

General Overview of the Safety Evaluation of Chemicals

PAUL ILLING

1 Introduction

Society demands that chemicals be 'safe'. This requirement is exercised through legislation concerned with the safety of chemicals, usually introduced *post hoc*, in response either to disasters or to perceived inadequacies in the previous legislation. Public perceptions of risk drive priorities and legislative agendas. Generally legislatures are concerned with setting frameworks and setting up bodies to develop detailed requirements, evaluate the evidence and enforce the legislation.

When human health is the issue, the aim is to prevent ill health. In order to do this it is necessary to have means of predicting ill health. Ethically, it is difficult to justify predictive testing in humans unless a substance is intended for administration to humans. Thus tests are normally conducted in laboratory animals as surrogates for humans. Even when testing in humans is possible, preliminary testing in animals is usually required. The bodies concerned with evaluating safety therefore set minimum testing requirements for those who wish to place on the market a chemical substance or to market it for a specified use. These requirements include:

- when and what tests should be conducted (testing strategies);
- the protocols setting out how the tests should be conducted (test methods); and
- the audit procedures to ensure that tests have been properly conducted and reported ('Good Laboratory Practice'; GLP).

Furthermore, the bodies may either evaluate the results of the tests or set guidance as to how the results should be interpreted.

In the context of safety testing, chemicals includes natural products as well as synthetic chemicals. While the general public may consider 'natural' as good (less harmful) and 'synthetic' as bad (more harmful), there is no scientific reason to differentiate between 'natural' and 'synthetic' chemicals. Ricin (from castor beans), the toxins produced by *Cl. botulinus* and a number of mycotoxins are examples of natural materials

of great potency as toxic agents. Generally, therefore, legislation does not differentiate between 'natural' and 'synthetic' chemicals.

One of the main tools used to predict what the potential ill health effects of chemicals might be is animal testing. A second societal demand, sometimes expressed with considerable force by groups within society who are against conducting work with animals, is associated with the welfare of animals used in experimental procedures. The aim is to prevent, or, at least, to minimize experiments on, usually, vertebrates. Generally, society has been less tolerant of animal testing for substances intended for use as cosmetics or cosmetic ingredients than for substances as a whole.

This clash between the demand for safety and the limitations on human experimentation is the reason why testing is undertaken in animals. The requirements concerning animal welfare dictate that, wherever possible, alternatives to animal testing should be sought.

2 Legislation and Regulatory Requirements

Predictive testing is best instigated when a licence or some other form of permission is required before a substance can be used. It may also be required in law prior to marketing even when there is no formal assessment of the data by a regulator. Thus, predictive testing is required for 'new substances' about to be placed on the market. It is also required prior to use for substances intended for specific uses, such as drugs, veterinary medicines, cosmetics/personal care products, food additives, plant-protection products and biocides. In these cases the person or corporate body desiring to market the substance carries out the required testing. Predictive testing almost always includes experiments in animals.

Legislation may also cover the presence of chemicals in air, water or soil. This legislation generally deals with evaluating existing situations rather than new ones and *post hoc* information from studies in humans should be available. Under these circumstances, predictive testing may be required for endpoints not easily addressed through studies in humans. However, in these cases it is usually Government that sponsors and finances this work.

Originally legislation concerning the safety of chemicals was based on the individual national government. Nowadays it is mostly based on regional (*e.g.* European Union [EU]) or international (Organization for Economic Co-operation and Development [OECD]; World Trade Organization [WTO] or United Nations [UN]) groupings.

Two areas of legislation are particularly pertinent to the current debate on alternatives to animal testing as the testing requirements associated with them are currently being revised. The first is that concerned with the placing on the market of chemicals generally. Currently this is governed by EU Council Directive 67/548/EEC.¹ A Council Directive,² implemented in the UK as the 'Notification of New Substances Regulations'³ deals with 'new substances', *i.e.* those placed on the market after 1981. Under the seventh amendment, a dossier, containing, *inter alia*, information on testing in animals, is required prior to placing the substance on the market. Testing is tiered, some testing being conducted before the substance is placed on the market and further testing being required as the amounts placed on the market increase. Currently testing

strategies are given in Directive 92/32/EEC⁴ and in a technical guidance document⁵ and test protocols are given in Annex V of Directive 67/548/EEC.¹ The test protocols are those set up by the OECD.⁶ Tests have to be conducted to these protocols and audited in accordance with the principles of good laboratory practice.⁷ Animal welfare considerations, set out in Directive 86/609/EC,⁸ should be taken into account when deciding on the need for a specific test.

Existing substances are still classified and labelled on the basis of existing information.

The Registration, Evaluation, Authorization and restriction of CHemicals (REACH) proposals⁹ currently before the EU Council and Parliament, intend to bring the regulation of both new and existing substances into a common scheme, and to require that specific information is available under specified conditions. For existing substances this implies those placing on the market the substance or preparations containing it may need to set up consortia to conduct testing to fill data gaps. The REACH proposals also have a testing strategy contained within them, and the test protocols are those set out in the OECD Guidelines. Although currently the classification is that for Directive 67/548/EEC, it is likely that the UNECE Globally Harmonized System (GHS)¹⁰ will be substituted before implementation of the proposals. It is claimed that the REACH proposals should encourage the use of alternatives to animal testing.

The second is the area of ‘cosmetics’ – more properly personal care products. These are governed by Council Directive 76/768/EEC.¹¹ There are lists of acceptable ingredients and of banned or restricted ingredients. The regulator evaluates dossiers on certain ingredients, information on other ingredients and products are evaluated by the manufacturer/formulator placing the product on the market. The testing requirements have been set out in guidance from the Scientific Committee on Cosmetic Products and Non-Food Products (SCCNFP) intended for Consumers, the latest version being the fifth revision.¹² The tests protocols are based on those set out in Annex V of Directive 67/48/EEC, and are supplemented with SCCNFP opinions concerning studies on human volunteers¹³ and mutagenicity/genotoxicity testing.¹⁴ The seventh amendment to this Directive¹⁵ requires that testing of cosmetic products and ingredients in live animals to meet the requirements of the Cosmetics Directive should cease. It gives the European Commission powers to set deadlines after which testing of cosmetics ingredients in animals will not be permitted if the intention is to meet the requirements of the Cosmetics Directive. The Directive sets maxima of a 6 year time limit for short-term testing and a 10 year time limit for repeated dose toxicity, reproductive toxicity and toxicokinetics, with provision that shorter deadlines can be introduced if suitable test methods become available earlier.

One difference between the two sets of Directives is the attitude to animal experimentation. Perhaps society has taken the view that the use of cosmetics and the choice of preparation used is, compared to the generality of the exposure of chemicals, relatively voluntary, and that animal welfare considerations should be given greater prominence in deciding on the need for animal testing. A second difference concerns human exposure, and hence the ability to carry out studies in humans. Studies in humans have to comply with ethical guidelines laid out in the World Medical Association Declaration of Helsinki, as amended and clarified.¹⁶ Generally,

studies in humans on chemicals are not encouraged, and this is reflected in the regulatory requirements for testing of chemicals placed on the market. However, cosmetics (personal care products) are a group of chemicals intended for application to human skin. Thus it is normally possible, with care, to undertake certain tolerance studies for short-term end points in humans, and this is reflected in the regulatory guidance for their testing.

3 The Regulatory Paradigm

Safety evaluation is essentially risk evaluation, with the intent that the risk of ill-health occurring should be acceptably low (*i.e.* 'safe'). The concepts of risk assessment and risk management go back to two key documents published in 1983. The first is the US 'Red Book'¹⁷ and the second is a Royal Society Study Group report.¹⁸ Some definitions are required if the regulatory paradigm is to be understood. A recent set of definitions is that of the OECD and IPCS.¹⁹

3.1 Hazard and Risk

Probably the most important concept in risk analysis is the distinction between hazard and risk.

Hazard: 'The inherent property of an agent or situation having the potential to cause adverse effects when an organism, system or (sub)population is exposed to that agent.'

Risk: The probability of an adverse effect in an organism, system or (sub)population caused under specified circumstances to an agent'.

A good short statement relating these definitions is that 'risk is the possibility of suffering harm from a hazard'.²⁰ In the case of human health, the hazard is the interaction of a toxic agent with a receptor (the human), and the risk is the likelihood of sufficient interaction occurring such that ill-health results. Because the human is a fixed element, the properties of the interaction are taken as the hazardous properties of the agent, for our purposes the chemical substance.

