be detrimental to an individual's well-being.

Sensory irritation can be a direct effect of brief exposure to industrial emissions or polluted air containing substances that are able to stimulate the trigeminal nerve endings in the mucosa of the eye and upper respiratory tract. Even though the irritation may be relatively mild, manifested as itchy eyes or a tingling nose, it can affect general amenity and well-being if it happens often and perhaps in conjunction with odour. In this situation the effect should be considered adverse.

There is wide variation in the human population in the ability to sense this type of irritation. Much of the variability is due to the influence of psycho-cognitive factors on perception and detection of air-borne chemicals. Sensory irritation is dose-related: the higher the dose, the greater the intensity of the response and the greater the likelihood that more people will be able to sense the chemical. Many scientific studies on humans also show that the discomfort associated with irritation of the eyes, nose and throat can be markedly modified by co-exposure to odorants (Axel 1995; Barrow *et al*. 1989).

[28] For example, extrapolations of the order of 1 hr ↔8hr, 3mth ↔12 or 24 mth, 15 min ↔1hr would be acceptable. Regardless of compound kinetics, extrapolations such as 15min ↔>2hr or 1hr ↔>8hr would be undesirable, as they represent more than an 8-fold extrapolation. They should occur only under exceptional circumstances, and should be fully justified.

[29] For example hydrogen sulphide and formaldehyde (WHO 2000).

According to Rozman (2000), it is likely that many of the experimental situations showing an exception to Haber's Law are the result of failure appropriately to consider relationships between the toxicokinetics of the substance and the exposure periods. Haber (1924) formulated the relationship between C and t using observations of the lethality of chemical warfare agents in experimental animals. It was known that substances easily detoxified were not as apt to follow a fixed C x *t* relationship (Miller *et al*. 2000). Rozman (2000) stated that, for substances of very short kinetic or dynamic half-lives, the reliance on time of exposure for an effect is much less than that for concentration. This is the direct result of the much shorter time required for effective concentrations to reach steady state, and rapid elimination once exposure decreases.

Sensory irritation responses are receptor-mediated and their intensity is determined by the final steady state concentration of substance in mucosa and/or tear film. This concentration is, in turn, a function of the water-to-air partition coefficient (Hau *et al*. 1999). The time taken to reach steady state (ie equilibrium between the tear film and air concentrations) determines the onset of sensory irritation: for highly water-soluble compounds, equilibrium is very rapidly achieved; for lipophilic substances, it may take a little longer but is usually within about 10 minutes.

For most compounds there is no evidence of facilitation of sensory irritation; indeed, accommodation generally occurs on continued exposure to low concentrations that cause mild to moderate sensory irritation. Given that the intensity of sensory irritation is dependent upon the concentration of the substance in the biological fluid bathing the eye, it follows that, once steady state between tear film and any given air concentration is reached, then the maximum, or near maximum, level of sensory irritation is also reached and longer exposures at the same concentration will not produce proportionally greater intensity of effect.

These theoretical considerations predicting rapid onset of sensory irritation to effective concentrations of sensory irritants and an early plateau effect of response with continued exposure are supported by empirical observation with formaldehyde (Sauder *et al*. 1987), octene (Hempel-Jorgensen *et al*. 1999) and Stoddard solvent (Hastings *et al*. 1984).

The theoretical considerations have been succinctly summarised by Calabrese and Kenyon (1991). In describing an approach for the development of scientifically defensible ambient air level guidelines (AALG), they noted several features of sensory irritation that differentiate it from other health endpoints. The localised nature of the response as a result of direct contact (rather than being due to absorption and distribution to a distant site of action) means that onset of the response is relatively rapid and the response tends to cease quickly when the offending agent is removed. 'Therefore, sensory irritant responses to air-borne chemicals tend to be concentration dependent, and the injury or effect is generally noncumulative in nature. The implication of this characteristic for AALG derivation is that adjustment for continuous exposure is not warranted.'

For substances and health end points that have rapid onset of action, the averaging time attached to the air standard value needs to be comparably short. As discussed above, it is inappropriate to extrapolate from a 5 – 10 minute experimental exposure time necessary to

elicit sensory irritation[30] responses, to a 24 hour averaging time. Furthermore, the probability of short 1 – 2 minute high concentration spikes within a long averaging period (relative to biological response time) increases as the standard averaging time increases.

Summary

- Averaging times attached to health-based air standards should be sympathetic with or, where practical, shorter than the elicitation time for the effect of concern.
- The starting point for setting standard averaging times should be the experimental data.
- If standard averaging times are established on considerations other than deliberations of biological and kinetic mechanisms giving rise to the health effects, it should be clearly stated that the averaging time does not reflect exposure times associated with health effects, and the reasons for this should be given.
- Adjustment of experimental exposure times to match a nominated averaging time can be achieved using the power version of Haber's Law (Equation 2).
- Reasons for not using experimental exposure times as the averaging time for the standard need to be fully justified and explained.
- Time span constraints are recommended within which it would be considered appropriate to use Haber's Law.
- For downward extrapolation from experimental exposures to a standard averaging time shorter than the experimental exposures, a value of n = 3 should be used for the exponential in the general form of Haber's Law. This is considered suitably precautionary for public health purposes.
- For upward extrapolation from short experimental exposure times to a longer standard averaging time, a value of n = 1 should be used for the exponential.
- Sensory irritation, and other adverse health responses that occur within very short exposures of effective concentrations of pollutants, require short averaging times. Haber's Law and/or adjustment for discontinuous experimental doses do not apply.
- The possibility of short-term high concentration spikes of a pollutant within an averaging period should be addressed and, if necessary, the concentration should be adjusted using default or observed peak-to-mean ratios.

5.7 USE OF EPIDEMIOLOGICAL STUDIES IN SETTING AIR QUALITY STANDARDS

The strength of human epidemiology derives from its direct relevance to human health. Used in conjunction with other scientific approaches, epidemiological studies can

[30] The considerations of rapid effect-response time directing a short guideline averaging time canvassed in this section are not limited to the endpoint of sensory irritation. Exercising asthmatics who are SO_2 responders experience adverse health effects within a few minutes of exposure to effective bronchoconstrictor concentrations of SO_2 (Horstman et al. 1988, Balmes et al. 1987, Roger et al. 1985) and longer periods of SO_2 exposure during exercise do not lead to a statistically significant worsening compared with the initial response (Kehrl et al. 1987, Horstman & Folinsbee 1986).

significantly contribute to the air quality standards-setting process. One example of this is the dose response modelling (DRM) that is derived from epidemiological studies. The DRM is important for quantifying mortality and morbidity associated with the respective air pollutants on a population level in the risk characterisation procedure. This step in the standards-setting procedure is represented in Figure 1.1.

The positive attributes and pragmatic uses of current epidemiological studies have been documented in Section 3.2.2. To date, though, Australian epidemiological studies have had a limited role in standard setting.

In the previous air quality standard-setting process, Australia did not have access to studies with the sophistication of design and statistical modelling that are currently available. The standards-setting process that is currently underway with the Environment Protection and Heritage Council and the National Environment and Protection Council will be informed by contemporary human epidemiological studies that can provide advantages including:

- being able to examine health endpoints that are considered unethical to produce by experimentation;
- being able to measure effects on populations exposed in the context of all the variables of usual exposure. Despite the imprecision, confounders and variability, well-conducted epidemiological studies are highly relevant to the exposures and responses of the population;
- avoiding the limitation of animal studies in that the dose ranges examined are in keeping with those likely to be encountered by humans. Animal toxicology involves high dose administration, and at these levels both the toxicokinetics (eg metabolism and pattern of metabolites) and toxicodynamics (mechanism of toxic effects) may differ from those associated with environmentally relevant doses; and
- being capable of identifying a broader range of health effects than is available with animals. These include more subtle effects such as sub-clinical effects, psychological disturbances and malaise.

As discussed briefly in Section 3.2.2, human epidemiological studies also have a number of limitations. The principle of these is their imprecision, particularly in establishing dose-response relationships across relatively shallow gradients of exposure. Even the best epidemiological studies are limited by factors such as:

- imprecise measurement of exposure (with consequent difficulty in establishing dose-response relationships);
- secondary or surrogate measures of outcome, such as emergency room attendances;
- difficulty in disentangling the influences of extraneous confounding factors such as co-exposures;
- difficulty in isolating effects in sensitive subgroups; and
- difficulty in establishing effect-modification by factors such as temperature or physical exercise.

