

***Framework for the Integration
of Human and Animal Data
in Chemical Risk Assessment***

Technical Report No. 104

ISSN-0773-8072-104
Brussels, January 2009

ECETOC TECHNICAL REPORT No. 104

© Copyright – ECETOC AISBL

European Centre for Ecotoxicology and Toxicology of Chemicals
4 Avenue E. Van Nieuwenhuyse (Bte 6), B-1160 Brussels, Belgium.

All rights reserved. No part of this publication may be reproduced, copied, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without the prior written permission of the copyright holder. Applications to reproduce, store, copy or translate should be made to the Secretary General. ECETOC welcomes such applications. Reference to the document, its title and summary may be copied or abstracted in data retrieval systems without subsequent reference.

The content of this document has been prepared and reviewed by experts on behalf of ECETOC with all possible care and from the available scientific information. It is provided for information only. ECETOC cannot accept any responsibility or liability and does not provide a warranty for any use or interpretation of the material contained in the publication.

Framework for the Integration of Human and Animal Data in Chemical Risk Assessment

CONTENTS

SUMMARY	1
1. INTRODUCTION	2
1.1 Background	3
1.2 Purpose	6
1.3 Scope	6
2. OVERVIEW OF DIFFERENT TYPES OF HUMAN DATA	7
2.1 Non-experimental data	7
2.1.1 <i>Case reports and series</i>	7
2.1.2 <i>Routinely collected data</i>	8
2.1.3 <i>Non-experimental epidemiology studies</i>	8
2.2 Experimental data	10
3. QUALITY ASPECTS OF HUMAN DATA	11
3.1 Study design	11
3.1.1 <i>Appropriateness of the study design</i>	11
3.1.2 <i>Appropriateness of the comparison group</i>	11
3.1.3 <i>Time between exposure and appearance of health effects</i>	12
3.1.4 <i>Adjustment for risk factors</i>	12
3.1.5 <i>Bias</i>	13
3.1.6 <i>Appropriate sample size</i>	13
3.1.7 <i>Appropriate statistical analysis</i>	14
3.2 Exposure information	14
3.3 Health effect data	16
3.4 General acceptability of the conclusions	17
3.5 Specificity of the association	17
4. HUMAN DATA QUALITY CRITERIA FOR RISK ASSESSMENT	19
4.1 Hazard identification	20
4.1.1 <i>Bradford Hill guidelines</i>	20
4.1.2 <i>Meta-analysis</i>	21
4.1.3 <i>Pooled analysis</i>	21
4.1.4 <i>Summary</i>	21
4.2 Lead or critical effect	22
4.3 Dose-response assessment	22

4.4	Scheme for scoring human data with regard to its quality	23
4.4.1	<i>Pre-requisites</i>	24
4.4.2	<i>Assessing the intrinsic quality of the data</i>	25
4.4.3	<i>Considering the nature of the health effect and deriving a quality score</i>	27
4.4.4	<i>Adjustments to the scheme: Positive versus null human data</i>	28
4.4.5	<i>Consideration of the risk assessment stage</i>	28
5.	QUALITY ASPECTS OF ANIMAL STUDIES	30
5.1	Assessment of reliability	30
5.2	Relevance and utility	31
5.3	Information gained from animal studies	32
6.	RELEVANCE OF ANIMAL DATA IN HUMAN RISK ASSESSMENT	34
6.1	Forming a mode of action hypothesis	35
6.2	Dosimetry	35
6.3	Considering relevance to man	36
6.4	Dose-route extrapolation	37
7.	INTEGRATION OF HUMAN AND ANIMAL DATA	38
8.	PROPOSED FRAMEWORK FOR THE USE OF HUMAN AND ANIMAL DATA	40
8.1	Implementation of the framework for selected case studies	40
8.1.1	<i>Scoring the quality of human data</i>	40
8.1.2	<i>Relevance of animal data</i>	42
8.1.3	<i>Integration of human and animal data</i>	43
8.2	General observations on the case studies	44
8.3	Concordance or lack of concordance between human and animal data	46
9.	CONCLUSIONS AND RECOMMENDATIONS	49
	BIBLIOGRAPHY	51
	LIST OF SPECIAL ABBREVIATIONS	63
	APPENDIX A: TERMS OF REFERENCE	64

APPENDIX B: EXAMPLES REPRESENTING DIFFERENT RISK ASSESSMENT AND RISK MANAGEMENT APPLICATIONS	65
B.1 Stepwise approach	65
B.2 Welding fumes	71
B.3 Wood dust	73
B.4 Sulphanilic acid	76
B.5 Reactive dye	79
B.6 Benzanthrone	82
B.7 Mectins	85
B.8 Glycidol	88
B.9 Methylene chloride (dichloromethane)	91
B.10 <i>bis</i> (Chloromethyl)ether	95
B.11 Hydrogen fluoride	97
B.12 Human monoclonal antibody ‘TeGenero TGN1412’	101
B.13 Dimethylcarbamoyl chloride	105
B.14 Carbon disulphide	108
B.15 Toluene	111
APPENDIX C: CRITERIA FOR RELIABILITY CATEGORIES (FOR ANIMAL DATA)	115
APPENDIX D: DERIVING EXPOSURE LIMITS	116
MEMBERS OF THE TASK FORCE	122
MEMBERS OF THE SCIENTIFIC COMMITTEE	123

SUMMARY

Human data form the most direct evidence for an association between health effects and exposure to chemicals. The availability and quality of human data vary greatly from one chemical to another; this may be strongly related to the prevalence of exposure and to concern about potential health effects. Guidance is currently available on the evaluation and use of animal toxicological data and human exposure data in the risk assessment process. However, such specific guidance is not available for human health effects, despite the fact that most international authorities recognise that the incorporation of human data would improve the utility and robustness of the risk assessment process.

Consequently, ECETOC identified the need to review and evaluate the different types of human data that are available, and to provide guidance on how such data could be used best in the risk assessment process. A multidisciplinary Task Force was thus assembled to address the problem and to consider in particular, when and where human data could be used to support risk assessment and risk management decisions, and how human and animal findings could be integrated and used in tandem.

Quality aspects play an important role in the choice of data sources regarding the leading health effect that will be crucial in the risk assessment process. Thus, quality aspects of human data, as well as of animal data, have been extensively addressed in this report.

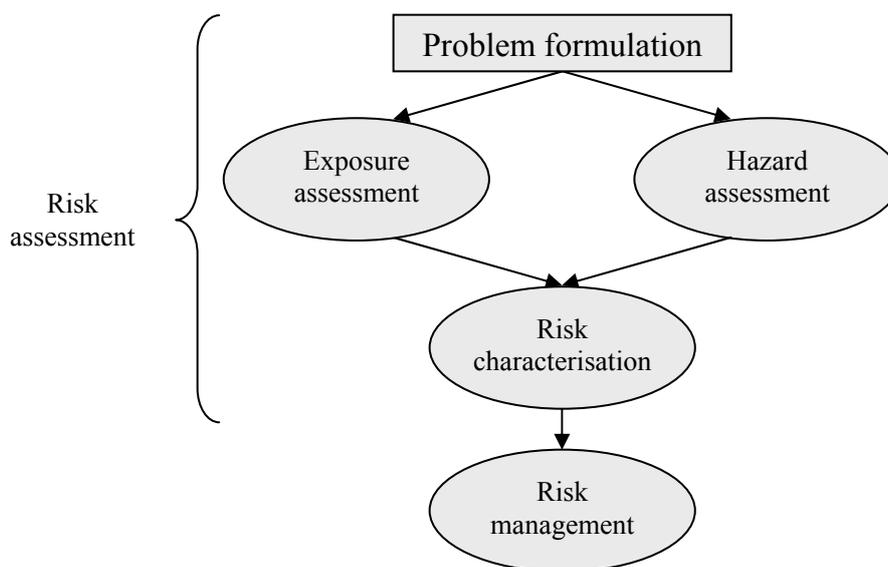
Following the description of the quality aspects of the human and animal data, a framework for the integration of these data and their use in the risk assessment process is proposed. The framework takes into account human as well as animal data; it is strongly encouraged to use both sources in a combined approach. Ideally, human data and animal data will be complementary and should confirm each other (i.e. both indicate excess risk, or both indicate the absence of risk). In cases where they are in apparent contradiction, efforts should be made to develop a better understanding of the biological basis for the contradiction. This will often be informative and result in a more reliable basis for risk assessment.

With this report, ECETOC provides guidance on how human data can be used and integrated into chemical risk assessment and management processes. The proposed framework is illustrated by a number of examples.

1. INTRODUCTION

The principles of risk analysis have been well described by a number of bodies and there is now broad consensus on the steps that constitute this process of risk assessment and risk management (NRC, 1983). Figure 1 summarises the key elements. The hazard assessment portion of the risk assessment can be further sub-divided into identification of the hazard and leading (critical) effect, and determination of the dose-response relationship.

Figure 1: The risk analysis process



All risk assessments must have a purpose. The first step in the risk assessment process is the formulation of the question that the risk assessment is intended to answer. For example, is a particular substance hazardous for human health? If it is, then are present-day exposures such that they constitute a risk? How might the hazards be judged, such that acceptable levels of exposure can be defined? How might associated risks be judged, such that acceptable levels of exposure can be defined? These questions serve to shape the way that any risk assessment is best executed, including its complexity and the extent to which different types of data (such as toxicity tests, exposure data and human experience data) can and should be included in the assessment.

1.1 Background

An effective risk assessment process should aim to reflect real-life experience. Human data, whether relating to exposures and/or effects, play a pivotal role and are reflective of the variability introduced by various factors including mixed exposures, lifestyle habits and use of medication, and other factors.

Whether or not human (effects) data are available to support any risk assessment will vary according to circumstances. At early stages these will be primarily animal data; at later stages human data may also be used. No animal or human data will be available for a new chemical. As a first step, animal data will be generated, and a risk assessment undertaken based on the findings. If the new chemical is a drug, the animal data will be used to determine the dose at which the first human volunteer testing should be conducted. Subsequently, more human and animal data will be generated before the product is launched. After the product reaches the market, in-use experience will continue to generate human safety data. For commodity chemicals, relevant human data can be derived from experiences in workers involved in the production of the chemical, from consumers who come in contact with the product or, in some instances, from the general public exposed to emissions from the production facility. At each stage of the process, all available data should be used.

In contrast to new chemicals, many existing substances possess relatively few animal data, but may be associated with data on human safety experience, arising from the use of the substance, from accidental exposure to the substance, and in some cases from formal epidemiology studies.

With both new and existing chemicals, however, the challenge is how best to use all available information to address a specific risk assessment.

Specific guidance is available on the manner in which animal toxicity and human exposure data should be evaluated and used in the risk assessment process. Such specific guidance is in general not available for human health effects, despite the fact that most international authorities recognise that the incorporation of human data into the risk assessment process can improve the utility and robustness of the subsequent findings (Health Canada, 1994; IPCS, 1999; ECB, 2003; EPA, 1993, 2005; IARC, 2006). In particular, the US Environmental Protection Agency considered that “Although for most chemicals there is a lack of well-controlled cohort studies investigating non-cancer endpoints, in some cases an epidemiologic study may be selected as the critical data (e.g. in cases of cholinesterase inhibition)” (EPA, 1993), while for cancer it advised that “positive or negative results from a properly controlled prospective study should weigh heavily in the risk assessment process” (EPA, 2005). The International Programme on Chemical Safety (IPCS, 1999) concluded that “Well-documented observational and clinical epidemiology studies have the clear advantage over studies in animals in providing the most relevant

information on health effects in the species of interest, thus avoiding extrapolation from animals to humans”, whilst the European Commission (EC) goes so far as to state that “Well reported relevant human data for any given endpoint are to be given preference for the risk assessment” (ECB, 2003). Thus, while conceptual guidance on the importance of human data is provided, detailed guidance on concepts such as ‘properly controlled’ or ‘well reported’ is lacking.

Human data, in their different forms, enable new health effects to be identified, human effect levels to be defined with more certainty, and the results of animal findings to be calibrated against human experience. Within Europe and the USA, the incorporation of human data into the risk assessment process is specifically encouraged. In part, this arises because of the recognition that many health effects of concern cannot be identified from animal studies (e.g. respiratory sensitisation, nausea and headaches). However, it also arises because of the intrinsic value that human data *per se* bring to the risk assessment process. More recently, this desire has been further reinforced by animal welfare considerations; the need to initiate animal tests is questionable if ‘Good’ quality human data are available.

There are many forms of human data. The findings of epidemiology studies of various designs contribute most commonly to the risk assessment process. Other forms of human data can range from case reports and accumulated experiences from poison centres, to results of workplace health surveillance programmes and human volunteer studies (IPCS, 1999). The quality and reliability of the data vary significantly. When compared with the data from animal studies, human data are often more challenging to interpret, but these difficulties are no reason to exclude human data as being of an inherent lesser quality or a less reliable basis for risk assessment. Human and animal data should be seen as complementary, i.e. one emanating from controlled exposures (usually to a single substance) using experimental study design and a relatively homogeneous surrogate population, the other reflecting the changes observed in a heterogeneous target population from mixed (and varying) exposure conditions using non-experimental study design.

In 1993, the EC established a programme under its Existing Substances Regulation to undertake comprehensive risk assessments for major commodity chemicals (EC, 1993). The EC risk assessment process is intended to follow a common form and to apply consistent interpretation criteria, as laid down in the relevant Technical Guidance Document (EC, 1994). For each substance that is evaluated, an assessment is made of the exposure arising from different uses of the material together with an assessment of its hazard potential for all endpoints. The risks arising from the manufacture, supply and use of the substance are then characterised. Available animal and human data should be considered at all stages of the process. Indeed, the current EC guidance clearly states that human data are to be given preference for the risk assessment, although no guidance is provided in relation to how different qualities and forms of human data can contribute to different elements of the risk assessment process. However, despite the

Technical Guidance and the expectation that EC risk assessment reports should follow a common format, in practice the procedures adopted by EU Member States for reporting and interpreting relevant human findings vary significantly (Money, 2007). At the same time, the value of using human data within the risk assessment process continues to be recognised and advocated (Dourson *et al*, 2001; Bridges, 2003; Alonzo and Laborde, 2005).

Given that there is little formal guidance on the interpretation and use of human data for risk assessment, and there appears to be different starting points relating to the inherent desirability and usefulness of human data for risk assessments, it is perhaps not surprising that human data are not consistently addressed in chemical risk assessments (Graham, 1995; EPA, 2004). ECETOC, in collaboration with IPCS, the Organisation for Economic Co-operation and Development (OECD) and the EC Joint Research Centre, therefore held a workshop with the aim of discussing and bringing consensus to the topic (ECETOC, 2004). In order to develop and focus its discussion, the workshop reviewed the proposals of Swaen (2006) for systematically evaluating human epidemiology data for use in chemical risk assessments. The workshop concluded that all forms of human data can be useful when making risk assessment decisions. It was also noted that ‘Good’ quality human data are often not readily accessible, while their usefulness is frequently constrained by a lack of information on exposure. The workshop recommended that this situation could be accounted for by developing a framework that would enable different sources of human data to be collected, evaluated and applied as part of the risk assessment process. Subsequent to the 2004 workshop, IPCS (2005) held a follow-up activity that examined how the information held by poison centres could be developed and assembled, in order to be more useful for risk assessors. This included a consideration of data quality and defining areas where data might be routinely sought from poison centres to aid in risk assessment decisions.

The process of risk assessment continues to play a central role in the development of chemicals regulation and policy. A new European Regulation on the Registration, Evaluation and Authorisation of Chemicals (REACH) will create additional challenges for the successful use of human data (EU, 2007). For example, whereas the previous regulatory focus was on major commodity chemicals (where human data might reasonably be expected to be available for most substances), REACH embraces a much wider range of chemicals, including those which are marketed at low volumes and/or for specific (limited) uses. This means that the range (and quality) of human data that might be expected to be available for such substances will probably be even more variable than previously. Recognising that the risk assessment process under REACH will be best served by the incorporation of all relevant data, it is clear that there is a need to develop guidance that will enable available human data to be assessed together with the findings of animal studies.

1.2 Purpose

Recognising that the lack of a clear interpretive framework continues to constitute a major constraint to the incorporation and use of human data in risk assessment, ECETOC established a Task Force with the objective of developing an approach which would enable different types and qualities of existing human data to be integrated with animal or other data for use in different risk assessment and risk management applications, particularly in the context of the current globally harmonised system for the classification and labelling of chemicals (GHS) and REACH proposals. Terms of Reference of the Task Force are given in Appendix A.

Although previous proposals have addressed how certain human data might be incorporated into 'weight of evidence' frameworks, these have focused largely on hazard classification for irritation and sensitisation (ECETOC, 2002a), on cancer (and other chronic endpoints) and on the role that human data have in deriving quantitative estimates of risk (Graham, 1995). No attempt has previously been made to evaluate, in the wider sense, the means by which available human data should be used throughout the different parts of the risk analysis process. With a few exceptions, for instance dioxin (Greene *et al*, 2003) and trichloroethylene (Lewandowski and Rhomberg, 2005), human and animal data have rarely been considered together for quality and relevance to human risk.

1.3 Scope

ECETOC was conscious that there was a need to initiate a discussion among the broader scientific and regulatory communities towards achieving the integration of human and animal data in the risk assessment process. With this in mind, the ECETOC Task Force, consisting of representatives of ECETOC member companies and other organisations, aimed to identify a potential solution, the merits of which could then be debated at relevant scientific and regulatory fora. This process has involved dialogue with European institutions and other international organisations.

The resultant document represents the consensus of these various inputs and is, by nature, intended to offer both a considered view of the available science and to serve as a catalyst for stimulating discussion on some of the broader issues presented by increasing the use of human data in the process of chemical risk assessment and risk management.

The report does not deal with the ethical considerations that need to be addressed when undertaking and communicating any form of human studies. Indeed the lack of a formalised framework for tackling such aspects would appear to be an issue that may require wider discussion amongst the scientific and public health communities.

2. OVERVIEW OF DIFFERENT TYPES OF HUMAN DATA

It is ethically and scientifically possible under certain circumstances to perform experiments in humans with chemical substances. Human data relating to the possible health effects of chemicals can be characterised *inter alia* as observational. Such data comprise essentially the ongoing human experience from the production and regular (sometimes inappropriate) use of chemicals. Observational data can be collected and analysed retrospectively, prospectively or cross-sectionally. In general, prospective data are preferred, since the independent parameters can be clearly defined *a priori* and are not biased by already existing knowledge on outcomes. However, if dealing with chronic exposures or long disease latency or induction periods, retrospective assessment of exposure or personal characteristics may be indispensable; conducting prospective studies would be too time-consuming in those circumstances. Cross-sectional studies are more useful for acute effects, since exposure and effects are measured simultaneously.

Human data can also be categorised depending on the type of outcome described (e.g. acute in skin irritancy or chronic in carcinogenicity). Other ways of categorising human data are by study design (cohort, case-control, case-crossover), by units of data collection (individual, ecological), methods of data collection (registry, questionnaire-based), or whether randomisation was employed (experimental, observational).

In this chapter, human data are considered as either non-experimental or experimental.

2.1 *Non-experimental data*

One way to differentiate between the various types of non-experimental data is by the manner in which the data are obtained or collected. These include anecdotal reports of one or several cases (case reports or series) that are considered as particularly interesting or unexpected data collected routinely by surveillance of more or less well-defined populations, and non-experimental epidemiology studies. The latter generally aim at establishing some association between one or more (potential) hazard(s) and health endpoints.

2.1.1 Case reports and series

Such reports describe a single case or a series of cases of disease/health effects that are of particular interest. They usually refer to an effect associated specifically with a particular exposure or to a number of cases with an unusually high incidence indicating that a certain exposure situation might be responsible. Vandenbroucke (2001) emphasised the potential roles of case reports and case series for recognition and description of new and rare manifestations of

diseases, detection of side effects, study of mechanism of disease and medical education and audit. Comprehensive and precise descriptions of the cases are crucial to identify similarities.

Case reports can be useful not only for the detection and characterisation of chemical exposures resulting in obvious acute effects (e.g. corrosive effects on the skin), but also for the recognition of associations with rare outcomes (e.g. cancers such as haemangio-endothelial sarcoma of the liver after vinyl chloride exposure) or common outcomes with increased risk (e.g. asthma after exposure to high concentrations of isocyanates). Vandenbroucke (2001) quotes the ‘rule of four’ as follows: “A number of cases (say, at least four) with an underlying relative risk of at least 4.0 must be seen over a relatively short time by a single physician or agency to permit the spontaneous discovery of occupational toxicity (Lee *et al*, 1985)”.

2.1.2 Routinely collected data

Human data are collected in various settings and by different organisations or bodies. Sometimes either exposure or health data are available, but rarely both. For example, biomonitoring for specific compounds or metabolites represents ‘pure’ exposure data, while cancer registers may contain only information on a specific health endpoint without data on exposure/risk factors.

Data can be collected for the whole population of a specific geographic area (cancer registry or national poison centre), but may also relate to certain population groups under surveillance for various reasons (e.g. employees of a specific company or members of a specific health insurance scheme). Data are often only available for a (selected) part of the base population, for certain (different) time periods/time points and in a range of quality, detail and completeness.

The data collection/surveillance may serve a variety of purposes ranging from responding to clearly defined and specific questions to simple documentation or descriptive statistics.

2.1.3 Non-experimental epidemiology studies

As pointed out by Dourson *et al* (2001), non-experimental epidemiology studies are routinely based on inadvertent exposures such as those that may occur in the workplace, by unusual natural contamination, or as the aftermath of an accident such as an explosion or industrial release. These studies may include experimental data, especially routinely collected data.

As mentioned previously, one way of categorising such studies is by design (i.e. prospective, retrospective or cross-sectional studies); all have characteristic advantages and limitations.

As already explained, prospective data collection is generally preferable to retrospective, but may not be feasible for studies in which long-time exposures or latency periods play a significant role. Rothman and Greenland (1998) defined a prospective study as a study in which exposure and covariate measurements were made before the cases of illness had occurred, and a retrospective study as one in which these measurements were made after the cases had occurred. Retrospective studies can be either case-control or cohort in design.

In a cross-sectional study, information on exposure and health endpoints is collected simultaneously. Cross-sectional data are useful for risk assessment if exposure and effect occur more or less simultaneously or if exposure does not change over time. If this is not the case, a cautious appraisal is needed to assess the various potential sources of bias before any conclusions are drawn on potential causal effects.

The principal types of non-experimental epidemiology studies are cohort and case-control studies.

In a cohort study, two or more groups of people usually free from the disease/symptoms under study are defined, based on their exposure to a potential cause of the disease/symptoms. The cohorts may differ with regard to exposure levels, or they may be either exposed or non-exposed to the potential cause (e.g. a cohort of chemical company workers exposed to a specific chemical compared to workers from the same company but not exposed to this specific chemical). The primary result is the comparison between the incidence rates of the disease/symptoms in the study cohorts.

In a case-control study, the proportion of people with a particular exposure versus those without this exposure is compared between a group of cases (people with a certain disease/symptoms) and a group of controls (people without the disease/symptoms). Therefore, it is critical that cases and controls are sampled from a well-defined source population and that sampling of the controls is independent from their exposure status. An example would be a comparison of the distribution of the exposure under question in a group of people with Parkinson's disease and a group without this disease.

Generally speaking, cohort studies can be useful for analysing relatively common health effects in a defined group of persons with a specific, often even rare, exposure. Case-control studies are preferable for studying uncommon health effects. However, if the exposure in question is uncommon, case-control studies may be inefficient.

Two other types of studies are frequently performed but require careful consideration if used in risk assessment, namely ecological and 'hypothesis-screening' studies.

Ecological studies are not based on individuals but usually represent values (e.g. averages) for a whole population group. Health endpoints and often disease incidences or mortality rates in population groups (e.g. nations, cities, companies or professions) are used for quantification. Exposures may be based, for example, on the average annual cigarette consumption, on sales of organophosphate pesticides, or readings from an air monitor located within a city.

Many ecological studies belong to the type of ‘hypothesis-screening’ studies which Rothman and Greenland (1998) refer to as “preliminary studies of limited validity or precision”. They emphasise that associations between exposures and health endpoints detected in such relatively easy and inexpensive tests should be subject to more rigorous testing with a more valid study design.

2.2 *Experimental data*

Human experimental data with regard to chemicals are usually obtained from small volunteer studies resulting in minor, reversible effects.

Where ethically possible, human experimental data are, in theory, preferable to non-experimental ones. According to Rothman and Greenland (1998) many scientists consider an experiment as a set of observations, conducted under controlled circumstances, in which the scientist manipulates the conditions to ascertain what effects such manipulation has on the observations. Because of ethical concerns, experiments to assess potential beneficial effects (e.g. pharmacological treatments) can usually only be performed if adverse health effects can be excluded. Etiological research by human experimentation is seldom possible. There are, however, rare cases in which substances may be administered to volunteers under the circumstances of a controlled clinical study up to doses from which only minor, reversible health effects are expected. The term ‘controlled’ as used here means that those factors believed to affect the outcome are kept constant or that their variation is at least well measured. This can be possible for many factors but certainly not for all factors as would be required for an ideal experiment.

Examples of experiments that can be used for risk assessment purposes include air chamber studies to determine a threshold level for an irritant effect. Another example is the administration of a substance in concentrations below the no observed effect level (NOEL) to obtain further information on toxicokinetics (e.g. identify metabolites in body fluids or compartments).

Intervention studies are those in which known but usually accepted exposures are reduced or eliminated by ‘experimental’ protective measures. These can be considered quasi-experimental, provided conditions are well controlled and exposures and outcomes determined precisely.

3. QUALITY ASPECTS OF HUMAN DATA

As described in Chapter 2, there are many types of human data varying from simple data points to complex databases. Human data also vary with respect to quality. The study design and the methodology applied to collect information on risk factors and the health endpoints can vary in respect of reliability, quality and validity. The statistical analysis techniques applied to the database can be more or less valid, or more or less appropriate, and the conclusions drawn from the study can be more or less appropriate. All these factors will have their impact on the study quality and ultimately on its weight of evidence.

The overall quality of human data will depend on the quality of the following components: Study design, exposure information, health outcome data, and this determines the general acceptability of the conclusions. These aspects are discussed below.

3.1 *Study design*

The quality of the study design depends on several factors including its appropriateness to study the hypothesis under investigation, the appropriateness of the comparison group, the time between exposure and appearance of health effects, an adjustment for other risk factors, the statistical power of the study and the appropriateness of the statistical analysis.

3.1.1 **Appropriateness of the study design**

The study should be designed in such a way that adequate data can be collected to test the hypothesis. In general terms, the more specific the hypothesis, the more focused the study will be with respect to the type of data that need to be collected. The hypothesis should specify which dependent and independent variable(s) will be investigated.

3.1.2 **Appropriateness of the comparison group**

A typical characteristic of epidemiology research is the use of comparisons. Disease incidence or prevalence in an exposed population is compared to that in a non-exposed population. The frequency of past exposure in a group of cases is compared to that of a control group free from the disease. The comparison group is used to estimate the occurrence of disease in the exposed population, had there not been any exposure to the risk factor under investigation. Particularly in studies of multi-causal diseases, it is crucial that the disease occurrence in the comparison group is a proper reflection of the background disease incidence from which the exposed population has been sampled. If certain other risk factors are thought to be important, a matching procedure

can be used to improve comparability between the exposed and non-exposed population or adjustment for confounding factors can take place in the statistical analysis. The representativeness of the comparison group can be compromised if the response rates in the study population are low or different.

3.1.3 Time between exposure and appearance of health effects

The time between exposure and appearance of a health effect is important. It is known that many types of effects do not occur immediately after exposure. Tissue damage may need to accumulate over time before it is expressed in clinically observable disease. Other long-term effects, such as cancer, are thought to have a latency period in which an initiated effect on the cellular level needs to go through a promotion phase in order to produce cancer. Cancer cells need to multiply to form tumours and infiltrative disease.

3.1.4 Adjustment for risk factors

Many diseases are known to be associated with multiple risk factors. Erroneous conclusions can be drawn if the risk factors are not similarly distributed amongst the study groups. This phenomenon is called confounding and can be either handled in the statistical analysis (by appropriate adjustment techniques) or in the sampling phase (by using selection criteria or matching procedures). Adjustment during the selection phase has the disadvantage that interactions with the matching or selection factor cannot be investigated. Matching can only be applied to a small set of confounding factors because of logistical reasons. Over time, adjustment in the statistical analysis has become the preferred approach, partly because adjustment techniques have become readily available through computer programs. The ability to completely adjust for the confounding effect of a risk factor depends on the accuracy of the data on that risk factor. If the data are not completely accurate, and there is a certain degree of misclassification, some residual confounding may still exist even after adjustment for that risk factor.

Confounding occurs when a second risk factor for a health effect is unevenly distributed among exposed and non-exposed. This is a phenomenon that can occur regardless of the study design. For instance, exposed persons can be heavier smokers than non-exposed persons. In such a situation, an adjustment must be made for the confounding factor. If data on the confounding factor are available on an individual basis, adjustment can be made in the statistical analysis.

A risk factor cannot be a confounding factor unless it is statistically associated with the disease or with the factor under investigation. In addition, simple adjustment for that factor will give a false representation if the risk factor is capable of modifying the association between the factor and the disease under investigation. In case of effect modification by a second risk factor, the strength of

the association should be expressed in terms of the relative risks for the specific categories of the effect modifying factor and not in terms of one overall relative risk.

3.1.5 Bias

The general acceptability and validity of epidemiology studies can be severely hampered by bias. Bias occurs when the quality of exposure information is associated with the disease status. For instance, bias occurs when individuals with the disease have a better recollection of past exposure than the controls, or if disease detection is different between exposed and non-exposed individuals. In these instances, the results of the study are unreliable, i.e. it is not known if bias has occurred or not and no adjustment can be made in the statistical analysis. Sensitivity analysis can often be useful in gauging the possible size of bias and thus, whether or not it is important.

Generally, bias cannot be detected or ruled out in an epidemiology study, whereas confounding by another risk factor can be adjusted for, provided the necessary data have been collected.