3.2 Risk Assessment and Risk Management

The OECD/IPCS definitions for risk assessment and risk management¹⁹ are given below.

Risk assessment: A process intended to calculate or estimate the risks to a given target organism, system or (sub) population, including the identification of attending uncertainties, following exposure to a particular agent, taking into account the inherent characteristics of the agent of concern as well as the characteristics of the specific target system.

'The risk assessment process includes four steps: hazard identification, hazard characterization (related term: dose-response assessment), exposure assessment and risk characterization. It is the first component in a risk analysis process'.

Risk management: '[A] Decision-making process involving considerations of political, social, economic and technical factors with relevant risk assessment information relating to hazard so as to develop, analyse, and compare regulatory

and non-regulatory options and to select and implement [an] appropriate regulatory response to that hazard.

Risk management comprises three elements: risk evaluation; emission and exposure control; and risk monitoring’.

Safety evaluation of chemicals involves all the stages of risk assessment and the stage of risk evaluation. These are:

Hazard identification: ‘The identification of the type and nature of adverse effects that an agent has the inherent capacity to cause an organism, system or (sub)population’.

Hazard characterization: ‘The qualitative, and wherever possible, quantitative description of the inherent properties of an agent or situation having the potential to cause adverse effects. This should, where possible, include a dose-response assessment and its attendant uncertainties’.

Exposure assessment: ‘Evaluation of the exposure of an organism, system or (sub) population to an agent (and its derivatives)’.

Risk characterization: ‘The qualitative and, whenever possible, quantitative determination, including attendant uncertainties, of the probability of an occurrence of known and potential adverse effects of an agent to a given organism, system or (sub)population under defined exposure conditions’.

Risk evaluation: ‘Establishment of a qualitative or quantitative relationship between risks and benefits of exposure to an agent, involving the complex process of determining *the significance of the identified hazards and estimated risks to the system concerned or affected by the exposure*, as well as the significance of the benefits bought about by the agent’ [my italics].

In the case of human health the ‘given organism’ is human. As risk evaluation involves assessment of the ‘significance of the identified hazards and estimated risks’ to ‘those concerned with or affected by the exposure’ it involves more than a technical assessment of the toxicological and exposure data. ‘Significance to those concerned with or affected by the exposure’ implies that the way in which those concerned with or affected by the exposure perceive the risks, and hence the sociological and psychological factors affecting how people perceive risk become important. Sociological and psychological opinions concerning risk have to be taken into account when evaluating risks.

3.3 The Risk Evaluation Framework

The Royal Society study group¹⁸ first put forward a risk evaluation framework. This framework (the ‘tolerability of risk’ concept) has been restated and slightly developed in ‘Reducing risks, protecting people’.²¹ It was originally developed to handle engineering risk, but it is equally applicable to risks to human health. Illing²² and the Health and Safety Executive (HSE)²¹ have discussed the application of this framework to the evaluation of health risks arising from exposure to chemicals.

Criteria for reaching decisions can be classified according to three ‘pure’ criteria.²¹ These are:

- An **equity based** criterion, which starts from the premise that all individuals have unconditional rights to certain levels of protection. This leads to standards,

applicable to all, held to be usually acceptable in normal life. In practice this leads to fixing a limit to represent the maximum level of risk above which no individual can be exposed. If the risk characterization indicates that the risk is above this limit the risk is held to be unacceptable, whatever the benefits.

- A **utility based** criterion, which applies to the comparison between incremental benefits of measures to prevent the risk of injury or detriment [for health effects, ill health], and the cost of the measures. The utility based criterion compares the relevant benefits (*e.g.* statistical lives saved, life-years extended, reduced ill-health and better quality of life) obtained by adoption of a particular risk prevention measure with the net cost of introducing it, and requires that a balance be struck between the two. This balance can be deliberately skewed towards benefits by ensuring gross disproportion between costs and benefits.
- A **technology based** criterion, which essentially reflects the idea that a satisfactory level of risk prevention is attained when ‘state of the art’ control measures (technological, managerial, organizational) are employed to control risks, whatever the circumstances.

These criteria underlie the regulatory process first outlined by the Royal Society.¹⁸ The scheme is based on:

- an upper limit of risk which should not be exceeded for any individual [‘unacceptable’];
- further control, so far as is reasonably practicable, making allowances if possible for aversions to the higher levels of risk or detriment [‘tolerable’]; and
- a cut-off in the deployment of resources below some level of exposure or detriment judged to be trivial [‘broadly acceptable’]. The scheme is outlined in Figure 1.

This approach to risk evaluation can be applied to health effects. For many health effects, the risk evaluation is concerned only with determining what constitutes a ‘broadly acceptable’ risk, and hence with the equity criterion. This is the case if any exceedance of an equity criterion for ‘safe’ (the ‘broadly acceptable’ level of risk), such as a residue level in a foodstuff, results in its immediate withdrawal from the market. However, the current approach to human medicines clearly makes use of a utility criterion, in that the risks and benefits associated with the use of a drug in patients in general are evaluated by a licensing body. In a specific patient they are evaluated by the person prescribing the drug. In other fields it can also be seen that utility/technology based criteria are employed. ‘As low as is reasonably practicable’ and ‘best available technology’ incorporate such criteria. In addition, the use of reducing limits for air quality indicates that there has been exceedance of the ‘broadly acceptable’ risk, and that time is needed to restore the exposures to those considered to be the maximum ‘broadly acceptable’ risk.

Historically, engineers and others have tried to deal with the application of these criteria using numerical estimates of ‘risk of death’ (risk of foreshortened life due to the exposure), *i.e.* a technical or actuarial approach to risk. Health scientists have usually preferred other more broadly based, but linkable approaches to developing the information to be set against these criteria.^{21–24} These approaches include use of

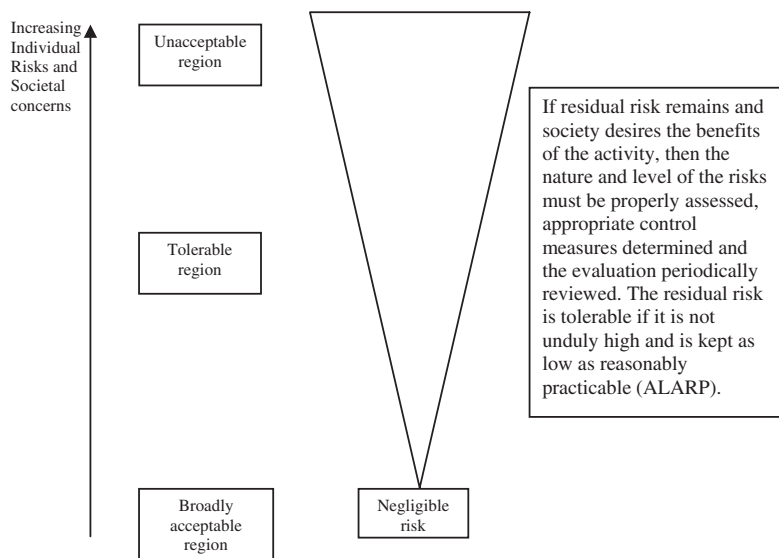


Figure 1 Outline of the Royal Society approach to risk management/'Tolerability of risk' framework (based on Health and Safety Executive²¹)

much lesser parameters of ill health than death, and can include parameters indicative of effects not considered to constitute clinical illness. The uncertainty factors approach and the margin of safety approach, both of which are outlined below, use minimal ill health rather than death as the end point.

3.4 Some Problems with the Engineering/Actuarial Approach to Risk

Until recently, the framework given above was seen as a framework for a technical or actuarial analysis of risk based on premature death as the end point. Engineers usually worked in terms of numbers derived from probabilities and consequences, and actuarial material in terms of causes of death. The framework can also be applied to the health risks associated with chemicals, especially when suitable historical data from human epidemiological studies are available. In 'Reducing risk, protecting people', the HSE²¹ believes that the boundary between 'broadly acceptable' and 'tolerable' risk is an individual risk of death of one in a million per annum for both workers and the public. This is the original criterion given in the Royal Society's 1983 report. A less clear boundary of one in a thousand per annum for workers and one in ten thousand for the general public is considered as the 'just tolerable' risk for any substantial body of people. HSE first stated these numbers in 1992.²⁵ If this approach is pursued for death due to the toxicity caused by exposure to an agent, extremely sensitive epidemiological studies are required to give certainty to the numbers found. The critical numbers for much lesser forms of ill health are ill defined, but will be lower, and it might be possible to measure exposures at which these occur directly.