To a large extent, these limitations are a function of costs, which are often the limiting factor in applying epidemiological studies to sensitive sub-populations. Epidemiological

studies have been very effective in situations of extremes of exposure. Such studies include those of the 1952 London fog episode, where much was learned about the influence of inhaled sulfur dioxide and particulate matter (Hunt *et al*. 2003; Logan 1953) as a result of the extreme conditions encountered. Similar insights can sometimes be obtained through studies of highly exposed occupational populations.

When large variations of exposure are not available, epidemiological studies are much more limited, as any underlying effect may be masked by variability as the signal-to-noise ratio is reduced. Their best contribution may be to exclude large effects on a health outcome at the highest available exposure under study.

Another issue that arises from the use of epidemiological data in standard setting is when the outcome is expressed in terms of a unit risk per concentration (eg x% increase in mortality relative risk per ug/m^3; see Brunekreef & Holgate 2002). Use of these types of DRM risk estimates has featured strongly in the considerations around the adverse health effects of fine ($PM_{2.5}$) or coarse (PM_{10}) particulates in the US EPA standard-setting process (US EPA 2005c). The implication is that there is no threshold for the adverse effect, at the levels of exposure commonly encountered in cities. It is important in any standards-setting process, to acknowledge realistic and experienced levels of exposure for the population concerned.

Having no identified threshold not only complicates the derivation of a health-based air quality standard (it requires the assumption of a finite level of risk that is 'acceptable'), but also places a greater onus on the transparency of the standard-setting process to highlight the fact that the health risks can only be minimised, but not eliminated.

An alternative expression of the DRM can be in terms of the years of life lost, the number of premature or attributable deaths per year (Roosli *et al*. 2005; AIRNET WG 4 2004), or quality- or disability-adjusted life years (QALY or DALY) per unit exposure (Stieb *et al*. 2005).

It is often assumed that such relationships are linear and that extrapolation can provide an estimate of the excess risk at particular, usually lower, levels of exposure than were characterised by the epidemiological study. Stieb *et al*. (2002) pointed out that while linear regression may be suitable for data where the daily number of deaths is large, a Poisson regression is more appropriate where the number of deaths is smaller:

$$\text{Mortality} = y = e^{(\alpha' + \beta x)} \text{ (Poisson regression)}$$

$$\text{Mortality} = y = \alpha' + \beta x \text{ (linear regression)}$$

where $\alpha' = \alpha + \Sigma\, \beta_1 x_1$ and β_1 are vector or parameter estimates, x_1 is a vector of covariates, and x corresponds to the concentration of the air pollutant of interest.

The per cent excess mortality (PEM) = $100*(RR_{\Delta x}-1)$, where Δx is the representative change in pollution concentration, and $RR_{\Delta x}$ is the change in relative risk (= $e^{\beta \Delta x}$ for a Poisson distribution, or = $1 + \beta \Delta x/y$, where y is the mean daily number of deaths).

In a standard-setting process, the problem with either of these regression methods arises in deciding when the excess risk represents a significant or meaningful increase above some baseline measure of the prevalence of the health effect. The setting of an appropriate *target*

risk level for the standard is a process that involves not only science[31], but also consideration of economic, social, political and risk communication factors.

Stieb *et al*. (2005) suggested that such data be converted to an Air Quality Index, which provides a tool for communicating the impacts of short-term changes in air pollutant concentrations.

Specific guidance—use of epidemiological studies

What criteria should be used to determine which individual epidemiological study results should be included (and which should be excluded) in the hazard assessment process for Australian standards setting for air pollutants?

While epidemiological studies have some disadvantages, they may also constitute the bulk of the scientific database for any particular hazard. The validity of epidemiological studies for aetiological inference varies according to study designs. The generally accepted ranking, from highest to lowest, is:

- randomised clinical trial
- prospective cohort study
- retrospective cohort study
- nested case-control study
- time-series analysis
- cross-sectional study
- ecologic study
- cluster analysis
- case study
- anecdote.

[31] The issue is analogous to that with genotoxic carcinogens in that a target cancer risk is needed in order to calculate the numerical value of the standard. The assumption is that in the low end of the dose response the relationship is linear. Hence the 'standard' is simply calculated as 'target risk' ÷ 'unit risk factor' = standard. The unit risk factor is the cancer risk per unit of concentration (eg risk per $\mu g/m^3$), and is traditionally derived from linear multistage DRM of animal data, or from studies of occupational cohorts. WHO (2000A) usually articulates its advice with respect to risk from air-borne genotoxic carcinogens as an air concentration that corresponds to a lifetime risk of 1 in 10,000 ($1x10^{-4}$), 1 in 100,000 ($1x10^{-5}$) or 1 in a million ($1x10^{-6}$). This enables a particular jurisdiction to choose the level of protection, or risk, they wish to implement in their standard.

6 GUIDANCE FLOWCHARTS

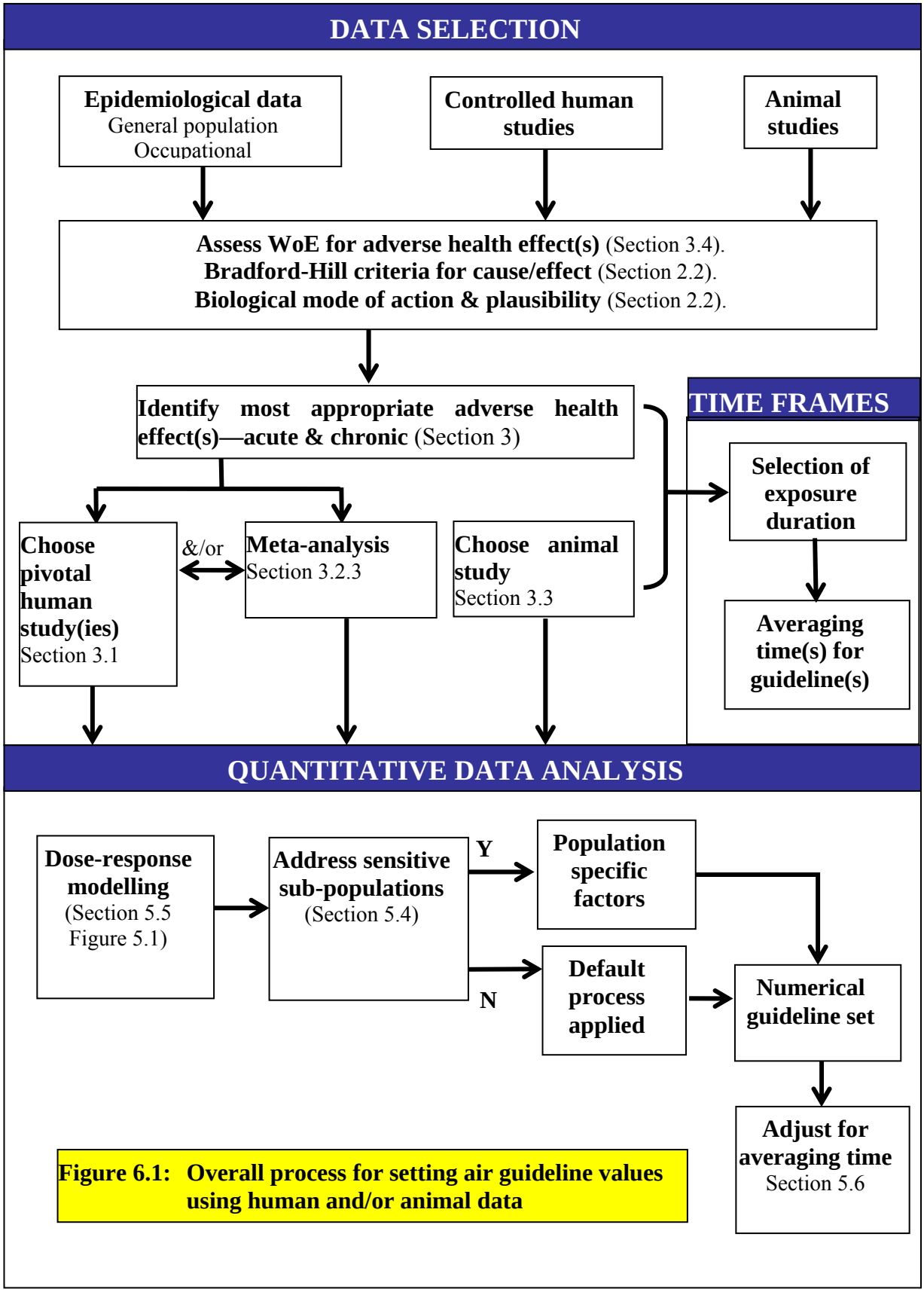

Figure 6.1: Overall process for setting air guideline values using human and/or animal data

7 OTHER CONSIDERATIONS

The regulatory process of setting air quality standards and standards in Australia and overseas involves consideration of a range of scientific, social, economic and political issues. In many cases the final standards are a balance of all of these issues and are set with an inherent level of human health risk associated with them. Air quality guidelines differ from air quality standards in that they do not have any statutory basis and are generally used to provide guidance as to whether air pollution may pose a risk to public health. Consequently, guidelines may consider only the scientific evidence and not the social, economic or political issues. For example, the WHO air quality guidelines (WHO 2000) provide guidance as to what issues must be considered by individual countries in converting the numerical guidelines into standards[32].