3.1.6 Appropriate sample size

Essentially two types of erroneous conclusions can be drawn from studies if the sample size is not appropriate. Firstly, a non-existent association can be declared (false positive or type 1 error). Secondly, a true association can be declared not to exist (false negative or type 2 error). The sample size of a study should be large enough to warrant sufficient power so that if an association in fact exists, the study will confirm this with appropriate statistical significance. Before the study has begun, the power of the study should be estimated. The required strength of the association (in term of relative risks to be expected from the exposure) and the desired significance level need to be determined. Next, an estimate of the background disease incidence (or in case-control studies the background exposure prevalence) is required. These three factors allow the statistical power of the study to be calculated; usually a power of 80% or more is considered satisfactory. If the power is below 80%, there is a risk that it will not be possible to conclude from the study that a real effect is statistically significant (type 2 error). It is generally regarded as incorrect to (re)calculate the power of a study after the study has been completed; as such an exercise ignores the actual findings. The confidence limits of the study are reflected in its sample size and that is a sufficient indicator of the precision of the risk estimate after a study is completed.

In general, larger studies provide greater weight than smaller studies in determining the presence or absence of risk. In addition, one small positive study may be more useful than several small null studies. [This is because a small positive study normally has a lower confidence limit still above 1.0, indicating the presence of risk even in the light of greater statistical uncertainty,

whereas a small null study usually has an upper confidence limit above 1.0, which cannot exclude the possibility of the presence of risk in the light of statistical uncertainty.] Thus, more than one small null study is necessary to be confident in an overall interpretation consistent with no effect.

3.1.7 Appropriate statistical analysis

Statistical analysis is carried out on the data collected according to the procedures described in the study protocol. The principal aim is to describe the data in terms of meaningful entities such as means, medians, standard deviations and risk metrics. Next, the statistical significance of differences between the means of groups is tested or the risk parameters (e.g. odds ratios, standard mortality ratios or relative risks) are assessed to exclude, with a certain probability, the likelihood of a chance finding. A third objective of the statistical analysis is to adjust for confounders resulting in adjusted risk metrics.

Epidemiology data can be diverse and complex. Extensive knowledge is required to select the appropriate statistical technique for a given database. An important part of the statistical analysis is to inspect the distributions of the variables to be analysed. Certain statistical techniques require that the variables should be normally distributed. If not, the data need to be transformed into a normal distribution; otherwise certain techniques (e.g. linear regression) cannot be used. If the data are not (approximately) normally distributed, other techniques such as Poisson regression may be useful.

The first step of statistical analysis involves checking the data for any errors, outliers and inconsistencies. Next, simple comparisons are made and crude association metrics between the variables are calculated to determine which potentially confounding variables should be taken into account. Once the potential confounding variables are identified the appropriate statistical model suited for the analysis can be determined and executed. The reliability of the statistical analysis can itself be studied by a sensitivity analysis. All statistical analyses are based on assumptions on how variables are distributed. In a sensitivity analysis the underlying assumptions are varied and the effect on the results of these varying assumptions is evaluated.

3.2 *Exposure information*

Money and Margary (2002) described a number of core principles to derive reliable and robust exposure assessments. The authors distinguish three types of exposure data: Actual data, analogous data and personal exposure data collected in a systematic manner. All three types of data can vary in quality and reliability.

Information on exposure can have many forms. It can vary from categorical information indicating the likelihood of exposure (e.g. as might be provided by the knowledge derived from the successful completion of a course in the safe handling of pesticides, a positive response to questions relating to use/contact with a specific substance, or employment at a company where a specific chemical was handled) to extensive individual exposure data (e.g. based on the systematic collection of air samples). If the exposure data are qualitative, the study results will be relevant for hazard identification, but not for dose-related quantification of risk. In many longitudinal epidemiology studies the construction of a job-exposure matrix has been shown to be a valuable means of handling exposure information, but it is only useful to construct a job-exposure matrix if (semi-)quantitative exposure information is available. The job-exposure matrix is based on homogeneous exposure groups, consisting of those jobs that are thought to be characterised by comparable exposure conditions. For each homogeneous exposure group, the exposure intensity is estimated. Historical changes in the production process or work practices, resulting in changes in exposure, are taken into account and form a dimension of the matrix. The job-exposure matrix allows the calculation of cumulative exposure, but can also serve to stratify the study groups into sub-groups with certain exposure characteristics (such as always exposed over a certain concentration). Exposure measurement error can lead to misclassification and this may have a significant effect on the results of epidemiology studies. Although arguments have been put forward that exposure misclassification intrinsically leads to relative risk estimates closer to 1, some authors are of the opinion that this is not always the case. In view of that, it cannot be concluded that misclassification of exposure always leads to an underestimation of the risk (Jurek *et al.*, 2006).

Exposure circumstances can vary substantially. The degree of variability of the exposure conditions largely determines the extensiveness of the information required to adequately describe exposure conditions. If exposure is stable and without variation (e.g. over the workday, the season or between time periods), a few sample points may be sufficient to characterise the exposure situation. However, in reality, exposures vary from location to location and from task to task, and change over time due to differences in production processes, exposure reduction measures, and use of personal protection equipment.

The type of exposure information can also vary. Sometimes the only exposure parameter available is that a person has been employed in a particular industry. More specific information would include the type of job the person has performed in that industry over a particular time period. A more specific exposure characterisation is only possible if industrial hygiene measurements have been made. Industrial hygiene measurements are carried out for various purposes (e.g. to identify sources or tasks with high exposure). In such case, the results are likely to constitute an over-estimate of general exposure at the workplace. Industrial hygiene measurements can also provide a reliable picture of the exposure conditions at a specific work place.

Exposure measurements collected by means of a systematic approach (e.g. including clarification of how, why and where samples were taken) are the most valuable. Exposure patterns can be characterised according to their temporal variability, spatial variability and variability due to individual behaviour. Work rosters and task/activity schedules can have a large impact on exposure.

True exposure is not only determined by the number of measurements that are made but also by the variability of exposure. The internal and external validity of the exposure data is important for the final interpretation of the findings and should describe adequately the actual exposure situation. Internal validity depends on the sampling strategy and sampling frequency, while external validity relates to the comparability between the exposure conditions under investigation and the exposure conditions in other situations. Nevertheless, poor exposure data can still be useful in the hazard identification stage of the risk assessment process.

The improper classification of exposed individuals as non-exposed (or vice versa) is an example of misclassification. If this misclassification is not associated with the disease it is called non-differential misclassification. Erroneously, non-differential misclassification has been thought to only cause a bias towards the null. Jurek *et al* (2005) have pointed out that non-differential misclassification can also result in over-estimation of the relative risk.

In summary, human exposure information can be available with varying degrees of detail and reliability. The quality of the data will to a large extent determine its value for quantitative risk assessment.

3.3 Health effect data

Health effects can occur in various ways and the types of effects for which human data exist can vary from acute to chronic (long-term) effects. The occurrence of health effects can be determined in various ways. Data can be collected from self-completed questionnaires or from databases (e.g. cancer registries). Occasionally, specific diagnostic procedures are performed to establish disease status. The quality and completeness of health effect data establish their overall reliability. In many epidemiology studies, the occurrence of disease is expressed in a relative measure, i.e. the incidence of the disease in an exposed population is divided by the incidence in a non-exposed population resulting in a relative risk metric (van den Brandt *et al*, 2002). The occurrence of (chronic) disease as such is not regarded as informative and the advantage of the relative risk metric is that it reflects the strength of the association between exposure and effect.

The quality of the health effect data depends on the collection methods, e.g. whether or not the data are standardised and validated or whether diagnostic techniques with satisfactory sensitivity

and specificity have been used. It is essential that the reliability of health effect data collection techniques is the same for the exposed and non-exposed groups. If not, bias can be introduced, which makes the study useless.

Even if a reliable diagnostic procedure is used, the health effect data may not be reliable, because of differences in completeness. Ideally, completeness would be 100% for exposed and for non-exposed groups. However, completeness is hardly ever achieved. Identified cases may be missed because of poor tracking or cases may not be diagnosed. The risk metrics (odds ratios and relative risks) are relative and, provided the completeness in the exposed group is the same as in the non-exposed, even incomplete health effect data will result in correct risk estimates. For instance, it is clear that case reports are unlikely to represent all occurring cases of that disease. However, they can still be important in the hazard assessment process and if the health effect is specific, this information can still be important in risk assessment.

3.4 General acceptability of the conclusions

Conclusions from a study can only be drawn if they are substantiated by the statistical analysis of the data. In this respect, not only the statistical significance, but also the internal consistency of the data play a major role. For instance, it should be established whether all the analyses supported an association, whether there was a dose-response relationship, whether the association was found for all sub-groups, and whether known confounding factors were taken into account.

General acceptability indicates the extent to which the results of a study may also be applicable to larger populations. In the case of descriptive studies, such as surveys, the study population needs to be representative of the larger base population. However, in case of analytical studies producing relative risk metrics, the study population does not necessarily need to be representative of the general population (Rothman and Greenland, 1998).

3.5 Specificity of the association

To a certain extent, the quality requirements of a study depend on the type of association under investigation. If a health effect is specific, i.e. it can only be caused by the exposure under investigation, no adjustments need be made for potential confounding factors. On the other hand, if it is known that other factors can play an important role in the aetiology of a non-specific health effect (such as smoking and pulmonary disease), the results of a study are not considered reliable, unless an adjustment for the other risk factor is made. Cross-sectional studies and even case reports on specific, acute health effects can provide sufficient and reliable information for risk assessment, whereas non-specific long-term health effects require complex research designs including adjustment for potential confounding factors.

Finally, all quality requirements are rarely met in epidemiology studies and it must be acknowledged that generally some doubt will remain about the validity. This complicates the interpretation of the results, but does not render them worthless. It is therefore important to have more than one study available, and preferably of different study designs. A causal association becomes more likely if these different studies show consistency in results.

4. HUMAN DATA QUALITY CRITERIA FOR RISK ASSESSMENT

Once the human data available on a substance have been gathered and evaluated, the relevance to risk assessment has to be assessed. This is often referred to as evaluating the ‘weight of evidence’ of the data. Weight of evidence techniques have long been used by epidemiologists when evaluating whether a substance is inherently capable of causing a health effect. The best-known guidelines (often referred to as ‘criteria’) are those of Bradford Hill (Hill, 1963). These guidelines apply to a specific step in the risk assessment process, that of ‘hazard identification’ or ‘causation’. This is a necessary step, before proceeding with subsequent steps in the risk assessment process i.e. only if it is plausible that substance X causes effect Y, should dose response, exposure assessment, and risk characterisation steps be considered for substance X.

The Bradford Hill guidelines were developed to draw causal inferences in general associations between an exposure and a health effect. The applicability of these guidelines for more specific exercises (e.g. to answer the question of a possible threshold for a certain health effect) is less clear. Somewhat surprisingly, in comparison to the hazard assessment step (which lends itself to the Bradford Hill guidelines), relatively little attention has been given to guidelines applicable to the dose-response step of the risk assessment process.

Related to both the issues of hazard identification and dose-response assessment is the topic of identifying the ‘lead effect’ or the ‘critical effect’ for a given substance. Often the lead or critical effect is the most sensitive. This implies that some information on the dose-response of a substance must exist in order to assess whether a given concentration or cumulative exposure produces an effect more readily than a competing effect. However, the ‘relative’ potency of the substance for different effects may require less exacting data than determining a critical concentration that produces an effect or a dose-response curve for a broad range of exposures.

Thus, criteria for evaluating human data need to consider the question being asked, or the stage of the risk assessment, i.e.:

- Hazard identification, or causality assessment;
- identification of the lead or critical effect;
- dose-response assessment [dose-response assessment subsumes identification of a no observed or lowest observed effect level (NOEL or LOEL), and a dose-response curve].

This is discussed further in Sections 4.1, 4.2 and 4.3. A scheme for scoring human data with regard to its quality is presented in Section 4.4.

The above considerations do not apply to human data alone. Ideally, human and animal data should be considered in an integrated fashion, to identify whether an effect can be caused by a given substance, to establish the critical or lead effect, and the dose that produces a given incidence (including no discernable excess incidence) of the critical effect in humans.

4.1 Hazard identification

The techniques described below can be used to arrive at causal interpretations or hazard identification regarding a body of data on a specific compound or substance.

4.1.1 Bradford Hill guidelines

The Bradford Hill guidelines frequently serve as the basis for ‘qualitative reviews’ concerning causality. While it is not the purpose of this report to review these guidelines in detail, a brief review is offered below.

Hill (1965) suggested nine considerations that should be assessed in determining whether a given association is causal. These are:

- Strength of the association [strong associations are less apt to be explained by some other factor than weaker associations, if all other aspects are considered equal. Strength of the association is usually expressed as a relative risk estimate].
- Consistency of the association [i.e. whether the association is seen consistently in other populations under different circumstances, or among sub-populations within a study].
- Specificity of the association [i.e. whether a suspected cause leads to a single effect rather than a myriad of effects].
- Temporality [the cause of the disease must precede its effect. This is the only aspect that must be satisfied].
- Biological gradient [i.e. the presence of an increasing risk with increasing exposure].
- Plausibility [i.e. the biological underpinnings of a potential cause-effect relationship].
- Coherence [i.e. a given cause/effect hypothesis does not conflict with the known natural history or biological underpinnings of the disease. Coherence is often assessed together with biological plausibility].
- Experimental evidence [this is sometimes interpreted as evidence from human or laboratory animal experiments, but it appears that Hill’s original intent was to indicate the removal of the causal factor and subsequent decreases in disease incidence].
- Analogy [i.e. analogous information from other causal associations, or more elaborate hypotheses that are consistent with the association being evaluated].

These considerations provide a good general framework on which to assess causality. Invariably, however, interpreting a body of evidence against the Bradford Hill guidelines still involves some judgement. With the possible exception of ‘temporality’ (which is a criterion rather than a guideline), it is apparent that failure to satisfy one of the other nine guidelines (e.g. specificity) will not invoke a non-causal interpretation of the data.

4.1.2 Meta-analysis

Meta-analysis is the systematic synthesis of a body of data (in the form of the aggregated study results) to improve insights on the causal nature of a suspected relationship. It can improve the precision of risk estimates, and can also identify potential determinants of differing results between several studies (or why studies differ); the latter can have a substantial effect on risk management actions. For example, if several studies of particulate matter showed a weak association with cardiovascular mortality, but meta-analysis revealed a stronger association with cardiovascular disease in high sulphate particulates, risk management actions to reduce sulphates (rather than other forms of particulate matter) might be more effective.

Meta-analysis is time consuming and only relevant in the case of ‘data-rich’ chemicals. It is recommended to use ‘quality indicators’ (e.g. sufficient follow-up and ‘Good’ exposure data) rather than ‘quality scores’ (Greenland, 1994).

4.1.3 Pooled analysis

In rare cases (e.g. radiation, benzene) pooled analysis of data on an individual level would aid in determining causality and describing the dose-response relationship. Pooled analysis differs from meta-analysis in that all individual data are available, rather than only those from study reports. By pooling raw data from several studies, greater insight can be obtained than through meta-analysis alone, since the investigators are not limited by data contained in the study report. Time, resources, and data heterogeneity will often be prohibitive for pooled analyses to be conducted or considered, unless such data are available before embarking on the risk assessment process.

4.1.4 Summary

In summary, human data can be used for hazard identification when:

- The available data are of sufficient quality.
- The positive data are free from strong biases or a strong potential for confounding.

- There exists a plausible biological basis for the positive association suggested. This often must come from animal or mechanistic studies.
- These concepts can be categorised into definite, probable, indeterminate, or improbable causal relationships.

4.2 *Lead or critical effect*

While there are no analogous guidelines for considering a critical or lead effect, this implies that a minimal amount of dose-response information, often in the form of no observed and/or lowest observed adverse effect level (NOAEL and/or LOAEL), can be compared for different potential effects due to exposure. Determining a critical effect can also depend on the severity of the effect, its reversibility or whether it is deemed to be 'adverse'. A detailed discussion of this is not presented in this report. The interested reader is referred to a review by ECETOC (2002b) and paper by Lewis *et al* (2002).

4.3 *Dose-response assessment*

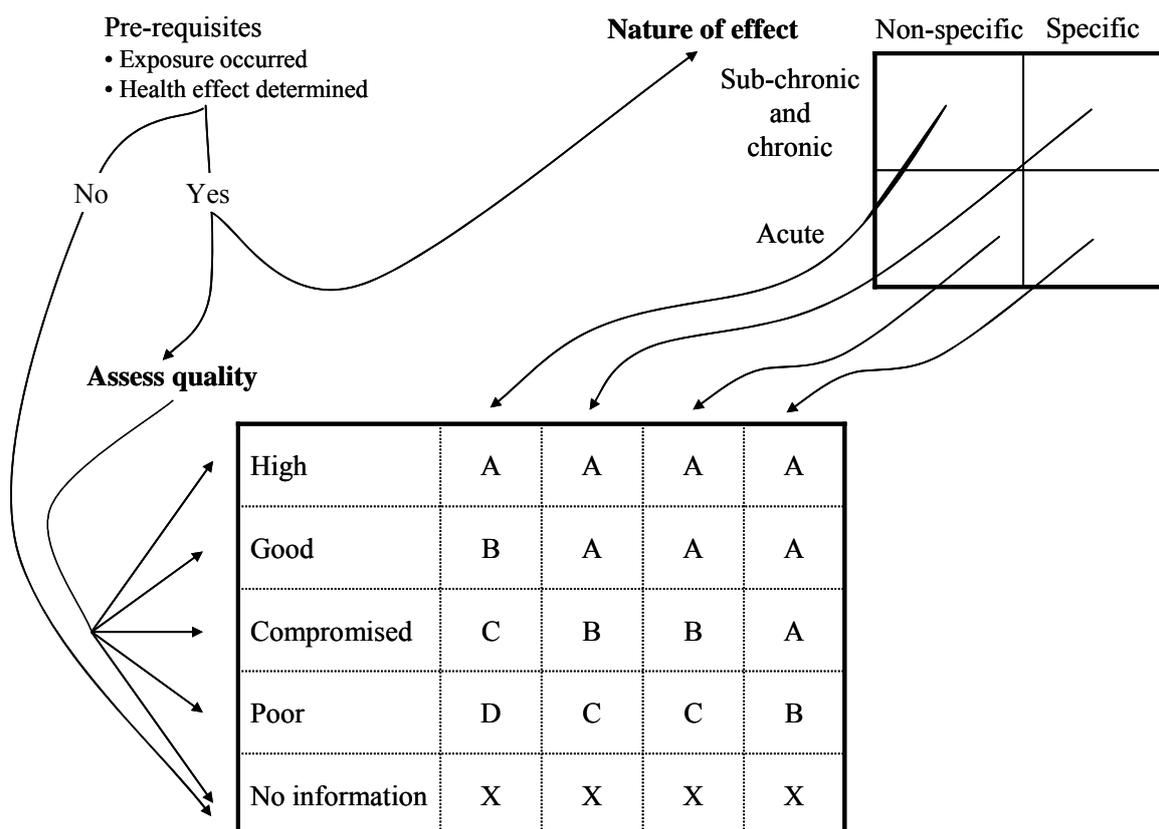
Estimating the incidence or risk of health effects for different levels of exposure or dose, can contribute information on the overall dose response of a substance. In this report, the terms 'dose response', and 'exposure response' are used synonymously. Dose-response data allow an estimate of the LOEL, NOEL, or dose-response curve. It is obvious that to characterise the dose response (in terms of NOEL, LOEL or more complete dose-response functions), the body of human data must be stronger and of better quality than data that are only sufficient for hazard identification. Additional necessary features include stronger (usually quantitative) data on dose or exposure to support inferences or extrapolations to populations outside the existing studies.

Hertz-Picciotto (1995) was one of the first to attempt to outline the characteristics of human data for use in dose-response assessment. The author suggested that the four features were necessary, namely a moderately strong dose-response association, a lack of strong bias, an absence of a high probability for confounding, and quantified exposures linked to individuals (rather than groups of individuals, as is the case for census-based studies). These guidelines can be advanced as a good starting point. However, the guidelines were written mainly for long-term epidemiology studies of chronic disease and need to be re-considered for acute, or more immediate effects (e.g. situations involving accidental ingestion and estimated doses). In addition, there are scenarios in which the nature of the effect is specific to the exposure. In such cases, less exacting data are often sufficient to show an effect at a given concentration (as well as a lack of an effect at lower concentrations).

4.4 Scheme for scoring human data with regard to its quality

In this section, a simple scoring scheme for human data is proposed, based on the quality of the data and the nature and specificity of the effect. The aim is to characterise whether the human data are a reliable source for the dose-response assessment (or another) stage of a risk assessment. This is done by creating a small number of categories to characterise human data. (This will be combined at a later stage with animal or *in vitro* data, to allow a transparent way to justify the basis of a risk assessment when there are both human and animal data.) The choice is driven by the inherent strengths and weaknesses of each data source, rather than the results (positive or null) of a particular body of data. If human data cover all ranges of quality for a given effect (e.g. from case reports to experimental studies), the scheme would categorise the highest quality human data (Figure 2).

Figure 2: Scoring the quality of human data: Pre-requisites, assessing intrinsic quality, considering the nature of the effect



For human data, the scheme involves the following steps. There are certain pre-requisites that have to be met (Section 4.4.1). When these are met, the intrinsic quality of the data is assessed in

detail and assigned to a category (High, Good, Compromised, Poor or No information) (Section 4.4.2). The nature of the health effect is then considered, e.g. specific or non-specific, acute or (sub-)chronic, and combined with the category to produce a quality score on a scale of A (strongest) to D (weakest), or X (no information) (Section 4.4.3). In certain circumstances, adjustments may be made to the quality score based on whether the findings are positive or null (Section 4.4.4).

4.4.1 Pre-requisites

There are two pre-requisites for human data that must (usually) be satisfied across all scenarios and situations. These can be summarised as: Exposure must have occurred and the health effect should be determined adequately (Figure 2). These pre-requisites can be further divided into the following three requirements:

1. Exposure to the substance in question should be present.
Ideally, exposure (not necessarily the degree) should be documented or inferred with near certainty based on the scenario in which exposure occurred. If the study population contains a mix of exposed and non-exposed subjects, the non-exposed group should not mask any effects of exposure (e.g. $> 2/3$ of the population). A study that examines an exposed subgroup would satisfy this pre-requisite.
2. Exposure should precede the effect.
3. The health effect should be determined on the basis of published criteria that are generally accepted as valid disease endpoints or symptom definitions. Alternative definitions of health effects need to be validated against the published criteria.
Studies that use official records that document disease occurrence (e.g. hospital records, incident reports, death certificates) would satisfy this requirement. Studies that measure earlier indicators of adverse effects (e.g. low blood counts, cells containing micronuclei, mutant genes) should specify detailed analytical protocols for determining the effects, to allow replication in other laboratories. Self-reports of disease or adverse symptoms should usually be validated (e.g. observation by qualified medical personnel or records systems).

In rare circumstances, when a disease is chronic, relatively common, and well-accepted alternative causes of the disease exist, a further requirement may come into play:

4. Other causal factors that have a large (e.g. $RR > 5$) influence on disease occurrence should be taken into account.

For example, when the effect of cigarette smoking on lung cancer incidence has not been measured or accounted for at all, it may be appropriate to account for this as a major alternative cause. As another example, squamous skin cancer could be ascertained through a registry. [However, if the exposed population lives at low latitude and the non-exposed at high, the significant effect of differential sunlight exposure on squamous cell skin cancer must be taken into account. Accounting for such an effect would require measurement and control of the factor in a statistical analysis, ‘matching’ on the factor or an attempt to exclude the factor from the study (i.e. in the above example, limiting the population to higher latitudes). Furthermore, a sensitivity analysis can be employed that argues that the maximum plausible disparity of the alternative causal factor in question would not lead to a different interpretation of the data.]

4.4.2 Assessing the intrinsic quality of the data

The following five categories that encompass different degrees of intrinsic quality for human data are proposed (it is assumed that the pre-requisites are already met).

High quality

Data are considered of ‘High’ quality if there are several studies of a commonly recognised epidemiology design (e.g. cohort, case control and/or experimental) with a large majority (2/3 or greater) showing consistent results with one another, and consistency with biological evidence. In particular, the studies will have:

- Quantified exposure data linked to individuals;
- measurement and control of confounding factors, or
- evidence and/or convincing argumentation that the effect of the potentially confounding factors does not affect the interpretation of the study.

A relevant selection bias, which may affect the validity of case-control studies in particular, needs to be excluded. For positive data, this means that the exposure can biologically and plausibly result in the disease, and that a monotonic dose-response gradient exists either for the body of data (supported by pooled or meta-analyses), or in at least 2/3 of the studies. For null data, there should be a lack of a monotonic dose-response in at least 2/3 of the studies. In some cases, especially for experimental design studies, one particularly strong, large study may suffice, especially if there are no reasonably strong conflicting human data. Since the studies will already have been completed, the width of the confidence interval (which takes both study size and the results into account) is important; the study power *per se* is no longer relevant. Confidence

intervals for the highest quality human data should generally span less than one order of magnitude to be considered in this 'High' quality category.

Good quality

Data quality should be considered as 'Good' rather than 'High' if most of the above conditions are met and no more than three of the following limitations apply:

- Consistency between studies is not high, yet still suggestive (> 50% of the studies are concordant);
- exposure data are not always quantifiable or linked to individuals;
- there is no good biological understanding to underpin the results (e.g. of how exposure could result in disease);
- not all strong confounding factors can be ruled out in the majority of the studies;
- health outcome measurements are not well-validated;
- confidence intervals are more than an order of magnitude;
- a monotonic dose-response relationship (or lack thereof, for null data) exists for the majority (but not all) of the studies.

Compromised quality

This category of data is either of sufficient design but lacking in more than three of the above criteria, or of a compromised design (e.g. cross sectional, time series, case reports for chronic, multi-causal endpoints, possible selection bias), but possessing at least three of the following items:

- Consistency;
- strong exposure data;
- biological gradient;
- biological plausibility, and/or
- adequate control of intervening variables.

Poor quality

This is the weakest category for suggesting a causal association or safe versus harmful concentrations. Case reports for chronic disease usually fall into this category because there is no study design. Reports from regularly collected data (e.g. mortality trends, poison centres) are likely to fall into this category, unless corroborating information is found subsequently.

No information

No valid data, or a large body of clearly discordant data.

4.4.3 Considering the nature of the health effect and deriving a quality score

The nature of the health effect has a large influence on quality requirements. An effect that is chronic and has a long latency period requires stronger and more complex human data than an acute effect (Swaen, 2006). Effects that are systemic and chronic in nature (e.g. pulmonary fibrosis, bladder cancer, renal failure) may take years to develop. In the development of disease, other attendant lifestyle factors and time-related confounders may also play a significant role and may also vary in time. This makes it difficult to observe causal associations without well designed and conducted studies. Exceptions to this are chronic health effects specifically caused by the chemical in question, such as angiosarcoma caused by vinyl chloride.

A sophisticated epidemiology study is not necessary for acute local effects. There is, for instance, an immediate awareness of the cause of symptoms such as coughing and wheezing in dusty environments. Also effects that are rare and/or specific can be associated with exposures without a formal study design (e.g. relatively few *Pneumocystis carinii* cases from government surveillance systems were sufficient to signal the start of the AIDS epidemic in the USA).

The nature of the health effect as well as the intrinsic data quality have been taken into account by the Task Force in assigning an overall quality score to human data (Table 1).

Table 1: Combining intrinsic data quality and the nature of the health effect to produce a human data quality score

Intrinsic human data quality	(Sub-)chronic effect		Acute effect	
	Non-specific (e.g. diesel exhaust - lung cancer)	Specific (e.g. vinyl chloride - angiosarcoma)	Non-specific (e.g. solvents - headache)	Specific (e.g. reactive dye - respiratory sensitisation)
Highest	A	A	A	A
Good	B	A	A	A
Compromised	C	B	B	A
Poor	D	C	C	B
No information	X	X	X	X

Thus, for specific (usually rare) effects or acute effects, quality categories A, B and C are possible often without any sophisticated epidemiology study design. It is also expected that, in such cases, it is unlikely that a large body of data will need to be evaluated with respect to consistency since hazardous concentrations will be more readily detectable.

4.4.4 Adjustments to the scheme: Positive versus null human data

In general, identical considerations should go into evaluating both positive and null human data. The one exception is the amount of data (or study size) as mentioned previously in Section 3.1.6. A small positive study may be all that is needed to unequivocally suggest causality or, perhaps potency (e.g. consider a lethal concentration of a substance). In contrast, the absence of effects usually requires larger population sizes. Fortunately, there is a unifying concept, i.e. the confidence interval (and the associated upper and lower limits) around the risk estimate. The related concept of study power should not be used because it considers only the design of the study (i.e. it will allow a prediction of the level of risk a study is apt to detect, given the number of subjects). Once a study is completed, this prediction is no longer relevant or appropriate. Instead, the risk observed needs to be accounted for, and the confidence interval can account for the size of the study population. All completed human studies should naturally be evaluated based on their actual results rather than their predicted results.

If a null study is of 'High' quality, yet has a fairly wide confidence interval (upper limit > 2) a downgrade from A to B is indicated. Similarly, a few small null studies would need to be downgraded. Thus, for null results that cannot exclude a doubling of risk (i.e. upper 95% confidence limit > 2), it is suggested that an overall quality score of B is the highest that can be obtained. Taking into account the relative qualities of the human and animal data, consistently null human data can be used to provide an upper bound for human risk assessment (Bukowski *et al*, 2001). Consistently 'High' quality null data of sufficient size, especially if generated in different situations or by a number of investigators, should still be able to achieve a quality score of A.

4.4.5 Consideration of the risk assessment stage

Risk assessment usually proceeds from hazard identification, dose-response assessment and exposure assessment to risk characterisation, and risk management (Chapter 1). Human data are applicable to every stage, although the requirements for each stage are (sometimes obviously) different.

Hazard identification

In this stage of risk assessment, the inherent capability of the substance to cause a given effect is assessed. Thus, the exposure level at which this effect does or does not occur is not important, e.g. for classification and labelling. As such, the quality criteria for human data do not require quantified exposure data.

This was recognised by Hertz-Picciotto (1995) in advancing the use of epidemiology studies in risk assessment. The author defines ‘Study category 3’ as contributing to weight of evidence determination whether or not the agent is a health hazard (i.e. hazard identification), and ‘Study category 2’ as useful for checking the plausibility of animal-based risk assessment. Thus, study category 2 or 3 data may also be useful for hazard identification.