In the 1983 Royal Society study group report, this was described as ‘objective (statistical) risk’ and they differentiated it from ‘perceived risk’. They recognized that there may be a perplexing disparity between these two.

Social scientists have outlined a number of difficulties with this approach to risk.^{26–29}

Jasanoff²⁶ suggests that scientists adopt three propositions in developing the technical or actuarial approach to determining the risk criteria representing ‘unacceptable’, ‘tolerable’ or ‘broadly acceptable’ risk, namely:

- given enough data, experts will generally agree with each other in their risk assessment;
- the only scientific way to think about risk is essentially in actuarial terms; and
- any other way of thinking about risk is possibly wrong, certainly unscientific and perhaps even antiscientific.

She did, however, admit that biological scientists have generally been more willing to recognize the need for qualitative judgements, and hence the possibility of expert disagreements in risk assessment. She also stated that chemical products, in particular, may be rejected not because they are ‘unsafe’ in any conventional sense, but because the public is insufficiently persuaded that they serve a legitimate social need.

The chapter on risk perception in the 1992 report from the Royal Society Study Group²⁷ states that:

‘Risk perception involves people’s beliefs, attitudes, judgements and feelings, as well as the wider social or cultural values and dispositions that people adopt, towards hazards and their benefits. ... Furthermore, the perception of risk is multidimensional, with a particular hazard meaning different things to different people (depending, for example, upon their underlying value systems) and different things in different contexts. In some circumstances, important aspects of risk perception and acceptability involve judgements not just of the physical characteristics and consequences of an activity but also social and organizational trustworthiness of risk management and regulatory institutions. What is clear is that risk perception cannot be reduced to a single subjective co-ordinate of a particular mathematical model of risk, such as the product of probabilities and consequences, because this imposes unduly restrictive assumptions about what is an essentially human and social phenomenon’.

Slovic²⁹ emphasizes the subjective and value-laden nature of risk assessment. Even a relatively simple measure, such as the end point for expressing the risk, can be complex and judgmental (Table 1). Which is the correct measure?

Clearly, the choice of measure will influence how a risk is perceived and evaluated. A classic case is where deaths per tonne of product drop following increased production through mechanization, but deaths per 100,000 employees remains the same or increases. The former might be a measure preferred by the employer, the latter by an employee/trade unionist. Thus, even what constitutes ‘objective risk’ is open to interpretation; it is framed within a societal judgement concerning the appropriate measure.

Table 1 Ten ways of expressing risk of death

-
- Per million people (employed in the industry)
 - Per million people within a given distance from the source
 - Per unit of concentration
 - Per site or industrial plant
 - Per tonne of toxic substance released to air
 - Per tonne of toxic substance released to air and absorbed by people
 - Per tonne of substance produced (placed on the market)
 - Per million pounds (sterling) of substance produced
 - Loss of life expectancy associated with exposure to hazard
 - Loss of quality assessed life years of expected life
-

Note: The first nine ways are based on Slovic.²⁹ In reality, once born, death is inevitable. Thus all except the last two methods measure the occurrence of premature death due to the substance, usually deaths that occur within a short time of the exposure. The last two measures are estimates of how premature that death may be.

There are also other sociological aspects of risk that affect risk analysis. Anthropologists and cultural sociologists have suggested that social responses to risk are determined by prototypes of cultural belief patterns – clusters of related convictions and perceptions of reality. Renn²⁸ summarizes four such prototypes:

- *Entrepreneurial* – who perceive risk taking as an opportunity to succeed in their personal goals and are less concerned about equity issues and wish government to refrain from extensive regulation and risk management efforts.
- *Egalitarian* – who emphasize cooperation and equality rather than competition and freedom, focus on long-term effects of human activities, are more likely to abandon an activity (even if perceived as personally beneficial) than to take chances and are particularly concerned with equity.
- *Bureaucrat* – relies on rules and procedures to cope with uncertainty and believe that, as long as risks are managed by a capable institution and coping strategies have been provided for all eventualities, there is no need to worry about risks.
- *Atomized or stratified individuals* – principally believe in hierarchy, but miss group identity and a system of social bonding; these people only trust themselves, are often confused about risk issues and take high personal risks while opposing any risk they feel is being imposed on them, and see life as a lottery.

These cultural prototypes have different ‘world-views’. Slovic²⁹ has identified five ‘world-views’:

- fatalism (‘I feel I have very little control over risks to my health’);
- hierarchy (‘Decisions about health risks should be left to the experts’);
- individualism (‘In a fair system, people with more ability should earn more’);
- egalitarianism (‘If people were treated more equally, we would have fewer problems’); and
- technological enthusiasm (‘A high technology society is important for improving our health and social well-being’).

He also suggests that worldview is strongly linked to public perception of risks.

The most obvious inference from this is that there is no common position concerning what constitutes a 'broadly acceptable' risk. What constitutes a 'broadly acceptable' risk varies with cultural prototype and worldview.

Factors that can influence the public perception of what constitutes a 'broadly acceptable' risk have been enumerated repeatedly since first being stated by Otway and von Winterfeldt.³⁰ A recent publication suggests six indicators that can be correlated with public concern.³¹ This set of indicators (see Table 2) was chosen as being manageable, transparent, and representative indicators of public concern which, from the available psychometric research, would correlate well with almost any other set that is likely to be proposed.

Equity of the consequences of the risk implies that the risk of harm and the benefits (rewards) are distributed fairly, either because the same people are affected or because those facing the risk of harm are properly compensated. Control of the risks refers to the perceived level of control that people feel that they have over the risk. The individual may feel the risk is voluntary and under his/her control, or that the risk is imposed and completely outside his/her control. These factors are important when determining what is a 'broadly acceptable' risk. For example, personal control of the risk is very limited when dealing with the imposed risks associated with air quality standards, but considerable when associated with activities where there is considerable choice, such as whether to use personal care products and which product to use.

In order to try to achieve some convergence between 'objective' and 'perceived' risk, a societal input is required when carrying out a risk evaluation. Although the Royal Society study group indicated that there was a risk evaluation phase in risk analysis in 1983, only recently has it been more widely recognized. If risk evaluation is to be undertaken then society, or at least 'stakeholders', parties who are concerned about or affected by the risk management problem, should have input into the evaluation. Some bodies, such as the UK Health and Safety Commission, had this input historically. However, wider recognition of the need for 'stakeholder' involvement is a relatively recent origin.^{32,33} Consequently, until recently, bodies consisting of technical experts have been taking decisions on behalf of society as a whole without adequate societal input. This lack of input into the risk evaluation has probably exacerbated the differences between the technical risk assessors' and the public's perception of the adequacy of the risk management.

Table 2 *General (negative) attributes of hazard that influence risk perception*

<i>Item no</i>	<i>Description</i>
1	Familiarity and experience of the risk
2	Understanding of the cause-effect mechanism
3	Equity of the consequences of the risk
4	Fear of the risk consequences ('Dread')
5	Control of the risk
6	Trust in risk management

In conclusion, toxicologists are attempting to provide information capable of being used to describe a risk, *i.e.* information concerned with 'objective' risk. However, there is a difference between 'objective' and 'perceived' risk that needs bridging. Understanding how risks are perceived and communicating risk information, as applied to toxicological risks arising from chemicals, are still in their infancy. Nevertheless, it is essential that communication between all those involved in the risk assessment and management process takes place. 'Stakeholder' involvement, transparency and openness when managing risks are all attempts to involve and inform those concerned with or affected by decisions. There is room for 'pressure groups' to gather individuals from cultural prototypes that are dissatisfied to exert influence on decisions concerning risk. This includes groups concerned about the effects of chemicals on the environment.

4 The Interface between Toxicology and Risk Assessment

4.1 Introduction

Health scientists have, perhaps at times unwittingly, used the paradigm described by the Royal Society,¹⁸ but allied it to less mathematical approaches to setting toxicological data against the paradigm. This was recognized by Jasanoff,²⁶ who admitted that scientists concerned with biological effects have generally been more willing [than scientists trained in engineering risk] to recognize the need for qualitative judgements, and hence the possibility of expert disagreements, in risk assessment. When 'gate-keeping' (authorizing/licensing), health science based regulators have taken decisions concerning the safety of chemicals (other than chemicals used as medicines and workplace chemical exposures), often they have done so on a 'broadly acceptable' criterion framed within an equity only paradigm. They have also minimized the likelihood of public criticism by skewing decisions based on predictive testing against 'safe' chemicals exhibiting possible minor concern during testing in order to prevent a licence being granted to a dangerous substance not exhibiting much cause for concern (conservatism).^{22,24} This could be called a precautionary approach, and may have saved them from the more extreme criticisms associated with the technical (actuarial) approach. However, the approach employed is more dependent on judgement and hence less transparent.