There are different schools of thought as to where quality of life impacts should be taken into consideration in the standard-setting process. If such impacts are considered to be truly adverse health effects, they should be taken into consideration during the stages where data are being assessed to set the appropriate effect levels. Alternatively, some suggest that quality of life considerations are more appropriately considered as part of the risk-management process, along with other factors that need consideration by risk managers. Figure 7.1 depicts the two possible sites where quality of life factors could be considered.

[32] For example, the development of an air quality standard has to consider a range of issues of which health considerations, although very important, are only one component. In Australia the NEPC Act requires an impact statement to be prepared to accompany the development of a NEPM or change to a NEPM, to assess what the potential costs and benefits of any changes to the NEPM (including changes to an air quality standard) may be. These costs need to be balanced with any benefits, including health benefits, that might be associated with the change or implementation of any new standards.

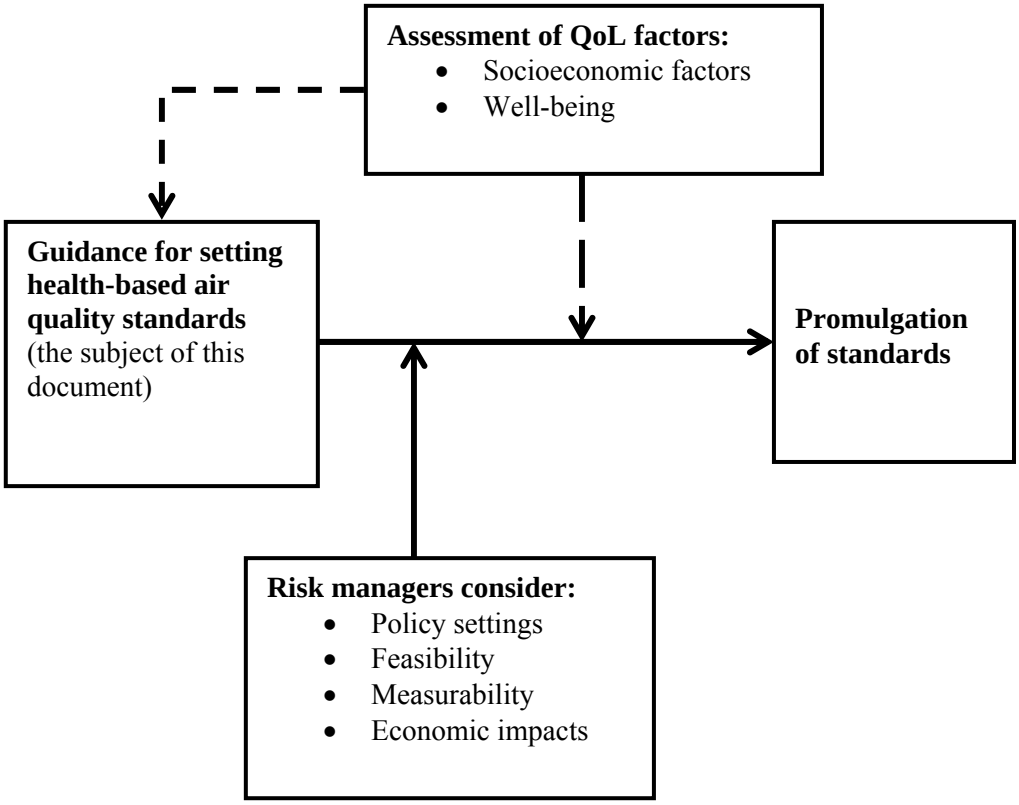

Figure 7.1: Relationship between setting health based air standards, risk management and quality of life (QoL) issues

SUBMISSIONS RECEIVED

The following organisations and associations provided submissions as part of the targeted consultation process:

1. National Research Centre for Environmental Toxicology;
2. Office of Environmental Health Hazard Assessment of California EPA;
3. The Australian Lung Foundation;
4. NSW Health;
5. Queensland University of Technology;
6. Thoracic Society of Australia and New Zealand; and
7. Australian Faculty of Public Health Medicine, The Royal Australasian College of Physicians.

The information provided in the submissions was collated and incorporated in the report by the NHMRC Working Committee for Air Standards Health Advice and the technical writer.

REFERENCES

AIRNET WG 4. 2004. Air Pollution and the Risks to Human Health. Amsterdam. http://airnet.iras.uu.nl/.

American Conference of Governmental Industrial Hygienists [ACGIH], 2005. TLVs and BEIs based on the documentation of the threshold limit values and biological exposure indices. Cincinnati, Ohio.

Angerer J, Bird MG, Burke TA, Doerrer NG, Needham L, Robinson SH *et al*. 2006 Strategic biomonitoring initiatives: Moving the science forward. Toxicol. Sci. 93(1):3-10

ATS. 1985. Guidelines as to what constitutes an adverse respiratory health effect, with special reference to epidemiologic studies of air pollution. Am Rev Respir Dis 131(4):666-668.

ATS Board of Directors. 2000. What Constitutes an Adverse Health Effect of Air Pollution? Am J Respir Crit Care Med 161:665-673.

ATSDR. 2001. Guidance Manual for the Assessment of Joint Toxic Action of Chemical Mixtures. Washington, D.C.: U.S. Department of Health and Human Services, Public Health Service, Agency for Toxic Substances and Disease Registry.

Axel R. 1995. The molecular logic of smell. Sci Am 273(4):154-159.

Balmes JR, Fine JM, Sheppard D. 1987. Symptomatic bronchoconstriction after short-term inhalation of sulfur dioxide. Am Rev Respir Dis 136(5):1117-1121.

Barnett AG, Williams GM, Schwartz J, Neller AH, Best TL, Petroeschevsky AL, *et al*. 2005. Air pollution and child respiratory health: a case-crossover study in Australia and New Zealand. Am J Respir Crit Care Med 171(11):1272-1278.

Barrow GS, Buckley LA, James RA, Steinhagen WH, Chang JCF. 1989. Sensory Irritation: Studies on correlation of pathology, structure activity, tolerance development and prediction of species differences to nasal injury. In: Toxicology of the Nasal Passages (Barrow GS, ed). Washington: Hemisphere Pub. Corp.

Bruckner JV. 2000. Differences in sensitivity of children and adults to chemical toxicity: the NAS panel report. Regul Toxicol Pharmacol 31(3):280-285.

Brunekreef B, Hoek G. 2000. Beyond the body count: air pollution and death. Am J Epidemiol 151(5):449-451.

Brunekreef B, Holgate ST. 2002. Air pollution and health. Lancet 360(9341):1233-1242.

Burin GJ, Saunders DR. 1999. Addressing human variability in risk assessment--the robustness of the intraspecies uncertainty factor. Regul Toxicol Pharmacol 30(3):209-216.

Burnett RT, Brook J, Dann T, Delocla C, Philips O, Cakmak S, *et al*. 2000. Association between particulate- and gas-phase components of urban air pollution and daily mortality in eight Canadian cities. Inhal Toxicol 12 Suppl 4:15-39.

Burney P, Chinn DJ, Luczynska C, Lai E. 1996. Variations in the prevalence of respiratory symptoms, self-reported asthma attacks, and use of asthma medication in the European Community Respiratory Health Survey. Eur Respir J 9:687-695.

Calabrese EJ, Kenyon EM. 1991. Chapter 3: An approach for the rapid development of scientifically defensible AAIs. In: Air toxics and risk assessment (Calabrese EJ, ed). Chelsea, Mich.: Lewis Publishers.