Dose-response assessment

The Hertz-Picciotto categories have been developed primarily for the dose-response stage of risk assessment (Section 4.3), a stage that requires the most stringent quality criteria, as exposure is of paramount importance. Exposure assessment must often be retrospective, especially for chronic diseases with long latency periods. It is a challenging proposition to assess exposure adequately for long periods of time. Furthermore, to guard against bias, quantification of exposure should be linked to specific study subjects, and not assessed on a group or geographical basis.

Exposure assessment

The use of human data for exposure assessment purposes has not been addressed in this report. Discussion of this aspect is covered by Money and Margary (2002).

Risk characterisation and risk management

Examples of diverse situations are included in Appendix B. An example of a risk not being managed well is illustrated by poison centre data for hydrogen fluoride. In this instance, confirmation that hydrogen fluoride exposure was still occurring in the general population did not come from a well-designed epidemiology study, but from isolated case reports; such reports were likely to have been classified as of poor intrinsic quality. These data should, however, be considered of ‘High’ quality (overall score A) since the effects were local, acute in nature and specific to hydrogen fluoride (Appendix B.11).

5. QUALITY ASPECTS OF ANIMAL STUDIES

For risk assessment purposes animal data and human data should be integrated. However, in cases where good human data are lacking, animal data are used and animal studies in particular are considered in this chapter.

5.1 *Assessment of reliability*

An animal study has certain intrinsic quality characteristics. What is important to the risk assessor is not only the quality of the study *per se*, but also the quality of the information provided to the reviewer of the study. For this reason, assessors should use the full description of the study. Many good studies are let down by the poor reporting of information available, for example within a brief description in a review article.

Criteria for systematically reviewing the reliability of reported animal studies are routinely used by some assessment authorities. Those of Klimisch *et al* (1997), for example, may be used in determining whether existing data are of sufficient quality or whether further testing is needed in the context of the OECD Existing Chemicals Programme; the programme is intended to ensure that all high production volume chemicals have sufficient quality data for a set of minimum toxicity endpoints to be established.

The OECD has described three terms used by Klimisch when referring to data quality i.e. reliability, relevance and adequacy (OECD, 2007). Reliability refers to the quality of a test report or publication and takes into account whether standardised methodologies were used to generate the report, as well as the manner in which the experimental procedure and results were described. Relevance refers to the extent to which data and tests are appropriate for a particular hazard; and adequacy refers to the usefulness of the data for hazard and/or risk assessment purposes.

Criteria for assigning animal studies into four reliability categories are shown in Appendix C, based on Klimisch *et al* (1997). The reliability criteria and categories provide a practical systematic approach for quality evaluation of experimental toxicological data, considering such factors as the use of a standardised test method, the availability of a description of the relevant details of the study design.

5.2 Relevance and utility

Some aspects of quality are neither intrinsic to the study *per se*, nor a function of the information available about the study. These additional quality aspects are concerned with the relevance and utility of the study for addressing the question in hand. For example, a study on an atypical test species may be performed to the highest standard, but may not have much value in risk assessment due to the lack of baseline data for that species and lack of experience of its utility in human risk assessment. Another example is where results from an *in vitro* screen are available, but the equivalent *in vivo* study (the results of which, if available, would usually supersede the *in vitro* study) has not been conducted. In this case, one quality consideration would be the status of the *in vitro* study in the validation process of the European Centre for the Validation of Alternative Methods (ECVAM)^a (Lilienblum *et al*, 2008). Another consideration would be the reliability of the *in vitro* study in terms of its false positive and false negative rates. In principle the same considerations apply to all animal studies, in that the ability of all *in vivo* animal protocols to predict human responses should be considered, though the data to do this are rarely available. The value of screening results and full animal studies may differ, for example a result in an Ames screen carries less weight than a result in a 2-year rat cancer bioassay.

One obvious advantage of animal studies over most human studies is that in the animal studies an exposure condition to a single known chemical can be created under a controlled dose regimen. However, it is only the applied dose that is generally known; there is usually no information on systemic exposure or exposure at the site of toxicity. This can complicate the interpretation of studies performed with dose-routes not relevant to human exposure. The question of relevance of dose-routes is a complex one. For example, rodent studies where agrochemicals are administered in the diet are routinely used for worker risk assessments, despite the fact that the worker (spray operator) exposure is predominantly dermal. In this case, systemic exposure is used as a common basis for inter-relation, i.e. systemic exposure from dietary dosing in rodents is compared to systemic exposure from dermal dosing in man. Dose-route extrapolation requires additional data, which may not always exist. Without such additional data it can be particularly difficult to interpret animal data where dosing is via routes such as sub-cutaneous, intraperitoneal and intravenous. For example, if human exposure is dermal, and there is negligible dermal absorption, then animal data from intravenous dosing may be of no relevance for risk assessment and arguably without relevance even for hazard identification. More unusual dosing routes include direct injection into the brain or cerebro-spinal fluid, which bypasses biological barriers, making the studies hard to interpret in a human risk assessment context. If two studies exist, one using a relevant dose-route and one not, that are otherwise of comparable quality, then the study using the relevant route should be deemed to be of a higher 'quality' for risk assessment.

^a ECVAM co-ordinates the validation of *in vitro* methods and promotes their role in regulatory processes (<http://ecvam.jrc.it>). The US equivalent is the Coordinating Committee on the Validation of Alternative Methods (ICCVAM) (<http://iccvam.niehs.nih.gov/home.htm>).

A general consideration for the relevance of an animal study is how closely toxicity in the test species might predict toxicity in man. In general, large animals (such as dogs) might be expected to show more toxicity, due to their slower clearance and metabolism, than smaller animals (such as rodents); in some cases the opposite is true, and there is great variation. For example, Vermeire *et al* (1999) compared rat, mouse and dog NOAELs, while Doe *et al* (2006) compared rat and dog NOAELs. Genetic relatedness is another factor to consider. This is the reason, in principle, why data from non-human primates are considered more relevant than data from other species, although such data are rarely generated for chemicals due to ethical reasons.

The relevance and utility of a study is context dependent, so this is a matter to be revisited when the uses of animal data in risk assessment are considered in the next chapter.

5.3 *Information gained from animal studies*

Testing for statistical significance at the 5% level, as is conventional, will on average identify 1 in 20 comparisons as being significant, even if there is no biological effect. In animal studies, a large number of biological endpoints are evaluated. If large numbers of statistical tests are made, it is inevitable that some will be statistically significant without having any biological relevance (false positive). Therefore, it is extremely important to evaluate them further (e.g. in relation to dose response, or temporality) to judge whether they can be considered real (i.e. not due to chance) or not. If they are considered real, then a judgement must be made as to whether or not they are considered to be biologically adverse. For example, in a large study a small increase in liver weight (10%) may be statistically significant, but is considered an adaptive response rather than an adverse effect. The 'weight of evidence' approach of Lewis *et al* (2002) provides detailed guidance on these matters.

The following information can be derived from an animal study:

- NOAELs and/or LOAELs for various biological effects, judged to be real and adverse;
- identity of target organ(s) and, if applicable, a pattern or syndrome of toxicity (potentially including pathology, clinical chemistry, and clinical signs);
- dose-response relationship;
- information on reversibility of effects;
- judgement of severity of effects.

The next step is to review all available animal studies, and to select one or more critical endpoints, i.e. those considered suitable to form the basis for determining a reference dose (if that is the objective of the assessment). The following factors should be considered:

- Information gleaned from each animal study, as described in the preceding section.
- Identity of target organ(s), and any pattern or syndrome of toxicity potentially including pathology, clinical chemistry, and clinical signs. [Although there may be information on this from individual studies, it is more likely that this information will emerge from a review of multiple studies.]
- Concordance or lack of concordance between results of comparable studies.
- Concordance or lack of concordance between results in different sexes, species and strains.
- Quality and relevance aspects of each study, described in the preceding sections. This is particularly important if there are several similar studies available, especially if their results differ.

A particular issue is what action to take if animal data are discordant. If two studies using the same species, sex and dosing conditions (e.g. duration and route) give different results, then the Klimisch criteria can be used to determine which data are of higher quality; the study of higher quality should be used in the assessment. If the studies are of the same quality, then the worst-case result should be used. However, it should be noted that a lack of concordance between sexes or species, or even between strains of the same species, could provide invaluable information about the mode of action (MoA) of the substance. The importance of MoA when considering relevance to man will be discussed in detail in Chapter 6. If the discordance between animal studies cannot be rationalised in terms of MoA and the animal studies are both Klimisch category 1 or 2, then the worst-case data should be used in the assessment, while also taking quality aspects into account.

6. RELEVANCE OF ANIMAL DATA IN HUMAN RISK ASSESSMENT

Irrespective of the quality of animal studies, the process of extrapolating the results to man is an uncertain one. Olson *et al* (2000) for example reviewed animal toxicity data for 150 pharmaceuticals for which human toxicity had been observed. Forty-three percent of the human toxicities were observed in rodent studies and 71% in either rodent or non-rodent studies. The highest predictability was in the case of haematological, gastrointestinal and cardiovascular toxicities, and the lowest for dermal toxicity. The nature of the animal toxicity data package required for pharmaceuticals is of course not the same as that for commodity chemicals.

The US Environmental Protection Agency compared reference doses and concentrations (RfD, RfC) derived from human data with those derived from animal data for the same substances. For only 39% of substances was the human-based RfD/RfC within a factor of 3 of the animal-based value. This illustrates the fact that using only human data will generally produce different results from using only animal data. The human-based RfD/RfC was lower than the animal-based RfD in 36% of cases (Dourson *et al*, 2001), illustrating that animal-based assessments are not always more protective than human-based assessments. The size of the overall assessment factor used to derive an RfD can be thought of as an indicator of uncertainty, a higher assessment factor indicating more uncertainty. The overall assessment factors used to derive the animal-based RfDs may be expected to be 10 times greater than those used to derive the human-based RfDs, but often the differences seen by Dourson *et al* (2001) were greater than 10, reflecting a variety of additional uncertainties in extrapolating from the animal data to the human situation. Assessment factors with no relation to uncertainty can also be related to regulatory policies.

Due to the uncertainty of extrapolating from animal data to the human situation, it is often suggested that RfDs based purely on animal data are a regulatory level chosen with the intention of protecting the public, rather than a scientific prediction of the NOAEL for a vulnerable human sub-population.

Many animal studies use a dose level at or above the maximum tolerated dose to evaluate potential adverse effects. In general, information derived in this manner is irrelevant if the health of the animals is compromised. However, the data from these studies can contribute to hazard identification (the first stage in the risk assessment process) in that once an effect has been identified, lower doses can be used to characterise the dose-response for this effect. Classically, the most sensitive adverse endpoint in the most sensitive animal species is then used for the purposes of human risk assessment. The steps involved in interpreting a body of animal studies in the context of human risk are described in more detail in Sections 6.1 to 6.4.

6.1 Forming a mode of action hypothesis

The MoA has been defined to be “a plausible hypothesis, supported by observations and experimental data, regarding events leading to a toxic endpoint” (Meek *et al*, 2003). Mode of action is different to mechanism of action, in that the latter involves a detailed understanding of the molecular basis of the toxic effect (Schlosser and Bogdanffy, 1999). The importance of MoA in risk assessment is discussed by ECETOC (2006a). It may not be possible to formulate a hypothesis about the MoA, due to lack of data or fundamental understanding of the biology. On the other hand, there may be more than one credible hypothesis about the MoA. Any pattern or syndrome of toxicity may enable a MoA hypothesis to be formulated, and the concordance (or lack of it) between results in different sexes, species and strains often provides useful information, as does information on different chemicals causing the same pattern of toxicity. Valuable information can be obtained by taking the MoA hypothesis to a more formal level and a non-standard narrowly-focussed investigative study, designed specifically to explore the MoA, can be particularly helpful. This involves the identification of a chain of key cellular and biochemical events that result in the observed toxicity in animals. The evidence for these key events can then be reviewed against an adapted form of the Bradford Hill guidelines (Seed *et al*, 2005; Boobis *et al*, 2006) to determine if the weight of evidence is sufficient to establish a MoA in animals. Quality considerations for these types of study may be of less importance than for studies which form the basis for selecting critical NOAELs.

6.2 Dosimetry

In conjunction with the MoA, consideration should be given to dosimetry. The available evidence on the following points is relevant:

- Identity of the toxicant, i.e. compound as dosed or a metabolite;
- identity of target organ;
- dose-response, in terms of dose given to the animal or concentration of toxicant at the target organ;
- temporal aspects of the dose-response. [For example, is a single dose sufficient to produce an effect or is chronic exposure necessary; is the effect best related to the peak exposure concentration or to the area under the concentration-time curve?]

Both animal data *per se* and a variety of kinetic modelling techniques applied to animal data contribute to the analysis of dosimetry.

6.3 Considering relevance to man

Whether or not a MoA hypothesis can be formed or formally established in animals, it is necessary to consider the relevance of the animal findings to human health. In practice this is often an ongoing process. The default position is that the animal findings are considered relevant to man unless it can be demonstrated otherwise. Knowledge of the dosimetry associated with the critical effect can be an important component of this consideration.

In the absence of human data, it is difficult to discount relevance to man of animal findings as the burden of proof is high. Such arguments might however be made in situations where the target organ has no equivalent in man, such as the rat forestomach or Harderian gland. A formal procedure termed the IPCS Human Relevance Framework has been established to make judgements about the relevance to man of findings in animal studies for both cancer endpoints (Boobis *et al*, 2006), and non-cancer endpoints (Seed *et al*, 2005; Boobis *et al*, 2008). This procedure involves describing key events leading to the toxicity observed, and establishing the MoA in animals (Section 6.1). Each key event in animals is then evaluated for its plausibility in man. The MoA can be judged as not relevant to humans and reasonably excluded on the basis of:

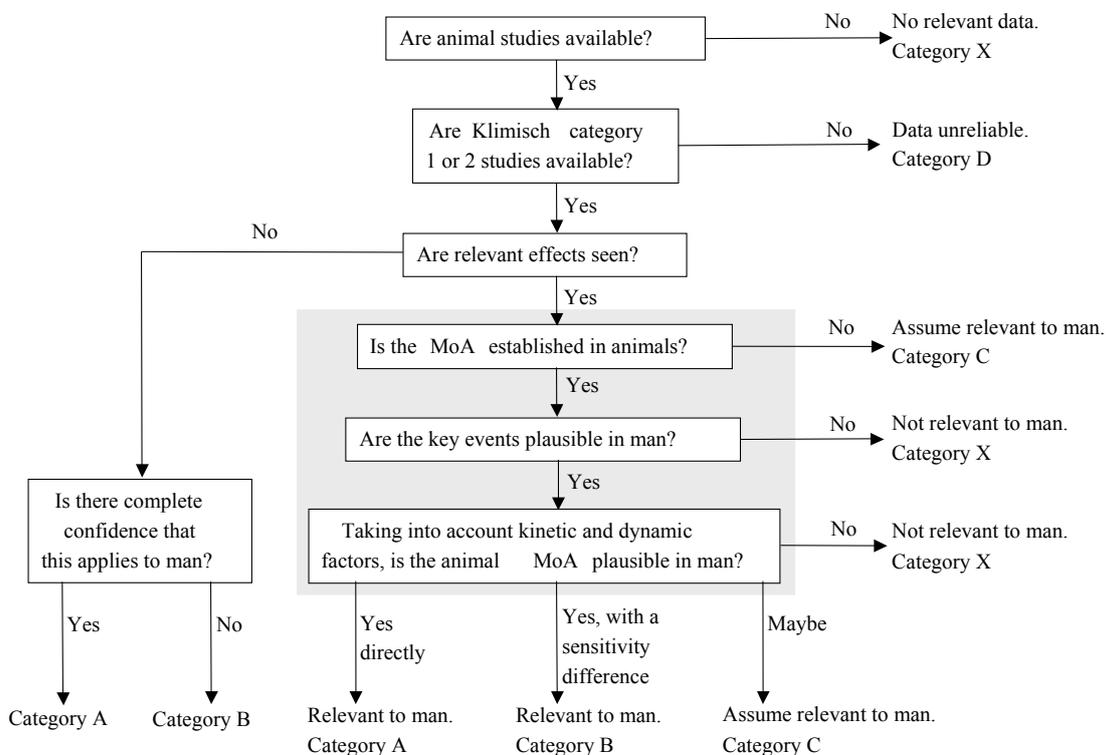
- Fundamental qualitative differences between the animal and humans, or
- quantitative differences in either kinetic or dynamic factors between animal and humans. (This involves a detailed consideration of the comparative dosimetry in man and animals).

The Task Force has extended the IPCS Human Relevance Framework by adding criteria for data availability and quality. The revised framework will result in the animal data being allocated to one of five categories in terms of the weight of evidence provided for human risk assessment (Figure 3). The categories are:

- A. Reliable animal data directly relevant to man. These could include reliable null findings, where there is complete confidence that these will apply to man.
- B. Reliable animal data relevant to man, with the inclusion of a correction for a sensitivity difference between animals and man, or reliable null data where there is no complete confidence that they will apply to man (default position for null data).
- C. Reliable animal data, with uncertain but assumed relevance to man.
- D. Unreliable animal data.
- X. No relevant animal data, or animal data that are not relevant to man.

If a toxic effect corresponding to a critical endpoint is discounted for its relevance to man, it will generally be necessary to revisit the animal data package to select a different toxic effect and endpoint for risk assessment purposes.

Figure 3: ECETOC scheme for categorising the relevance of animal data for use in human risk assessment^a



^a The shaded area represents the IPCS human relevance framework

6.4 Dose-route extrapolation

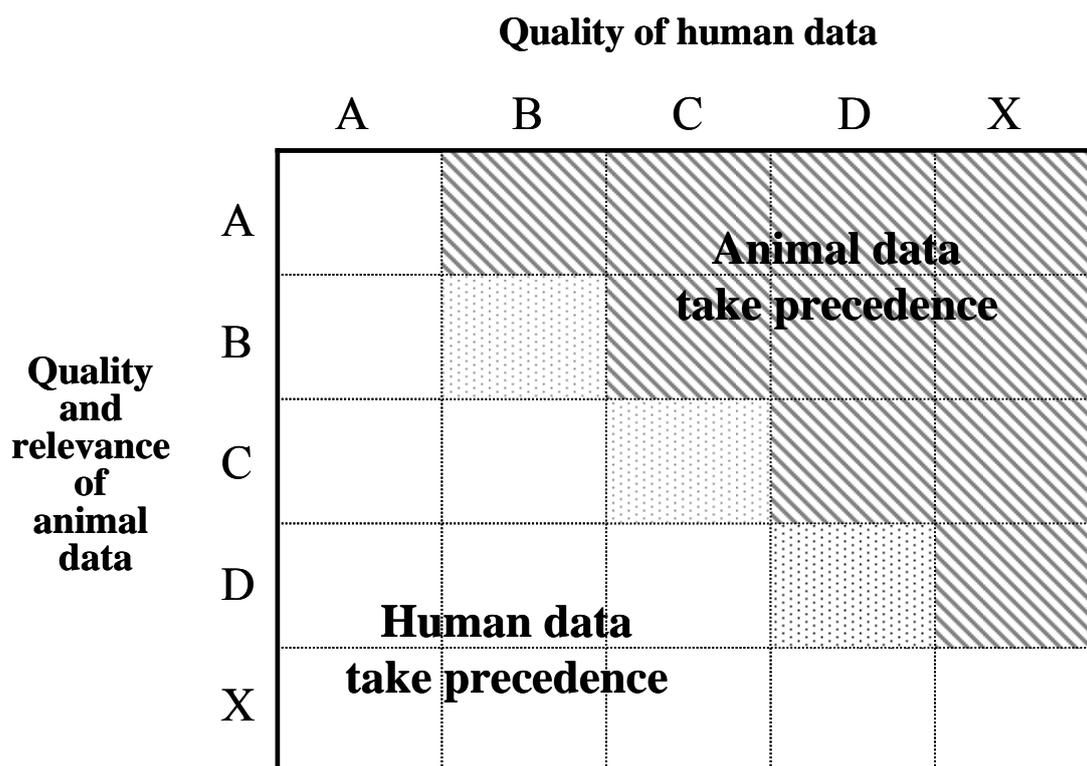
The route of human exposure being considered may differ from that in the critical animal study from which the NOAEL has been derived. This is not an ideal situation, and for this reason the relevance of the dose-route was considered to be a quality aspect of animal studies. However, this situation is often unavoidable and some form of correction is necessary to interpret the animal NOAEL in terms of the human exposure route. Defaults are sometimes used, but it is greatly preferable that this correction is done on the basis of data (e.g. data on the systemic exposure resulting from exposure via the routes concerned). Dosimetry considerations (Section 6.2) can assist with dose-route extrapolation; in some cases exposure modelling techniques are used. Modelling is a powerful tool for integrating human exposures from multiple routes, expressing it in terms of systemic dose or exposure at the target organ. Human exposure via multiple routes is common (e.g. inhalation and dermal, or dietary and dermal) and such modelling techniques are increasingly used in chemical risk assessments.

7. INTEGRATION OF HUMAN AND ANIMAL DATA

Figure 4 shows the components of a matrix that allows for the comparison and integration of human and animal data. Although Figure 4 could be seen to be inflexible in nature, the integration of human and animal data is not an exact process. The intention behind the matrix, and the overall framework, is not to support a prescriptive process but, rather, to serve as a transparent, well-founded and consistent basis for the comparative interpretation of human and animal findings. In this respect, two basic principles were adopted by the Task Force.

- The stronger evidence should be used to form the basis of any decision, whether that stronger evidence comes from human or animal data.
- Where human data of ‘High’ quality are available, these should take precedence in all circumstances.

Figure 4: Matrix for integrating human and animal data



Positive data take precedence (be it animal or human). If data are not in agreement, the data with a steeper slope or lower safe level should be used, but should be moderated by the upper risk level of the ‘less positive’ data (see text).

‘High’ quality human data (category A) should always be used as the basis for risk assessment, regardless of the quality of the animal data. Where it is considered that human data achieve this criterion, this indicates that the best methods, procedures and data sources have been used to assess risk in a relevant population.

Where human data and animal data are scored at an equivalent level (e.g. animal data are valid but not scaled correctly, and human data are not consistent or have other shortcomings) then it might be argued that one could justify the use of either data. A distinction does, however, need to be made dependent on the nature of the findings. When the human and animal data are concordant (e.g. both suggest a hazard, or both suggest similar safe exposure levels), it may not be important (e.g. in hazard identification) to distinguish which data source is used; in essence both or either can be used. However, when using the data to interpret, for example, dose-response characteristics or the magnitude of an LOAEL, the data that need fewer adjustment or assessment factors are preferred. In most cases this will be the human data.

On the other hand, where there are identical quality scores for human and animal data but the data are not concordant (apart from in the instance of category A human data which because of its quality are given precedence regardless), the framework confers a protective approach, consistent with normal practice in regulatory risk assessment. This operates as follows:

- Hazard assessment: Where human and animal data scores are identical (i.e. B/B, C/C, or D/D), the data that suggest a hazard should generally take precedence.
- Dose-response assessment: Where the animal data suggest a lower safe level (i.e. they are ‘more positive’) than the human data, or *vice versa*, and the scores are the same (i.e. B/B, C/C, or D/D), the data resulting in the ‘most positive’ (i.e. lower) safe level should take precedence; the other data source should be used as the upper limit. In this way, a range that is consistent with both data sets is obtained, while allowing the most positive data to be used in a protective manner (e.g. methylene chloride case study, Appendix B.9).

It is recognised that it is not possible to construct an integrative matrix that will give appropriate outcomes in all circumstances, and it is considered acceptable to deviate from the procedures outlined above, provided there are soundly-based and well-supported scientific reasons for so doing. If the quality of the data from the human or from the animal studies is considered to be no better than category C, hazard and risk assessments need to be considered with great caution, particularly if the data are not concordant.

8. PROPOSED FRAMEWORK FOR THE USE OF HUMAN AND ANIMAL DATA

An overall framework for the use of human and animal data is presented which builds upon previous chapters and uses all available human and animal data. The overall framework involves three steps shown schematically in Figure 5.

1. Assessment of the collective weight of evidence of the human data to the human risk question being considered. This results in a weight of evidence score for the quality of human data, based on a 5-point scale (A - D plus X) (Section 4.4).
2. Assessment of the collective weight of evidence of the animal data to the human risk question being considered. This results in a weight of evidence score for the quality and relevance of animal data, on a 5-point scale (A - D plus X) (Section 6.3).
3. Integrating the available evidence from human and animal data sources. This is undertaken using the matrix detailed in Chapter 7.

8.1 *Implementation of the framework for selected case studies*

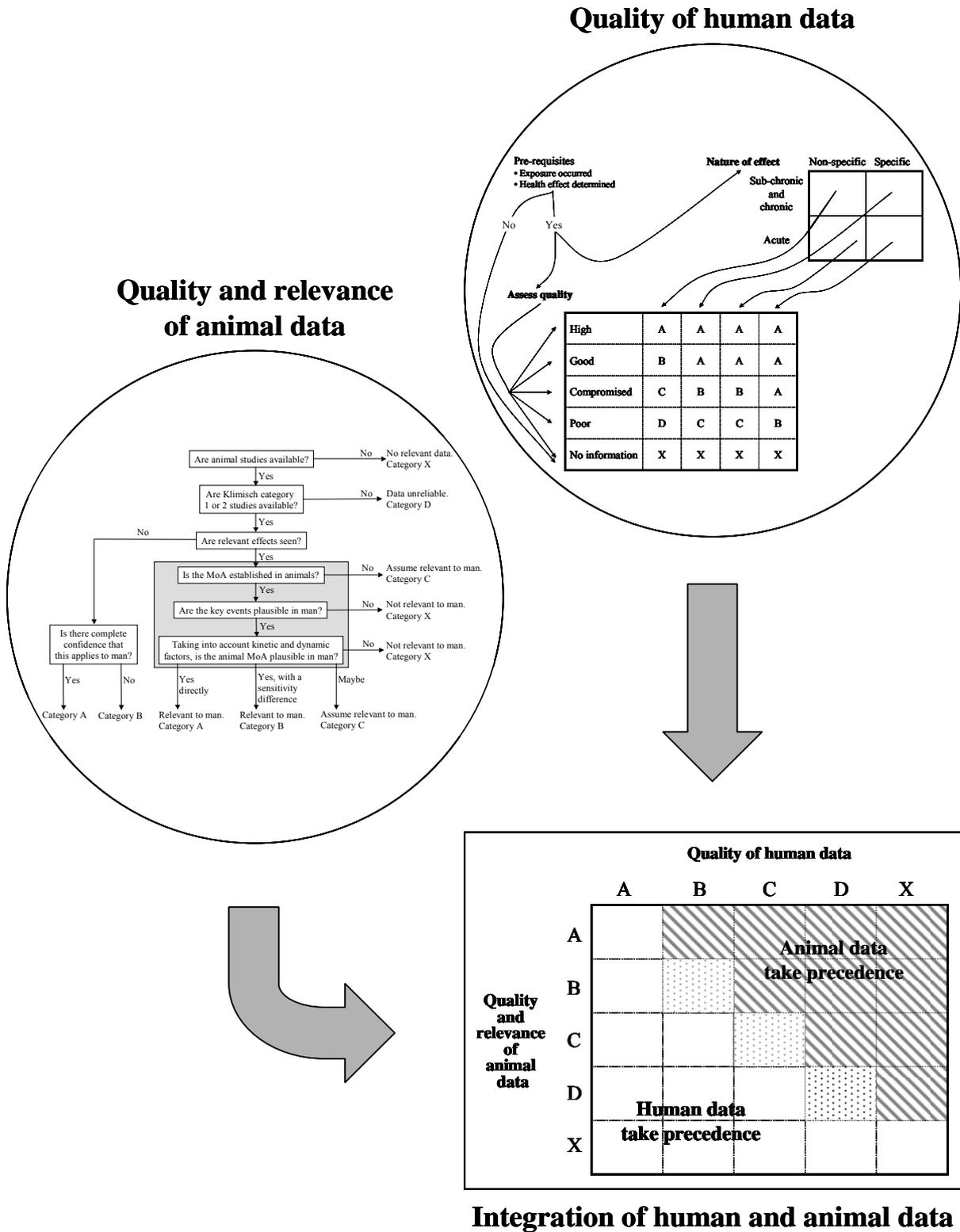
The overall framework (Figure 5) has been developed, tested and refined against a series of case studies. The Task Force sought to identify a suitable range of examples that were reflective of the types of situations that characterise where and how animal and human data need to be evaluated, compared and resolved. The case studies served to evaluate the utility and veracity of the approach, and helped to define the boundaries within which the framework can be applied reliably. These case studies are summarised in this section and are described in greater detail in Appendix B.

8.1.1 Scoring the quality of human data

Based on the considerations outlined in Section 4.4, the human data in each case study were evaluated in terms of a weight of evidence score, this being a function of the inherent data quality and the nature of the effect. Applying the principles of Section 4.4 to the case studies, a range of human data qualities were identified:

- A. Mectins for acute neurotoxicity (mectins NT), TeGenero antibody (TGN1412), hydrogen fluoride, welding fumes, wood dust, *bis*(chloromethyl)ether (BCME), toluene, carbon disulphide.
- B. Methylene chloride, reactive dye, sulphanilic acid, mectins for developmental toxicity (mectins DT).
- C. Benzanthrone.
- D. Dimethyl carbamyl chloride (DMCC), glycidol.
- X. No relevant example.

Figure 5: Proposed ECETOC framework for the use of human and animal data in chemical risk assessment



In the case of the case study on TGN1412, the use of animal and *in vitro* data is at issue, though the human clinical data are certainly category A.

The differences in nature of effects for the case studies are shown in Figure 6.

Figure 6: Distribution of nature of effects for case studies

	Non-specific	Specific
Sub-chronic and chronic	Methylene chloride DMCC BCME Toluene Wood dust Glycidol Mectins DT	Welding fumes Carbon disulphide
Acute	Mectins NT Benzanthrone	Sulphanilic acid TGN1412 Reactive dye Hydrogen fluoride

8.1.2 Relevance of animal data

Having categorised the available human data, the next step is to evaluate the animal data according to the process described in Chapter 6, i.e. following an initial assessment of inherent study quality (Chapter 5), through the categorisation resulting from the application of the extended IPCS human relevance framework (Section 6.3, Figure 3). Following the application of this step of the process, the animal data within the case studies were categorised according to the following scores:

- A. Mectins NT, hydrogen fluoride, BCME, glycidol, DMCC.
- B. Methylene chloride, TGN1412, benzanthrone, toluene.
- C. Carbon disulphide, sulphanilic acid, mectins DT.
- D. No example.
- X. Reactive dye, welding fumes, wood dust.