There are two main sets of circumstances for interpretation where risk is being examined: setting maximum level of exposure deemed still 'broadly acceptable' and determining the 'margin of exposure'. The former (the uncertainty factors approach) requires hazard information (hazard identification and hazard characterization) and a standardized risk evaluation process and determines the maximum exposure considered 'broadly acceptable' for a particular set of exposure circumstances. The latter requires both hazard information (hazard identification and characterization) and exposure information (frequency, level and duration of exposure).

When the 'broadly acceptable' exposure level cannot be attained without risk management measures, a consideration will be needed of the extent to which risk management measures can reduce exposure, either to the 'broadly acceptable' level or to a level deemed 'tolerable'. In essence, the 'margin of safety' is deemed too small to be 'broadly acceptable' without further consideration, and that further

consideration is an iterative process of risk evaluation, emission and exposure control and risk monitoring. In REACH this is the evaluation, authorization and restriction element of the proposals.

4.2 *The Uncertainty Factors Approach to 'safe' for Threshold Effects*

The usual procedure for predictive health risk assessment has been based on animal testing, the 'No Adverse Observed (or Observable) Effects Level' (or surrogate) and 'Uncertainty Factors'. The aim is to develop a numerical value for the level of exposure or detriment judged to be trivial (*i.e.* judged to be a 'broadly acceptable' level of risk, or a safe level of exposure). This dose is normally referred to as the 'Reference dose' when limit setting, although the International Programme on Chemical Safety (IPCS) in its 1994 document³⁴ called it the 'tolerable intake'. The procedure, and the factor of 100 for use in extrapolating from the no-observed-adverse-effect-level (NOAEL) in animal tests to the reference dose, was first formally enunciated in the 1950s for use with food additives and contaminants.³⁵

The 'Reference dose': 'An estimate of the daily exposure dose that is likely to be without deleterious effects even if continued exposure occurs over a lifetime'.

The 'Uncertainty factors': Reductive factor by which an observed or estimated NOAEL is divided to arrive at a criterion or standard that is considered safe or without appreciable risk'.¹⁹

The 'No observed adverse effect level': The highest observed dose or concentration of a substance at which there is no detectable adverse alteration in morphology, functional capacity, growth, development or life span of the target'.^{34,36}

In this approach the hazards are characterized and then set against pre-established risk evaluation criteria to determine the maximum exposure consistent with a 'trivial' or 'broadly acceptable' risk, *i.e.* a 'safe' exposure.

The REACH proposal seeks to obtain a 'Derived no effect level' (DNEL). The 'DNEL' is defined as: 'The level of exposure to a substance above which humans should not be exposed'.

It is derived using the NOAEL and applying uncertainty factors to it.⁹ The DNEL is therefore a version of the reference dose. However, the Annex does acknowledge there may be a need for several DNELs to represent different exposure circumstances.

This approach mixes risk characterization with risk evaluation.

4.3 *The 'Margin of Exposure' ('Margin of Safety' [OECD First Definition]) Approach*

Health scientists also use the 'margin of exposure' approach to examining whether a chemical is 'safe'.

'Margin of exposure': 'The ratio of the 'NOAEL' for the critical effect to the theoretical, predicted or estimated exposure dose or concentration'.¹⁹

This definition is also one of the two definitions used for the 'margin of safety'.¹⁹ In this approach the risk characterization is separated from the risk evaluation. The development of the numerical value for the margin of exposure is the risk characterization. The decisions concerning its adequacy, either generally or for a specified use and either

without further risk management measures or following application of risk management measures, are the risk evaluations.

Within the European Community, the present Technical Guidance Document⁵ for use with new and existing chemicals and biocides and the SCCNFP notes for guidance for use with cosmetics¹² use this ‘margin of exposure’ for risk assessment purposes, although the latter calls it a ‘margin of safety’.

4.4 *Margin of Safety (OECD Second Definition)*

The OECD¹⁹ defines two meanings for the term ‘margin of safety’.

- The first definition is that ‘margin of exposure’ and ‘margin of safety’ are synonymous.
- The second definition is that the ‘margin of safety’ is the margin between the reference dose and the actual exposure dose.

Annex 1 of the REACH proposal⁹ includes use of the ‘DNEL’, obtained from the no observed adverse effect level and uncertainty factors, and then sets exposure data against this DNEL. It therefore uses the ‘margin of safety’ (OECD second definition) approach and rolls risk characterization and risk evaluation together.

4.5 *Carcinogenicity (and other Non-threshold [also called Deterministic or Stochastic] Effects)*

In the UK, health scientists have deemed certain effects, such as genotoxic carcinogenicity, to be of a type where the hazard is such that there is no ‘safe’ level of exposure. Exposure is eliminated or reduced ‘as low as is reasonably practicable’.³⁷ Substitution of a different level of effect to the NOAEL, such as ‘serious effect’ into the uncertainty factors process described above results in a different maximum level of exposure, and hence level of risk. This is only possible when dealing with circumstances where utility/technology criteria are tolerated. If a sufficiently serious level of effect is used then the level of exposure would be that for the cut-off of the upper limit of risk, which should not be exceeded for any individual, and banning/prevention of exposure would be required. If the level of risk lies within the ‘tolerable’ region then the ‘as low as is reasonably practicable’ approach is feasible. In the US and elsewhere, extrapolation is undertaken using numerical approaches based on mathematical modelling, *i.e.* the approach so strongly criticized in the Royal Society study group²⁷ and by Jasanoff.²⁶ UK Regulators, whose views are based on a 1991 set of Guidelines,³⁸ do not generally accept such approaches. A more detailed discussion of this point is available.^{36,39} This is an area where further developments are needed in order to develop a single, common approach.

In the EU, under Directive 67/548/EEC¹ carcinogens are classified into three categories according to the nature and the quality of evidence. Potency is not specifically considered. Under the proposed REACH regulation, carcinogens in categories 1 and 2 will enter the evaluation, authorization and restriction process. The seventh amendment to the cosmetics directive bans the use of substances that are classified

as carcinogenic, mutagenic or toxic to reproduction and listed in Annex 1 to Directive 67/548/EEC, unless, for category 3 only, the SCCNFP has found the use acceptable. Thus the UK approach is being followed, namely that, if the evidence that the substance is carcinogenic is adequate, exposure is eliminated or reduced 'as low as is reasonably practicable'.

4.6 *Additional Comments*

Two of the three approaches (uncertainty factors and margin of safety [OECD second definition] approaches) are intended to examine the question of whether a risk is 'broadly acceptable', and hence an exposure 'safe' directly. These approaches include a standardized risk evaluation. Only the 'margin of exposure' approach is a risk characterization, with the risk evaluation (the decision concerning acceptability) clearly differentiated from the risk characterization. It therefore separates risk characterization and risk evaluation. Although the 'margin of exposure' approach is used for cosmetics, the SCCNFP indicates that it is generally accepted that the margin should be at least 100 to declare a substance safe for use. The SCCNFP is therefore providing a standardized risk evaluation, a statement identifying what margin of exposure is generally accepted by society as being associated with the maximum 'broadly acceptable' level of risk.

In principle, there should be some form of 'stakeholder' involvement both in the selection of the generalized approach to what constitutes a 'broadly acceptable' risk and in its applicability to the specific substance and exposure circumstance. In the past this has not been a feature of risk evaluation; the technical experts have taken the decision on behalf of society.

Historically a single approach to a 'reference dose' and use of common default uncertainty factors has been considered by many toxicologists to be appropriate for all exposure circumstances. However, the critical effect may depend on the type of exposure encountered (intermittent, continuous, short-term, long-term) and may need to be re-calculated according to the exposure circumstances being investigated. In addition, although the Royal Society study group identified the risk evaluation stage in its 1983 report, this stage has only become recognized by health scientists generally following the analysis carried out when developing the OECD definitions published in 2003. The setting of a 'reference dose' (for example, an occupational exposure limit) or a 'DNEL' implies a risk evaluation, and therefore societal judgements. Thus, use of a single set of defaults for all circumstances may be inappropriate.