California EPA. 1999. Determination of Acute Reference Exposure Levels for Airborne Toxicants (California Office of Environmental Health Hazard Assessment (OEHHA), ed): California Environmental Protection Agency.

Centers for Disease Control and Prevention (CDC). National Center for Environmental Health. 2005. *Third national report on human exposure to environmental chemicals.* <http://www.cdc.gov/exposurereport/3rd/>.

Cohen AJ, Ross Anderson H, Ostro B, Pandey KD, Krzyzanowski M, Kunzli N, *et al*. 2005. The global burden of disease due to outdoor air pollution. J Toxicol Environ Health A 68(13-14):1301-1307.

Department of Environment and Heritage. 2001. State of Knowledge Report: Air Toxics and Indoor Air Quality in Australia. Canberra: Australian Federal Government. Available: http://www.deh.gov.au/atmosphere/airquality/publications/sok/index.html [accessed 23/05/06].

Dockery DW, Luttmann-Gibson H, Rich DQ, Link MS, Mittleman MA, Gold DR, *et al*. 2005. Association of air pollution with increased incidence of ventricular tachyarrhythmias recorded by implanted cardioverter defibrillators. Environ Health Perspect 113(6):670-674.

Dockery DW, Pope CA, 3rd. 1994. Acute respiratory effects of particulate air pollution. Annu Rev Public Health 15:107-132.

Dominici F, McDermott A, Zeger SL, Samet JM. 2002. On the use of generalized additive models in time-series studies of air pollution and health. Am J Epidemiol 156(3):193-203.

---. 2003. Airborne particulate matter and mortality: timescale effects in four US cities. Am J Epidemiol 157(12):1055-1065.

Donaldson K, Stone V, Seaton A, MacNee W. 2001. Ambient particle inhalation and the cardiovascular system: potential mechanisms. Environ Health Perspect 109 Suppl 4:523-527.

Dourson M, Charnley G, Scheuplein R. 2002. Differential sensitivity of children and adults to chemical toxicity. II. Risk and regulation. Regul Toxicol Pharmacol 35(3):448-467.

Dybing E, Soderlund EJ. 1999. Situations with enhanced chemical risks due to toxicokinetic and toxicodynamic factors. Regul Toxicol Pharmacol 30(2):S27-30.

ECETOC 2005 (European Centre for Ecotoxicology and Toxicology of Chemicals). Guidance for the Interpretation of Biomonitoring Data. Document No. 44. Brussels, Belgium: ECOTOC.

enHealth. 2004. Environmental health risk assessment : guidelines for assessing human health risks from environmental hazards. Canberra, ACT: Dept. of Health and Ageing. Available: http://www.health.gov.au/pubhlth/strateg/envhlth/risk [accessed 14 June 2006].

Festing MF. 2001. Experimental approaches to the determination of genetic variability. Toxicol Lett 120(1-3):293-300.

Fletcher CM, Peto R. 1977. The natural history of chronic airflow obstruction. British Medical Journal 1:1645-1648.

Gauderman WJ, McConnell R, Gilliland F, London S, Thomas D, Avol E, *et al*. 2000. Association between air pollution and lung function growth in southern California children. Am J Respir Crit Care Med 162(4 Pt 1):1383-1390.

Gaylor DW. 2000. The use of Haber's law in standard setting and risk assessment. Toxicology 149(1):17-19.

Ginsberg G, Hattis D, Sonawane B, Russ A, Banati P, Kozlak M, *et al*. 2002. Evaluation of child/adult pharmacokinetic differences from a database derived from the therapeutic drug literature. Toxicol Sci 66(2):185-200.

Guyatt GH, Berman LB, Townsend M, Pugsley SO, Chambers LW. 1987. A measure of quality of life for clinical trials in chronic lung disease. Thorax 42(10):773-778.

Haber F. 1924. Zur geschichte des gaskrieges (On the history of gas warfare). In: Fünf Vorträge aus den Jahren 1920–1923 (Five Lectures from the Years 1920–1923). Berlin: Springer, 76–92.

Haber LT, Maier A, Gentry PR, Clewell HJ, Dourson ML. 2002. Genetic polymorphisms in assessing interindividual variability in delivered dose. Regul Toxicol Pharmacol 35(2):177-197.

Hastings L, Cooper GP, Burg W. 1984. Chapter 19: Human sensory response to selected petroleum hydrocarbons. In: Applied toxicology of petroleum hydrocarbons, Advances in Modern Environmental Toxicology, vol 6 (MacFarland HN, ed). Princeton, N.J.: Princeton Scientific Publishers, 225-270.

Hau KM, Connell DW, Richardson BJ. 1999. Quantitative structure-activity relationships for nasal pungency thresholds of volatile organic compounds. Toxicol Sci 47(1):93-98.

HCN. 2002. Exposure to combinations of substances: a system for assessing health risks U 505/EvV/wyl/442-C3. Amsterdam: Health Council of the Netherlands.

Hempel-Jorgensen A, Kjaergaard SK, Molhave L, Hudnell HK. 1999. Time course of sensory eye irritation in humans exposed to N-butanol and 1-octene. Arch Environ Health 54(2):86-94.

Higgins JPT, Green S. 2005. Cochrane handbook for systematic reviews of interventions. 4.2.5. Available: http://www.cochrane.dk/cochrane/handbook/hbook.htm [accessed 14 June 2006].

Hill AB. 1965. The Environment and Disease: Association or Causation? Proc R Soc Med 58:295-300.

Holguin F, Tellez-Rojo MM, Hernandez M, Cortez M, Chow JC, Watson JG, *et al*. 2003. Air pollution and heart rate variability among the elderly in Mexico City. Epidemiology 14(5):521-527.

Horstman DH, Folinsbee LJ. 1989. Sulfur dioxide-induced bronchoconstriction in asthmatics exposed for short durations under controlled conditions. In: Susceptibility to inhaled pollutants (Frank R, Utell MJ, eds). Philadelphia, PA: ASTM, 244 p.

Horstman DH, Seal E, Jr., Folinsbee LJ, Ives P, Roger LJ. 1988. The relationship between exposure duration and sulfur dioxide-induced bronchoconstriction in asthmatic subjects. Am Ind Hyg Assoc J 49(1):38-47.

Horvath I, Hunt J, Barnes PJ, Alving K, Antczak A, Baraldi E, *et al*. 2005. Exhaled breath condensate: methodological recommendations and unresolved questions. Eur Respir J 26(3):523-548.

Hunt A, Abraham JL, Judson B, Berry CL. 2003. Toxicologic and epidemiologic clues from the characterization of the 1952 London smog fine particulate matter in archival autopsy lung tissues. Environmental Health Perspectives 111(9):1209-1214.

IPCS. 1994. Environmental Health Criteria 170: Assessing Human Health Risks of Chemicals: Derivation of Guidance Values for Health-Based Exposure Limits. Geneva: International Programme on Chemical Safety, World Health Organization.

---. 2004. Principles for modelling dose-response for the risk assessment of chemicals DRAFT Geneva. :World Health Organization, . Available: http://www.who.int/ipcs/methods/harmonization/draft_document_for_comment.pdf [accessed June 2006].

---. 2005. Chemical-Specific Adjustment Factors for Interspecies Differences and Human Variability: Guidance Document for use of Data in Dose/Concentration - Response Assessment. Harmonization Project Document No.2. Geneva: International Programme on Chemical Safety, World Health Organization.

Johnston FH, Kavanagh AM, Bowman DM, Scott RK. 2002. Exposure to bushfire smoke and asthma: an ecological study. Med J Aust 176(11):535-538.

Jones-Otazo HA, Diamond ML, Richardson GM. 2005. An interagency comparison of screening-level risk assessment approaches. Risk Anal 25(4):841-853.

Jones PW. 1993. Measurement of health-related quality of life in asthma and chronic airways disease.: Dordrecht: Boston: Kluwer Academic Publishers, 301-320.

Katestone Scientific. 1998. Peak-to-mean ratios for odour assessments. Sydney: A report from Katestone Scientific to Environment Protection Authority of New South Wales.

Katsouyanni K, Touloumi G, Samoli E, Gryparis A, Monopolis Y, LeTertre A, *et al*. 2002. Different convergence parameters applied to the S-PLUS GAM function. Epidemiology 13(6):742-743.

Kehrl HR, Roger LJ, Hazucha MJ, Horstman DH. 1987. Differing response of asthmatics to sulfur dioxide exposure with continuous and intermittent exercise. Am Rev Respir Dis 135(2):350-355.