8.1.3 Integration of human and animal data

Applying the evaluations from Sections 8.1.1 and 8.1.2 for each case study to Figure 4 provides the basis for a transparent comparison of the relative strengths of the human and animal data. In this way, Figure 7 represents the distribution of the findings from all the case studies in terms of the influence of the available data (the specific circumstances of each case study are described in Appendix B). It can be seen that the examples are reasonably representative of the range of conditions where both human and animal data might be expected to have a bearing on the resolution of a risk assessment question.

Figure 7: Distribution of data qualities of case studies

		Quality of human data				
		A	B	C	D	X
Quality and relevance of animal data	A	Mectins NT Hydrogen fluoride BCME/CCME	BCME		Glycidol DMCC	
	B	Toluene TGN1412	Methylene chloride	Benzanthrone		
	C	Carbon disulphide	Sulphanilic acid Mectins DT			
	D					
	X	Wood dust Welding	Reactive dye			

8.2 *General observations on the case studies*

The case studies exhibit a number of characteristics, several of which are worth highlighting.

The reactive dye case study is useful in helping to understand how both the inherent study quality and the nature of the effect under examination are important in gauging human data quality. In essence, a few instances of human respiratory sensitisation were reported. Judgement of causality is facilitated by the acute nature of the respiratory sensitisation, the fact that the effect occurred at the site of exposure to the dye and was associated with its use (actually misuse), together with the specific diagnosis via immunological testing. The intrinsic quality of the human data would be rated as 'Good'. However, when Table 1 (Section 4.4.3) is used to account for the nature of the effect, an overall quality score of 'A' is obtained. Other data may also help in making the decision. In the case of the reactive dye, there are related dyes that were already known to be human respiratory sensitisers (Appendix B.5). [If, instead, the reactive dye had been under question as a possible cause of testicular cancer, then a much better quality of data would have been required to make such a determination, probably supported by other data (e.g. a known association of testicular cancer to related chemical structures, or data from animals)].

Sulphanilic acid tested positive for skin sensitisation in a well-conducted guinea pig maximisation test (GPMT). Under the scheme shown in Figure 3 (Section 6.3) the animal findings would be deemed category C. If only a small amount of poor quality human data had existed showing no skin sensitisation, then it would be reasonable to classify sulphanilic acid as a skin sensitiser. However, in many years of routine (and systematic) health monitoring of workers exposed to significant amounts of sulphanilic acid, no cases of skin sensitisation were observed. The human data would thus be rated as 'Good' for intrinsic quality. When Table 1 (Section 4.4.3) is used to account for the nature of the effect, the data would have an overall quality score of A. The quality of data should be sufficient to take precedence over the animal findings, with a resultant conclusion that sulphanilic acid should not be classified as a skin sensitiser (Appendix B.4).

The case of mectins (a group of chemicals including ivermectin) explores how human experience can aid in the classification of the substance in terms of its suitability for use during pregnancy and lactation. On the basis of animal findings, ivermectin was originally labelled as being unsuitable for use in pregnancy and lactation when used to treat onchocerciasis (category C under the scheme shown in Figure 3). However, epidemiology data provided an indication of its safety, based on accidental use in pregnancy. Under the framework, the human data would be rated as B. Thus the human data would take precedence over the animal findings. Indeed, in practice, once the human data were considered to be of sufficient quality and quantity, the warning advising restriction against its use in pregnancy and lactation was removed (Appendix B.7).

Clearly, an inappropriate dose was selected in the TGN1412 case study. [Initially, only animal and *in vitro* data (using material from animal, human or microbial sources) are generated on a novel pharmaceutical product. Subsequently, based on these data, a risk assessment for dosing in the first human trial is conducted.] For TGN1412, the first dose in man was based on a safe dose in animals and an assessment factor. This did not take into account the specific affinity of the antibody for human, relative to animal, target proteins, and this resulted in severe toxicity in the clinical trial (Appendix B.12). Several lessons can be learned from the TGN1412 case:

- Using animal data and applying an assessment factor is not always protective for humans.
- All human data (including data from human *in vitro* test systems) should be evaluated. In this case, *in vitro* data were available that showed that the chosen dose was excessive, but these data were not properly included in the risk assessment process.
- There are no good animal models for all human toxicities. In this case, the lack of significant cross-reactivity of TGN1412 for the animal equivalent of the human target protein meant that no animal species could be considered truly relevant for human risk assessment. This is a problem for all biopharmaceuticals that are designed to bind specifically to a particular human protein. There are three alternative solutions. Firstly, to use a molecule for which there is some cross-reactivity to the equivalent protein in some animal species (typically a primate). Secondly, to genetically engineer the human protein into a rodent strain and use this to generate animal data. Thirdly, to proceed cautiously into human clinical trials using data from *in vitro* studies in human test systems and modelling.

Finally, the methylene chloride case illustrates an approach where taking direct numerical account of apparently discordant human and animal data can improve the quality of a risk assessment. In this case, a well-conducted 2-year mouse study suggested a high dose effect for lung and liver cancer. However, a well-conducted epidemiology study, supported by detailed exposure monitoring histories, did not suggest an excess risk for either cancer site in humans. Subsequently, it became clear that the metabolic differences in mice versus other species were likely to be important in the observation of the mouse lung and liver tumours (ECETOC, 1989). In addition, while the epidemiology data were null, the size of the study was not big, therefore the confidence interval from the study could be used to account for the study size and results. Taking both these factors into account suggested that there was a region of consistency between the original animal and human data. Thus, the apparently discordant data may be better attributed to a difference in species sensitivity and ultimately considered as concordant with a difference in species sensitivity (Appendix B.9). The issue of study concordance is discussed in greater detail in Section 7.4.

8.3 *Concordance or lack of concordance between human and animal data*

There is concordance when animal and human data yield similar results in either hazard identification or hazard characterisation (i.e. dose-response curve characteristics, or safe level determinations). It should be noted that concordance is defined based on study outcome, i.e. the findings resulting from the human versus animal data set, not by the quality scores obtained.

By first assessing data quality, as in the scheme advanced by ECETOC, the issues surrounding the concordance between human and animal findings are only relevant when the human and animal data result in the same quality score (excepting the 'High' quality human data, category A). Otherwise, the higher quality data take precedence. When human and animal data are of a similar quality (i.e. the diagonal cells from the upper left hand corner to the lower right hand corner in Figure 4), it becomes important to assess concordance in terms of study outcome, in order to determine which should be given precedence. The only cell in which this is not recommended is the upper left, when both human and animal data quality are highest (A/A). In this case, the risk assessment should be based on the human data.

If human and animal data are concordant in terms of study outcome, it means that the whole dataset is consistent with the theory that the same biological mechanisms are present in animals and humans and, if the exposures are relevant, comparable toxic effects will develop. There may be a sensitivity difference between animals and humans in either direction or human exposure may not be sufficient for the toxicological effect to be manifest. This does not imply that there is no concordance in terms of outcome. If there is concordance, the choice between using human and animal data in the risk assessment is not as fundamental as when there is discordance, since both human and animal datasets indicate the same risk.

If there is discordance between the animal and the human data, the reasons for this difference should be considered. If the difference can be understood on the basis of the underlying biological mechanisms, then confidence in the risk assessment will increase. If the difference cannot be rationalised or understood, then the risk assessment may be less certain, especially if the weight of evidence of the data is weak (e.g. C/C).

When the quality of both sets of data is the same, but there is discordance in terms of effect, it is advisable to examine the data with respect to its variability using error bars or confidence intervals. Two point estimates of risk may be judged not to be concordant. However, if error bars or confidence intervals are placed around point estimates, a region of overlap will often emerge. In this case, the basis for the risk estimate should be within this region of overlap (Bukowski *et al*, 2001). The ECETOC framework uses this approach for safe level assessment (i.e. dose-response assessment). However, this approach is not possible for hazard assessment as

a decision needs to be made on whether or not a hazard is present. As such, the body of data suggesting that the hazard is present is used to set a safe level protective of human health.

In practice, the distinction between concordance and discordance in terms of effect is not always clear and any system, such as that proposed by ECETOC, that attempts to reconcile such differences must be flexible (Table 2). This is also reflected in the case studies.

Table 2: Considerations for data concordance

Are animal and human data concordant?	Illustrative case studies
Yes	Welding fumes, wood dust, benzanthrone, BCME, hydrogen fluoride, carbon disulphide
Yes, but there is a difference in the sensitivity between animals and humans	Mectins NT, methylene chloride, TGN1412
No	Sulphanilic acid
Not possible to judge	Reactive dye, glycidol, DMCC, toluene, mectins DT

Alternative evaluations are possible, depending on the database used. For example, Table 2 indicates a lack of concordance for the sulphanilic acid data. This would be overturned on the basis of a later evaluation considering the negative findings in the local lymph node assay and considered as a ‘yes’. Similarly, it would also be possible for TGN1412 to be considered ‘yes’, if the chemical reactivity shown by *in vitro* studies were to be examined in detail, or ‘no’ if this analysis were to be undertaken at a more simplistic level.

If there is concordance between the human and animal data, the risk assessment should use all the data, with both serving to reinforce the other. Similarly, where there is a sensitivity difference between humans and animals, then the same holds true provided the species differences are taken into account (e.g. mectins NT).

When data quality scores are weaker than ‘A’, it is often possible to recommend further studies that would upgrade the score. If the higher quality data so generated are at odds with the previous lower quality data, the higher quality data should be used, even in circumstances where they do not, in contrast to previous positive data, suggest a hazard.

For example, a more in-depth investigation revealed subsequently that the toxicity seen in the TGN1412 clinical trial should not have been unexpected. Further investigation of the skin sensitising potential of sulphanilic acid resulted in a study in a second animal model, the results of which matched the human experience of safety. Developing a better understanding of the

MoA can lead to new insights on species differences, and these insights will deepen as the science of human biology and genetics progresses. The relevance to man, for example, of the sensitivity to mectins of CF-1 mice lacking *p*-glycoprotein will ultimately be addressed using data on human genetics and the consequences for sensitivity variation in the human population assessed.

It is interesting to note that, if the case studies included in this report are at all typical, it seems to be unusual for there to be a lack of concordance between human and animal data that cannot be rationalised scientifically. In most cases the data support each other and, when used together, reduce uncertainty thus resulting in a more robust risk assessment than if either animal or human data alone are used.

9. CONCLUSIONS AND RECOMMENDATIONS

Both human and animal data have an important role to play in chemical risk assessment. Each has its own merits and each should be seen as a critical component of any well-conducted risk assessment. The aim of this report is to provide an approach whereby both can be evaluated and analysed with a comparable degree of rigour, and used in tandem to improve decision making in the context of risk assessment.

Up till the present time, an awareness of the limitations of epidemiology or observational studies has been used frequently to dismiss or reduce a potentially useful body of information. Too often emphasis has been given to the statistical power of a study at the expense of other characteristics. Regulators have clear guidance on acceptable animal toxicity studies and consequently animal data have received more attention in regulatory risk assessment than human data. There is no similar guidance available for human studies and there have been few efforts to systematically evaluate human data for risk assessment purposes. Yet, as indicated in the introduction, many regulatory bodies have expressed a clear preference for the use of human data for risk assessment.

This report proposes an integrative framework for human and animal data that assesses the quality of each database with respect to a given chemical, drug or exposure scenario (Figure 5). The aim of the ECETOC framework is to build on international collaborative efforts for the better integration of human relevance and MoA into good assessment practice and, in so doing, to broaden the application of these efforts and find practical applications. The Task Force also believes that the schemes proposed to score human data quality and categorise animal data will be beneficial *per se* (Figures 2 and 3).

Any attempt to systematise reporting, conduct, or classification of data is likely to be criticised. The objection is well-founded; if a classification framework is too rigid it can stifle creativity and if it is too lax, it may only provide the veneer of an evaluation. The Task Force believes that the primary benefit of the proposed ECETOC framework will be an evolving improvement towards the transparent evaluation and integration of human and animal data in the risk assessment process. Ultimately, that may result in improved study designs or better reporting of studies. The ECETOC integration framework should not be used overly prescriptively, but rather of as a guide to enhance the use of all study results in regulatory decision making.

One possible objection to the integrative matrix (Figure 4) as proposed in the ECETOC framework is that the same number of categories is proposed for both human data and animal data, and the implication is that the same non-validated scale is used for quality assessment. It is acknowledged that this could be a potential weakness. However, care was taken not to imply that the scores (A - D) be used in a summary fashion. For example, a chemical with human and animal data scores of B and B is not equivalent to a different example with scores A and C. But

there is more confidence that within each database, the strength of data for quality score A is superior to that for quality score B. There is less confidence that the quality scores of B for human, and B for animal, imply a strictly equivalent quality. However, the ECETOC framework has been developed and successfully tried by using these scores across both sets of data in a number of selected case studies. The Task Force believes that further testing via a wider variety of case studies will add confidence (or suggest revisions) to the existing criteria for scoring both sets of data.

Rothman and Poole (2007) have commented on guidelines for strengthening the reporting of observational studies in epidemiology (as published by Von Elm *et al*, 2007). They suggested that guidelines be published with expiry dates to ensure that they are not used past a time where they are reflective of current practice. For risk assessment, this is particularly germane, since the field is likely to change appreciably in the future (NRC, 2007). In view of that, the Task Force recommends that the proposed framework for the integration of human and animal data in chemical risk assessment be considered current for a period of 5 years and then be revisited.

In the meantime, the use of the framework and the collection of experience should be promoted among the regulatory community. More specific guidance can be developed in specific situations, such as derived no effect levels (DNELs) in Europe and the GHS globally that rely on or give weight to human data,.

BIBLIOGRAPHY

- ACGIH. 2001. Glycidol, bis(chloromethyl)ether. In *Documentation of the threshold limit values and biological exposure indices*, 7th edition. American Conference of Governmental Industrial Hygienists, Cincinnati, Ohio, USA.
- ACGIH. 2003. Dimethyl carbamoyl chloride. In *Documentation of the threshold limit values and biological exposure indices*, 7th edition, 2003 supplement. American Conference of Governmental Industrial Hygienists, Cincinnati, Ohio, USA.
- ACGIH. 2006. Hydrogen fluoride, toluene. In *Documentation of the threshold limit values and biological exposure indices*, 7th edition, 2006 supplement. American Conference of Governmental Industrial Hygienists, Cincinnati, Ohio, USA.
- Albert RE, Sellakumar AR, Laskin S, Kuschner M, Nelson N, Snyder CA. 1982. Gaseous formaldehyde and hydrogen chloride induction of nasal cancer in the rat. *J Nat Cancer Inst* 68:597-603.
- Alonzo C, Laborde A. 2005. Priority setting for risk assessment – the benefit of human experience. *Toxicol Appl Pharmacol* 207S2:692-696.
- Antonini JM, Lewis AB, Roberts JR, Whaley DA. 2003. Pulmonary effects of welding fumes: Review of worker and experimental animal studies. *Am J Ind Med* 43:350-360.
- Basketter DA, Scholes EW, Cumberbatch N, Evans CD, Kimber I. 1992. Sulphanilic acid: Divergent results in the guinea pig maximization test and the local lymph node assay. *Contact Dermatitis* 27:209-213.
- Basketter DA, Gerberick GF, Kimber I. 1998. Strategies for identifying false positive responses in predictive skin sensitization tests. *Food Chem Toxicol* 36:327-333 [<http://altweb.jhsph.edu/publications/llna/basketter.htm>].
- Boobis AR, Cohen SM, Dellarco V, McGregor D, Meek ME, Vickers C, Willcocks D, Farland W. 2006. IPCS framework for analysing the relevance of a cancer mode of action for humans. *Crit Rev Toxicol* 36:781-792.
- Boobis AR, Doe JE, Heinrich-Hirsch B, Meek ME, Munn S, Ruchirawat M, Schlatter J, Seed J, Vickers C. 2008. IPCS framework for analysing the relevance of a non-cancer mode of action for humans. *Crit Rev Toxicol* 38:87-96.

Bridges J. 2003. Human health and environmental risk assessment: The need for a more harmonised and integrated approach. *Chemosphere* 52:1347-1351.

BUA. 2004. Benzanthron. BUA-Soffbericht 251. Gesellschaft Deutscher Chemiker, Beratergremium für Altstoffe. S. Hirzel Verlag, Stuttgart, Germany.

Bukowski JA, Schnatter AR, Korn L. 2001. Using epidemiological studies to check the consistency of the cancer risks predicted by high-dose animal experiments: A methodological review. *Risk Analysis* 21:601-612.

Chandra SV, Butler WH, Magos L. 1972. The effect of carbon disulphide on the myocardium of the rat. *Exp Mol Pathol* 17:249-259.

Chang SJ, Chen CJ, Shih TS, Chou TC, Sung FC. 2007. Risk for hypertension in workers exposed to carbon disulfide in the viscose rayon industry. *Am J Ind Med* 50:22-27.

Chouanière D, Wild P, Fontana JM, Hery M, Fournier M, Baudin V, Subra I, Rousselle D, Toamain JP, Saurin S, Ardiot MR. 2002. Neurobehavioral disturbances arising from occupational toluene exposure. *Am J Ind Med* 41:77-88.

Department of Health. 2006. Expert Scientific Group on phase one clinical trials: A consultation. Department of Health, London UK
[\[www.dh.gov.uk/en/Consultations/Closedconsultations/DH_4139038\]](http://www.dh.gov.uk/en/Consultations/Closedconsultations/DH_4139038).

DFG. 2004. Dimethylcarbmidsäurechlorid. In Henschler D, ed, *Gesundheitsschädliche Arbeitsstoffe, Toxikologisch-Arbeitsmedizinische Begründung von MAK-Werten*, 39th ed. Wiley-VCH, Weinheim, Germany.

DFG. 2007. Schwefelkohlenstoff, Toluol. In Greim H, ed, *MAK- und BAT-Werte-Liste 2007. Senatskommission zur Prüfung gesundheitsschädlicher Arbeitsstoffe. Mitteilung 43*. Deutsche Forschungsgemeinschaft, Wiley-VCH, Weinheim, Germany, pp 116-123

Dixie Chemical. 1992. Communication to ACGIH from Dixie Chemical Company, Houston, Texas, USA.

Dixie Chemical. 1993. Communications provided to ACGIH for chemical substances. Dixie Chemical Company, Houston, Texas.

Docker A, Wattie JM, Topping MD, Luczynska CM, Newman Taylor AJ, Pickering CA, Thomas P, Gompertz D. 1987. Clinical and immunological investigations of respiratory disease in workers using reactive dyes. *Brit J Ind Med* 44:534-541.

Doe J, Boobis A, Blacker A, Dellarco V, Doerr N, Franklin C, Goodman J, Kronenberg J, Lewis R, McConnell E, Mercier T, Moretto A, Nolan C, Padilla S, Phang W, Solecki R, Tilbury L, van Ravenzwaay B, Wolf D. 2006. A tiered approach to systemic toxicity testing for agricultural chemical safety assessment. *Crit Rev Toxicol* 36:37-68.

Dourson ML, Andersen ME, Erdreich LS, MacGregor JA. 2001. Using human data to protect the public's health. *Regul Toxicol Pharmacol* 33:234-256.

EC. 1993. Council Regulation (EEC) 793/93 of 23 March 1993 on the evaluation and control of the risks of existing substances. *OJ EC* L84:1-75.

EC. 1994. Technical guidance on risk assessment of existing substances in the context of Commission Regulation (EC) No. 1488/94 in accordance with Council regulation (EEC) No. 793/93 on the evaluation and control of existing substances. Commission of the EC, doc XI/919/94-EN. European Commission Joint Research Centre, European Chemicals Bureau, Ispra, Italy.

ECB. 2003. Technical guidance document on risk assessment. Document EUR 20418 EN. European Chemicals Bureau. European Commission Joint Research Centre, Ispra, Italy [<http://ecb.jrc.it/Documents/tgd/>].

ECETOC. 1989. Methylene chloride (dichloromethane): An overview of experimental work investigating species differences in carcinogenicity and their relevance to man. Technical Report No. 34. European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels, Belgium.

ECETOC. 2002a. Use of human data in hazard classification for irritation and sensitisation. Monograph 32. European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels, Belgium.

ECETOC. 2002b. Recognition of, and differentiation between, adverse and non-adverse effects in toxicology studies. Technical Report No. 85. European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels, Belgium.

ECETOC. 2003. Derivation of assessment factors for human health risk assessment. Technical Report No. 86. European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels, Belgium.

ECETOC. 2004. Workshop on the use of human data in risk assessment, 23-24 February 2004, Cardiff. Workshop Report No. 3. European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels, Belgium.

ECETOC. 2006a. Toxicological modes of action: Relevance for human risk assessment. Technical Report No. 99. European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels, Belgium.

ECETOC. 2006b. Guidance for setting occupational exposure limits: Emphasis on data-poor substances. Technical Report No. 101. European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels, Belgium.

Ehrenberg L, Hussain S. 1981. Genetic toxicity of some important epoxides. *Mutat Res* 86:1-113.

EPA. 1985. Health assessment document for dichloromethane (methylene chloride). EPA/600/8-82/004FF. Final Report. Office of Health and Environmental Assessment, US Environmental Protection Agency, Washington, DC, USA.

EPA. 1993. Reference dose: Description and use in human risk assessments. US Environmental Protection Agency, Washington, USA [www.epa.gov/iris/rfd.htm].

EPA. 2004. Risk assessment principles and practices. An examination of EPA risk assessment principles and practices. Staff Paper, Office of the Science Advisor. US Environmental Protection Agency, Washington DC, USA [www.epa.gov/OSA/pdfs/ratf-final.pdf].

EPA. 2005. Guidelines for Carcinogen risk assessment. Document EPA/630/P-03/001F. US Environmental Protection Agency, Washington, USA [<http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=116283>].

EU. 2007. Corrigendum to Regulation (EC) 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation Authorisation and Restriction of Chemicals (REACH). *OJ EU* L136:3-280 [<http://eur-lex.europa.eu/JOHtml.do?uri=OJ:L:2007:136:SOM:EN:HTML>].

FDA. 1996. Approval package for application number 050742, trade name: Mectizan, generic name: Ivermectin, sponsor: Merck Research Laboratories. Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Beltsville, Maryland [www.fda.gov/cder/foi/nda/96/050742ap.pdf].

- Goldman PJ. 1972. Allergene und Hautkrankheiten in der werksärztlichen Praxis eines chemischen Industriebetriebes. *Arbeitsmed Sozialmed Arbeitshyg* 7:98-100.
- Goldschmidt BM, Langseth L, Mercado G, Sivak A. 1968. Alpha-haloethers: A new type of alkylating carcinogen. *Arch Environ Health* 16:472-476.
- Graham JG, ed. 1995. The role of epidemiology in regulatory risk assessment. Excerpta Medica International Congress Series no 1092. Elsevier, Amsterdam, Netherlands.
- Greene JF, Hays S, Paustenbach D. 2003. Basis for a proposed reference dose (RfD) for dioxin of 1-10 pg/kg-day: A weight of evidence evaluation of the human and animal studies. *J Toxicol Environ Health B*:115-159.
- Greenland S. 1994. Invited commentary: A critical look at some popular meta-analytic methods. *Am J Epidemiol* 140:290-296.
- Guzelian PS, Victoroff MS, Halmes NC, James RC, Guzelian CP. 2005. Evidence-based toxicology: A comprehensive framework for causation. *Hum Exp Toxicol* 24:161-201.
- Health Canada. 1994. Human health risk assessment for priority substances. Publications, Health Canada, Ottawa, Ontario, Canada
[\[http://www.hc-sc.gc.ca/ewh-semt/pubs/contaminants/approach/index-eng.php\]](http://www.hc-sc.gc.ca/ewh-semt/pubs/contaminants/approach/index-eng.php).
- Hearne FT, Grose F, Pifer JW, Friedlander BR, Raleigh RL. 1987. Methylene chloride mortality study: Dose-response characterization and animal model comparison. *J Occup Med* 29:217-228.
- Hertz-Picciotto I. 1995. Epidemiology and quantitative risk assessment: A bridge from science to policy. *Am J Public Health* 85:484-491.
- Hill AB. 1963. Medical ethics and controlled trials. *BMJ* 1:1043-1049.
- Hill AB. 1965. The environment and disease: Association or causation? President's address. *Proc R Soc Med* 9:295-300.
- Hudak A, Ungváry G. 1978. Embryotoxic effects of benzene and its methyl derivatives: Toluene, xylene. *Toxicology* 11:55-63.
- IARC. 1987. bis(Chloromethyl)ether. In *Monographs on the evaluation of carcinogenic risks to humans* supplement 7. International Agency for Research on Cancer, Lyon, France.

IARC. 1999. Re-evaluation of some organic chemicals, dimethylcarbamoyl chloride. In *Monographs on the evaluation of carcinogenic risks to humans* vol 71. International Agency for Research on Cancer, Lyon, France, pp 531-544.

IARC. 2000. Some industrial chemicals, glycidol. In *Monographs on the evaluation of carcinogenic risks to humans* vol 77. International Agency for Research on Cancer, Lyon, France, pp 469-486.

IARC. 2006. Preamble to the IARC Monographs. International Agency for Research on Cancer, Lyon, France [<http://monographs.iarc.fr/ENG/Preamble/index.php>].

IGHRC. 2003. Uncertainty factors: Their use in human health risk assessment by UK Government. Interdepartmental Group on Health Risks from Chemicals. Institute for Environment and Health, Leicester, UK.

IPCS. 1999. Principles for the assessment of risks to human health from exposure to chemicals. Environmental Health Criteria No. 210. International Programme on Chemical Safety, World Health Organisation, Geneva, Switzerland. [www.who.int/pcs/risk-assessment-ehc/docs/ehc210_exposure.htm#3.2].

IPCS. 2005. Workshop on poison centres and the use of human data in consumer product risk assessment. Report. WHO in collaboration with EC JRC, EAPCCT, ECETOC and the Federal Institute of Risk Assessment, Germany held at the Federal Institute for Risk Assessment, Berlin, Germany, 9 May 2005. International Programme on Chemical Safety, World Health Organisation, Geneva, Switzerland. [www.who.int/ipcs/publications/methods/workshop_report.pdf].

Iregren A. 1982. Effects on psychological test performance of workers exposed to a single solvent (toluene) - a comparison with effects of exposure to a mixture of organic solvents. *Neurobehav Toxicol Teratol* 4:695-701.

Irwin RD, Eustis SL, Stefanski S, Haseman JK. 1996. Carcinogenicity of glycidol in F344 rats and B6C3F₁ mice. *J Appl Toxicol* 16:201-209.

Jurek AM, Greenland S, Maldonado G, Church TR. 2005. Proper interpretation of non-differential misclassification effects: Expectations vs observations. *Int J Epidemiol* 34:680-687.

Jurek AM, Maldonado G, Greenland S, Church TR. 2006. Exposure-measurement error is frequently ignored when interpreting epidemiologic study results. *Eur J Epidemiol* 21:871-876.

- Klimisch HJ, Andreae M, Tillmann U. 1997. A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. *Regulat Toxicol Pharmacol* 25:1-5.
- Kotseva K. 2001. Occupational exposure to low concentrations of carbon disulfide as a risk factor for hypercholesterolaemia. *Int Arch Occup Environ Health* 74:38-42.
- Lee JA, Vaughan TL, Diehr PH, Haertle RA. 1985. The recognition of new kinds of occupational toxicity. In Castellani A, ed, *Epidemiology and quantitation of environmental risk in humans from radiation and other agents*. Plenum, New York, USA, pp 307-337.
- Lewandowski TA, Rhomberg LR. 2005. A proposed methodology for selecting a trichloroethylene inhalation unit risk value for use in risk assessment. *Regul Toxicol Pharmacol* 41:39-54.
- Lewis RW, Billington R, Debruyne E, Gamer A, Lang B, Carpanini F. 2002. Recognition of adverse and nonadverse effects in toxicity studies. *Toxicol Pathol* 30:66-74.
- Lilienblum W, Dekant W, Foth H, Gebel T, Hengstler JG, Kahl R, Kramer PJ, Schweinfurth H, Wollin KM. 2008. Alternative methods to safety studies in experimental animals: role in the risk assessment of chemicals under the new European Chemicals Legislation (REACH). *Arch Toxicol* 82:211-236.
- Liou SH, Cheng SY, Lai FM, Yang JL. 1996. Respiratory symptoms and pulmonary function in mill workers exposed to wood dust. *Am J Ind Med* 30:293-299.
- Lund K, Ekstrand J, Boe J, Sørstrand P, Kongerud J. 1997. Exposure to hydrogen fluoride: An experimental study in humans of concentrations of fluoride in plasma, symptoms, and lung function. *Occup Environ Med* 54:32-37.
- Lund K, Refsnes M, Sandstrøm T, Sørstrand P, Schwarze P, Boe J, Kongerud J. 1999. Increased CD3 positive cells in bronchoalveolar lavage fluid after hydrogen fluoride inhalation. *Scand J Work Environ Health* 25:326-334.
- Luo JCJ, Hsu KH, Shen WS. 2006. Pulmonary function abnormalities and airway irritation symptoms of metal fumes exposure on automobile spot welders. *Am J Ind Med* 49:407-416.
- Määttä J, Lehto M, Leino M, Tillander S, Haapakoski R, Majuri MJ, Wolff H, Rautio S, Welling I, Husgafvel-Pursiainen K, Savolainen K, Alenius H. 2006. Mechanisms of particle-induced pulmonary inflammation in a mouse model: Exposure to wood dust. *Toxicol Sci* 93:96-104.

Mandryk J, Alwis KU, Hocking AD. 1999. Work-related symptoms and dose-response relationships for personal exposures and pulmonary function among woodworkers. *Am J Ind Med* 35:481-490.

Marquart H, Smid T, Heederik D, Visschers M. 1989. Lung function of welders of zinc-coated mild steel: Cross-sectional analysis and changes over five consecutive work shifts. *Am J Ind Med* 16:289-296.

Mayor S. 2006. Severe adverse reactions prompt call for trial design changes. *BMJ* 332:683 [[doi:10.1136/bmj.332.7543.683](https://doi.org/10.1136/bmj.332.7543.683)].

Meek ME, Bucher JR, Cohen SM, Dellarco V, Hill RN, Lehman-McKeeman LD, Longfellow DG, Pastoor T, Seed J, Patton DE. 2003. A framework for human relevance analysis of information on carcinogenic modes of action. *Crit Rev Toxicol* 33:591-653.