It has long been recognized that occupational exposure limits developed separately and tended to use lower numerical values for uncertainty factors.^{22-24,34,40-43} As the approach to occupational exposure limits developed separately from the other arenas, and employed 'stakeholder involvement', it is possible that the sociological factors affecting what constitutes a 'broadly acceptable' risk, and hence the risk evaluation were different from those for other risk evaluations. This emphasizes the need to clearly differentiate between risk characterization and risk evaluation and to understand the non-technical factors being taken into account in the risk evaluation.

The European Commission REACH proposal recognizes that several DNELs may be required.

The actual levels of risk considered ‘tolerable’ or ‘unacceptable’ involve a societal judgement that should be taken through a political process. They therefore vary according to the circumstances surrounding the exposure,^{22,24,36} and that variation also has to be incorporated into the risk evaluation. The evaluation, authorization and restriction process within REACH is intended to consider the circumstances when a generally applicable ‘broadly acceptable’ risk is exceeded.

5 Interpreting Toxicology

5.1 Introduction

There are two main sets of circumstances for interpretation where risk is being examined: setting maximum level of exposure deemed still ‘broadly acceptable’ and determining the ‘margin of safety’. The former requires hazard information (hazard identification and hazard characterization) and a standardized risk evaluation process, and determines the maximum exposure considered ‘broadly acceptable’ for a particular set of exposure circumstances. The latter requires both hazard information (hazard identification and characterization) and exposure information (frequency, level and duration of exposure).

When the ‘broadly acceptable’ exposure level cannot be attained without risk management measures, a consideration will be needed of the extent to which risk management measures can reduce exposure, either to the ‘broadly acceptable’ level or to a level deemed ‘tolerable’. In essence, the ‘margin of safety’ is deemed too small to be ‘broadly acceptable’ without further consideration, and that further consideration is an iterative process of risk evaluation, emission and exposure control and risk monitoring.

Medicines are a special case and have to be considered separately. In population terms there is a ‘margin of exposure’ style of approach to the problem. These substances are intended to have effects deemed beneficial, and the exposure level is the intended therapeutic dose level. Risk and benefit are considered together, often with reference to the requirements of the patient, which may be somewhat different from those of the normal individual.

There is a third set of circumstances involving the interpretation of toxicology. However, hazard, not risk, is being examined. Classification (and labelling) depends on the hazard identification and characterization, and involves setting the hazard information against criteria and deciding on the appropriate classification.

5.2 Types of Information

Hazard information relevant to establishing the safety of a chemical in terms of potential human ill health includes:

- physicochemical data;
- structure activity relationships;
- toxicity studies *in vitro* and *in vivo*; and
- human studies (case reports, experimental studies and epidemiology).

In vivo toxicity studies are studies carried out in animals or humans. Although the best species for studying effects in humans is humans, ethical considerations¹⁶ limit the type and amount of human testing that can be undertaken. Thus animal studies are required, either as a preliminary to tests in humans or, particularly in the case of longer-term studies, because studies involving deliberate administration to humans would be unethical and/or impractical. Toxicity studies may cover some or all of the following end points:

- acute toxicity, including skin and eye irritancy;
- genotoxicity;
- repeated dose toxicity;
- reproductive toxicity;
- carcinogenicity;
- neurotoxicity; and
- immunotoxicity (including skin sensitization testing).

The need for and timing of the toxicity studies required for regulatory testing to determine safety predictively will depend on the regulatory regime. Thus, the testing strategy for a particular regulatory regime will need to identify appropriate studies and when they should be conducted.

5.3 *The Key Pieces of Information*

Traditionally predictive risk assessment for ill health effects has been based on the results of animal studies. For acute toxicity the key information is either (for older studies) the dose (or concentration) resulting in death of 50% of the animals exposed (LD₅₀ or LC₅₀), or, for more recent studies, 'evident toxicity' or predetermined thresholds.

For repeat dose toxicity studies the key piece of information is the NOAEL or NOAEC, Lowest Observed Adverse Effect Level (LOAEL) or 'Benchmark dose' (BMD). Essentially these are aimed at obtaining similar information, a measure of the borderline between ill health effects being seen/not seen. The NOAEL has been defined above. When exposure is by inhalation the concentration (for a specified time) may be substituted for the dose level, in which case the term becomes 'No Observed Adverse Effect Concentration'.

The 'LOAEL': 'The lowest concentration or amount of a substance, found by experiment or observation, which causes an adverse alteration of morphology, functional capacity, growth development or life span of the organism distinguishable from a normal (control) organism of the same species and strain under the same defined conditions of exposure'.

The 'BMD': 'The lower confidence limit of the dose calculated to be associated with a given incidence (e.g. 5 or 10% incidence) of effect estimated from all toxicity data on that effect within that study'.³⁴

Protocols usually indicate a 'limit dose', a maximum quantity that should be dosed, in order to avoid testing at unduly unrealistic dose levels.

The information available will depend on the type of substance being investigated. For 'existing chemicals' there should be a body of knowledge, either in the scientific

literature or the company archives, already extant. That knowledge may have been acquired from studies not to current standards and may refer to end points required under REACH. It may also be knowledge based on studies of relevant analogues and structure-activity relationships. It may also include human information. If that material is relevant then Annex IX to REACH indicates how it can be incorporated into the risk assessment/evaluations in order to minimize new animal testing. For 'new substances' including new substances for specified uses, it is unlikely that there will be pre-existing knowledge. Thus, predictive information has to be gathered. That information generally includes information from studies in animals. The methods for characterizing and/or evaluating the risk make use of information obtained from animal tests.

For hazard characterization (*i.e.* classification and labelling) the toxicity information is set against pre-determined criteria to yield a classification. The REACH scheme is likely to inherit the system used for new and existing chemicals. The criteria used for acute toxicity, including irritancy, and repeated dose toxicity take dose into account the dose–effect relationship through the use of more than one classification criterion. Studies on sensitization, genotoxicity, reproductive toxicity and carcinogenicity are usually aimed at obtaining 'yes/no' answers. In these cases the result at the 'limit dose' defines whether the substance does or does not require classification in respect of that particular form of toxicity. For skin sensitization, the move from guinea pig tests to the mouse local lymph node assay offers the opportunity of moving from yes/no answers to a consideration of potency.

The aims behind obtaining these data are:

- to classify the substance and establish whether the substance is dangerous; and/or
- to obtain a 'reference dose' and hence to set standards or determine whether exposure is acceptable.

When dealing with 'broadly acceptable' risk ('safe' exposure), three approaches may be used. They are:

- the 'uncertainty factors' approach;
- the 'margin of exposure' approach; and
- the 'margin of safety' (OECD second definition)/DNEL approach.

5.4 The 'Uncertainty Factors' (UFs) Approach

When dealing with the standard setting approach to risk evaluation, the hazard data are set against pre-determined evaluation criteria to obtain a numerical value for the maximum exposure considered 'safe' – the OECD 'reference dose' (RfD) or the European Community 'DNEL'.

$$\text{RfD (DNEL)} = \frac{\text{NOEL (or LOAEL or BMD, etc.)}}{\text{UFs}}$$

The way in which the reference dose is established involves defining what is the critical effect and which study is the pivotal study. The procedure then followed is set out by IPCS.^{34,36}

The critical effect is the ill-health effect that yields the lowest reference dose. The pivotal study is the study used to determine the no observed adverse effect level for that effect. It may be necessary to examine several effects in order to determine which effect is the critical effect. Critical effect may also depend on route and duration of exposure. The critical effect for an airborne exposure following a major accident is likely to be an effect following acute inhalation exposure, one for food contaminants may be associated with long-term low level oral exposure.

Uncertainty factors cover the following:

- inter-species extrapolation (default value 10);
- inter-individual variability in humans (default value 10);
- route-to-route extrapolation (if required; generally discouraged);
- LOAEL to NOAEL (default 3, 5 or 10);
- duration of study (for transfer from a 28-day study to a 90-day reference exposure duration the factor used in the EU is 3);
- nature of toxicity/severity of effect (for severe and irreversible phenomena such as teratogenicity and non-genotoxic carcinogenesis an extra factor of up to 10 is used); and
- adequacy of database (usually 1, but up to 100).^{22,24,34,36,42,43}

The historic 100 factor is based on 10 for inter-species variation and 10 for inter-individual variation. All the other defaults are set at 1. It is based on an adequate database of studies, no route-to-route extrapolation, an adequate NOAEL from a study of the appropriate duration and a conventional threshold effect not deemed a severe/ irreversible effect.