Kharitonov SA, Yates D, Robbins RA, Logan-Sinclair R, Shinebourne EA, Barnes PJ. 1994. Increased nitric oxide in exhaled air of asthmatic patients. Lancet 343(8890):133-135.

Kimmel CA. 2002. Use of Toxicological Data in Estimating Reference Values for Risk Assessment. Human and Ecological Risk Assessment 8(6):1303-1313.

Krewski D, Bakshi K, Garrett R, Falke E, Rusch G, Gaylor D. 2004. Development of acute exposure guideline levels for airborne exposures to hazardous substances. Regul Toxicol Pharmacol 39(2):184-201.

Leisenring W , Ryan L. 1992 Statistical properties of the NOAEL. Reg. Toxicol. Pharmacol. 15(2):161-171.

Lewandowski TA, Rhomberg LR. 2005. A proposed methodology for selecting a trichloroethylene inhalation unit risk value for use in risk assessment. Regulatory Toxicology and Pharmacology 41(1):39-54.

Logan WPD. 1953. Mortality in the London Fog Incident, 1952. The Lancet:336-338.

Lucas RM, McMichael AJ. 2005. Association or causation: evaluating links between 'environment and disease'. Bull World Health Organ 83(10):792-795.

Marks GB, Dunn SM, Woolcock AJ. 1992. A scale for the measurement of quality of life in adults with asthma. J Clin Epidemiol 45(5):461-472.

Maynard R, Krewski D, Burnett RT, Samet J, Brook JR, Granville G, *et al*. 2003. Health and air quality: directions for policy-relevant research. J Toxicol Environ Health A 66(16-19):1891-1904.

McDonnell WF, Horstman DH, Hazucha MJ, Seal E, Jr., Haak ED, Salaam SA, *et al*. 1983. Pulmonary effects of ozone exposure during exercise: dose-response characteristics. J Appl Physiol 54(5):1345-1352.

McDonnell WF. 1993 Utility of controlled human exposure studies for assessing the health effects of complex mixtures and indoor air pollutants. Environ Health Perspect. Dec;101 Suppl 4:199-203

Meek ME, Bucher JR, Cohen SM, Dellarco V, Hill RN, Lehman-McKeeman LD, *et al*. 2003. A Framework for Human Relevance Analysis of Information on Carcinogenic Modes of Action. Critical Reviews in Toxicology 33(6):591-653.

Miller FJ, Schlosser PM, Janszen DB. 2000. Haber's rule: A special case in a family of curves relating concentration and duration of exposure to a fixed level of response for a given endpoint. Toxicology 149(1):21-34.

Ministry for the Environment (NZ). 2002. Review of Odour Management in New Zealand. Ministry for the Environment, New Zealand,. Available: http://www.mfe.govt.nz/publications/air/odour-tr-aug02.pdf [accessed 14 June 2006].

Morgan G, Lincoln D, Lumley T, Shepeard V, Bears J, Jalaludin B. 2003. Time scale effects of particulate air pollution and mortality in Sydney, Australia, 1994-2000. Epidemiology 14(5):S111.

National Research Council (2006) Committee on Human Biomonitoring for Environmental Toxicants, Human Biomonitoring for Environmental Chemicals. Available: http://www.nap.edu/catalog/11700.html

NHC. 2004. A Guide to Health Impact Assessment: A Policy Tool for New Zealand. Public Health Advisory Committee, National Health Council of New Zealand, March 2004. Available: http://www.nhc.govt.nz/PHAC/publications/GuideToHIA.pdf [accessed 14 June 2006].

NHMRC. 1999. Toxicity Assessment for Carcinogenic Soil Contamination. Canberra: National Health and Medical Research Council, Commonwealth of Australia. Available http://www.nhmrc.gov.au/publications/_files/eh21.pdf

NHMRC. 2000. How to review the evidence: Systematic identification and review of the scientific literature. Canberra: National Health and Medical Research Council, Commonwealth of Australia; Available http://www.nhmrc.gov.au/publications/_files/cp65.pdf

NICNAS. 2001. PEC 21 benzene. Canberra: Department of Health. Available: http://www.nicnas.gov.au/Publications/CAR/PEC/PEC21.asp [accessed 9/6/06].

NRC. 1994. Science and Judgement in Risk Assessment. Washington, D.C.: National Research Council, National Academy Press.

---. 2001. Standing operating procedures for developing acute exposure guideline levels for hazardous chemicals. Washington, D.C.: Subcommittee on Acute Exposure Guideline Levels, Committee on Toxicology, Board on Environmental Studies and Toxicology, Commission on Life Sciences, National Research Council, National Academy Press.

NSW Department of Environment and Conservation. 2005. Approved methods for the modelling and assessment of air pollutants in New South Wales. Sydney: New South Wales, Department of Environment and Conservation,.

OEHHA. 1999. Acute toxicity summary determination of an acute reference exposure level for airborne toxicants. Sacramento: Office of Environmental Health Hazard Assessment, California EPA

Oxman AD, Cook DJ, Guyatt GH. 1994. Users' guides to the medical literature. VI. How to use an overview. Evidence-Based Medicine Working Group. JAMA 272(17):1367-1371.

Paustenbach D, Galbraith D. 2006. Biomonitoring: is body burden relevant to public health? Regul Toxicol Pharmacol. Apr;44(3):249-61.

Pellegrino R, Viegi G, Brusasco V, Crapo RO, Burgos F, Casaburi R, *et al*. 2005. Interpretative strategies for lung function tests. Eur Respir J 26:948–968.

RATF. 2000. Report. Canberra: National Environmental Protection Council. Available: http://www.ephc.gov.au/pdf/Air_Quality_NEPM/RATF/RATF_aug_2000_ draft.pdf.

Renwick AG, Lazarus NR. 1998. Human Variability and Noncancer Risk Assessment: An Analysis of the Default Uncertainty Factor. Regulatory Toxicology and Pharmacology 27(1):3-20.

Rich DQ, Mittleman MA, Link MS, Schwartz J, Luttmann-Gibson H, Catalano PJ, *et al*. 2006. Increased risk of paroxysmal atrial fibrillation episodes associated with acute increases in ambient air pollution. Environmental Health Perspectives 114(1):120-123.

Roger LJ, Kehrl HR, Hazucha M, Horstman DH. 1985. Bronchoconstriction in asthmatics exposed to sulfur dioxide during repeated exercise. J Appl Physiol 59(3):784-791.

Roosli M, Kunzil N, Braun-Fahrlander C, Egger M. 2005. Years of life lost attributable to air pollution in Switzerland: dynamic exposure-response model. Int. J. Epidemiol. 34: 1029-1035.

Rozman KK. 2000. The role of time in toxicology or Haber's c x t product. Toxicology 149(1):35-42.

Rusch G. 2006. Chapter 2: The development and application of acute exposure guideline levels for hazardous substances. In: Inhalation toxicology (Salem H, Katz SA, eds). Boca Raton, FL: CRC/Taylor & Francis, 1034 p.

Sakai T. 2000. Biomarkers of lead exposure. Ind Health. 2000 Apr;38(2):127-42.

Salem H, Katz SA. 2006. Inhalation toxicology. Boca Raton, FL: CRC/Taylor & Francis.

Samet JM, Dominici F, Zeger SL, Schwartz J, Dockery DW. 2000. The National Morbidity, Mortality, and Air Pollution Study. Part I: Methods and methodologic issues. Res Rep Health Eff Inst(94 Pt 1):5-14; discussion 75-84.

Sandrini A, Ferreira IM, Gutierrez C, Jardim JR, Zamel N, Chapman KR. 2003. Effect of montelukast on exhaled nitric oxide and nonvolatile markers of inflammation in mild asthma. Chest 124(4):1334-1340.

Sauder LR, Green DJ, Chatham MD, Kulle TJ. 1987. Acute pulmonary response of asthmatics to 3.0 ppm formaldehyde. Toxicol Ind Health 3(4):569-578.

Scherer G. 2005. Biomonitoring of inhaled complex mixtures - Ambient air, diesel exhaust and cigarette smoke. Experimental and Toxicologic Pathology 57:75-110.

Schwartz J. 2000. Harvesting and long term exposure effects in the relation between air pollution and mortality. Am J Epidemiol 151(5):440-448.

Simpson R, Denison L, Petroeschevsky A, Thalib L, Williams G. 2000. Effects of ambient particle pollution on daily mortality in Melbourne, 1991-1996. J Expo Anal Environ Epidemiol 10(5):488-496.