Money CD. 2007. The use of human experience data in the EU risk assessment process. *Risk Analysis* 27:387-396 [[doi:10.1111/j.1539-6924.2007.00891.x](https://doi.org/10.1111/j.1539-6924.2007.00891.x)].

Money CD, Margary SA. 2002. Improved use of workplace exposure data in the regulatory risk assessment of chemicals within Europe. *Ann Occup Hyg* 46:279-285.

Nilsson R, Nordlinder R, Wass U, Meding B, Belin L. 1993. Asthma, rhinitis, and dermatitis in workers exposed to reactive dyes. *Brit J Ind Med* 50:65-70.

NIOSH. 1978. Dimethylcarbamoyl chloride current intelligence bulletin 11, revised. National Institute for Occupational Safety and Health, Atlanta, Georgia, USA [DHEW (NIOSH) publication 78-127; NTIS publication PB-83-105-080].

NRC. 1983. Risk assessment in the Federal Government: Managing the process. US National Research Council, Washington DC, USA.

NRC. 2007. Toxicity testing in the twenty-first century: A vision and a strategy. Committee on Toxicity Testing and Assessment of Environmental Agents, US National Research Council. National Academies Press, Washington DC, USA.

NTP. 1986. Toxicologic and carcinogenesis studies of dichloromethane in F344/N rats and B6C3F1 mice (inhalation studies). NTP-TRS-306. US National Toxicology Program, USA.

NTP. 1990. Toxicology and carcinogenesis studies of glycidol (CAS No. 556-52-5) in F344/N rats and B6C3F1 mice (gavage Studies). NTP technical report 374. NIH publication 90-2829. National Toxicology Program, US Department of Health and Human Services, Public Health Service, National Institutes of Health, Research Triangle Park, North Carolina, USA.

OECD. 2007. Manual for investigation of HPV chemicals. Organisation for Economic Co-operation and Development, Paris, France
[\[www.oecd.org/document/7/0,3343,en_2649_34379_1947463_1_1_1_1,00.html\]](http://www.oecd.org/document/7/0,3343,en_2649_34379_1947463_1_1_1_1,00.html).

Olson H, Betton G, Robinson D, Thomas K, Monro A, Kolaja G, Lilly P, Sanders J, Sipes G, Bracken W, Dorato M, Van Deun K, Smith P, Berger B, Heller A. 2000. Concordance of the toxicity of pharmaceuticals in humans and in animals. *Regul Toxicol Pharmacol* 32:56-67.

Ott MG, Skory LK, Holder BB, Bronson JM, Williams PR. 1983. Health evaluation of employees occupationally exposed to methylene chloride: Mortality. *Scand J Work Environ Health* 9S1:8-16.

Park HW, Kim, DI, Sohn SW, Park CH, Kim SS, Chang YS, Min KU, Kim YY, Cho SH. 2007. Outcomes in occupational asthma caused by reactive dye after long-term avoidance. *Clin Exp Allergy* 37:225-230.

Peters D, Miethchen R. 1996. Symptoms and treatment of hydrogen fluoride injuries. *Journal of Fluorine Chemistry* 79:161-165.

Popp W, Vahrenholz C, Schmieding W, Krewet E, Norporth K. 1991. Investigations of the frequency of DNA strand breakage and cross-linking and of sister chromatid exchange in the lymphocytes of electric welders exposed to chromium- and nickel-containing fumes. *Int Arch Occup Environ Health* 63:115-120.

Press E, Done AK. 1967. Solvent sniffing. *Pediatrics* 39:451-461.

Robert LG, Nicolich MJ, Schreiner CA. 2007. Developmental and reproductive toxicity evaluation of toluene vapor in the rat II, developmental toxicity. *Reprod Toxicol* 23:521-531.

Rothman KJ, Greenland S. 1998. Types of epidemiological studies. In Rothman KJ, Greenland S, eds, *Modern epidemiology*. Lippincott Williams and Wilkins, Philadelphia, USA, pp 133-134.

Rothman KJ, Poole C. 2007. Some guidelines on guidelines: They should come with expiration dates. *Epidemiology* 18:794-796.

Schlosser PM, Bogdanffy MS. 1999. Determining modes of action for biologically based risk assessments. *Regul Toxicol Pharmacol* 30:75-79.

Schlunssen V, Schaumburg I, Ebbe T, Mikkelsen AB, Sigsgaard T. 2002. Respiratory symptoms and lung function among Danish woodworkers. *J Occup Environ Med* 44:82-98.

Seeber A, Schäper M, Zupanic M, Blaszkewicz M, Demes P, Kiesswetter E, van Thriel C. 2004. Toluene exposure below 50 ppm and cognitive function: A follow-up study with four repeated measurements in rotogravure printing plants. *Int Arch Occup Environ Health* 77:1-9.

Seed J, Carney EW, Corley RA, Crofton KM, DeSesso JM, Foster PMD, Kavlock R, Kimmel G, Klaunig J, Meek ME, Preston RJ, Slikker W, Jr, Tabacova S, Williams GM, Wiltse J, Zoeller RT, Fenner-Crisp P, Patton DE. 2005. Overview: Using mode of action and life stage information to evaluate the human relevance of animal toxicity data. *Crit Rev Toxicol* 35:663-672.

Sellakumar AR, Laskin S, Kuschner M, Rusch G, Katz GV, Snyder CA, Albert RE. 1980. Inhalation carcinogenesis by dimethyl carbamoyl chloride in Syrian golden hamsters. *J Environ Pathol Toxicol* 4:107-115.

Sellakumar AR, Snyder CA, Solomon JJ, Albert RE. 1985. Carcinogenicity of formaldehyde and hydrogen chloride in rats. *Toxicol Appl Pharmacol* 81:401-406.

Singh GB, Zaidi SH. 1969. Preliminary clinical and experimental studies on benzanthrone toxicity. *J Ind Med Assoc* 52:558-560.

Srivastava LP, Misra RB, Joshi PC. 1990. Photosensitising potential of benzanthrone. *Food Chem Toxicol* 28:653-658.

Stein HP. 1976. Meeting on dimethyl carbamoyl chloride (DMCC) and diethyl carbamoyl chloride (DECC) at New York University Medical Center, October 27-28. Memorandum. Health Services and Mental Health Administration, Washington DC, USA.

Stern RM, Hansen K, Madsen AF, Olsen KM. 1988. *In vitro* toxicity of welding fumes and their constituents. *Environ Res* 46:168-180.

Sulsky SI, Hooven FH, Burch M, Mundt K. 2002. Critical review of the epidemiological literature on the potential cardiovascular effects of occupational carbon disulfide exposure. *Int Arch Occup Environ Health* 75:365-380.

Swaen G. 2006. A framework for using epidemiological data for risk assessment. *Hum Exp Tox* 25:147-155.

SZW. 2006. De nationale MAC lijst. 2006. Ministerie van Sociale Zaken en Werkgelegenheid. Sdu, Servicecentrum Uitgevers, Den Haag, The Netherlands.

Takeuchi Y, Onon Y, Hisanaga N. 1981. An experimental study on the combined effects of *n*-hexane and toluene on the peripheral nerve of the rat. *Brit J Ind Med* 38:14-19.

Tan X, Peng X, Wang F, Joyeux M, Hartemann P. 2002. Cardiovascular effects of carbon disulfide: Meta-analysis of cohort studies. *Int J Hyg Environ Health* 205:473-477.

Uemitsu N, Shimizu Y, Hosokawa T, Nakayoshi H, Kobayashi M, Minato S, Ogawa T, Tsutsumi S. 1984. Inhalation toxicity study of welding fumes: Effect of single or repeated exposure on the lungs of rats. *Toxicology* 30:75-92.

Ungváry G, Mányay S, Tátray E, Szeberényi S, Cseh RJ, Molnár J, Folly G. 1980. Effect of toluene inhalation on the liver of rats - dependence on sex dose and exposure time. *J Hyg Epidemiol Microbiol Immunol* 24:242-52.

Van den Brandt P, Voorrips L, Hertz-Picciotto I, Shuker D, Boeing H, Speijers G, Guittard C, Kleiner J, Knowles M, Wolk A, Goldbohm A. 2002. The contribution of epidemiology. *Food Chem Toxicol* 40:387-424.

Van Duuren BL, Goldschmidt BM, Langseth L, Mercado G, Sivak A. 1968. Alpha-haloethers: A new type of alkylating carcinogen. *Arch Environ Health* 16:472-476.

Vandenbroucke JP. 2001. In defense of case reports and case series. *Ann Intern Med* 134:330-334 [www.annals.org].

Vermeire T, Stevenson H, Pieters M, Rennen M, Slob W, Hakkert BC. 1999. Assessment factors for human health risk assessment: A discussion paper. *Crit Rev Toxicol* 29:439-490.

Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. 2007. The strengthening of reporting of observational studies in epidemiology (STROBE) statement: Guidelines for reporting observational studies. *Lancet* 370:1453-1457.

VROM. 2001. Hydrogen fluoride, CAS-no: 7664-39-3, EINECS-no: 231-634-8, summary risk assessment report. Special publication I.00.129. Netherlands Ministry of Spatial Planning, Housing and the Environment, The Hague, the Netherlands. European Commission, Joint Research Centre, Institute for Health and Consumer Protection, European Chemicals Bureau Ispra, Italy.

WHO. 2002. Toxicological evaluation of certain veterinary drug residues in food: Doramectin. Joint FAO/WHO Expert Committee on Food Additives (JECFA). WHO Food Additives Series 49. World Health Organisation, Geneva, Switzerland
[\[www.inchem.org/documents/jecfa/jecmono/v49je02.htm\]](http://www.inchem.org/documents/jecfa/jecmono/v49je02.htm).

LIST OF SPECIAL ABBREVIATIONS

BCME	<i>bis</i> (Chloromethyl)ether
CMME	Chloromethyl methyl ether
CNS	Central nervous system
DMCC	Dimethylcarbamoyl chloride
DNA	Deoxyribonucleic acid
DNEL	Derived no effect level
GHS	Globally harmonised system for the classification and labelling of chemicals
GLP	Good laboratory practice
GPMT	Guinea pig maximisation test
IPCS	International Programme on Chemical Safety
kgbw	Kilogramme body weight
LOAEL	Lowest observed adverse effect level
LOEL	Lowest observed effect level
MAK	Maximale Arbeitsplatzkonzentration
MoA	Mode of action
NOAEL	No observed adverse effect level
NOEL	No observed effect level
OECD	Organisation for Economic Co-operation and Development
OEL	Occupational exposure limit
QSAR	Quantitative structure-activity relationship
REACH	Registration, evaluation, authorisation and restriction of chemicals
RfC	Reference concentration
RfD	Reference dose
TLV	Threshold limit value
TWA	Time-weighted average

APPENDIX A: TERMS OF REFERENCE

Objective: Facilitate the development of a framework/approach which enables different types and qualities of existing human data to be integrated with animal or other data for use in different risk assessment and risk management applications, particularly in the context of current GHS and REACH proposals.

1. Review and evaluate:
 - Different types of observational human data (epidemiology, human volunteer, case report, surveillance and clinical data);
 - different qualities (positive and null) of human data, both within and across types, for specific clinical endpoints;
 - means by which available human data might be evaluated by weight of evidence;
 - means by which human and animal data can be integrated or used in tandem in an overall weight of evidence scheme;
 - when and where available human data can be used to support risk assessment and risk management decisions;
 - whether one single framework can accommodate all principle applications or whether an integrated approach is required.
2. Test the proposed framework/approach by a range of examples representing different risk assessment and risk management applications.
3. Provide advice to enable human data collection systems to be used more effectively in the risk assessment and risk management of chemicals.

These Terms of Reference were adopted by the Scientific Committee on 17-18 October 2006.

APPENDIX B: EXAMPLES REPRESENTING DIFFERENT RISK ASSESSMENT AND RISK MANAGEMENT APPLICATIONS

This appendix illustrates, by means of selected examples, the means by which the proposed ECETOC framework (Chapter 8, Figure 5) can be implemented in risk assessment and risk management processes (Chapter 1, Figure 1). It should be noted that the examples have been chosen to highlight the utility of the framework and are not intended as a synopsis of the issues in question.

B.1 Stepwise approach

A stepwise approach has been devised to facilitate comparison by the reader. It is outlined here and explained in more detail under sections B.1.1 to B.1.9.

- B.1.1 Formulation of the problem
- B.1.2 Summary of key human and animal data
- B.1.3 Quality aspects of the data
- B.1.4 Relevance to humans and data integration
- B.1.5 Identification of critical effect (hazard)

Sections B.1.6 to B.1.9 are optional and depend on the data available.

- B.1.6 Dose-response and safe levels
- B.1.7 Determination of NO(A)EL
- B.1.8 Application of assessment factors
- B.1.9 Current situation

B.1.1 Formulation of the problem

Every risk assessment or risk management process should begin with the formulation of the question that needs to be addressed. The appropriate question will depend on the stage of the risk analysis process, i.e. whether this involves, for example, hazard identification, exposure assessment, risk characterisation, or risk management, and if there is a need for establishing a dose-response relationship and a NO(A)EL (or DNEL). The most important question to be resolved for risk assessment is whether there is any hazard relevant to humans and, if possible, what is the leading (critical) health effect in humans?

Generally, questions are formulated along the following lines:

- Is a particular agent hazardous to human health? (Hazard identification).
- Are the eventual health effects that are identified in humans and/or animals caused by the agent under consideration? (Causal relationship).
- If several effects can be identified, what is the lead or critical health effect to animals and/or humans? (Identification of the critical health effect).
- Is the critical health effect identified in animals in the end also relevant to humans? Alternatively, is the critical health effect in humans, if not identified already, different from the one identified in animals? (Relevance to humans).

If this last question can be answered on the basis of the current (human and/or animal) data, then the following questions arise:

- Does this hazard constitute a real risk to humans? This can only be judged on the basis of exposure data, if available (Risk characterisation).
- Can a dose-response relationship be identified on the basis of the available exposure data? (Identification of dose-response relationship.)
- If a dose-response relationship can be identified, is it possible to identify a NOAEL?
- If this is possible, a safe level of no effect can be derived by applying assessment factors in such a way that an acceptable exposure level can be identified for humans e.g. in the form of a derived no effect level (DNEL)?
- If a DNEL is identified, can compliance with this DNEL be obtained? If compliance is not met, can the cause of this non-compliance be identified and which measures are to be taken? (Risk management.)

B.1.2 Summary of key human and animal data

Data to be gathered include human, animal, *in vitro* and *in vivo* data on the substance *per se*, and also on structurally-related substances. Once such data are collected, a summary is made of human as well as animal data. Even if the critical study is derived from animal data there will still be a place for human data or *vice versa*. It is important to include *in vitro* studies as well as studies that try to establish dose-response relationship, or attempt to give an idea of current exposure.

Human data may be in the form of case reports or case series, data collected from routine health surveillance in more or less well-defined populations (such as occupational health surveillance data) and data from non-experimental epidemiology (mostly case-controls and cohort) studies trying to assess some association between one or more potential health hazards and health endpoints.

It is important to note that the human data should document exposure to the substance in question, and that health effects should be measured adequately and other confounding factors accounted for, if they are to be included in the summary. Furthermore, the nature of the effect (local, chronic, systemic) will determine whether or not the available human data are strong enough or need to be supplemented by other data.

B.1.3 Quality aspects of the data

Once the data package is gathered and summarised, a quality review is necessary. Quality aspects of human data are described in Chapter 3 and refer to study design, quality of exposure data and quality of effects data. Many human data fall short of these quality standards but may nevertheless be of crucial importance. Five categories of human data have been identified in Chapter 4, encompassing different degrees of quality from A ('High' and including several well-conducted studies) to D ('poor'). The latter is the weakest category for suggesting a causal relationship according to the Bradford Hill guidelines (and just above the 'No information' category X).

Sometimes, human data indicate several possible health effects. In such cases, it is useful to group the data according to each potential health effect and then to combine these with the quality aspects, in order to assess which health effects are likely to be associated with the substance under investigation.

'Good' human data are ideal for risk assessment, but often are not available or not of sufficient quality, and reliance has to be placed on animal data. An assessment of the intrinsic quality of the animal data can be made by assigning them to one of 5 categories on the basis of the Klimisch criteria of reliability (Chapter 5). However, the essential question remains of whether the available animal data can be extrapolated to man. This part of the process is frequently confronted with a lot of uncertainty.

Additional quality aspects are not related to a study as such, but are concerned with the relevance and utility of the study for addressing the question formulated under B.1.1. Thus the quality aspects of animal studies may be context dependent, e.g. a high quality animal study is of limited value if the exposure route is not appropriate.

B.1.4 Relevance to humans and data integration

Most animal studies contribute to hazard identification. Classically, the most sensitive adverse health endpoint in the most sensitive animal species is considered for the purpose of risk assessment and one or more critical endpoints are identified. It may be possible to formulate a

MoA hypothesis, particularly where there is concordance between results of comparable animal studies or when different chemicals show the same pattern of toxicity. It is essential to identify the toxicant and the target organ as well as the dose-response curve of the considered effect and its temporal relationship. If the different key events leading to toxicity and a MoA hypothesis can be identified, it is sometimes possible to evaluate the plausibility of these events to man. The route of exposure in animal studies is frequently different from the relevant one in humans; this needs careful consideration, and sometimes requires the application of an assessment factor. Moreover, once a critical animal NOAEL is identified and considered relevant to humans, various assessment factors need to be applied in order to derive a reference dose protective of human health.

An integrated approach is developed, by bringing together animal and human data in an overall ‘weight of evidence’ framework, for each case study addressing a human risk question for a chemical substance.

After careful consideration of the quality and applicability/relevance of all human and animal studies (*in vivo* and *in vitro*), the collective weight of evidence of the human data with respect to the human risk question being considered is scored on a 5-point scale (A - D plus X) featuring on the X-axis of an integrative matrix (Chapter 7, Figure 4). A similar assessment is made of the collective weight of evidence of the animal data with respect to the human risk question being considered. This results in the animal data being scored in five categories (A - D plus X) along the Y-axis of the matrix. If the weight of evidence score for the human data is equal or greater than the weight of evidence score for the animal data then the human data takes precedence as the primary tool to address the human health risk. Otherwise the animal data takes precedence.

If both human and animal data are available it is particularly important to consider the issue of concordance. Four situations may occur:

- Concordance of effect: Both animal and human data give similar indications of effects at similar concentrations;
- concordance of no effect: Neither animal nor human data show effects at the concentrations studied;
- discordance of effect in animals with effects in humans;
- discordance of effect in humans with effects in animals.

If animal and human data are concordant it might be assumed that the whole dataset is consistent with the hypothesis that the same processes are going on in both animals and humans and that, in the event of appropriate exposure, similar toxicities might result. The question of whether or not to use animal or human data then becomes less important. Sensitivity differences between

animals and humans may exist; animals may be more or less sensitive than humans or human exposure may never be sufficient for the toxicological effect to become manifest.

In the case of discordance with effect in humans, human data should take precedence in risk assessment. Data from human studies might also be classified in this category, where effects occur at lower concentrations than in animals.

In the case of discordance with effects in animals, some well-defined human data might be discarded.

In every situation of non-concordance, the reasons for the lack of concordance should be examined. If the reason is found to be on the basis of the underlying biology, then confidence in the risk assessment will increase. If the reason cannot be understood or explained, risk assessment may be less secure, especially if the weight of evidence of the data is weak.

In general, positive data take precedence over null data.

B.1.5 Identification of critical effect (hazard)

Once the data are 'weighted' according to the proposed framework, it is usually possible to identify the body of evidence that takes precedence. When the data (be it human or animal) are grouped according to the potential health effect, information on exposure should be used to determine which health effect can be expected to occur at the lowest concentration. This is usually the critical health effect. If the critical effect is prevented from occurring, it can be expected that other potential adverse effects will not occur, since they have only been reported at higher concentrations. However, if animal data indicate more subtle mechanistic changes, these data might determine the critical effect. The identification of the critical effect will require reliable information on the dose-response relationships of all involved health effects.

B.1.6 Dose-response and safe levels

Once the critical effect has been identified, a detailed description of the dose-response relationship will serve as a basis for the estimation of a NOAEL. Dose-response curves for human data are typically given for exposure that occurred in practice in groups of individuals. Combinations of job activities and workplace exposure are frequently used to classify individuals according to their exposure profile, ultimately aiming at identifying exposure groups containing individuals with a similar exposure. A large number of similar exposure groups will increase the specificity of the dose-response curve. However, smaller sample size may increase the inaccuracy of the estimates of the occurrence of the effect.

B.1.7 Determination of NO(A)EL

Most human data are generated by non-experimental studies. Hence, the exposure categories are usually based on the exposure conditions under investigation. In many cases the width of the exposure groups is so large that it can only be concluded that the NO(A)EL lies within one of these exposure categories. The exposure category in which the NO(A)EL lies is the lowest exposure category in which effects are observed. The steepness of the dose-response curve can provide information on where, within the lowest category with an observed effect, the NO(A)EL is expected to lie.

B.1.8 Application of assessment factors

If the key human data are of poor quality, giving rise to uncertainty regarding the NOAEL, the use of several assessment factors may be required. Also, if an exposure limit is to be derived for a more sensitive population than the population from which the data were derived, an additional assessment factor will be necessary. An assessment factor is also needed for situations in which there remains a residual uncertainty about the severity of the effect, and when a NOAEL is determined from null data. Other commonly-used assessment factors, apart from those already identified, are related to:

- Extrapolation from animal to human;
- extrapolation for intraspecies and interspecies differences;
- extrapolation from LOAEL to NOAEL (in the absence of LOAEL);
- extrapolation from shorter-term to longer-term studies (in the absence of a study of desired duration);
- extrapolation from one exposure route to another;
- assessment factors for an incomplete data set or for a steep dose-response curve.

These assessment factors are detailed in Appendix D.2.

B.1.9 Current situation

Finally, the outcome of the risk assessment process can be put into perspective and compared with the current evaluation by others (e.g. current or former decisions of regulatory bodies, literature on former evaluation exercises, new case reports, and newly emerging data) to assess the possible impact (or not) of this risk assessment exercise on the current state of the art.

B.2 Welding fumes

B.2.1 Formulation of the problem

Fumes are emitted during welding processes. Such fumes are difficult to characterise since the composition varies by welding equipment and type of metal to be welded. They include oxidised and pure metal fumes, shielding gas, haemolytic products and ozone; UV light is formed and emitted. There is substantial evidence that welding fumes, produced indoors without proper ventilation and personal protection can cause pulmonary effects.

Do the available data enable the identification of a NOAEL and is it possible to derive a DNEL?

B.2.2 Summary of key human and animal data

Human data

Epidemiology studies have found pulmonary effects including respiratory symptoms, pulmonary function decline, fibrotic effects and possibly increased lung cancer rates in welders. Differences in respiratory function between pre- and post-shift examinations have been observed on several occasions (Antonini *et al*, 2003).

Elevated prevalences of DNA-protein cross-links have been observed in lymphocytes of welders (Popp *et al*, 1991).

Luo *et al* (2006) found a dose-response relationship between the daily average number of welding hours and respiratory symptoms in long-term welders.

Exposure data in the above studies, reporting health effects in welders, are limited.

Pulmonary effects are seen in workers currently exposed to around 5 mg/m³ as total dust welding fume. For instance, zinc-coated mild steel welders showed reduced pulmonary function following exposure to 0.91 mg/m³ (geometric mean) total dust with highest levels between 5 and 8 mg/m³ (Marquart *et al*, 1989).

Animal data

Short-term rat studies showed no significant histopathological changes (Uemitsu *et al*, 1984).

No appropriate chronic exposure studies exist, because of the difficulty in maintaining a steady exposure situation over long periods of time.

In vitro studies indicate that mild steel welding fumes are weakly mutagenic (Stern *et al*, 1988).

B.2.3 Quality aspects of the data

The available human data are of good quality, although the exposure concentrations in the studies are sparse or have been poorly evaluated.

Only short-term animal data are available, but these are of good quality.

B.2.4 Relevance to humans and data integration

The human data from several studies clearly show an effect on pulmonary function (Category A). The results from short-term studies in rats, indicating pulmonary effects, are in concordance with these findings (Category B) (Figure B.1). An inflammatory process is thought to be the most likely mechanism involved in the respiratory effects observed.

B.2.5 Identification of the critical effect (hazard)

The critical effect is a decrease in pulmonary function.

B.2.6 Dose-response and safe levels

It can be concluded that exposure concentrations of 5 mg/m³ as total dust are in the effect range.

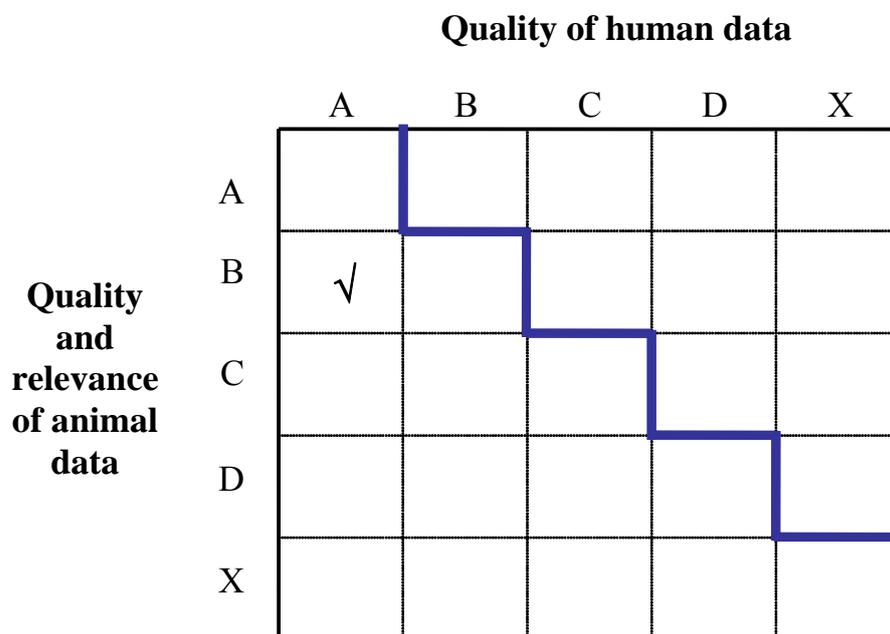
B.2.7 Determination of NOAEL

The human data are insufficient to derive a NOAEL.

B.2.8 Application of assessment factors

Since no NOAEL can be derived, no assessment factors are applicable.

Figure B.1: Integration of human and animal data for welding fumes



B.3 Wood dust

B.3.1 Formulation of the problem

Dust is emitted during many processes in which wood is machined. Occupational exposure to wood dust is ubiquitous; the type of wood dust and the exposure intensity and particle size will vary from process to process. Wood dust may cause effects on the upper respiratory tract. Animal studies are in line with results from human studies, indicating that the respiratory tract is the target organ.

As there is some debate about the carcinogenic potential of exposure to wood dust, it is considered important to derive a DNEL for wood dust exposure in the workplace to prevent any type of respiratory effect.

B.3.2 Summary of key human and animal data

Human data

There is extensive literature on the respiratory effects of exposure to wood dust, but many studies lack specific information on exposure, or are of a relatively small sample size. However, several studies present adequate results which are useful for DNEL derivation.

A cross-sectional study in Chinese mill workers exposed to wood dust at levels between 4.4 and 22.4 mg/m³ total dust (2.4% to 50.2% being of a inhalable size) showed a clear increase in respiratory symptoms and a significantly reduced pulmonary function, in smokers as well as non-smokers. Pulmonary function decrease showed a significant association with exposure levels (Liou *et al*, 1996).

A cross-sectional study of Australian sawmill workers exposed to approximately 1.6 mg/m³ of inhalable dust, also reported work related symptoms and decreased pulmonary function (Mandryk *et al*, 1999).

A large cross-sectional study of 2,423 Danish wood workers with exposures averaging 1.2 mg/m³ of inhalable dust showed minimal effects on pulmonary function with few respiratory symptoms (Schlunssen *et al*, 2002).

The human data presented above show clear evidence of respiratory effects from inhalable wood dust in the exposure range of 1.6 mg/m³ and above.

Animal data

Animal studies have focused mainly on the underlying mechanism of the respiratory effects of wood dust exposure. For example, Määttä *et al* (2006) showed that particle induced pulmonary inflammation was associated with an increase in cytokines and chemokine receptors in the lung.

B.3.3 Quality aspects of the data

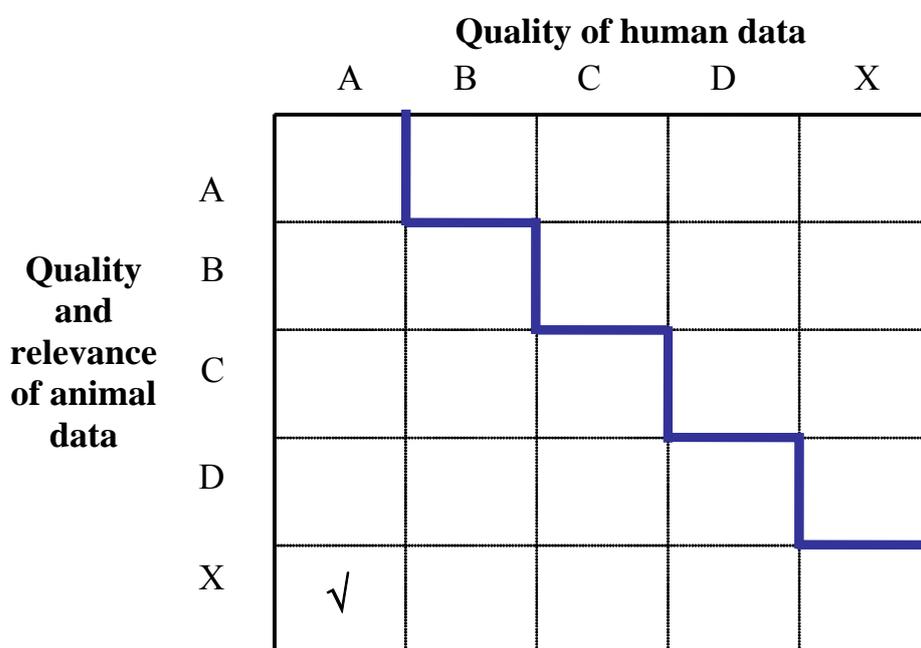
The human data are of relatively good quality. Other confounding factors such as smoking were taken into account. The studies are adequate to detect the respiratory effects noted in earlier reports and show consistent findings.