If information on toxicokinetics or toxicodynamics is available, it may be used to modify the defaults for inter-species and inter-individual variation. For inter-individual differences each counts equally, *i.e.* the factor of 10 is divided into two factors of 3.2. For inter-species variation the factors used are weighted differently, with a factor of 4 being used for toxicokinetics and a factor 2.5 for toxicodynamics.

A summary of this approach is set out diagrammatically in Figure 2.

Illing^{22,24} has made the case that there is a societal input into uncertainty factors. The evidence for this factor is associated with risk assessment for occupational exposure, which developed separately from other toxic risk assessments, and where considerably lower overall uncertainty factors have been deemed acceptable.^{40,41,43} Part of the reasoning for this difference lies in the population at risk, which excludes children, the elderly and the ill. Another part of the reasoning is associated with how the risk is seen:

- the exposure can be seen as relatively (compared with food and drinking water contaminants and general ambient air quality) voluntary – there is a choice concerning whether to work and where to work; and
- the risk has to fit into a mixed equity/tolerability system, whereas food and general ambient air quality seek equity alone, and therefore can be more conservative in their approaches.^{22,23}

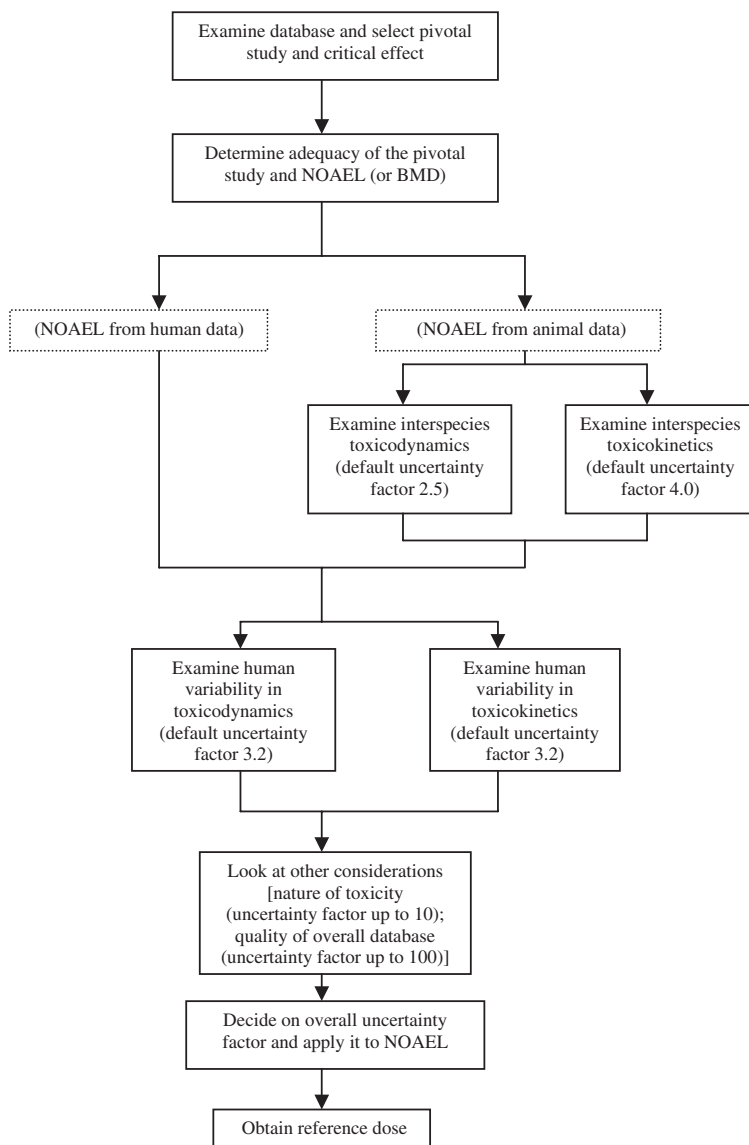


Figure 2 Procedures for the derivation of uncertainty factors^{22,34}

This conservatism acted as a ‘comfort factor’ for the technical experts taking decisions concerning the acceptability of the risk, a decision taken by the technical experts concerning how the proposed exposure level would be viewed by society.

Historically, with the exception of the Health and Safety Commission and Advisory Committees, UK Government committees advising on human health issues associated with chemicals were made up of technical experts. At about the

same time as the 1998 report of the Royal Commission on Environmental Pollution, additional non-technical (lay) members were appointed to these bodies. Thus stakeholders are now included within the UK-based decision-making process. At international level, stakeholder involvement is achieved by making the technical body recommend to a political body that accepts the recommendation on behalf of the states who have a stake in the decision.

5.5 *The Margin of Exposure Approach*

The margin of exposure (MoE) is:

$$\text{MoE} = \frac{\text{NOEAL (or LOAEL or BMD, etc.)}}{\text{Theoretical, predicted or estimated exposure.}}$$

The MoE therefore requires exposure information as well as hazard information. Although actual exposure information is preferable to theoretical or predicted information, it is frequently not available. Thus modelled information is used. In the case of cosmetics the ‘systemic exposure dosage’ (SED), the amount expected to enter the blood stream (and therefore be systematically available) per kg body weight per day is set against the NOAEL. As the NOAEL is usually derived from repeated dose, carcinogenicity or reproductive toxicity tests conducted using oral administration, with an assumption of complete oral uptake, it also is given also in amount/kg bw/day.

The value of this approach is the clear distinction between risk assessment and risk evaluation. The principal problem is that a suitable exposure model is required. Such models are difficult to generate. Manufacturers often are not completely aware of how downstream users and consumers are handling the chemicals and how the chemicals will enter waste streams. Often, several conservative assumptions are included in the process used to develop an exposure assessment. These assumptions need to be validated as they can easily render an exposure assessment unrealistic.

5.6 *The Margin of Safety (OECD second definition) Approach*

The Margin of safety (MoS) is calculated from the equation:

$$\text{MoS} = \frac{\text{NOEAL (or LOAEL or BMD, etc.)}}{\text{UFs} \times \text{Theoretical, predicted or estimated exposure.}}$$

or:

$$\text{MoS} = \frac{\text{RfD (DNEL)}}{\text{Theoretical, predicted or estimated exposure.}}$$

This approach is a mixture of both the processes given above and has the problems associated with each of them.

5.7 *Medicines*

Although the ‘margin of exposure’ approach is used for therapeutic agents, it is modified somewhat to allow for the special circumstances surrounding administration of

therapeutic substances. In the case of medicines, risk and benefit (the therapeutic effect) are seen in the same individual. Furthermore, the therapeutic agent may itself be toxic or carcinogenic (as with many anti-cancer agents) or (as with anti-diabetic agents) the therapeutic effect (lowering blood glucose) may be appropriate in the patient but totally inappropriate in the normal individual. Thus more sophisticated evaluations are required, frequently including considerations concerning the extension of life expectancy and quality of life for the individual. The risks may only be tolerable in the circumstances associated with the individual. There is a two-step evaluation process. A regulatory body deals with the general principle concerning acceptability/tolerability and licences the agent and, where the risks are significant, the patient's medical practitioner advises the patient and prescribes the medicine.

5.8 Classification (and Labelling)

Classification is essentially an attempt at a simple system of hazard characterization. The classification system attached to Directive 67/548/EEC¹ is used to classify substances. Although this system is the system currently applied in the REACH proposal, it is likely that the GHS,¹⁰ which, although similar, has some minor differences, will be substituted before REACH is implemented. Classification may be carried out by the EU, in which case any requirement to label the substance will be contained in Annex 1 of Directive 67/548/EEC. Otherwise manufacturers/importers are expected to classify and label the substance in line with requirements set out in Annex VI to Directive 67/548/EEC (the current version is Commission Directive 2001/59/EEC).⁴⁴

The classification in terms of severity of effect covers acute toxicity, including skin and eye irritation, and repeated dose toxicity. Acute toxicity is divided into four classes: not classified, harmful, toxic and very toxic, on the basis of the dose at which effects are seen. Classification on grounds of skin and eye irritancy includes three classes: not classified, irritant and corrosive/severe irritant. Likewise classification on grounds of repeated dose toxicity includes three classes: not classified, harmful and toxic. Other effects (mutagenicity, carcinogenicity and reproductive toxicity) are classified on grounds of nature and quality of information, with category 1 requiring evidence in humans, category 2 good evidence in animals and category 3 limited evidence. As the appropriate carcinogenicity and reproductive toxicity test methods include limit doses, there is a cut-off above that at which any effects seen do not result in classification.