Simpson R, Williams G, Petroeschevsky A, Best T, Morgan G, Denison L, *et al*. 2005a. The short-term effects of air pollution on hospital admissions in four Australian cities. Aust N Z J Public Health 29(3):213-221.

---. 2005b. The short-term effects of air pollution on daily mortality in four Australian cities. Aust N Z J Public Health 29(3):205-212.

Simpson RW, Williams G, Petroeschevsky A, Morgan G, Rutherford S. 1997. Associations between outdoor air pollution and daily mortality in Brisbane, Australia. Arch Environ Health 52(6):442-454.

Smith AD, Cowan JO, Brassett KP, Herbison GP, Taylor DR. 2005. Use of exhaled nitric oxide measurements to guide treatment in chronic asthma. New England Journal of Medicine 352(21):2163-2173.

Stavem K, Aaser E, Sandvik L, Bjornholt JV, Erikssen G, Thaulow E, *et al*. 2005. Lung function, smoking and mortality in a 26-year follow-up of healthy middle-aged males. Eur Respir J 25(4):618-625.

Stieb DM, Judek S, Burnett RT 2002. Meta-analysis of the time series studies of air pollution and mortality: Effects of gases and particles and the influence of cause of death, age and season. J. Air Waste Manage. Assoc. 52: 470-484

Stieb DM, Doiron MS, Blagden P, Burnett RT. 2005. Estimating the public health burden attributable to air pollution: An illustration using the development of an alternative air quality index. J Toxicol. Env. Hlth Pt A 68: 1275-1288.

Tashkin DP, Detels R, Simmons M, Liu H, Coulson AH, Sayre J, *et al*. 1994. The UCLA population studies of chronic obstructive respiratory disease: XI. Impact of air pollution and smoking on annual change in forced expiratory volume in one second. Am J Respir Crit Care Med 149(5):1209-1217.

ten Berge WF, Zwart A, Appelman LM. 1986. Concentration-time mortality response relationship of irritant and systemically acting vapours and gases J Hazrd Mater 13:301-309.

The International Study of Asthma and Allergies in Childhood (ISAAC) Steering Committee. 1998. Worldwide variation in prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and atopic eczema: ISAAC. Lancet 351:1225-1232.

Tunnicliffe WS, Burge PS, Ayres JG. 1994. Effect of domestic concentrations of nitrogen dioxide on airway responses to inhaled allergen in asthmatic patients. Lancet 344(8939-8940):1733-1736.

U.K. Department of Environment. 2000. The air quality strategy for England, Scotland, Wales and Northern Ireland. Working together for clean air. London: Department of Environment, Transport and the Regions, in partnership with the Scottish

Executive, The National Assembly of Wales and the Department of the Environment for Northern Ireland. Section 9.

US EPA. 1989. Review of the national ambient air quality standard for ozone: Assessment of scientific and technical information EPA450-2/92. . Washington, D.C.: United States Environmental Protection Agency.

---. 2000. Supplementary Guidance for Conducting Health Risk Assessment of Chemical Mixtures EPA/630/R-00/002. Washington, D.C.: Risk Assessment Forum, United States Environmental Protection Agency.

---. 2002. A review of the reference dose and reference concentration processes. EPA/630/P-02/002F. Washington, D.C.: Risk Assessment Forum, United States Environmental Protection Agency.

---. 2003. Benchmark Dose Software version 1.3.2. Washington, D.C.: National Center for Environmental Assessment, Available at http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=20167 [accessed 9/6/06].

---. 2004. Introduction to Air Toxics Risk Assessment EPA-453-K-04-001A. Washington, D.C.: United States Environmental Protection Agency,.

---. 2005a. Draft guidelines for carcinogen risk assessment EPA/630/P-03/001B. Washington, D.C.: United States Environment Protection Agency.

---. 2005b. Particulate Matter Health Risk Assessment for Selected Urban Areas EPA 452/R-05-007A. Washington, D.C.: United States Environment Protection Agency.

---. 2005c. Review of the National Ambient Air Quality Standards for Particulate Matter: Policy Assessment of Scientific and Technical Information EPA-452/R-05-005a. Washington, D.C.: Office of Air Quality Planning and Standards Staff Paper, United States Environmental Protection Agency,.

---. 2005d. Human Exposure Measurements: National Human Exposure Assessment Survey (NHEXAS) [online]. Available: http://www.epa.gov/heasd/edrb/nhexas.htm.

---. 2006a. The Development of Acute Exposure Guideline Levels (AEGLs). Washington, D.C.: United State Environment Protection Agency. Available: http://www.epa.gov/oppt/aegl/.

---. 2006b. A framework for assessing health risks of environmental exposure to children. DRAFT FOR EXTERNAL REVIEW. DO NOT CITE OR QUOTE. Washington, D.C.: National Centre for Environmental Assessment, United States Environmental Protection Agency.

---. 2006c. Review of the Process for Setting National Ambient Air Quality Standards. Washington, D.C.: United States Environmental Protection Agency,.

Voigt T, Bailey M, Abramson M. 1998. Air pollution in the Latrobe Valley and its impact upon respiratory morbidity. Aust N Z J Public Health 22(5):556-561.

WA Department of Health. 2006. Health risk assessment in Western Australia. Perth: Western Australia Department of Health.

Ware JE, Kosinski M, Keller SD. 1994. SF-36 Physical and Mental Health Summary Scales: A User's Manual. Boston, Massachusetts: The Health Institute, New England Medical Centre.

Weed DL. 2005. Weight of evidence: a review of concept and methods. Risk Anal 25(6):1545-1557.

Whalan JE, Foureman GL, Vandenberg JJ. 2006. Chapter 1: Inhalation Risk Assessment at the Environmental Protection Agency. In: Inhalation toxicology (Salem H, Katz SA, eds). Boca Raton, FL: CRC/Taylor & Francis, 1034 p.

WHO. 1948. Preamble to the Constitution of the World Health Organization as adopted by the International Health Conference, New York, 19-22 June, 1946; signed on 22 July 1946 by the representatives of 61 States (Official Records of the World Health Organization, no. 2, p. 100) and entered into force on 7 April 1948.

---. 2000a. Air quality guidelines for Europe, Second Edition, WHO regional publications. European series, no. 91. Copenhagen: World Health Organization, Regional Office for Europe. Available: http://www.euro.who.int/document/e71922.pdf.

---. 2000b. Quantification of the health effects of exposure to air pollution. Report of a WHO Working Group. EUR/01/5026342. Available http://www.euro.who.int/document/e74256.pdf

---. 2000c. Evaluation and use of epidemiological evidence for environmental health risk assessment. Copenhagen: World Health Organization, Regional Office for Europe. Available: http://www.euro.who.int/document/e68940.pdf.

---. 2001. Environmental Health Criteria 222: Biomarkers in risk assessment: Validity and Validation. Geneva: International Programme for Chemical Safety, World Health Organization.

---. 2003. Health aspects of air pollution with particulate matter, ozone and nitrogen dioxide. Bonn, Germany: World Health Organization.

---. 2005. Effects of Air Pollution on Children's Health and Development, A Review of the Evidence. Geneva: World Health Organization Europe.

---. 2006 Air Quality Guidelines global update 2005 (in press)

Witschi H. 1999. Some notes on the history of Haber's law. Toxicol Sci 50(2):164-168.

