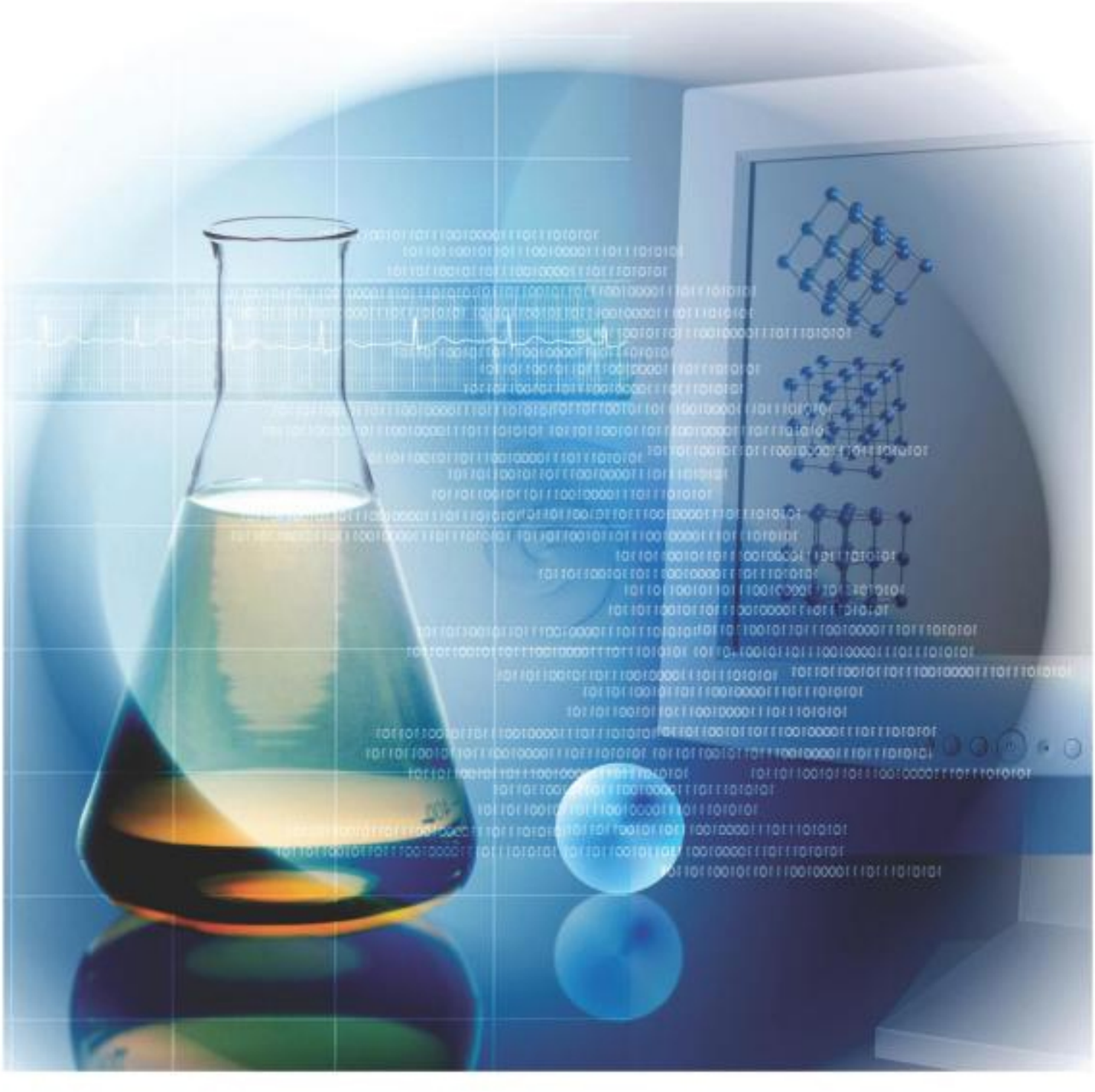
Issues in Environmental Science and Technology