5.9 Non-Standard Information

Currently the general practice for the Directives concerned with new and existing chemicals has been to retain a formal animal test and to use physicochemical, structure activity and *in vitro* information to decide whether the formal regulatory test is required. With the exception of genotoxicity, information from 'alternatives' to the regulatory test has been used for elimination of animal-based studies only where the decision is that an outcome requiring classification would result if the animal test were conducted. Information from 'alternatives' can also be taken by the regulator as reasons to indicate that an animal study should be conducted earlier (*i.e.* at a lower threshold tonnage) than

would otherwise be the case. Thus, currently, the general case is that ‘alternatives’ to animal tests are only acceptable when a positive outcome is predicted.

There are two exceptions to this generalization concerning ‘alternatives’. First, it has been possible to use ‘alternatives’ as part of a package to ‘read across’ a test on a related substance to the substance being considered, and thus, on a case-by-case basis to delay or eliminate some animal testing under the current schemes for chemicals placed on the market. The second case is genotoxicity. For genotoxicity, a full set of negative *in vitro* tests is regarded as indicative that the substance is not genotoxic.

In Annex IX to REACH⁹ the case-by-case approach to use of structure-activity relationships and ‘read across’ is formalized. It also introduces the possibility of exposure-driven toxicity testing, *i.e.* the possibility that, if there is minimal or no exposure, some tests in animals can be omitted. Nevertheless, the ultimate arbiter concerning the acceptability of this material is the regulatory authority. The REACH proposals are intended to minimize animal testing by restricting the tests undertaken until the substance is placed on the market in substantial tonnage or has exhibited some alert as to potential toxicity.

The seventh amendment to the Cosmetics Directive¹⁵ is aimed at avoiding animal testing entirely when developing personal care products. Possibly, society has also taken the view that the use of cosmetics and the choice of preparation used is relatively voluntary, and that animal welfare considerations should be given greater prominence in deciding on the need for animal testing. For cosmetics it is possible to obtain some information in humans. Thus the decision has been taken that there will be no need to conduct detailed tests in animals. A timetable is given for dates after which animal testing for various end points will no longer be permitted if the substance is for use as a cosmetic. Physicochemical parameters, structure–activity relationships, and *in vitro* testing will yield adequate information to permit short-term testing in humans and to eliminate enough of the substances likely to cause effects following prolonged dosing (carcinogenicity, mutagenicity and reproductive toxicity). This is an alternative approach not yet considered acceptable for substances generally.

All of these proposals lead away from the precision implied in the current animal test based models for toxic risk assessment/evaluation. Research is needed into whether the detail obtained by the regulatory test in animals is required in all circumstances and how alternatives to animal tests can be taken into the regulatory system. In addition, it is possible that less sophisticated risk assessment and evaluation processes may be all that is required for adequate risk management. For example, one approach to the control of substances hazardous to health (COSHH) in the workplace, set out in COSHH Essentials⁴⁵ has moved towards ‘easier to understand’ systems for risk management. These systems are based on the classifications into the categories of hazard used for classification of chemicals under Directive 67/548/EEC, coupled with assessments of exposure. This system is nominally more transparent than the previous system.

5.10 Comments

A more extensive discussion of how toxicology information is set within risk frameworks and/or classification frameworks is available.^{39,46}

Generally, physicochemical data and data from structure–activity relationships gives preliminary suggestions as to possible ill health. While some end points are now accepted as being amenable to *in vitro* studies, generally regulators still see a need for animal studies. In doing so, regulatory authorities are acting as ‘gatekeepers’ and are cautious. As part of the need to minimize the likelihood of public criticism they are likely to be precautionary and request more tests in order to ensure that any identified hazards are defined as closely as possible. This tends to involve more and more detailed animal studies, so there is a clear need to take animal welfare into consideration. Where exposure is more ‘voluntary’, as with cosmetics/personal care products, regulators are prepared to consider the elimination of testing in animals. This could be because they can accept that, as exposure is voluntary and some testing can be conducted in humans, there is less need for certainty in the non-human element of the risk assessment.

5.11 Some Problems with Animal Testing

All testing has to have objectives. Ideally, the best species for studying human health effects is humans. However, there are ethical requirements that mean testing in humans can be unacceptable. Animal testing is required either as a preliminary to testing in humans or as a substitute for testing in humans. Animal tests are used as general (and therefore coarse) screens for various types of toxicity, or detailed testing for specific end points. Experience is used to guide what is tested for. This implies that the testing is directed at the known, and the unknown can be missed. Once a new form of potential toxicity has been identified, the need for and ability to test it have to be assessed and, if necessary, current test procedures modified or a new test developed and a criterion for decisions concerning the effects developed. At present, this is the situation for endocrine disrupters where no formal test procedure and criteria for classification have been developed.

There is conflict of interest between the current, animal test based toxic risk assessment/evaluation procedures and the ability to depend on alternatives to animal testing. At present, the risk assessment/evaluation procedures encourage more and more detailed testing in order to acquire more detailed information. As REACH is bringing existing chemicals into regulation, the information available for these chemicals includes exposure information and information on effects in humans not available or, at least, not as readily available when dealing with new substances. Annex IX is intended to deal with that situation.

There are also ethical considerations associated with human testing and ethical and animal welfare considerations associated with animal testing. It is probable that different cultural prototypes will have different attitudes to the ethics and/or the scientific value of human and animal experimentation. ‘Pressure groups’ are a valid way by which to demonstrate that there are ethical and welfare considerations associated with animal experimentation. However, there is a point at which the activities of the more extreme of these animal welfare groups interferes with the freedoms of those conducting what society as a whole considers legal and proper information gathering. The less extreme of these pressure groups seek alternatives to animal testing.

Research on alternatives to animal testing includes research aimed at the 3 Rs (reduce, refine and replace). Some progress has been made on all three Rs and both the

European Centre for the Validation of Alternative Methods (ECVAM) and the Fund for the Replacement of Animals in Medical Experiments (FRAME) are keen to develop and validate methods based on them. An assessment of ways to improve the OECD test guidelines has been published by the Fund for the Replacement of Animals in Medical Experiments (FRAME).⁴⁷ They conclude that opportunities for streamlining individual assays are very limited, but improvements could be achieved by

- only undertaking studies that provide relevant data;
- making greater use of screens and preliminary testing;
- applying some tests simultaneously to the same animals;
- using only one sex; and
- eliminating redundant tests.

This suggests that there is only limited scope for improvements to the REACH proposals for new substances. However, perhaps more importantly, Combes *et al.*⁴⁸ have proposed a strategy for the testing of chemicals. They suggest that there should be much more use of exposure (or tonnage)-driven testing and use of non-animal methods in an integrated testing scheme. They also indicate that such schemes can be developed. The option of reducing testing in animals and using exposure information, structure–activity relationships, toxicogenomics and early biomarkers in humans in place of animal testing is clearly more applicable to existing substances and needs to be examined closely.

Alternatives generally give less detailed information on health effects. Given the conservatism of regulatory authorities when developing testing strategies it can take a considerable number of years to achieve international agreement that a test is unnecessary or a newer test is better. This ‘drag’ should be minimized; thus it is appropriate that the pressure towards alternatives exists.

6 Conclusions

Traditionally, toxicological risk assessment and evaluation has relied on physico-chemical information, animal testing and human data, where available. When information is required prior to placing of a substance on the market or prior to its being marketed for a specific use, predictive testing generally includes substantial testing in animal models. Testing in humans is only ethically acceptable in restricted circumstances. Historically, the information derived from toxicity tests (and other data) has been combined with use of uncertainty factors to develop a ‘reference dose’. Uncertainty factors are used to identify a ‘safe’ level of exposure. Many toxicologists have considered that common default uncertainty factors are appropriate for all exposure circumstances. However, the setting of a ‘reference dose’ (for example, an occupational exposure limit) or a ‘DNEL’ implies a risk evaluation, and therefore societal judgements. This societal judgement has to be incorporated into the uncertainty factors used. The actual levels of risk considered ‘tolerable’ or ‘unacceptable’ depends on cultural prototype and attitude to risk. Uncertainty factors therefore vary according to the circumstances surrounding the exposure and that variation also has to be incorporated into the risk evaluation.

Understanding how risks are perceived and communicating risk information, as applied to toxicological risks arising from chemicals, are still in their infancy. Nevertheless, it is essential that communication between all those involved in the risk assessment and management process takes place. There is room for 'pressure groups' to gather individuals from cultural prototypes that are dissatisfied to exert influence on decisions concerning risk. This includes groups concerned about the effects of chemicals on the environment.