GLOSSARY

The items in this glossary are sourced or adapted primarily from the enHealth guidance document (enHealth 2004).

adverse health effect	The change in morphology, physiology, growth, development or life span of an organism that results in impairment of functional capacity or capacity to compensate for additional stress, or an increase in susceptibility to the harmful effects of other environmental influences. Some adaptive changes are not generally considered to be adverse (eg some changes in enzyme levels). In the context of this guidance document on air pollutants, adverse health effects listed by the American Thoracic Society are of particular relevance. These are: changes in lung function, clinical symptoms associated with respiratory function, asthma, chronic obstructive pulmonary disease (COPD), cardiovascular disease, and increased mortality.
agent	Any chemical, physical, biological or social substance or factor being assessed.
air pollutant	Any substance in air that could, in high enough concentration, harm humans, other animals, vegetation, or material.
air pollution	The presence, in the air, of contaminants or pollutant substances that interfere with human health or welfare or produce other harmful environmental effects.
air pollution episode	A period of abnormally high concentration of air pollutants, often due to low winds and temperature inversion, that can cause illness and death.
ambient air	Any unconfined portion of the atmosphere: open air, surrounding air.
asthma	A chronic inflammatory disease of the human respiratory system where the airways narrow, often in response to a 'trigger' such as exposure to an allergen, cold air, exercise, or emotional stress.
association	Statistical dependence between two or more events, characteristics, or other variables. An association is present if the probability of an event or characteristic, or the quantity of a variable, depends upon the occurrence of one or more other events, the presence of one or more characteristics, or the quantity of one or more other variables.
bias	A process resulting in a tendency to produce results that differ in a systematic way from the true values. Also known as

	systematic error.
BMD	Benchmark dose. The dose associated with a given incidence (eg 1%, 5% or 10% incidence) of effect, the benchmark risk, based on the best-fitting dose-response curve.
biomonitoring	Measurement of a contaminant or metabolite in body tissue, fluid, blood, expired air, breast milk and sweat. It is usually used as a marker or indicator of exposure to environmental chemicals.
biomarker	Any measurement reflecting an interaction between a biological system and an environmental agent, which may be chemical, physical or biological. Often used to describe measurements used in biomonitoring.
carcinogen	Chemical, biological or physical cancer-causing agent.
carcinogenesis	The origin, causation and development of tumours. The term applies to all forms of tumours (benign and malignant).
carcinogenicity	The ability to produce tumours (benign or malignant).
causality	The relating of causes to the effects they produce. Most of epidemiology concerns causality, and several types of causes can be distinguished; however epidemiological evidence by itself is insufficient to establish causality, although it can provide powerful circumstantial evidence.
critical health effect	A health effect that is of such significance, in terms of severity or concern, or the number of people likely to be affected, that it drives the risk assessment process leading to an air quality standard. A qualifying factor is that the data defining the dose-response relationships must be sufficiently robust to support the DRM component of the standard-setting process.
chronic obstructive pulmonary disease (COPD)	**A** general term for a group of respiratory tract diseases that are characterised by obstruction or limitation of air flow in the lungs.
confidence	Weight assigned by the evaluator to the quality of the information available to indicate that a chemical possesses certain toxicological properties.
confidence interval	A range of values determined by the degree of presumed random variability in a set of data, within which the value of a parameter (eg the mean) lies, with a specified level of confidence or probability (eg 95%).
confounding factor	A factor that distorts the apparent effect or magnitude of the effect of a study factor or risk. Such factors must be controlled for in order to obtain an undistorted estimate of a given effect.

critical effect(s)	The adverse effect judged to be the most important for setting an acceptable human intake or exposure. It is usually the most sensitive adverse effect (ie that with the lowest effect level) or sometimes a more severe effect, not necessarily having the lowest effect level.
default value	A pragmatic, fixed or standard value used in the absence of relevant data.
developmental toxicity	The ability to produce an adverse effect in embryo, foetus or immature organism, which is induced and/or manifest either prenatally or postnatally before sexual maturity.
disability adjusted life years (DALYs)	For a given health condition, the sum of the years of life lost due to premature mortality in the population and the years lost due to disability for incident cases.
dose	A stated quantity or concentration of a substance to which an organism is exposed over a given time (continuous or intermittent). It is most commonly expressed as the amount of test substance per unit weight of test animal (eg mg/kg body weight). The applied dose is the amount of chemical in contact with the primary absorption boundaries (eg skin, lungs, gastrointestinal tract) and available for absorption. The absorbed dose is the amount crossing a specific absorption barrier (eg the exchange boundaries of skin, lung, and/or digestive tract) through uptake processes. The delivered dose of an organ or cell is the amount of the chemical available for interaction by that organ or cell. The systemic dose is the dose to which the whole, or extensive parts, of the body is exposed. The absorbed dose may not be the systemic dose as substances absorbed in the digestive tract may be removed by the liver and not enter the systemic circulation.
dosage	A general term comprising the dose, its frequency and the duration of dosing. Dosage is properly applied to any rate or ratio involving a dose. Dosages often involve the dimension of time (eg mg/kg/day), but the meaning is not restricted to this relationship.
dose-response	The correlative association existing between the dose administered and the response (effect) or spectrum of responses that is obtained. The concept expressed by this term is indispensable to the identification, evaluation, and interpretation of most pharmacological and toxicological responses to chemicals. The basic assumptions that underlie

	and support the concept are: (a) the observed response is a function of the concentration at a site, (b) the concentration at a site is a function of the dose, and (c) response and dose are causally related. The existence of a dose-response relationship for a particular biological or toxicological response (effect) provides a defensible conclusion that the response is a result of exposure to a known substance.
dose-response assessment	Determination of the relationship between the magnitude of the dose or level of exposure to a chemical and the incidence or severity of the associated adverse effect.
dose-response modelling (DRM)	A dose-response model describes the probability of a specified response from exposure to a specified pathogen in a specified population, as a function of the dose. This function is based on empirical data, and will usually be given in the form of a mathematical relationship.
epidemiology	The study of the distribution and determinants of health-related states or events in specified populations, and the application of the study to the control of health problems.
endpoint	An observable or measurable biological event used as an indicator of the effect of a chemical on a biological system (cell, organism, organ etc.).
expert	An expert has (1) training and experience in the subject area resulting in superior knowledge in the field (2) access to relevant information, (3) an ability to process and effectively use the information, and (4) is recognised by his or her peers or those conducting the study as qualified to provide judgements about assumptions, models, and model parameters at the level of detail required.
exposure	Contact of a chemical, physical or biological agent with the outer boundary of an organism (eg via inhalation, ingestion or dermal contact).
exposure assessment	The estimation (qualitative or quantitative) of the magnitude, frequency, duration, route and extent (for example, number of organisms) of exposure to one or more contaminated media for the general population, for different subgroups of the population, or for individuals.
exposure pathway	The course a chemical or physical agent takes from a source to an exposed organism. An exposure pathway describes a unique mechanism by which an individual or population is exposed to chemicals or physical agents at or originating from a site. Each exposure pathway includes a source or release from a source, an exposure point, and an exposure route. If the exposure point differs from the source, a transport/exposure

	medium (eg air or, in cases of inter-media transfer, media) also is indicated.
exposure response	The quantitative relationship between the magnitude of the exposure to a chemical and the incidence or severity of the associated adverse effect. In the context of this document, the terms 'dose response' and 'exposure response' are used synonymously, so that 'dose' may be expressed in terms of an air concentration or some other exposure surrogate.
exposure route	The way a chemical enters an organism after contact; for example, by ingestion, inhalation, or dermal absorption.
extrapolation	For dose–response curves, an estimate of the response at a point outside the range of the experimental data. Also refers to the estimation of a response in different species or by different routes from that used in the experimental study of interest.
forced vital capacity (FVC)	A pulmonary function test that measures the total amount of air that can be forcibly blown out after full inspiration, measured in litres.
forced expiratory volume in 1 second (FEV_1)	A pulmonary function test that measures the amount of air that can be forcibly blown in one second, measured in litres. Along with FVC, it is considered one of the primary indicators of lung function.
generalised additive modelling (GAM)	Models that assume that the mean of the dependent variable depends on an additive predictor through a non-linear link function. Generalised additive models permit the response probability distribution to be any member of the exponential family of distributions. Many widely used statistical models belong to this general class, including additive models for Gaussian data, nonparametric logistic models for binary data, and nonparametric log-linear models for Poisson data.
genotoxic	Agents for which a direct activity is the alteration of the information encoded in genetic material.
genotoxic carcinogen	A chemical that induces tumours via a mechanism involving direct damage to DNA.
genotoxicity	A broad term describing the ability to produce damage to the genetic material (DNA) of cells or organisms.
guidance values	Values such as concentrations in air or water, that are derived after appropriate allocation of Tolerable Intake (TI) among the possible different media of exposure. Combined exposure from all media at the guidance values over a lifetime would be expected to be without appreciable health risk. The aim of a guidance value is to provide quantitative information from risk assessment for risk managers to enable them to make