The animal studies are also of satisfactory quality but are based on small numbers of animals, and focus on early effect parameters.

B.3.4 Relevance to humans and data integration

The animal data on wood dust are of limited value for DNEL derivation (Category X). The human data are relevant and form a consistent pattern, indicating that pulmonary effects can be anticipated from wood dust exposures in the 2 mg/m³ range and over. In some studies the effect of confounding factors such as smoking was taken into account (Category A) (Figure B.2).

Figure B.2: Integration of human and animal data for wood dust



B.3.5 Identification of the critical effect

The critical effect as observed in the epidemiology studies on wood dust is pulmonary function decline.

B.3.6 Dose-response and safe levels

The epidemiology studies lack detail for describing a full dose-response curve.

B.3.7 Determination of the NOAEL

The Danish study with exposures to wood dust in the range of 1.2 mg/m³ showed minimal effects and this exposure concentration could be regarded as a NOAEL.

B.3.8 Application of assessment factors

An assessment factor of 2 seems appropriate to extrapolate from the minimal effect level to a NOAEL, resulting in a value of 0.6 mg/m³ inhalable wood dust. It can be argued that an additional assessment factor is required because of the cross-sectional nature of the data. However, a selection bias (caused by affected workers leaving the industry before the studies were conducted) is not likely because of the relatively mild symptoms. In addition, it cannot be ruled out that exposures in these workers may have been higher in the past and that the dose-response curve based on current exposure levels is an over-estimation of the actual dose-response curve.

B.4 Sulphanilic acid

B.4.1 Formulation of the problem

The major uses of sulphanilic acid are as a fine chemical and as a corrosion inhibitor in some forms of Portland cement. No human cases of contact allergy have been identified amongst the workforce, despite routine dermal exposure that can occur during the manufacture of sulphanilic acid. Furthermore, no cases have been reported of sulphanilic acid causing (or being associated with) allergic contact dermatitis in cement workers or users. (It is noted that skin allergies caused by cement are frequently attributed to chromium. This could serve as a confounding factor and potentially lead to under-reporting of any sulphanilic acid-related cases).

In contrast, the EC has classified sulphanilic acid as a dermal sensitiser (R43) since it is a dermal sensitiser in the guinea pig maximisation test (GPMT).

B.4.2 Summary of key human and animal data

Human data

The main manufacturing site in Europe has produced sulphanilic acid for over 50 years. The annual production volume at the facility is 10,000 tonnes. The site's physician reported that no cases of dermal sensitisation have been recorded at the plant (or detected during the annual

medical examinations of over 250 employees, routinely employed at the plant over this period). However, no specific skin surveillance programme was in place as the substance had not previously been considered to present a significant dermal risk. The physician reported that hygiene practices at the site had not been high and that exposures are likely to have exceeded 10 mg/m³ in the past.

Animal data

Sulphanilic acid is a strong dermal sensitiser in the GPMT (Basketter *et al*, 1992).

B.4.3 Quality aspects of the data

The GPMT on sulphanilic acid represents a well-conducted animal study and would be treated as 'reliable with no restrictions' (Klimisch category 1).

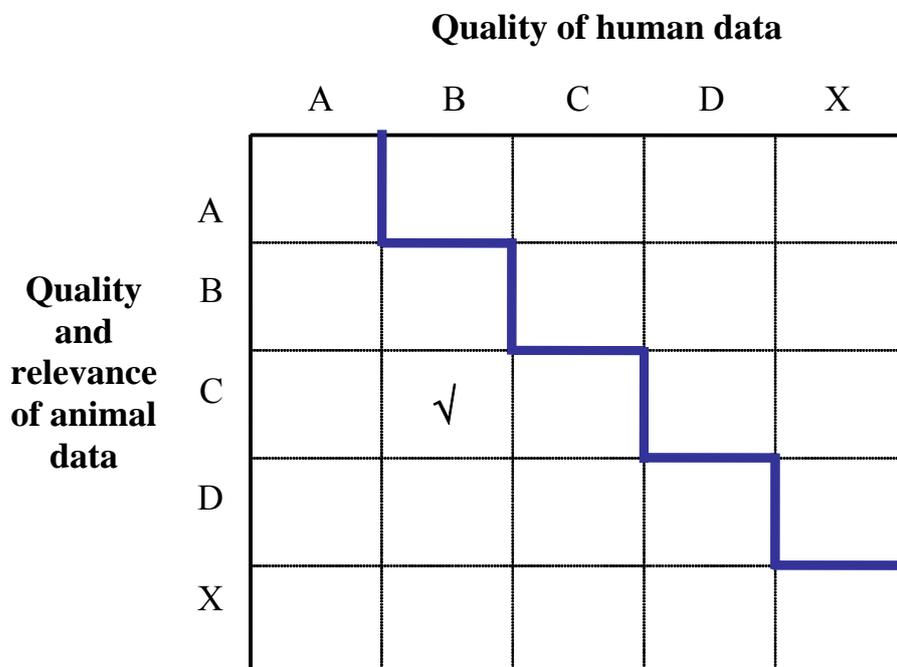
The fact that exposed workers at the manufacturing site had been routinely and systematically examined over an extended period is considered equivalent to a small epidemiology study. Although this has shortcomings (e.g. comparatively small size, lack of knowledge of exposure and lack of patient blinding), it should be reliable in its ability to determine the presence or absence of an obvious effect in the exposed population. For this reason, the collective human data would be classified as category B.

B.4.4 Relevance to humans and data integration

Both the animal and human data on sulphanilic acid must be considered as relevant. The issue is when data characterising the null human experience should take precedence over positive animal findings.

The animal data would be rated category C (presumed relevant to man in absence of mechanistic evidence). The sole absence of reported human cases (i.e. published case reports) would be rated as category D under the proposed scheme. Taking both the animal (Category C) and human (Category B) data together would indicate that the human data prevail. Thus, the substance should not be classified as a skin sensitiser (Figure B.3).

Figure B.3: Integration of human and animal data for sulphanic acid



B.4.5 Identification of critical effect (hazard)

No cases of contact dermatitis linked to sulphanic acid have been reported in the literature, despite widespread occupational exposures in the chemical and building industries. However, in contrast to the conclusion arrived at by following the framework in this report, in the regulation of sulphanic acid in Europe the available null human findings were not considered to be of a sufficient quality to over-ride the positive effects observed in an animal study. In other words, the available human evidence was considered insufficiently robust by the EC Classification and Labelling working group to allow any conclusion to be made regarding the potential of sulphanic acid to cause (or not to cause) dermal sensitisation in humans. The EC group concluded that the strongly positive guinea pig findings were sufficient to classify the substance as a skin sensitiser (R43).

B.4.6 Dose-response and safe levels

Not applicable in this example.

B.4.7 Current situation

The allergic contact potential of sulphanilic acid has subsequently been investigated in the guinea pig cumulative contact enhancement assay and in the murine local lymph node assay. In neither case has it been found to result in sensitisation (Basketter *et al*, 1992). These results reinforce the null human experience, suggesting that sulphanilic acid should not be classified as a contact sensitiser. [The possible reasons why the GPMT can yield ‘false positive’ findings have subsequently been discussed in more detail by Basketter *et al* (1998)]. Despite these developments and the continuing absence of human cases of contact dermatitis induced by sulphanilic acid, the substance remains officially classified as a contact sensitiser.

B.5 Reactive dye

B.5.1 Formulation of the problem

Reactive dyes are a group of substances, several of which are known respiratory sensitisers (Docker *et al*, 1987). No reliable animal model is available to test for respiratory sensitisation. Thus, identification (and confirmation) of cases can only reliably emanate from human experience.

As users of chemicals receive much of their hazard information through labels and safety data sheets, the manner in which the classification of a substance is communicated is important. In the system by which the EC classified chemicals before REACH, it could take several years to formally classify a substance (under the Dangerous Substances Directive 67/548/EEC, following submission of a dossier by a competent authority and its review by the EC Technical Committee on Classification and Labelling and, finally, its adoption under a Council Technical Progress Directive). Under both sets of circumstances (i.e. both before and after REACH), it is important for suppliers of these materials to respond to early indications of their possible allergenicity and to act responsively and judiciously.

The problem is the extent to which adverse human experience needs to accumulate before cautionary action (in this case, voluntary classification and labelling of the product) is taken.

B.5.2 Summary of key human and animal data

Human data

Two cases of respiratory allergy caused by a reactive dye were reported by a UK regional occupational health clinic to the dye manufacturer. The clinic suggested that the dye be

(voluntarily) labelled by the dye manufacturer as a respiratory sensitiser (R42, may cause sensitisation by inhalation). Both cases of sensitisation occurred in artists who were using a mouth-operated nebuliser to spray fabrics with an aqueous solution of the dye. There is no exposure information for such a method of application, an application that is neither endorsed nor foreseen by the manufacturer, the dye being intended to be used within industrial facilities. Indeed, within the overall context of the general cautionary advice contained on the product safety data sheet, such a method of application would constitute a ‘misuse’ of the substance. The dye had been marketed for 2 years without any other reports of allergy or other adverse effects.

The company selling the product operated an adverse health effects register, based on reports received from its customers (as obligated by the US Toxic Substances Control Act). The company put forward the argument that the dye would not result in allergy under normal conditions of use and hence should not be classified as a respiratory sensitiser. This was based on the fact that no other cases of respiratory allergy to the particular dye had been reported in either the literature or its own register, and that the reported cases most likely occurred due to exposures associated with clear conditions of misuse.

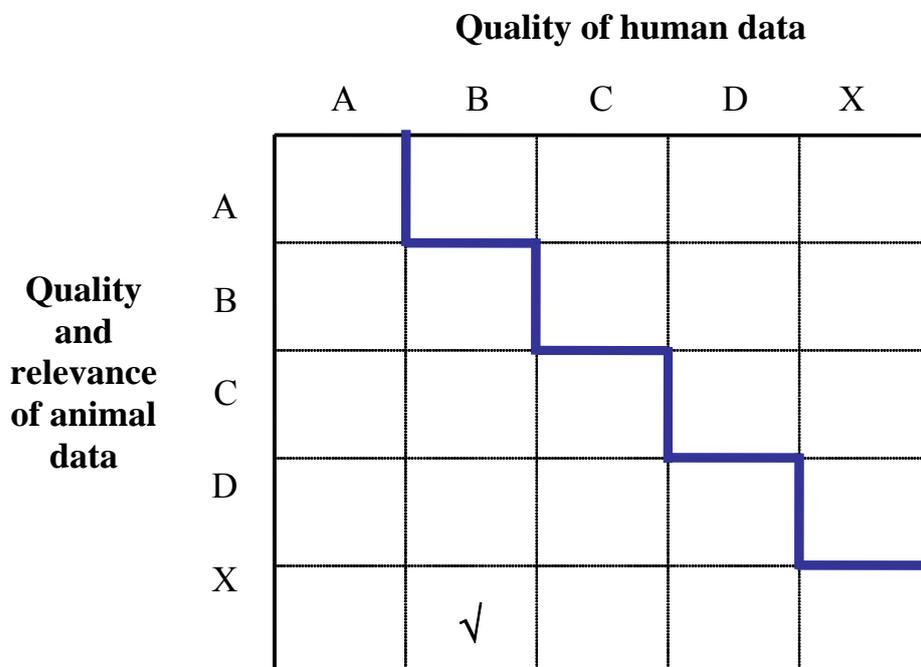
Animal data

No data are available.

B.5.3 Quality aspects of the data

The only data available on the reactive dye emanate from human experience. Firstly, there are two (diagnostically) confirmed cases, where the magnitude of the exposure that resulted in the induction of the sensitisation is unknown (but suspected of being substantial). Secondly, consideration should be given of the negative experience from an adverse effects register that relies on complaints from users.

Under the proposed ECETOC scheme, the animal data would be classified as category X, the positive human experiences as category B and the null human experiences as category D. The positive findings would take precedence in the evaluation of the hazard/classification of the substance and the product would therefore be considered a respiratory sensitiser (Figure B.4).

Figure B.4: Integration of human and animal data for a reactive dye

B.5.4 Relevance to humans and data integration

Although the data on the reactive dye are of variable quality, all are relevant as they derive from human experience. However, the extent to which the human data can be interpreted reliably is constrained by the limited amount of supportive exposure information available.

The question was the extent to which adverse human experiences needs to accumulate before cautionary action (in this case, voluntarily classification and labelling of the product) is taken. For example, would a single case be evidence of a concern or symptomatic of exposure to a particularly sensitive individual? The company response may have been affected by the possible loss of market that such a 'severe' classification could be expected to result in (when contrasted against competing products which are not classified in such a manner). But, when combined with the similar experience from other reactive dyes (Docker *et al*, 1987; Nilsson *et al*, 1993) and the comparatively short period in which the dye had been marketed, it would be reasonable to assume that despite the low 'power' of two cases, there would be sufficient cause for concern to justify (voluntarily) classifying the substance.

The company adverse effects register suggested that at more normal levels of exposure to the reactive dye, sensitisation is unlikely. But adverse effects registers that are solely reliant on

‘logging’ complaints from users are inherently unreliable. Moreover, as the dye had only been marketed for a limited period of time, such registers might only have been expected to have detected any problem if the product was a very potent sensitiser. As such, the confirmed clinic cases constitute a sound basis for concluding that the substance results in respiratory sensitisation despite their being associated with a likely misuse of the product. This argumentation was accepted by the company and the dye labelled accordingly.

B.5.5 Dose-response and safe levels

Such considerations are not applicable in the case of determining whether a substance should be classified as a respiratory sensitiser. Although theoretically an induction threshold exists for such materials, the exposure required to initiate a reaction once an individual has become sensitised is extremely low (Park *et al*, 2007). Indeed, sensitisation of this kind is not only likely to be physically debilitating, but also to affect adversely a person’s job prospects and lifestyle. Under such circumstances, the fact that confirmed cases have occurred is sufficient to justify a classification.

B.5.6 Current situation

The reactive dye remains voluntarily labelled as a respiratory sensitiser. In the 7 years following classification and labelling, 2 further cases have been confirmed among users (both from the dye industry). These findings indicate that the dye, while not a potent allergen, can cause respiratory sensitisation amongst some individuals under normal working conditions, confirming the original recommendations of the health clinic.

B.6 Benzanthrone

B.6.1 Formulation of the problem

Workplace exposure to benzanthrone results in phototoxic dermatitis with photo-allergic reactions having been reported in some instances. Among some worker groups, the consequence of exposure is skin pigmentation which is most noticeable in those with light skins. The effect is also seen in animals, although the reliability of the animal model for man is unknown. Benzanthrone has also been observed to cause skin irritation in mice (BUA, 2004).

Dermal exposure to benzanthrone needs to be well controlled because of the nature of the effects. This will encompass the control of airborne exposures resulting from both direct and indirect dermal contact (via settled material) and the prevention of contamination of workplace surfaces.

The establishment of suitable 'health values' is therefore a valuable tool in managing such dermal risks. However, such values are only useful if they can be shown to be protective i.e. by identifying NOAELs for the effects of interest.

B.6.2 Summary of key human and animal data

Human data

A number of human studies are available describing the effects of benzanthrone in man (Singh and Zaidi, 1969; Goldman, 1972). Most studies focus on the dermal effects of benzanthrone, arising as a consequence of its manufacture or use.

Animal data

Animal studies have been published addressing the irritancy, phototoxicity and sensitising potential of benzanthrone (BUA, 2004). These range from standard regulatory tests to protocols designed to characterise specifically the basis of the observed phototoxicity. The irritation potential differs according to species, whereas consistent results have been observed for phototoxicity. A GPMT observed a dose dependent hyper-pigmentation and established a NOAEL for contact allergy.

B.6.3 Quality aspects of the data

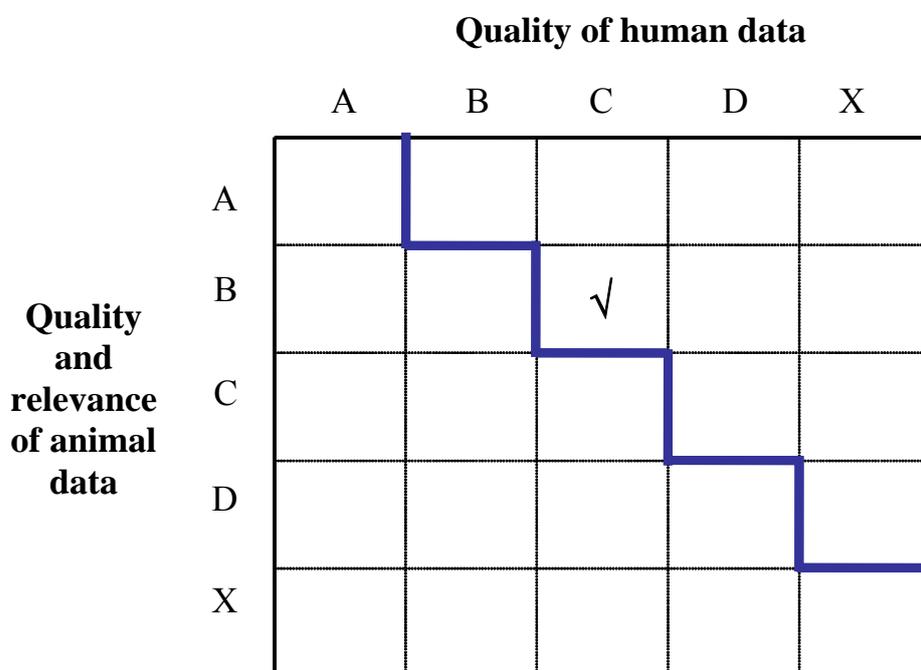
Although many of the animal studies on benzanthrone were undertaken prior to the development of current GLP practices (particularly concerning characterisation of test material), most of them are of an acceptable standard and would be classified as either Klimisch category 1 or 2. These studies would be rated category B (animal MoA plausible to man with sensitivity correction).

The reported human cases (i.e. published observational studies and case reports), whilst being sufficient to describe the incidence of skin conditions, are characterised by an absence of relevant exposure information (at best, some studies describe airborne concentrations) and hence, bearing in mind the importance of characterisation of the dose-response curve, these would only be rated as category C (poor).

B.6.4 Relevance to humans and data integration

Consideration of the animal (Category B) and human (Category C) findings on benzanthrone indicates that the animal findings should prevail. From this it can be concluded that the NOAEL observed in the GPMT (Srivastava *et al*, 1990) could be used as a key input for the development of a dermal hygiene standard (assuming additional assessment factors are applied to account for residual uncertainties e.g. the apparent differences in species sensitivity) (Figure B.5).

Figure B.5: Integration of human and animal data for benzanthrone



B.6.5 Hazard (critical effect) identification

Both the animal and human studies on benzanthrone are potentially relevant to the problem as identified. However, in the absence of any information describing the nature of dermal contact/exposure, the human studies do not adequately address the nature of relevant exposures associated with the observed effects. Thus while they are useful for hazard identification, they are of limited use in risk assessment.

B.6.6 Dose-response and safe levels

Not applicable.

B.7 Mectins

B.7.1 Formulation of the problem

The three mectins considered here are ivermectin, abamectin and doramectin. They are structurally similar, and have similar properties and toxicity. Being ion-channel disruptors, they are variously used as veterinary de-wormers and acaricides (doramectin and ivermectin), agricultural insecticides (abamectin), and pharmaceutically to control onchocerciasis ('river blindness' caused by a parasitic worm) in man (ivermectin).

The problem considered here is whether to use the animal or human data as the basis for regulating acute neurotoxicity and developmental toxicity in humans.

B.7.2 Summary of key human and animal data

A comprehensive review by the US Food and Drug Administration (FDA) of human and animal data for ivermectin is available (FDA, 1996). A comparative overview of the toxicology of mectins has been published by WHO (2002).

Human data

There are a few well-controlled clinical trials available which support the use of ivermectin for onchocerciasis control; the treatment is 0.15 - 0.2 mg/kgbw given as a single dose. One trial showed low transfer to infants via breast milk.

Most experience is from widespread onchocerciasis control programmes (e.g. > 23 million people treated in 2000). There are no reports of neurotoxicity caused by ivermectin in this context.

Accidental poisoning incidents (children eating pet-worming pills) show neurotoxicity at doses comparable to that in non-human primates.

Ivermectin has been found to be safe when accidentally used during pregnancy, and may now be used during breastfeeding. It should not be directly dosed to children (< 1 year old).

No human data are available in relation to chronic exposure.

Animal data

Comparable neurotoxic symptoms occurred for all three mectins, in all species, after acute or chronic dosing. Mectins activate a chloride ion current by a γ -aminobutyric acid-like opening of chloride ion channels: Secondly, they also open glutamate-gated chloride channels.

In the CF-1 mouse, an acute NOEL of 0.1 mg/kgbw was established. Large species differences in sensitivity have been reported in that rodents are most sensitive and non-human primates least sensitive. Some CF-1 mice, collie dogs and Murray Grey cattle lack *p*-glycoprotein (as a result of a genetic abnormality which normally shunts ivermectin out of the brain, and the gut lumen) and are therefore particularly sensitive to ivermectin neurotoxicity. (There are no reports of humans with an absence of *p*-glycoprotein.)

Ivermectin is a developmental toxin, with increased incidences of cleft palate in foetuses at maternally toxic doses. The CF-1 mouse is particularly sensitive owing to a lack of *p*-glycoprotein protection in the placenta. Neonatal rats are also especially sensitive to the toxic effects of ivermectin pre-weaning, due to a combination of a lack of *p*-glycoprotein expression, poorly developed blood-brain barrier, and high transfer rate from the mother via milk.

B.7.3 Quality aspects of the data

The acute neurotoxicity seen with mectins is an acute non-specific effect. Human studies of 'High' and 'Good' quality are available, so the human data are category A. Developmental toxicity is a chronic non-specific endpoint, and quality of the human data for this endpoint is 'Good'. Therefore the human data for developmental toxicity are Category B.

The animal data are of good quality. The neurotoxic MoA is established and is relevant to man, but with an expected sensitivity difference, especially between man and the CF-1 mice which lack *p*-glycoprotein. For the acute neurotoxicity endpoint, the animal data are rated category A if the CF-1 mouse data are considered as being directly relevant, or category B if they are not.

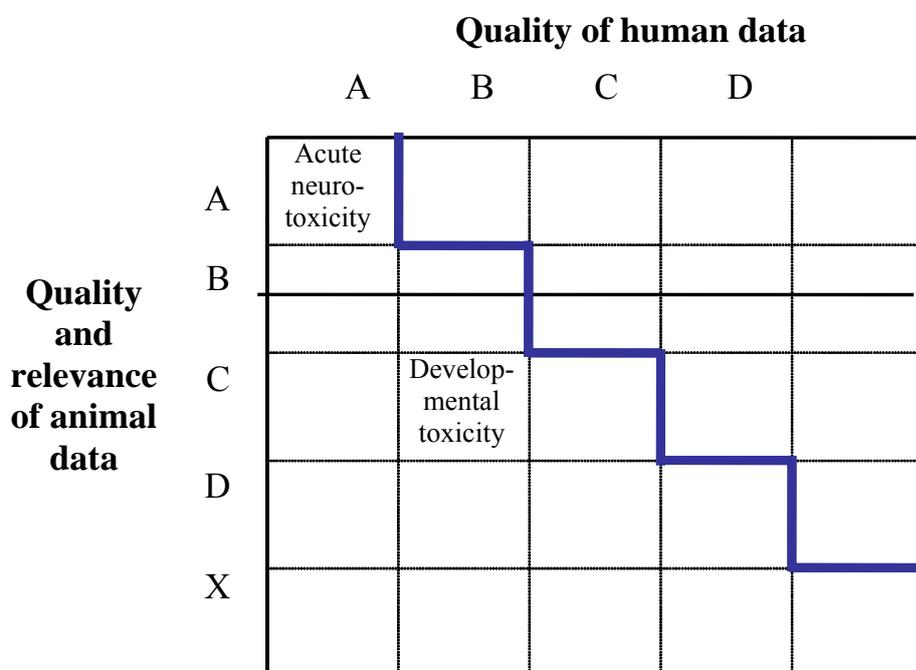
In the case of developmental toxicity, the animal data are rated category C. (An alternative view would be that the absence of a well-formed blood-brain barrier in neonatal rats, and absence of *p*-glycoprotein in the CF-1 mouse placenta are pre-requisites for developmental toxicity in animals. In this case, the animal data would be category X.)

B.7.4 Relevance to humans and data integration

Category A human data take precedence over the category A (or perhaps B) animal data for the acute neurotoxicity of mectins. The human and animal data are in concordance, but given the quality of the human data, for which no species extrapolation is necessary, they take precedence for risk assessment.

For developmental toxicity, the animal data are category C and the human data are category B, and so the human data take precedence (Figure B.6).

Figure B.6: Integration of human and animal data for mectins



B.7.5 Identification of critical effect (hazard)

The hazards of mectins being considered are acute neurotoxicity and developmental toxicity.

B.7.6 Dose-response and safe levels

For the acute neurotoxicity of mectins, there are human data to establish the safety of single doses of 0.2 mg/kgbw.

For developmental toxicity, a case could be made based on the human data, for the safety of single doses of 0.2 mg/kgbw. However, the human data are not supported by any statement about safety from chronic exposures. The judgement of the animal data for this endpoint centres on the complex issue of the relevance to man of data from neonatal rats, and from CF-1 mice lacking *p*-glycoprotein.

B.7.7 Current situation

The human safety assessment of mectins varies according to use (agricultural insecticide, veterinary and human pharmaceutical). This is partly due to the different risk/benefit balances in each case, and the different human exposures involved. Use of ivermectin as a human pharmaceutical in pregnancy and lactation was initially excluded based on the animal data, but later allowed, based on human experience and a favourable risk/benefit analysis. Regulation of the use of abamectin as a crop insecticide is based entirely on animal data, with authorities differing in their view of the relevance of the data in animals lacking *p*-glycoprotein, and the relevance of the toxicity seen in rat pups pre-weaning.

B.8 Glycidol

B.8.1 Formulation of the problem

Glycidol is primarily used as stabiliser in the manufacture of vinyl polymers. It is a known irritant of skin and mucous membranes. In view of the lack of reports of adverse effects on health in humans its carcinogenic potential should be assessed.

B.8.2 Summary of key human and animal data

Human data

No adverse effects on worker health were reported in 70 persons exposed annually to levels of 2 ppm glycidol (6.20 mg/m³) (Dixie Chemical, 1992; cited by ACGIH, 2001).

Animal data

There is clear evidence of increased neoplasia in rats and mice after oral exposure to glycidol for up to 2 years. In rats, there were gliomas and forestomach tumours in both sexes, mesotheliomas of the tunica vaginalis/peritoneum, intestine, skin, thyroid gland, and Zymbal gland tumours in males, clitoral gland, mammary gland, oral mucosa, and leukaemia in females. In mice, there

were Harderian gland tumours in both sexes, forestomach, lung, liver, skin in males, mammary gland and subcutaneous tissue in females (NTP, 1990; Irwin *et al*, 1996; both cited by IARC, 2000).

In addition, various genotoxicity studies have demonstrated that glycidol caused the induction of mutations and unscheduled DNA synthesis, as well as an increased number of sister chromatid exchanges and chromosomal aberrations (Ehrenberg and Hussain, 1981; cited by IARC).

B.8.3 Quality aspects of the data

The NTP 2-year gavage dosing carcinogenicity studies on glycidol in rats and mice can be considered as Klimisch category 1 (NTP, 1990).

B.8.4 Relevance to humans and data integration

In the absence of toxicity reports on glycidol in a relatively small group of workers, the data available in humans would be classified as category D with regard to carcinogenicity.

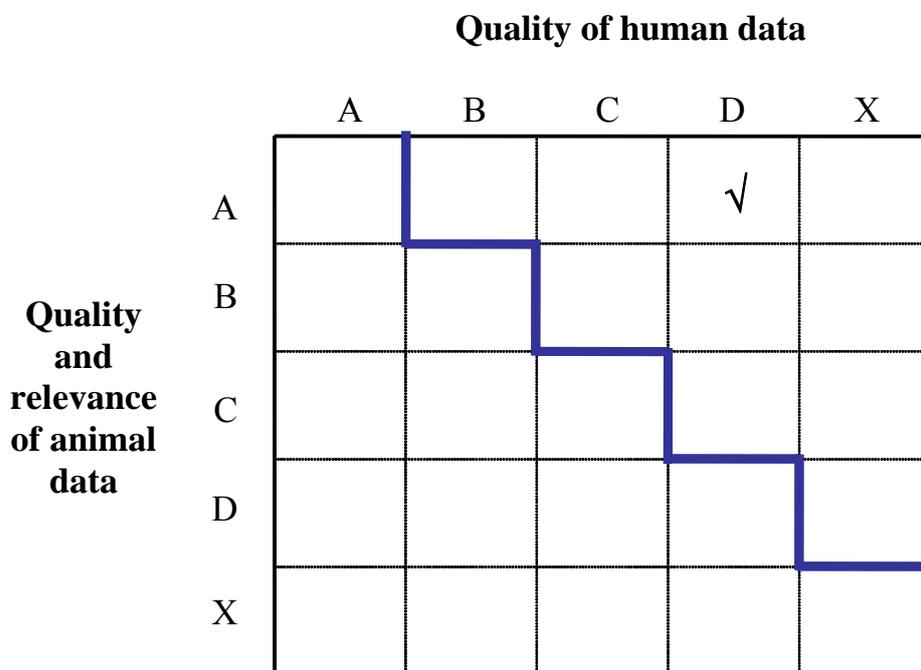
The animal studies would be rated category A (animal MoA plausible to man; there is no obvious reason to assume a different sensitivity of humans).

The human data are of negligible relevance with regard to the carcinogenicity of glycidol; the high quality animal data clearly take precedence (Figure B.7). [The data are not necessarily in conflict, since exposures at the primary site of glycidol manufacture in the USA seem to have been well controlled with a TLV-TWA of 2 ppm (Dixie Chemical, 1993; cited by ACGIH, 2001)].

B.8.5 Identification of critical effect (hazard)

While no human carcinogenicity data on glycidol are available, multiple tumours have been observed in rats and mice. The types of tumours observed do not point to a single target organ. Thus, (multiple) cancer(s) in general can be considered the critical effect(s).

Figure B.7: Integration of human and animal data for glycidol



B.8.6 Dose-response and safe levels

In the NTP studies described above, dose-related increases in the incidence of neoplasms in several tissues were recorded. Taking into account the positive genotoxicity studies, a linear dose-response is assumed.