The clash between the demand for safety and the requirements concerning animal welfare is one major reason for seeking alternatives to animal testing when seeking to prevent ill health due to exposure to chemicals. Given the conservatism of regulatory authorities when developing testing strategies it can take a considerable number of years to achieve international agreement that a test is unnecessary or a newer test is better. This 'drag' should be minimized; thus it is appropriate that the pressure towards alternatives exists.

The option of reducing testing in animals and using exposure information, structure-activity relationships, toxicogenomics and early biomarkers in humans in place of animal testing needs to be examined closely. All of these proposals lead away from the precision implied in the use of the NOAEL derived from current animal test-based models for toxic risk assessment/evaluation. Research is needed into whether the detail obtained by the regulatory test in animals is required in all circumstances and how alternatives to animal tests can be taken into the regulatory system.

References

1. Directive 67/548/EEC.
2. Originally the 'sixth amendment' directive, Directive 79/831/EEC; now the 'seventh amendment' Directive, Directive 92/32/EEC.
3. Originally SI 1982/1496, re-enacted as SI 1993/3050.
4. Directive 92/32/EEC.
5. Technical Guidance Document on Risk Assessment in support of Commission Directive 93/667/EEC on Risk Assessment for New Notified Substances, Commission regulation (EC) no 1488/94 on Risk Assessment for Existing Substances Directive 98/8/EC of the European Parliament and of Council concerning the placement of biocidal products on the market. Institute for Health and Consumer Protection, European Chemicals Bureau. EUR 2048 EN. Luxemburg: European Communities (available from www.ecb.jrc.it/).
6. OECD Guidelines for the testing of chemicals, 1981 and updates. Paris: OECD.
7. Directive 1999/11/EEC; OECD Principles of Good Laboratory Practice, 1982 and updates.
8. Directive 86/609/EC.
9. Proposal for a Regulation of the European Parliament and of the Council concerning the registration, evaluation, authorization and restriction of chemicals (REACH), establishing a European Chemicals Agency and amending Directive 1999/45/EEC and Regulation (EC) (on persistent organic pollutants) COM (2003) 644 final. Proposal for a Directive of the European Parliament and Council amending Council Directive 67/548/EEC in order to adapt it to

- Regulation (EC) of the European Parliament and of the Council concerning the registration, evaluation, authorization and restriction of chemicals. Luxembourg: Commission of the European Communities.
10. www.unece.org/trans/danger/publi/ghs/officialtexr.html.
 11. Directive 76/768/EEC.
 12. The SCCNFP's Notes of Guidance for the Testing of Cosmetics Ingredients and their Safety Evaluation. Fifth revision. Adopted by the SCCNFP during the 25th plenary meeting of 20 October 2003. SCCNFP/0690/03 Final.
 13. Opinion-Guidelines on the use of human volunteers in the testing of potentially cutaneous irritant cosmetic ingredients or mixtures of ingredients adopted by the plenary session of the SCCNFP of 25 November 1998. SCCNFP/0003/98 Final. Opinion concerning guidelines on the use of human volunteers in compatibility testing of finished cosmetic products – adopted by the Scientific Committee on Cosmetics and Non-food Products intended for consumers during the plenary session of 23 June 1999. SCCNFP/0068/99 Final. Opinion concerning basic criteria of the protocols for the skin compatibility testing of potentially cutaneous irritant cosmetic ingredients or mixtures of ingredients in human volunteers. Adopted by the SCCNFP during the plenary session of 8 December 1999. SCCNFP/0245/99 Final.
 14. SCCNFP Notes of Guidance. Recommended mutagenicity/genotoxicity tests for the safety testing of cosmetic ingredients to be included in Annexes to Council Directive 76/768/EEC adopted by the SCCNFP on 23 April 2004 by written procedure. SCCNFP/0755/03.
 15. Directive 2003/15/EEC.
 16. World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects, 1964. (available at www.wma.net/e/policy/b3.htm).
 17. National Research Council. *Risk assessment in the Federal Government: Managing the process*, National Academy Press, Washington, DC, 1983.
 18. Royal Society Study Group. *Risk assessment*, Royal Society, London, 1983.
 19. Descriptions of key generic terms used in chemical hazard/risk assessment. OECD Series on testing and Assessment no 44. ENV/JM/MONO(2003)15. Organization for Economic Cooperation and Development, Paris. Also: IPCS Risk Assessment Terminology, Harmonisation Project Document No. 1. Interagency Organisation for the Sound Management of Chemicals, WHO Geneva, 2004.
 20. J.J. Cochrane, V. T. Covello. *Risk Analysis: A guide to the principles and methods for analysing health and environmental risks*, National Technical Information Service, Springfield, VA, 1989.
 21. Health and Safety Executive. Reducing risks, protecting people. HSE Books, Sudbury, Suffolk, 2001.
 22. P. Illing, *Toxicity and Risk: Context, principles and practice*, Taylor and Francis, Basingstoke, 2001.
 23. HPA. Illing, *Human Exper. Toxicol.*, 1991, **10**, 215.
 24. HPA. Illing. *Regulatory Toxicol. Pharmacol.*, 1999, **29**, 300.
 25. Health and Safety Executive. *The tolerability of risk from nuclear power stations*, HSE Books, Sudbury, Suffolk, 1992.

26. S. Jasanoff, in *Risk management of chemicals*, M. L. Richardson (ed), Royal Society of Chemistry, Cambridge, UK, 1992, 337.
27. Royal Society Study Group. *Risk: Analysis, perception and management*, Royal Society, London, 1992.
28. O. Renn. *J. Risk Res.*, 1998, **1**, 49.
29. P. Slovic. *Risk Anal.*, 1999, **19**, 689.
30. H.J. Otway and D. von Winterfeldt. *Policy Sci.*, 1982, **14**, 247.
31. HM Treasury. *Managing Risks to the public: Appraisal guidance. Draft for consultation*, October 2004. HM Treasury, London, 2004.
32. Presidential/Congressional Commission on Risk Assessment and Risk Management. Framework for environmental health risk management. Final Report, Volume 1. Commission on Risk Assessment and Risk Management, Washington, DC, 1997.
33. Royal Commission on Environmental Pollution. Setting environmental standards. Twenty first report. The Stationary Office, London, 1998.
34. IPCS. Assessing human health risks of chemicals: Derivation of Guidance values for health based limits. Environmental Health Criteria 170. World Health Organisation, Geneva, 1994.
35. A.J. Lehman and O.G. Fitzhugh *Q, Bull Assoc Food Drug Officials*, 1954, **18**, 33.
36. IPCS. Principles for the assessment of risks to human health from exposure to chemicals. Environmental Health Criteria 210. World Health Organization, Geneva, 1999.
37. Assessment of chemical carcinogens: Background to general principles of a weight of evidence approach. CR8. Interdepartmental Group on Health Risks from Chemicals. Institute for Environment and Health, Leicester, 2002.
38. Committee on Carcinogenicity of Chemicals in Food, consumer Products and the Environment. Guidelines for the evaluation of chemicals for carcinogenicity. *Reps. Health Social Subjects*, 1991, **42**, 1.
39. H.P.A. Illing, Risk Management, in *Fundamental Toxicology for Chemists*, 2nd edn, J. H. Duffus and H.G.J. Worth (eds), Royal Society of Chemistry, Cambridge, UK, 2006.
40. H.P.A Illing. *Ann. Occup. Hyg.*, 1991, **35**, 569.
41. S. Fairhurst. *Ann. Occup. Hyg.*, 1995, **39**, 375.
42. Uncertainty factors: Their use in human health risk assessment by UK Government. CR9. Interdepartmental Group on Health Risks from Chemicals. Institute for Environment and Health, Leicester, 2003.
43. T. Vermeire, H. Stevenson, M. Pieters, M. Rennen, W. Slob and B.C. Hakkert. *Crit. Rev. Toxicol.*, 1999, **29**, 439.
44. Commission Directive 2001/59/EEC.
45. Health and Safety Executive. COSHH Essentials: Easy steps to control chemicals,. HSG193. HSE Books, Sudbury, Suffolk, 1999.
46. H.P.A. Illing, *Risk Assessment*. In 'Fundamental Toxicology for Chemists', 2nd edn, J.H. Duffus and H.G.J. Worth (eds), Royal Society of Chemistry Cambridge, (in press).
47. R.D. Combes, I. Gaunt and M.A. Balls, *ALTA*, 2004, **32**, 163.
48. R. Combes, M. Barratt and M. Balls, *ALTA*, 2003, **31**, 7.