	decisions concerning the protection of human health.
guidelines	These provide guidance on how standards or goals may be achieved (eg nutrient management strategies), or how specified environmental problems can be addressed (eg site contamination). Guidelines are not mandatory, they provide a basis for harmonised approaches, and they may stand alone or be part of another NEPM . Guidelines can be used in a number of ways. They can be part of a NEPM and set out the preferred approach to achieving or maintaining an environmental standard. This has advantages including the sharing of resources in the development of management strategies (eg in the control of motor vehicle emissions). As guidelines are not mandatory, they allow jurisdictions to experiment with other approaches, or for small jurisdictions to take a lower-cost but, in their terms, equally effective route.
harvesting	The idea that there is a pool of frail elderly individuals whose death is advanced by a few days following an air pollution event.
hazard	The identification, from animal and human studies, *in vitro* studies and structure identification activity relationships, of adverse health effects associated with exposure to an agent.
health	A state of complete physical, mental and social well-being, and not merely the absence of disease or infirmity (WHO 1948).
health risk assessment	The process of estimating the potential impact of a chemical, biological, physical or social agent on a specified human population system under a specific set of conditions and for a certain timeframe.
health risk management	The process of evaluating alternative actions, selecting options and implementing them in response to health risk assessments. The decision making will incorporate scientific, technological, social, economic and political information. The process requires value judgements, for example, on the tolerability and reasonableness of costs.
LOAEL	Lowest observed adverse effect level. The lowest concentration or amount of a substance found through experiment or observation to cause adverse alterations of morphology, functional capacity, growth, development or life span of target organisms.
lifetime	Covering the average life span of an organism (eg 70 years for humans).
maximum tolerated	The maximum dose that an animal species can tolerate for a

dose	major portion of its lifetime without significant impairment or toxic effect other than carcinogenicity.
meta-analysis	A statistical synthesis of the data from separate but similar (ie comparable) studies, leading to a quantitative summary of the pooled results.
metabolite	A substance that is the product of biochemical alteration of the parent compound in an organism.
mode of action (MoA)	The specific biochemical interaction through which an agent produces its toxicological effect. Description of a mode of action usually includes mention of the specific molecular targets to which the agent binds, such as an enzyme or receptor.
model	A mathematical representation of a biological system intended to mimic the behaviour of the real system, using empirical data and allowing predictions about untested states of the system.
mutagenicity	The ability to produce a permanent, heritable change in the amount or structure of genetic material of cells or organisms.
neurotoxicity	The ability to produce an adverse effect in the central or peripheral nervous system.
NOAEL	No Observed Adverse Effect Level. The highest dose or exposure to a substance administered to a group of experimental animals at which there is an absence of observable effects on morphology, functional capacity, growth, development or lifespan, that are observed at higher dose levels used in the study and considered to be toxic or 'adverse'. Thus dosing animals at the NOAEL should not produce any toxicologically significant differences between the group of chemically exposed animals and an unexposed control group of animals maintained under identical conditions. The definition of a No Observable Effect Level (NOEL) is equivalent, but with the removal of the term 'adverse'. Often the difficult issue in the use of the terms NOEL and NOAEL is in deciding whether a compound-related effect noted in a particular study is necessarily adverse. Alterations of morphology, functional capacity, growth, development or life span of the target organism may be detected that are judged not to be adverse. The NOAEL (and NOEL) are generally expressed in milligrams of chemical per kilogram body weight per day (mg/kg bw/day), but in an inhalational study, the exposure

	may be expressed in terms of the concentration in air.
non-genotoxic	A chemical that induces tumours via a mechanism which does not involve direct carcinogen damage to DNA.
PM_{10}	Particulate Matter 10μm. The fraction of particles passing an inlet with a 50% cut-off efficiency at an aerodynamic diameter of 10μm.
$PM_{2.5}$	Particulate Matter 2.5μm. The fraction of particles passing an inlet with a 50% cut-off efficiency at an aerodynamic diameter of 2.5μm.
public health	The science and art of preventing disease, prolonging life and promoting health through the organised efforts of society.
quality adjusted life years (QALYs)	The number of years of life that would be added by a particular intervention. Each year in perfect health is assigned the value of 1.0, down to a value of 0 for death.
reproductive toxicity	The ability to produce an adverse effect on any aspect of reproductive capacity, function or outcome. It includes effects on the embryo, foetus, neonate and prepubertal organism, and on adult reproductive and neuroendocrine systems.
reference concentration (R*f*C)	An estimate (with uncertainty factors spanning perhaps an order of magnitude) of the daily exposure (usually expressed in mg/m^3) of the general human population (including sensitive sub-groups) that is likely to be without an appreciable risk of deleterious effects during a lifetime of exposure. It is derived from the NOAEL, LOAEL or BMD (BMC) by application of uncertainty factors that reflect various types of data used to estimate the RfC. The term has the same meaning as Reference Dose (RfD), which is more generally applied to doses administered by routes other than inhalation.
risk	The probability that, in a certain timeframe, an adverse outcome will occur in a person, a group of people, plants, animals, and/or the ecology of a specified area that is exposed to a particular dose or concentration of a hazardous agent (ie it depends on both the level of toxicity of the agent and the level of exposure).
safety factor	A single factor or product of several single factors used to derive an acceptable intake, usually in relation to health-related concerns. Safety factors take into account adequacy of the study, interspecies extrapolation, inter-individual variability in humans, adequacy of the overall data base, nature and extent of toxicity, public health regulatory concern and scientific uncertainty.

standard	A measure of environmental quality. It may be a simple numerical standard (eg pollutant concentration >10ppm), area-specific (eg the pH must be within ± 1 units of the average background level), or more complex (eg species diversity index >10). A standard is a quantifiable characteristic of the environment that provides a surrogate for the environmental values that are to be protected. It is a necessary but not always sufficient indicator against which measured environmental quality can be assessed. Standards are used in each Australian jurisdiction to guide programs and assess their success. Participating jurisdictions must adopt the standard, design and implement programs to meet that standard, and follow the standard procedure (ie protocols) to monitor and report achievement.
threshold	The lowest dose or exposure level that will produce a toxic effect and below which no toxicity is observed.
tolerable intake	An estimate of the intake of a substance that over a lifetime is without appreciable health risk. An example is the Reference Dose.
toxicity	The quality or degree of being poisonous or harmful to plant, animal or human life.
uncertainty factor	A numerical factor applied to the no-effect level to derive an exposure level considered to be without appreciable risk to health (the no-effect level is divided by the uncertainty factor). The magnitude of the uncertainty factor depends on the nature of the toxicity observed, the quality of the toxicological data available, and whether the effects were observed in humans or animals.
unit risk	An expression of the incremental risk associated with increase in exposure by a single unit of exposure measure. It is derived from the slope of the linearised dose-response relationship. It is usually expressed in terms such as: incremental risk per $\mu g/m^3$.
variability	Measurable factors that differ (eg height is variable across populations). The major types of variability are temporal, spatial and inter-individual. Variability may be discrete (eg albinism) or continuous (eg body weight); and it may be readily identifiable (eg presence of albinism) or difficult to identify (eg ability to detoxify a particular chemical metabolite).
weight of evidence	Considerations in assessing the interpretation of published

(WoE)	information about toxicity, quality of testing methods, size and power of study design, consistency of results across studies, and biological plausibility of exposure-response relationships and statistical associations.

ACRONYMS

AALG	ambient air level guidelines
AEGL	acute exposure guideline levels
APHEA2	Air Pollution and Health: a European Approach
ATS	American Thoracic Society
ATSDR	Agency for Toxic Substances and Disease Registry
BMD	benchmark dose
COPD	chronic obstructive pulmonary disease
CSAF	chemical specific adjustment factor
CSIRO	Commonwealth Scientific and Industrial Research Organisation
DRM	dose-response modelling
FEV_1	forced expiratory volume in 1 second
fENO	fraction of exhaled nitric oxide
FVC	forced vital capacity
GAM	generalised additive models
IPCS	International Programme on Chemical Safety
LOAEL	lowest observed adverse effect level
*m*BMD	modified benchmark dose
MOA	mode of action
$\mu g/m^3$	micrograms per cubic meter
NEPC	National Environment Protection Council
NEPM	National Environment Protection Measure
NHMRC	National Health and Medical Research Council
NICNAS	National Industrial Chemicals Notification and Assessment Scheme
NOAEL	no observed adverse effect level
NO_2	nitrogen dioxide
NRC	National Research Council (United States)
OEHHA	Office of Environmental Health Hazard Assessment (California, USA)
PBPK	physiologically based pharmacokinetic
ppm	parts per million
PM	particulate matter
QALYs	quality adjusted life years
RATF	Risk Assessment Task Force
RIVM	The (Dutch) National Institute for Public Health and the Environment
SO_2	sulfur dioxide
US EPA	United States Environment Protection Agency
WHO	World Health Organization
WOE	weight of evidence