B.8.7 Determination of NOAEL

No NOAEL can be defined.

B.8.8 Current situation

Based on animal studies, ACGIH (2001) has proposed a TLV-TWA of 2 ppm glycidol (6.2 mg/m³) which minimises, but does not exclude the potential risk of cancer and genotoxicity.

B.9 Methylene chloride (dichloromethane)

B.9.1 Formulation of the problem

Methylene chloride is used as an ingredient of paint stripping and aerosol preparations. It is used as a solvent in a wide variety of industrial processes, such as film manufacture. Excess lung and liver tumours were observed in mice, but not in rats, exposed to methylene chloride in a 2-year bioassay (NTP, 1986). There are distinct metabolic and pharmacokinetic differences in humans, rats and mice. Ultimately, the carcinogenic dose-response curve was adjusted for these pharmacokinetic differences, resulting in a 9-fold reduction in the unit risk estimate. Before this adjustment was made, an epidemiologic study had been conducted. No excess lung or liver cancers were observed, but bounds on the risk due to methylene chloride could be made using this human study. Such bounds are possible because an upper confidence limit can be used as a bound on excess risk for the precise exposure estimates that were available in this cohort study.

The problem is to provide a dose-response assessment for the carcinogenic effects of methylene chloride in air.

B.9.2 Summary of key human and animal data

Human data

A well-conducted study on methylene chloride in 1,013 Kodak workers found no excess of lung and/or liver tumours (Hearne *et al*, 1987). The study had extremely detailed exposure data from years of monitoring methylene chloride concentrations in the workplace. Lung cancer showed 14 observed deaths versus 21 expected, while liver cancer showed no deaths versus 0.8 expected. For both sites combined, the risk ratio was 0.64 with a 95% confidence interval of 0.35 to 1.08. The confidence interval accounts for the size as well as the results of the study. [The power of the study would account only for its size, not its results.] Thus, it is appropriate to use the 95% confidence interval.

A study in Dow workers (Ott *et al*, 1983) found no statistical excess for these two cancer sites, though there were little exposure data.

Animal data

A 2-year study with methylene chloride in mice and rats showed excesses of lung and liver cancer, only in mice (NTP, 1986). Short-term exposure showed evidence of inflammatory cell infiltration in rats (NTP, 1986).

No excesses for cancers were observed in analogous studies in rats and/or hamsters (NTP, 1986). An initial slope factor was calculated from the mouse bioassay data (EPA, 1985).

Mode of action information

Two metabolic pathways have been demonstrated for methylene chloride. It is thought that the species difference in cancer occurrence is due to the preference in mice for a glutathione-S-transferase pathway.

Pharmacokinetic differences were subsequently accounted for which reduced the slope factor 9-fold. [This subsequent pharmacokinetic adjustment is not considered below].

B.9.3 Quality aspects of the data

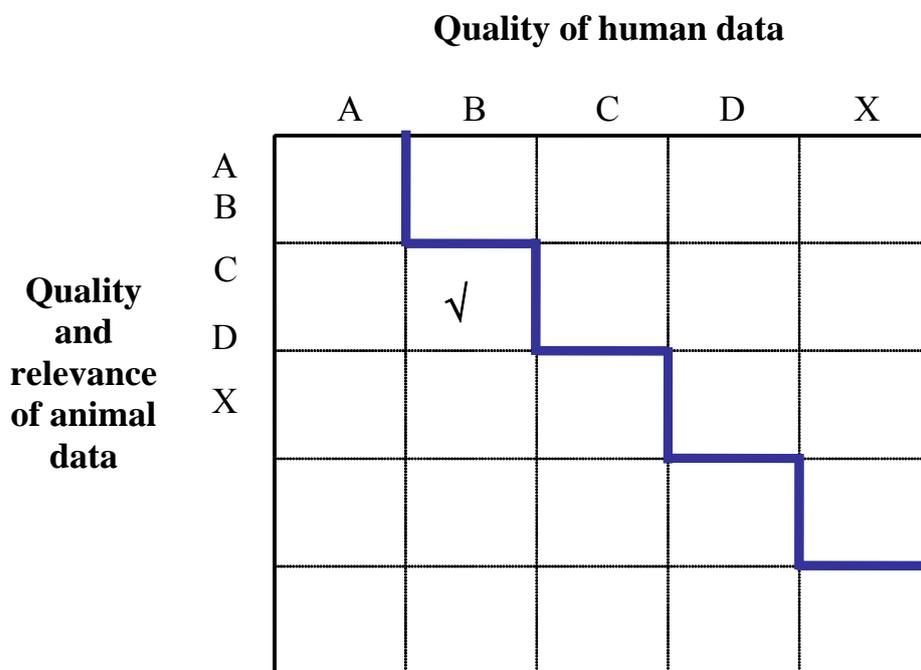
The animal data on methylene chloride followed standard bioassay protocols and were well documented. Consequently they attract the highest Klimisch reliability score. The relevance of the animal data to human is category B (i.e. lung and liver tumours are relevant to humans, with a sensitivity difference). [For illustrative purposes, no consideration is made of subsequent pharmacokinetic adjustments, which may have changed the animal quality score from category B to A.]

The quality of the human data on methylene chloride is difficult to judge. A category A would require several well-conducted studies, whereas only two exist, but category A also allows for one particularly strong, large study, especially if there are no conflicting data. In the case of methylene chloride, the study is strong; rarely are such carefully documented exposure data available. In addition, there are no conflicting data. However, the study is not particularly large, as it consists of only 1,013 workers. Thus, the judgment is that the human data fall into the high end of category B.

B.9.4 Relevance to humans and data integration

At the time of this assessment of methylene chloride, the human relevance score for the lung and liver tumours is category B, i.e. the data are relevant to man with a sensitivity difference. Subsequently, PBPK modelling accounted for the sensitivity difference, bringing the modern day assessment to category A (Figure B.8).

Figure B.8: Integration of human and animal data for methylene chloride



B.9.5 Identification of critical effect (hazard)

The animal data on methylene chloride (before PBPK modelling adjustments) are deemed to be category B. The human data fall just short of category A and are also considered category B. An assessment needs to be made as to whether or not the data are in concordance. If so, it would not matter which dataset was used to characterise the cancer dose-response curve. The animal data are as follows:

Table B.1: Tumour incidence in rats/mice exposed to methylene chloride (NTP, 1986)

Tumour type	Concentration		Tumour incidence
	(ppm)	(mg/m ³)	
Liver	0	0	3/45
	2,000	7,060	16/46
	4,000	14,100	40/46
Lung	0	0	3/45
	2,000	7,060	30/46
	4,000	14,100	41/46

From these data, a potency estimate of 1.4×10^{-2} was calculated, for the critical effects of lung and liver tumours.

The human data on methylene chloride consisted of follow-up among 1,013 workers exposed to an average of 26 ppm (91.8 mg/m³) for 22 years. They showed the following:

Table B.2: Standard mortality ratio for tumours in Kodak workers (Hearne *et al*, 1987)

	Standard mortality ratio Observed/Expected	Relative risk	95% confidence interval
Liver cancer	0/0.8	0	-
Lung cancer	14/21	0.67	-
Total	14/21.8	0.64	0.35 - 1.08

To assess whether there is concordance, steps are taken as in Bukowski *et al* (2001). In essence the animal findings are applied to the epidemiologic study, to assess whether the results fall into the 95% confidence interval. This accounts for the size (related to power) as well as the results of both sets of data.

The workers were exposed to an average of 2.5 ppm (8.82 mg/m³) each day for 20% of their lifetime (factoring in an 8- rather than a 24-hour day, as well as a 5 day/working week). A slope factor of 1.4×10^{-2} /day among 1,013 workers would predict 7.2 excess cancers. Thus, it is estimated that $2.5 \text{ ppm/day} \times 1.4 \times 10^{-2}/\text{ppm} \times 1,013 \text{ workers} \times 20\% \text{ partial lifetime} = 7.2 \text{ excess deaths due to methylene chloride}$.

There are 21.8 background cases of liver and lung cancer in the study. Additional 7.2 excess cases would result in an observed number of 29. This would produce a relative risk estimate of 29/21.8, or 1.3. Since the 95% confidence interval (0.35 - 1.08) does not contain 1.3, the data are lacking concordance.

The slope estimate consistent with the upper bound on risk in the epidemiology study has been calculated. Back calculation yielded a slope factor of 0.3 ppm/d (1.06 mg/m³/d) as consistent with the epidemiology data. Thus, the animal data-derived slope factor should be reduced by a factor of 4, to take into account both data sets. This is still a conservative, protective approach, since it is using the upper bound on risk from the human data set. The advantage is that neither set of data is ignored, since they are of similar quality.

B.9.6 Current situation

Subsequently, PBPK adjustments were made to reduce the slope factor 9-fold, which brings the animal data in concordance with the human data. This is supportive of the fact that a conservative, protective account of the human data was taken above.

B.10 bis(Chloromethyl)ether

B.10.1 Formulation of the problem

bis(Chloromethyl)ether (BCME) has been used in industry for chloromethylation. Exposure has occurred mainly as a contaminant of chloromethyl methyl ether (CMME). BCME has been used as monitoring indicator for CMME due to its greater stability in workroom air. Most of the human data cannot distinguish between effects caused by BCME or CMME. The available animal data are specific with regard to carcinogenic effects of BCME.

The question is whether the animal data are concordant with a carcinogenic effect of BCME in humans.

B.10.2 Summary of key human and animal data

Human data

IARC (1987) has reviewed several epidemiology studies and case reports, and concluded that workers exposed to CMME and/or BCME have an increased risk of lung cancer. Among the heavily exposed workers, the risk was about 10-fold or more. The risks increased with duration and cumulative exposure. The latency period was shortened among workers with heavier exposure who tended to develop primarily small cell carcinoma. Almost all the workers had been smokers. However, exposure to BCME without CMME seems to have occurred only in two of the studies; in the majority CMME exposure was higher, BCME being only a contaminant.

Animal data

BCME has been reported to be a highly potent alkylating carcinogen for mouse skin and an inducer of respiratory tract tumours in rats and mice (Goldschmidt *et al*, 1968; Van Duuren *et al*, 1968 cited by ACGIH, 2001). Increased numbers of pulmonary carcinomas occurred in mice after repeated daily exposure at 1 ppm (4.8 mg/m³). In rats, squamous cell carcinomas of the lung

and esthesio-neuroepitheliomas were reported at 0.1 ppm, but not at 0.01 or 0.001 ppm (Albert *et al*, 1982; Sellakumar *et al*, 1985; both cited by ACGIH, 2001).

B.10.3 Quality aspects of the data

The carcinogenicity studies in rats, mice, and hamsters are reliable (Klimisch category 1 or 2) and of 'High' quality. The animal studies would be rated Category A (animal MoA plausible to man; there is no obvious reason to assume a different sensitivity of humans).

The human data could be classified as category A for the co-exposure of CMME and BCME, taking into account the clear and consistent association in the various reports (although smoking was often a co-exposure which could not always be well controlled). However, for BCME alone in humans, the evidence is less clear, due to co-exposure to CMME in most of the cases; thus category B.

B.10.4 Relevance to humans and data integration

Although for BCME alone the human data are of slightly lower quality than the animal data, both types of data are relevant, especially if for humans the evidence regarding exposure to CMME and/or BCME is considered (Figure B.9).

B.10.5 Identification of critical effect (hazard)

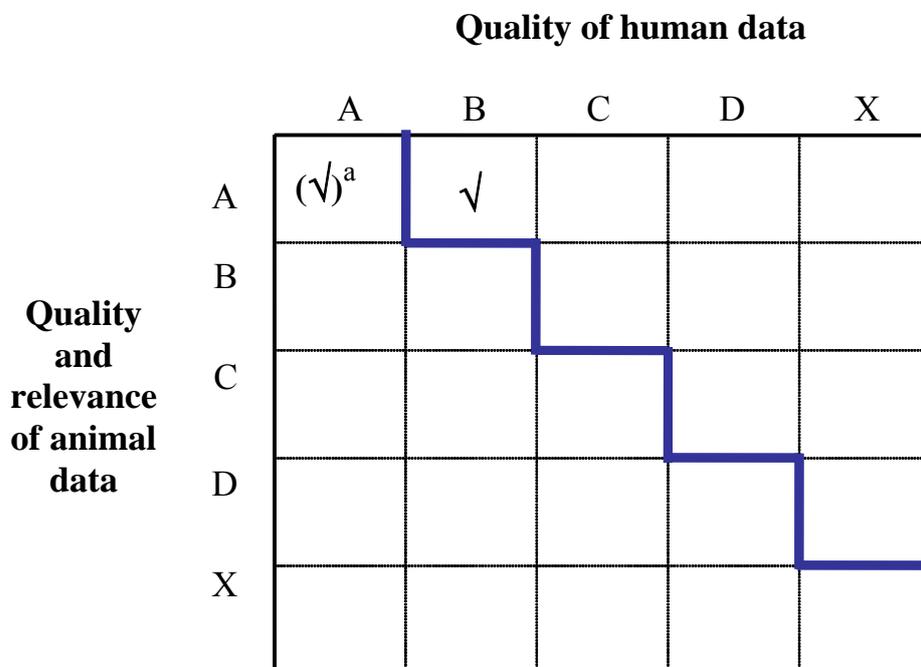
Both human and animal data point to lung cancer as the critical effect of BCME. Other sites of the respiratory tract may also be potential targets, but of less importance than the lung.

B.10.6 Dose-response and safe levels

Increased risks with duration and accumulation of exposure to BCME have been demonstrated in humans, with highest relative risks 15 to 20 years after first exposure. However, exposure data are insufficient for quantitative risk assessment.

In animals, carcinogenic effects in 98 of 111 rats have been observed at a concentration of 0.1 ppm BCME, but not at 0.01 ppm (0.48 and 0.048 mg/m³, respectively).

Figure B.9: Integration of human and animal data for BCME and/or CMME



^a Co-exposure with CMME

B.10.7 Current situation

BCME is no longer used for chloromethylation.

ACGIH (2001) has recommended a TLV-TWA of 0.001 ppm (0.0048 mg/m³).

B.11 Hydrogen fluoride

B.11.1 Formulation of the problem

Producers of hydrogen fluoride have stated that there is no longer consumer exposure since historic uses for aqueous hydrogen fluoride (e.g. in rust cleaning agents) have been discontinued. It was claimed that consumer use only occurred in accidental cases (e.g. workers using industrial products in private life or when products containing hydrogen fluoride are reformulated and/or marketed as consumer products). However, data obtained from a Belgian poison centre showed that the use of hydrogen fluoride in rust, stone and wood cleaning agents, all available and

marketed to consumers, was still common practice. Moreover the situation seemed to be similar in France.

The problem is apparently not the quality of the data *per se*, but the effectiveness of risk containment measures taken on the basis of the results of the (completed) risk assessment process (VROM, 2001). This is clearly an issue of risk management. Taking into consideration these incident reports, previous risk control measures (such as banning aqueous hydrogen fluoride use in consumer products) may be questioned with respect to its implementation and/or effectiveness.

B.11.2 Summary of key human and animal data

Animal as well as human studies are available for hydrogen fluoride (reviewed by VROM, 2001). Signs of acute fluoride effects in humans resemble those observed in animals. In accordance with EC guidelines, hydrogen fluoride is classified and labelled as T+ (very toxic).

Dermal contact with hydrogen fluoride, either as a liquid or a gas, produces severe dermal and mucosal lesions. Inhalation of hydrogen fluoride is highly damaging to the respiratory tract and can cause systemic (cardiac) effects including death. However, local, acute effects are most commonly seen following exposure to hydrogen fluoride.

Information obtained from a poison centre showed that a significant number of consumer accidents had occurred through the use of rust, stone or wood cleaning agents containing hydrogen fluoride. In the majority of cases these involved burns on the hands requiring medical attention. It was concluded that complementary risk reduction measures were needed because of the strong corrosive properties of the substance and the fact that apparently incidental exposure could not be completely excluded.

B.11.3 Quality aspects of the data

In some situations, the nature of the effect determines the quality of the data needed. In this case, because it is a local, acute, obvious effect, no sophisticated epidemiology study design is needed to provide reliable information. It could be considered that the data do not totally meet required quality standards and are not likely to be representative of all cases that occur. Nevertheless, these isolated incidents that were confirmed through at least one other poison centre were sufficient to detect a failure in the risk management strategy. Neither the risk assessment nor the data that have contributed to it are at fault. Rather the way in which the problem was dealt with in the risk management process is questionable and subject to improvement.

B.11.4 Relevance to humans and data integration

There is overall concordance between the animal and the human data in terms of effects. As such the human data in terms of exposure are adequate to assess this issue. Both datasets may be scored Category A (Figure B.10).

Since the human data take precedence, reliance can be placed on the incident reports to address this problem of risk management.

The human data confirm that exposure to hydrogen fluoride in consumer products is still occurring. Although these are not 'real' epidemiology data (i.e. generated on the basis of an epidemiology study design) and would likely be classified as isolated case reports, in this situation, where the effect of hydrogen fluoride is predominantly local and acute, the data are sufficient to detect a fundamental problem in risk management.

B.11.5 Identification of critical effect (hazard)

The main health effect with hydrogen fluoride solutions following oral, nasal, ocular, respiratory or gastro-intestinal exposure is an acute, corrosive effect at the site of exposure. This is the most sensitive effect shown in humans and animals. Any other competing effect is expected to be more readily overruled by this immediate critical effect, even at low concentrations.

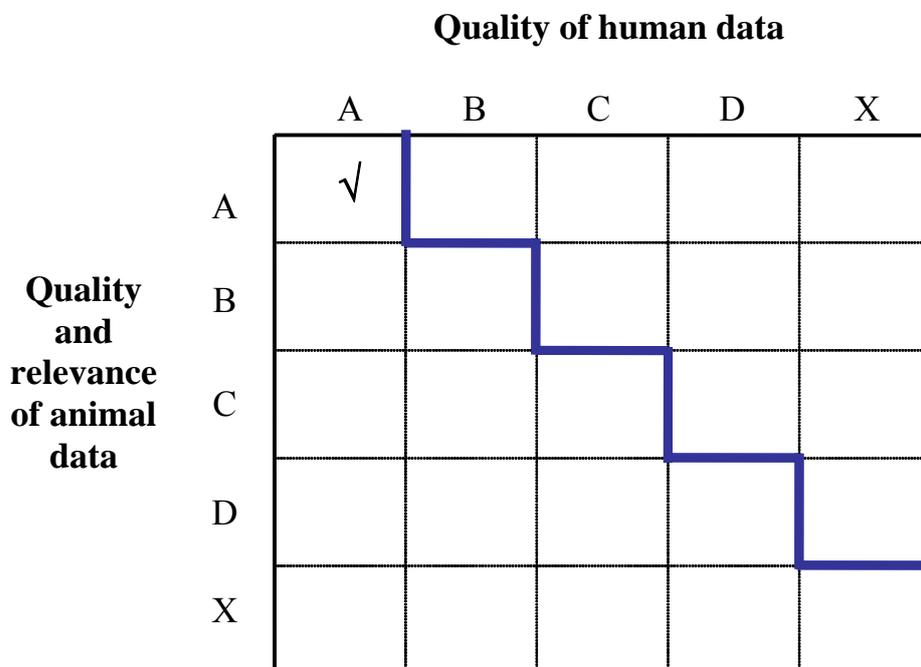
B.11.6 Dose-response and safe levels

Given the fact that the critical health effect of hydrogen fluoride is an acute, local, immediate severe irritation (i.e. corrosive damage of the skin and mucosae), a dose-response curve is not relevant as exposure to hydrogen fluoride generates a clear, direct and prolonged effect even in limited concentrations (Peters and Miethchen, 1996).

B.11.7 Determination of NO(A)EL

In the case of hydrogen fluoride, this step is not relevant.

Figure B.10: Integration of human and animal data for hydrogen fluoride



B.11.8 Current situation

Based on its local effects, hydrogen fluoride has been classified by the EC as very toxic and corrosive (risk phrases 26/27/28-35, “Very toxic by inhalation, in contact with skin and if swallowed” and “Causes severe burns”) under the Dangerous Substances Directive 67/548/EEC and its subsequent amendments. No NOAEL could be determined.

The EC Risk Assessment (VROM, 2001) concluded that irritation scores were increased significantly above 3.0 ppm (2.5 mg/m³) (Lund *et al*, 1997). This value may be considered a LOAEL, however further studies are deemed necessary to cover lower concentrations.

According to Lund *et al* (1999), sub-clinical effects occurred between 0.8 and 2.9 ppm (0.7 and 2.4 mg/m³) and above; no controls were used in this study.

In the USA, the ACGIH (2006) has set an OEL (TLV) for hydrogen fluoride (F hydrolysed) as an 8-hour TWA of 0.5 ppm with a ceiling value of 2 ppm (0.42 and 1.66 mg/m³, respectively).

The above data confirm the current situation i.e. that the main effect of hydrogen fluoride is irritation and corrosive damage. This is a specific and local effect, which demands an appropriate risk management strategy (such as avoiding any consumer exposure).

B.12 Human monoclonal antibody ‘TeGenero TGN1412’

B.12.1 Formulation of the problem

TeGenero TGN1412 is an immuno-modulatory human monoclonal antibody, designed to activate the T-cell response. It is intended for the treatment of chronic inflammatory conditions (multiple sclerosis, rheumatoid arthritis) and leukaemia.

Following animal testing, TGN1412 was subjected to a ‘phase one’ trial designed to test its safety in humans. Six treated volunteers fell ill and were eventually admitted to a hospital intensive care unit, two in a critical life-threatening condition, suffering from multiple organ failures due to a ‘cytokine storm’ (striking release of cytokines and profound lymphocyte proliferation).

A novel pharmaceutical begins its development with only animal safety data and data from *in vitro* models. A risk assessment is normally conducted before dosing the first human trial. Clearly, in this instance, errors occurred in the process. The problem was that the safe level for human dosing was derived by applying assessment factors to the animal data, while omitting consideration of some vital *in vitro* data that were available in the public domain. As a consequence, the exquisite affinity of the antibody for human relative to animal target was disregarded.

B.12.2 Summary of key human and animal data

Human data

A phase one (human volunteer) trial on TGN1412 was instigated using a dose of 0.1 mg/kgbw (> 5 mg/volunteer). The dose was calculated from studies in *Cynomolgus* monkeys (below), using a safety margin of 500. The trial involved a small number (8) of healthy volunteers, 2 of whom received a placebo, while the remaining 6 received the drug. This resulted in severe injuries to all 6 treated volunteers following a ‘cytokine storm’. Based on additional literature data, it was concluded that the starting dose was incorrect. It should have been about 1 µg/kgbw, in the light of the basic pharmacology of this human receptor (reviewed by Mayor, 2006).

TGN1412 binds to T-lymphocytes at a site on the receptor protein CD28 causing cytokine release, resistance to apoptosis and cell multiplication. Unlike the normal T-cell activation process, no second binding site on the T-cell receptor (CD2) is required, so the whole population of T-cells can be activated causing a cytokine storm. The constant K_d for TGN1412 binding to human CD28 (molecular weight 150,000) is 2×10^{-9} mol/l (0.3 mg/l). The distribution pharmacokinetics show that the immunoglobulin is extracellular and a dose of 0.1 mg/kgbw will lead to a plasma level of about 0.5 mg/l. It appears that the dose that was used in the trial is in the mid range of the amount expected to be pharmacologically active, since binding was sufficient for activation *in vitro*. This is contrary to the normal procedure, where a new molecule is given to volunteers at the active site at a dose well below the calculated active pharmacological/toxicological dose.

Furthermore, a similar anti-lymphocyte antibody, OKT3 has been proved to cause a 'cytokine storm' reaction at the usual human immunosuppressant dose of 5 mg. At this dose the number of bound antibodies/cell can be calculated and confirmed by *in vitro* studies of lymphocyte activation. This calculation can also be made for TGN1412 predicting its activation, but not its magnitude.

Animal data

Toxicology studies on TGN1412 used *Cynomolgus* monkeys, which have a CD28 receptor similar to human, but react quite mildly to 50 mg/kgbw doses. This dose was considered to be the NOAEL. It does give an indication of possible effects in terms of endpoints, but fails to integrate the fact that in the case of signalling molecules (such as hormones or receptor binding antibodies) the magnitude and direction of response is quite variable between tissues and species (Mayor, 2006).

B.12.3 Quality aspects of the data

The animal data on TGN1412 may be regarded as 'High' quality *per se*. The same pertains to the available human *in vivo* and *in vitro* data.

However, the quality of the data as such is not at stake here. Instead it is the way that these data were handled in the risk assessment process that was a failure, i.e. using extrapolation from animals to humans corrected by the necessary assessment factors, but disregarding the critical data from the *in vitro* systems.

B.12.4 Relevance to humans and data integration

The animal data on TGN1412 may be scored Category B, i.e. the data are relevant to man, but a sensitivity difference is expected between animals and man on the basis of knowledge of the specificity of the antibody obtained from human *in vitro* systems.

On the other hand, there was a lack of significant cross-reactivity of the TGN1412 antibody in the animal equivalent of the human target protein. This means that there was no animal species that could be considered truly relevant for human risk assessment. For this reason, the case could also be made for the animal data to be considered as of limited relevance (needing a high conversion factor), leading to Category X (Figure B.11).

B.12.5 Identification of critical effect (hazard)

Unlike most of the monoclonal antibodies developed up till now (which inhibit a specific protein target or triggering cell death), the novel pharmaceutical TGN1412 was designed to activate T-lymphocytes. The aim was to use TGN1412 to treat inflammatory disorders and a form of B-cell lymphocytic leukaemia where T-cell activity is impaired.

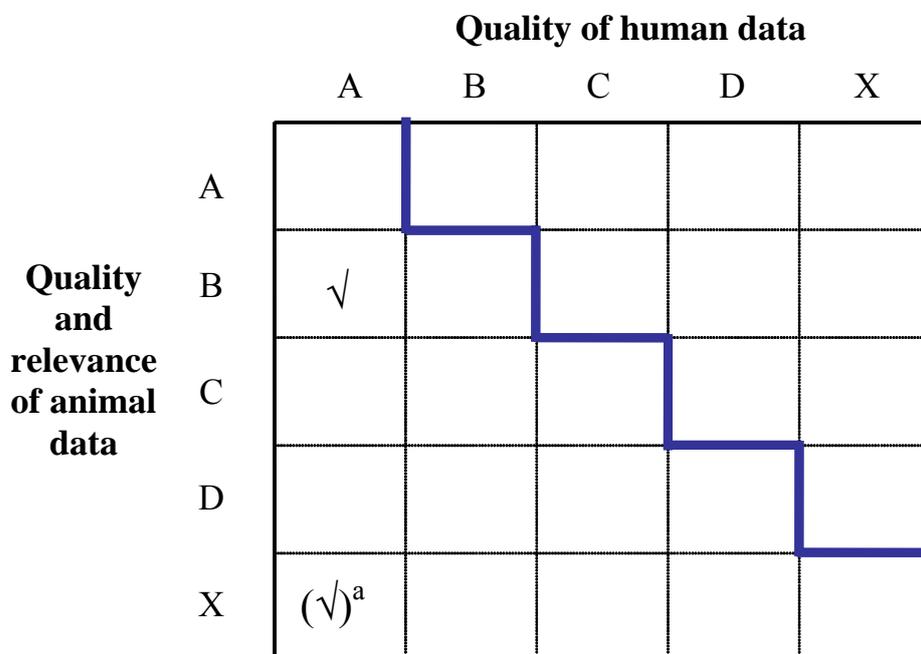
In both animal and *in vitro* studies the lead effect was aimed at an immuno-stimulatory binding to a CD28 receptor, since TGN1412 is an agonistic anti-CD28 monoclonal antibody which has been genetically engineered. Its postulated immuno-modulatory (anti-inflammatory) properties are related to its capacity to activate and expand the number of regulatory T-lymphocytes and induce the release of anti-inflammatory cytokines. As the precise cellular systems involved in the MoA of TGN1412 are specific to humans and non-human primates it was decided that *Cynomolgus* and rhesus monkeys were the most relevant species for toxicology studies.

The intended binding effect of TGN1412 was safely produced in animals, so the first human trial aimed at the same critical effect. Unfortunately, because of the unexpected specific human affinity for the target protein, the magnitude of the intended effect was exaggerated, causing a cytotoxic storm.

B.12.6 Dose-response and safe levels

In *Cynomolgus* monkeys, intravenous doses of up to 50 mg TGN1412/kgbw/wk for 4 weeks were well tolerated and there was no sign of adverse effects on the major physiological (cardiovascular, respiratory and central nervous) systems.

Figure B.11: Integration of human and animal data for TGN1412



^a If the animal data are considered irrelevant

B.12.7 Determination of NO(A)EL

In *Cynomolgus* monkeys TGN1412 had a blood serum half-life of around 8 days. Single doses of 5, 10, 25 or 50 mg/kgbw and a cumulative dose of 90 mg/kgbw failed to elicit toxicity even though enlargement of the lymph nodes was observed in both animals that had received the cumulative dose. The onset of lymph node swelling corresponded roughly with a 4-fold expansion in serum CD4⁺ T-cells. However, this was concluded to be a treatment related effect. Furthermore, in the 28-day studies it was observed that TGN1412 had an optimal T-cell proliferation at the 5 mg/kgbw dose level, suggestive of a departure from a strict dose-effect relationship, yet 50 mg/kgbw was considered to be the NOAEL.

B.12.8 Application of assessment factors

The application of an assessment factor of 500 might have been correct, provided that the available data from the *in vitro* studies were taken into account. This would have given a starting dose in the range of 1 µg/kgbw. Assessing toxicity in non-human primates and applying an assessment factor in this case proved totally misleading, since safety in the monkey did not necessarily predict safety in man.

B.12.9 Current situation

The volunteers survived the incident with TGN1412, but continued to have sequelae such as headaches and memory loss and one participant needed ongoing treatment for dry gangrene of his fingers and toes. In the aftermath of the trial, the UK government convened an Expert Scientific Group to recommend ways to improve the safety of first-in-man trials of high risk drugs (such as biological molecules with novel mechanisms of action, new agents with a highly specific action and new drugs targeting the immune system). The UK group recommended that for a high risk drug the starting dose should be calculated using of the minimal anticipated biological effect level rather than the NOAEL. [This would have yielded a starting dose of 5 µg TGN1412/kgbw]. Another recommendation proposed was infusion of such high risk drugs over a longer period of time and separate dosing between study participants with an appropriate observation interval (Department of Health, 2006). However, the fact still remains that bringing together relevant animal and *in vitro* data would have prevented the incident.