Edited by R E Hester and R M Harrison

Alternatives to Animal Testing



RSC Publishing

Alternatives to Animal Testing

ISSUES IN ENVIRONMENTAL SCIENCE AND TECHNOLOGY

EDITORS:

R.E. Hester, University of York, UK
R.M. Harrison, University of Birmingham, UK

EDITORIAL ADVISORY BOARD:

Sir Geoffrey Allen, Executive Advisor to Kobe Steel Ltd, UK, **A.K. Barbour**, Specialist in Environmental Science and Regulation, UK, **N.A. Burdett**, Eastern Generation Ltd, UK, **J. Cairns, Jr**, Virginia Polytechnic Institute and State University, USA, **P.A. Chave**, Water Pollution Consultant, UK, **P. Crutzen**, Max-Planck-Institut für Chemie, Germany, **S.J. de Mora**, International Atomic Energy Agency, Monaco, **P. Doyle**, Syngenta, Switzerland, **G. Eduljee**, SITA, UK, **M.J. Gittins**, Consultant, UK, **J.E. Harries**, Imperial College of Science, Technology and Medicine, London, UK, **P.K. Hopke**, Clarkson University, UK, **Sir John Houghton**, Meteorological Office, UK, **N.J. King**, Environmental Consultant, UK, **P. Leinster**, Environment Agency, UK, **J. Lester**, Imperial College of Science, Technology and Medicine, UK, **S. Matsui**, Kyoto University, Japan, **D.H. Slater**, Consultant, UK, **T.G. Spiro**, Princeton University, USA, **D. Taylor**, AstraZeneca plc, UK, **T.L. Theis**, Clarkson University, USA, **Sir Frederick Warner**, Consultant, UK.

TITLES IN THE SERIES:

1: Mining and its Environmental Impact
2: Waste Incineration and the Environment
3: Waste Treatment and Disposal
4: Volatile Organic Compounds in the Atmosphere
5: Agricultural Chemicals and the Environment
6: Chlorinated Organic Micropollutants
7: Contaminated Land and its Reclamation
8: Air Quality Management
9: Risk Assessment and Risk Management
10: Air Pollution and Health
11: Environmental Impact of Power Generation
12: Endocrine Disrupting Chemicals
13: Chemistry in the Marine Environment
14: Causes and Environmental Implications of Increased UV-B Radiation
15: Food Safety and Food Quality
16: Assessment and Reclamation of Contaminated Land
17: Global Environmental Change
18: Environmental and Health Impact of Solid Waste Management Activities
19: Sustainability and Environmental Impact of Renewable Energy Sources
20: Transport and the Environment
21: Sustainability in Agriculture
22: Chemicals in the Environment Assessing and Managing Risk

How to obtain future titles on publication

A subscription is available for this series. This will bring delivery of each new volume immediately on publication and also provide you with online access to each title via the Internet. For further information visit *www.rsc.org/issues* or write to:

Sales and Customer Care, Royal Society of Chemistry
Thomas Graham House, Science Park, Milton Road, Cambridge, CB4 0WF, UK
Telephone: +44(0) 1223 432360, Fax: +44(0) 1223 426017, E-mail: *sales@rsc.org*

ISSUES IN ENVIRONMENTAL SCIENCE AND TECHNOLOGY

EDITORS: R.E. HESTER AND R.M. HARRISON

23
Alternatives to Animal Testing

RSC Publishing

ISBN-10: 0-85404-211-3
ISBN-13: 978-0-85404-211-1
ISSN-1350-7583

A catalogue record for this book is available from the British Library

© The Royal Society of Chemistry 2006

All rights reserved

Apart from fair dealing for the purposes of research for non-commercial purposes or for private study, criticism or review, as permitted under the Copyright, Designs and Patents Act 1988 and the Copyright and Related Rights Regulations 2003, this publication may not be reproduced, stored or transmitted, in any form or by any means, without the prior permission in writing of The Royal Society of Chemistry, or in the case of reproduction in accordance with the terms of licences issued by the Copyright Licensing Agency in the UK, or in accordance with the terms of the licences issued by the appropriate Reproduction Rights Organization outside the UK. Enquiries concerning reproduction outside the terms stated here should be sent to The Royal Society of Chemistry at the address printed on this page.