B.13 Dimethylcarbamoyl chloride

B.13.1 Formulation of the problem

Dimethylcarbamoyl chloride (DMCC) is used as an intermediate in the manufacture of pharmaceuticals and pesticides. Although DMCC is a direct-alkylating agent, no cancer cases due to DMCC have been reported to date.

The question is whether the lack of reported human cancer cases is of any relevance for the risk assessment of DMCC.

B.13.2 Summary of key human and animal data

Human data

A study in a small group of workers with potential exposure to DMCC did not reveal any cancer cases of the respiratory tract. However, since these persons were protected by respiratory devices, they may not have been truly exposed to the compound. No other reports of cancer cases possibly associated with DMCC are available (DFG, 2004).

Animal data

Studies in various animal species have demonstrated a carcinogenic effect of DMCC. In a chronic inhalation study in rats, squamous cell carcinomas of the nose were observed at 1 ppm DMCC (4.47 mg/m³) in over 90% of the surviving animals. Similar results were obtained in hamsters. Different types of local tumours were seen in mice caused by DMCC following skin application, or subcutaneous or intraperitoneal injection (Stein, 1976; Sellakumar *et al*, 1980; both cited by ACGIH, 2003).

DMCC is considered to be a direct-alkylating agent with a wide spectrum of genotoxic activity, including in somatic cells *in vivo* (IARC, 1999).

B.13.3 Quality aspects of the data

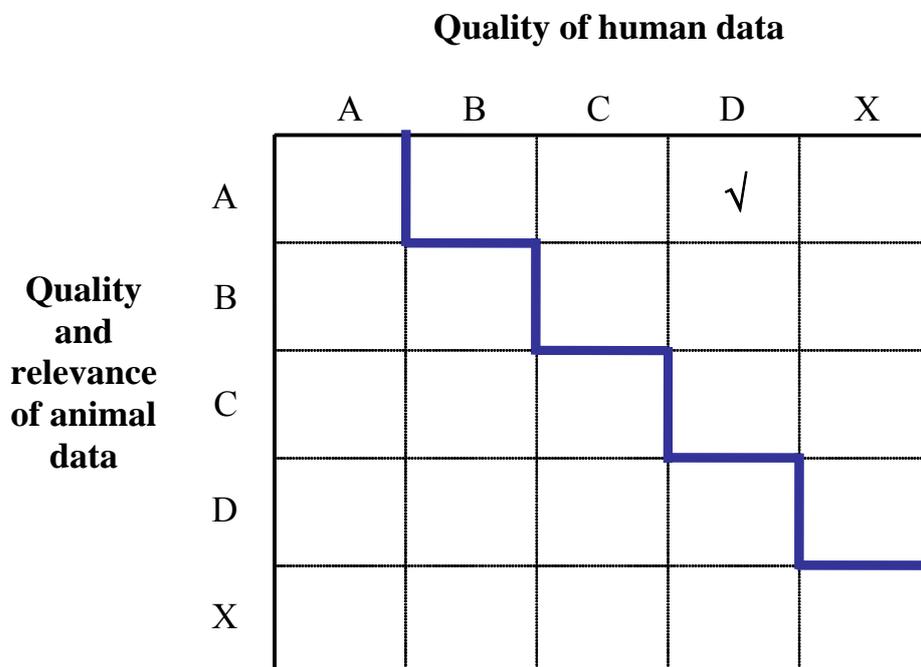
The only human data available concerns a small study in workers that may or may not have been exposed to DMCC. This information would be qualified as Category D at most.

The carcinogenicity studies on DMCC in rats, mice, and hamsters are reliable (Klimisch category 1 or 2) and of 'High' quality (Category A).

B.13.4 Relevance to humans and data integration

With regard to carcinogenicity of DMCC, the null human data carry little weight (Category D remains). The high quality animal data clearly take precedence, also because of their relevance (animal MoA plausible to man; there is no obvious reason to assume a different sensitivity of humans). The data are not necessarily in conflict, since exposures seem to have been well controlled and/or workers were protected (NIOSH, 1978). Thus, overall only the animal data are relevant (Figure B.12).

Figure B.12: Integration of human and animal data for DMCC



B.13.5 Identification of critical effect (hazard)

In rats, mice, and hamsters, local tumours and carcinomas of the nasal tract have been observed following exposure to DMCC. Thus, cancers, in particular those of the sites of direct contact (i.e. skin and respiratory tract), can be considered the critical effects of this compound. No human carcinogenicity data are available.

B.13.6 Dose-response and safe levels

In animals, effects have already been observed at the lowest experimental concentrations. Thus, no OEL can be recommended.

B.13.7 Current situation

DMCC is considered to be a proven carcinogen in animals. The ACGIH (2006) has established a TLV of 0.005 ppm (0.022 mg/m³) with a 'skin' notation. It is classified by this organisation as a probable human carcinogen and an upper respiratory tract irritant.

B.14 Carbon disulphide

B.14.1 Formulation of the problem

Carbon disulphide is used extensively as a solvent in the rubber industry. Due to its high vapour pressure, past exposures in the workplace were commonly above 50 ppm (158 mg/m³). Neurological and cardiovascular effects are the most prominent health effects from occupational exposure to carbon disulphide.

The question is whether an exposure limit can be established below which these effects will not occur.

B.14.2 Summary of key human and animal data

Human data

Long-term health effects described in the literature include retinopathy, neurological effects and cardiovascular effects. Case reports indicate symptoms of nausea, vomiting, dizziness, eye irritation following exposure to over 50 ppm carbon disulphide.

Retrospective cohort mortality studies have been conducted in the rubber industry, focusing on cardiovascular disease and carbon disulphide exposure. The studies have been reviewed by Sulsky *et al* (2002) who concluded that the epidemiology evidence for an association between carbon disulphide exposure and various cardiac risk indicators was equivocal. A meta-analysis of 11 cohort studies on cardiovascular disease mortality in carbon disulphide exposed workers showed a relative risk of 1.56 (95% CI: 1.12 - 2.1) (Tan *et al*, 2002).

An increased risk for hypertension in viscose rayon workers exposed to carbon disulphide has been reported (Chang *et al*, 2007). Another study reported an increased risk of hypercholesterolaemia in viscose rayon workers (Kotseva, 2001).

Animal data

Increased myocardial necrosis in rats exposed to carbon disulphide has been observed (Chandra *et al*, 1972).

No long-term studies on carbon disulphide in animals are available.

B.14.3 Quality aspects of the data

Retrospective cohort mortality studies on carbon disulphide were conducted in accordance with generally accepted principles. However, only a few studies were adjusted for smoking habits and in most there was concurrent exposure to hydrogen disulphide. Since studies in which adjustments for smoking were made essentially reports similar findings as studies without these adjustments it is concluded that the available epidemiology evidence for an increased risk of cardiovascular disease is substantial and meets the criteria for a Category A classification. Furthermore, human mechanistic studies showed supportive evidence of hyper-cholesterolaemia and hypertension.

The evidence in the form of animal studies is limited and is rated Category C.

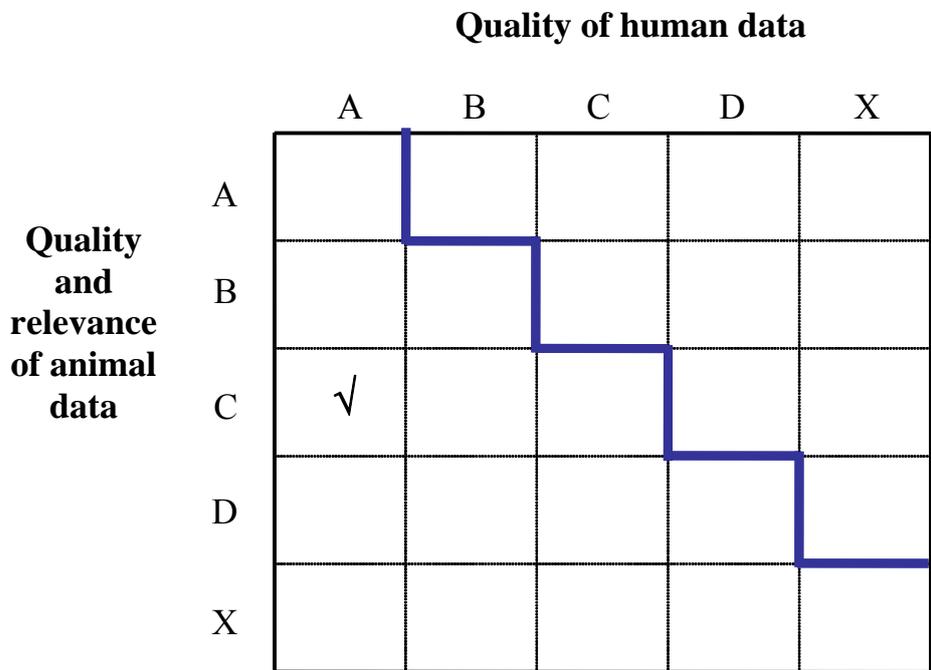
B.14.4 Relevance to humans and data integration

Epidemiology studies indicate an increased risk of cardiovascular disease in workers exposed to carbon disulphide; adjustment for differences in smoking patterns did not change these findings. Thus a quality score of Category A seems warranted. The animal data are less relevant, but are in support of such an association and a Category C is considered appropriate (Figure B.13).

B.14.5 Identification of critical effect (hazard)

Among the effects caused by carbon disulphide in humans, cardiovascular effects have been reported at exposure levels below 7 ppm (22 mg/m³). This is regarded as the critical effect.

B.13: Integration of human and animal data for carbon disulphide



B.14.6 Dose-response and safe levels

The cohort studies with the highest reported exposures show the strongest associations with the relative risks for cardiovascular disease mortality (Tan *et al*, 2002). TWA exposures in the 7 ppm range should be regarded as effect levels.

B.14.7 Current situation

The German MAK value of 5 ppm (16 mg/m³) is probably sufficient to protect against cardiovascular effects (DFG, 2007).

B.15 Toluene

B.15.1 Formulation of the problem

Toluene is a colourless, volatile, non-corrosive and flammable liquid. It is widely used as a solvent for oils, varnishes, resins, rubbers, and paints. Toluene can be combined with other solvents to produce mixtures with more specific physical properties.

Acute irritation has been observed in humans exposed to high concentrations of toluene. Toluene can also induce neurological effects at high exposure levels. In addition, there is equivocal evidence of effects on reproduction in animals.

The question is whether an exposure limit can be established below which these adverse effects are not expected to occur.

B.15.2 Summary of key human and animal data

Human data

Case reports on toluene indicate symptoms of nausea, vomiting, dizziness and eye irritation. Highly exposed toluene sniffers clearly showed symptoms of CNS intoxication (Press and Done, 1967).

Seeber *et al* (2004) reported the results of a 5-year follow-up study on neurological effects in 333 workers in the printing industry. Past exposures were high, in the 50 ppm (190 mg/m³) range, but had diminished to around 25 ppm (80 mg/m³) during the actual study. No clear evidence of neurological effects was seen.

A cross-sectional study of 129 workers exposed to up to 27 ppm toluene (103 mg/m³) as an 8-hour average found that some of the cognitive tests were related to exposure levels (Chouanière *et al*, 2002).

Several studies in human volunteers using small sample sizes showed no acute effects below 100 ppm toluene (380 mg/m³). A group of 34 printers with current exposure in the 50 ppm range and past exposures of around 150 ppm toluene (570 mg/m³) was studied. Only one of 11 psychological tests (nerve conduction velocity) appeared to be associated with exposure (Irgren, 1982).

Animal data

Male rats exposed to toluene for 6 months showed increased liver weight at all dose levels. No histological abnormalities were observed. The enzyme systems appeared to be activated in the 250 to 300 ppm range (957 - 1,150 mg/m³) (Ungvary *et al*, 1980).

Takeuchi (1981) exposed rats to concentrations between 200 and 2,000 ppm toluene (766 - 7,660 mg/m³) for 32 weeks. The observed haematological changes were regarded as transient and reversible.

Toluene caused reproductive effects in rats (Robert *et al*, 2007).

Hudak and Ungvary (1978) exposed pregnant mice daily to 133 and 400 ppm toluene (509 - 1,530 mg/m³). No reprotoxic effects were found.

In addition, toluene was not active in several mutagenicity and genotoxicity tests.

B.15.3 Quality aspects of the data

The human data on toluene are of 'Good' quality (Category B). The study conducted by Seeber *et al* (2004) is relatively large. Repeated measurements were available for 192 exposed persons; past exposure was taken into account. Other studies indicate that acute effects occur at high concentrations. The cross-sectional studies and follow-up study conducted in the 20 to 50 ppm range (77 - 190 mg/m³) show no clear effects.

The animal data are limited or of poor quality (Category B).

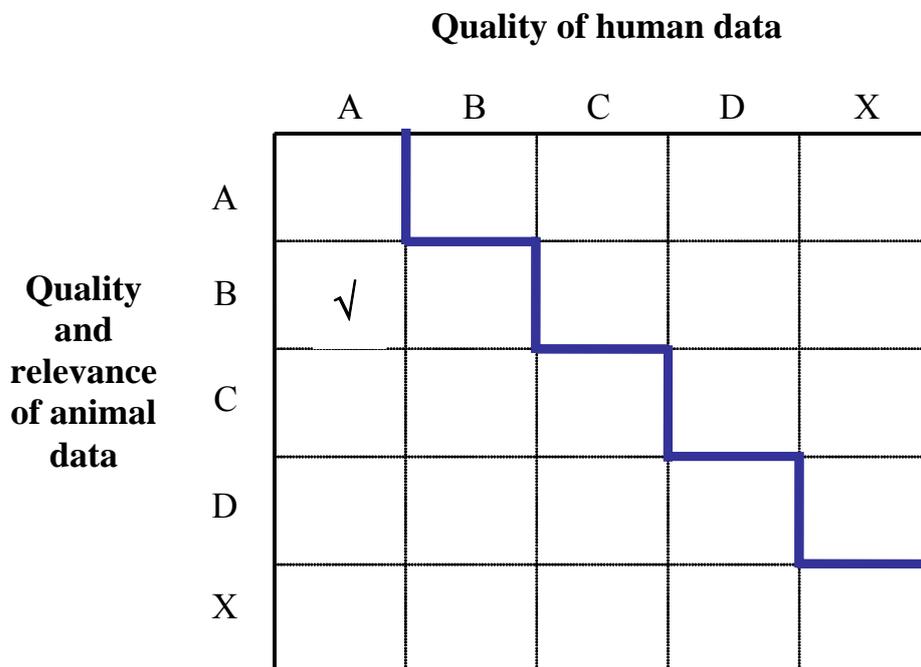
B.15.4 Relevance to humans and data integration

Several epidemiology studies report neurological effects in toluene exposed workers. Bias from other exposure or lifestyle factors seems unlikely. Therefore the human data merit a Category A qualification. The animal data are seem less relevant and could remain Category B (Figure B.14).

B.15.5 Identification of critical effect (hazard)

Neurological effects from exposure to toluene are the critical health effects.

Figure B.14: Integration of human and animal data for toluene



B.15.6 Dose-response and safe levels

Detailed information required for a full dose-response analysis is not available.

B.15.7 Determination of an NOAEL

The study by Seeber *et al* (2004) indicates that exposures below 50 ppm toluene do not result in adverse neurological effects. No assessment factors need be applied because of the good quality of the database and its satisfactory sample size. Additionally, the results are consistent with findings from other studies.

B.15.8 Application of assessment factors

Since exposures in the range of 50 ppm toluene are not reported to induce neurological effects in humans, no assessment factor is required.

B.15.9 Current situation

In many countries an OEL has been established for exposure to toluene. For instance in the Netherlands the OEL is 40 ppm (150 mg/m³) (SZW, 2006). The MAK in Germany is 50 ppm (190 mg/m³) (DFG, 2007). These values seem to adequately protect against neurological effects. The ACGIH (2006) has established a TLV of 20 ppm (77 mg/m³). Visual impairment and female reproductive effects are mentioned as the critical effects.

APPENDIX C: CRITERIA FOR RELIABILITY CATEGORIES (FOR ANIMAL DATA)

Adapted from Klimisch *et al* (1997).

Code of Reliability (CoR)	Category of reliability
1	Reliable without restriction
1a	'Good laboratory practice' guideline study (OECD, EC, EPA, FDA, etc.)
1b	Comparable to guideline study
1c	Test procedure in accordance with national standard methods (AFNOR, DIN, <i>etc.</i>)
1d	Test procedure in accordance with generally accepted scientific standards and described in sufficient detail
2	Reliable with restrictions
2a	Guideline study without detailed documentation
2b	Guideline study with acceptable restrictions
2c	Comparable to guideline study with acceptable restrictions
2d	Test procedure in accordance with national standard methods with acceptable restrictions
2e	Study well documented, meets generally accepted scientific principles, acceptable for assessment
2f	Accepted calculation method
2g	Data from handbook or collection of data
3	Not reliable
3a	Documentation insufficient for assessment
3b	Significant methodological deficiencies
3c	Unsuitable test system
4	Not assignable
4a	Abstract
4b	Secondary literature
4c	Original reference not yet available
4d	Original reference not translated
4e	Documentation insufficient for assessment

APPENDIX D: DERIVING EXPOSURE LIMITS

D.1 Background

The objective of developing an Occupational Exposure Limit (OEL) is to protect workers against potential harmful effects of chemicals under typical workplace exposure conditions.

More precisely OELs are maximum acceptable air concentrations that are used as reference parameters for the protection of workers from over-exposure to chemicals by inhalation (and/or dermal uptake – ‘skin notation’). So far as can be predicted from the current state of knowledge, repeated exposure to concentrations below these reference levels during an entire working life does not cause any significant adverse health effect to the majority of the exposed persons and/or their progeny. OELs are not intended to indicate levels that are safe for the total workforce (i.e. not necessarily safe for the particularly sensitive).

There are well-established procedures for setting OELs provided there is sufficient toxicological information available from animals and/or from humans.

The amount of information available on a chemical substance varies significantly but generally the OEL is based on the NOAEL or LOAEL for the most critical effect seen in one or more repeated dose animal studies. Various default assessment factors are used by scientists and risk assessors for extrapolating from animal data to humans. The derivation of risk assessment factors has been the subject of an ECETOC report (ECETOC, 2003).

No such OEL-setting procedures exist for substances with limited or no toxicological or human data. These substances may be designated as ‘data-poor’ substances as opposed to ‘data-rich’ substances for which a sufficient amount of data is available to be evaluated in the process of setting an OEL. Several ways of addressing OELs for data-poor substances are explored in another ECETOC report (ECETOC, 2006b). This report also addresses the use of assessment factors for mixtures of chemical substances.

Substances tested according to Annex VII and VIII of REACH with volumes up to 10 tonnes/year fall into this category of data-poor substances. A DNEL can be calculated as part of the output of a chemical safety assessment under the REACH regulation (assuming that the DNEL is more or less equivalent to the OEL).

However, REACH provides no guidance on deriving this value for data-poor substances. In the above-mentioned ECETOC report (ECETOC, 2006b) several methods have been proposed and evaluated. These include:

- Hazard banding;
- derivation of an OEL on the basis of the maximum tolerated dose in long-term studies;
- calculation on the basis of 4-hour lethal concentrations from animal inhalation studies;
- read across of data from substances with similar structures and known toxicity as an alternative for QSAR (quantitative structure-activity relationship) assessment;
- calculation on the basis of the respiratory dose in case of sensory irritation;
- a method based on the threshold of toxicological concern if less conservative assessment factors are applied.

These methods may be applied independently or as an integrated approach. However, only a provisional OEL can be proposed on such a basis. More reliable and careful interpretation of the limited results and the use of more conservative default assessment factors are necessary to compensate for the lack of data.

D.2 Determining assessment factors and deriving reference doses

Assuming that an animal NOAEL has been identified and considered relevant to human risk assessment, various factors are applied to this NOAEL in order to derive a dose protective of human health. These factors are variously described as assessment factors, uncertainty factors, adjustment factors, and modifying factors, but here the term assessment factors will be used. The literature on assessment factors is extensive, including an ECETOC Technical Report (ECETOC, 2003). The most commonly used assessment factors are:

- Extrapolation test animal to man: An assessment factor of 1 is used to extrapolate from rodents to humans in case of aerosols as well as in case of local effects of water-soluble gases and vapours, resulting in local respiratory effects. In case of oral exposure to substances with a systemic MoA allometric scaling based on metabolic rate (oxygen consumption per kilogramme body weight) is considered a sound default approach to account for interspecies extrapolation for systemic effects. In case of inhalation exposure to systemically acting substances intake is already related to the oxygen consumption and the interspecies assessment factor is equal to 1 for concentrations in the air.
- Extrapolation from the average man to a vulnerable subpopulation: An assessment factor of 5 is proposed to account for (intraspecies) variability within the general population, a factor of 3 for the homogenous worker population.
- Extrapolation from LOAEL to NOAEL (in the absence of an NOAEL: A default assessment factor of 3 is proposed).
- Extrapolation from shorter-term to longer-term studies (in the absence of a study of the desired duration) e.g.:
 - Subacute to chronic NOAEL (default assessment factor 6);

- sub-chronic to chronic NOAEL (default assessment factor 2);
- local effects by inhalation (default assessment factor 1).
- Extrapolation from one exposure route to another, which is only feasible for substances with a systemic MoA. In some cases the extrapolation implies a lower rate of dosing which may be considered as a built-in safety margin.
- Other assessment factors, such as a penalty for an incomplete dataset, for a particularly severe critical effect or for a steep dose-response curve.

Extrapolation is an uncertain activity, which is why assessment factors of 10 are often used. With a good animal data package many assessment factors are unnecessary, but extrapolations to account for species differences and for human variation are generally necessary. Regulatory practices and conventions for the determination of assessment factors vary between different classes of chemical and different use situations, often dividing between cases where different legislative frameworks apply (e.g. IGHRC, 2003).

An additional extrapolation that is also needed with animal data is from the high doses used in animal studies to the low exposures likely to be experienced by humans. This extrapolation does not attract a specific assessment factor, but is generally done in a conservative way. For example, at low doses, systemic exposure to a chemical is generally proportional to received dose. But at high doses metabolic and excretion pathways can become saturated, resulting in a sharp increase in systemic exposure over and above the exposure predicted by proportionality. When these results are extrapolated to lower doses, the presumed systemic exposure is over-estimated, resulting in a conservative factor in the risk assessment.

D.3 Proposed method to derive occupational exposure limits from animal and human data

In general each substance is evaluated individually. If the critical effect occurs via a non-threshold mechanism, no health based OEL can be established.

In general a decision is made as to whether or not the available human and animal data are sufficient to derive a health-based OEL. This is done by the following steps:

1. Collecting information on all hazards of the substance (physical, chemical, toxicological data from animals as well as from humans and including all available epidemiology data from all relevant sources);
2. evaluating all adverse effects amongst the available data and assessing which health effects are likely to be associated with the exposure;

3. establishing which adverse effect occur at the lowest exposure: This may be the key or critical effect for setting an OEL;
4. selecting relevant human and animal studies of sufficient quality, in which the critical effect has been shown;
5. establishing the MoA and mechanism, threshold or non-threshold;
6. evaluating the dose-response relationship for all relevant adverse effects;
7. establishing the NOAEL and the LOAEL;
8. recommending a numeric value for an OEL for a substance below the NOAEL, while applying appropriate assessment factors;
9. determining and validating the appropriate methods for air monitoring and analysis of the samples.

It must be noted that in studies on pharmaceuticals, the intended ('pharmacological') effect is not considered adverse. However, when it comes to assessing the effect of such molecules on the individuals in the general population or on workers, these pharmacological effects are undesirable. Therefore, the term 'adverse' must be interpreted appropriately and if necessary, the NOEL must be used rather than the NOAEL (ECETOC, 2002b; Lewis *et al*, 2002).

When using human data alone in setting an OEL for a substance the above mentioned procedure remains to a great extent valid but some particular steps need some more specific recommendations because of the fact that most human data have been generated by non experimental study designs. Possible sources of bias are to be expected from such studies in contrast to experimental animal studies in which conditions are likely to be controlled.

In the following section a short elucidation is given for the convenience of the reader for the 9 steps for OEL derivation. In principle the elucidations are repetitions and summaries of the more elaborate text in the report.

D.3.1 Evaluate the quality of the available human data

The quality aspects of human data refer to the applied research design and the quality of exposure and effects data (Chapter 2). Many human data do not completely meet all the described quality criteria and it should be recognised that imperfect data can still be of importance for OEL derivation.

D.3.2 Assess which health effects are likely to be associated with the exposure

Human data may indicate that a certain exposure is (or is not) associated with a number of health effects. The aim of step 2 (Section D.3) is to group the human data according to the potential health effect. Combined with the quality assessment from step 1 (Section D.3), an assessment can be made which health effects are (or are unlikely to be) associated with the chemical.

D.3.3 Identify the critical effect

Once the human data are grouped according to the potential health effect, data on exposure concentrations should be used to determine which health effect can be expected to occur at the lowest effect concentrations. This is usually the critical effect unless animal data have indicated a more subtle mechanism, in which case these data may determine the critical effect. If the critical effect is prevented from occurring it can be anticipated that other potential effects will not occur since they have only been reported at higher concentrations. The identification of the critical effects requires reliable information on the dose-response curves of all involved health effects.

D.3.4 Describe the dose-response curve of the critical effect

Once the critical effect has been identified, the most detailed description of the dose-response curve will form the basis for the estimation of the NOAEL. Dose-response curves for human data typically are given for actually occurring exposure groups of subjects. For instance in studies of occupational exposures combinations of job and workplace are generally used to group individuals according to their exposure experience. The aim is to create exposure groups containing individuals with a similar exposure, homogeneous exposure groups. A larger number of homogeneous exposure groups have the advantage that the dose-response curve is more specific, but the disadvantage of small groups is that the estimate of the occurrence of the effect can become imprecise because of the smaller sample size.

D.3.5 Determine the NOAEL

Since most human data have been generated by non-experimental study designs, the exposure categories used in the data analysis are not selected by the researcher but are mainly based on the exposure condition under investigation. In many instances the width of the exposure groups presented in the publication of the study is so large that it can only be concluded that the NOAEL must lie within one of the exposure categories. The exposure category in which the NOAEL must lie is the lowest exposure category in which effects are observed. The steepness of the

dose-response curve can provide information on where the NOAEL can be expected to lie. For instance if the prevalence of the effect in the lowest effect exposure category is 5% and the prevalence in the next exposure group is 20% it can be assumed that the NOAEL is closer to that exposure category than to the lower exposure category. On the other hand if the prevalence of the effect in the lowest exposure category is 15% and the prevalence in the higher category is 20%, it can be expected that the NOAEL is closer to the lower limit of the lowest effect category than to the higher limit.

D.3.6 Apply the required assessment factors

If the key human data are of poor quality giving rise to uncertainty regarding the NOAEL, it may be necessary to use an assessment factor. Additionally, if an exposure limit is derived for a more sensitive population than the population from which the human data were derived, an additional assessment factor should be used. Moreover any residual uncertainty not yet accounted for needs to be addressed, such as the severity of the effect, concordance or not with animal findings, and the associated lack of power of the study design.

MEMBERS OF THE TASK FORCE

A. Schnatter (Co-Chairman)	ExxonMobil Biomedical Sciences USA - Annandale
K. Travis (Co-Chairman)	Syngenta UK - Jealott's Hill, Bracknell
C. Money	ExxonMobil B - Brussels
L. Onyon	WHO-IPCS CH - Geneva
D. Pallapies ^a	BASF D - Ludwigshafen
P. Priem	Solvay B - Brussels
G. Swaen	Dow NL - Terneuzen
H. Vrijhof (Secretary)	ECETOC B - Brussels

Acknowledgement

The Task Force wishes to thank D. Weed (Weinberg Group, USA - Washington, DC) for his additional peer review.

^a Currently at BGFA - Forschungsinstitut für Arbeitsmedizin der Deutschen Gesetzlichen Unfallversicherung Institut der Ruhr-Universität Bochum, D - Bochum

MEMBERS OF THE SCIENTIFIC COMMITTEE

(Peer Review Committee)

J. Doe (Chairman) Head of Product Safety	Syngenta UK - Jealott's Hill, Bracknell
R. Bars Team Leader, Toxicology Research	Bayer CropScience F - Sophia Antipolis
P. Calow Professor, Dept. of Environmental, Social and Spatial Change	Roskilde University DK - Copenhagen
W. de Wolf Director of Health and Environmental Sciences	DuPont B - Brussels
D. Farrar Occupational Health Business Manager	Ineos Chlor UK - Runcorn
A. Flückiger ^a Head of Corporate Health Protection	F. Hoffmann - La Roche CH - Basel
H. Greim Director, Institute of Toxicology and Environmental Hygiene	Technical University München D - München
F. Lewis Head of Environmental Safety - EAME	Syngenta UK - Jealott's Hill, Bracknell
G. Malinverno European Public Affairs Manager	Solvay B - Brussels
S. Marshall SEAC	Unilever UK - Bedford
C. Money ^b Industrial Hygiene Adviser, Europe	ExxonMobil B - Brussels

^a Responsible for primary peer review

^b Steward

MEMBERS OF THE SCIENTIFIC COMMITTEE (cont'd)

D. Owen Regulatory and Science Issues Manager	Shell UK - London
M. Pemberton Global Product Integrity Manager	Lucite UK - Billingham
C. Rodriguez Principal Toxicologist, Corporate Central Product Safety	Procter and Gamble B - Strombeek-Bever
D. Salvito Environmental Scientist	RIFM USA - Woodcliff Lake
G. Swaen ^a Epidemiologist, Epidemiology, Health Services	Dow NL - Terneuzen
J. Tolls Director Environmental Safety Assessment	Henkel D - Düsseldorf
S. van der Vies Professor of Biochemistry	Vrije Universiteit Amsterdam NL - Amsterdam
B. van Ravenzwaay ^b Director, Experimental Toxicology and Ecology	BASF D - Ludwigshafen
E. von Keutz Vice President - Head of Toxicology	Bayer HealthCare D - Wuppertal
H.-J. Wiegand Product Stewardship, Corporate Environment, Safety, Health, Quality	Evonik D - Essen

^a Steward

^b Responsible for primary peer review