Published by The Royal Society of Chemistry,
Thomas Graham House, Science Park, Milton Road,
Cambridge CB4 0WF, UK

Registered Charity Number 207890

For further information see our web site at www.rsc.org

Typeset by Macmillan India Ltd, Bangalore, India
Printed by Henry Ling Ltd, Dorchester, Dorset, UK

Preface

Alternatives to Animal Testing is a topic which has attracted much public interest and continues to be of concern not only to those professionally involved with chemicals that present a hazard to human health but also to the wider public. All the contributors to this volume are experts in the field and their articles are authoritative and detailed, with appropriate reference to the extensive literature on the subject. However, they also have written for non-experts with a general interest in discovering what progress has been made in applying the three Rs: reducing the use of animals, refinement of procedures to reduce pain and distress, and replacement of animals with alternative tests.

The volume begins with a general overview of the safety evaluation of chemicals, written by Paul Illing who is an independent consultant and an honorary lecturer at the University of Manchester. His chapter examines the context in which toxicological risk assessment is undertaken for regulatory purposes. The next two chapters, on 'international validation and barriers to the validation of alternative tests' and on '*in vitro* testing for endocrine disruptors' are written by Michael Balls and Robert Combes, respectively, both of whom are at the Fund for the Replacement of Animals in Medical Experiments (FRAME). Although much progress has been made in using computational methods and molecular and cell biological methods to progressively replace animal testing, all new tests must be validated before they can be accepted into regulatory practice and the problems remain significant. Endocrine disrupting chemicals were the subject of an earlier volume in this series (Vol.12, published in 1999) and much of that material is brought up to date here in the context of an assessment of non-animal test methods.

Intelligent approaches to safety evaluation of chemicals are examined in Chapter 4 by Derek Knight, who is the Director of Regulatory Affairs at Safepharm Laboratories Ltd. He proposes a tiered approach to safety evaluation and testing which aims to inform commercial decisions on what to prioritise: minimising the cost, reducing the time to market, or achieving regulatory compliance for new chemicals. Finally, insights into the way in which a major manufacturer of cosmetic and personal care products is dealing with the relevant regulatory framework are provided by Carl Westmoreland's chapter on alternative tests and the 7th amendment to the EU Cosmetics Directive.

This volume of issues in Environmental Science and Technology provides an important statement of the current state of development of alternatives to animal testing of chemicals. It details the regulatory environment and charts the progress made towards the goal of reducing the use of experimental animals while improving our understanding of the health risks associated with chemicals. The volume will be of particular interest to those involved in the development and introduction of new

chemicals, to those concerned with regulatory controls such as REACH and the Cosmetics Directive, and to all with a concern for animal welfare.

Ronald E. Hester
Roy M. Harrison

Contents

General Overview of the Safety Evaluation of Chemicals **1**
Paul Illing
1 Introduction 1
2 Legislation and Regulatory Requirements 2
3 The Regulatory Paradigm 4
4 The Interface between Toxicology and Risk Assessment 11
5 Interpreting Toxicology 15
6 Conclusions 24
References 25

International Validation and Barriers to the Validation of Alternative Tests **28**
Michael Balls
1 Introduction 28
2 Background to the Validation Process 29
3 The Validation Process 30
4 Problems being Encountered in Implementing the Validation Process 34
5 The Need for an Invalidation Process 46
6 Concluding Remarks 50
References 50

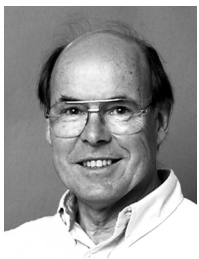
***In Vitro* Testing for Endocrine Disruptors** **53**
Robert D. Combes
1 Introduction 53
2 *In Silico* Approaches 57
3 *In Vitro* Tests for EDs 59
4 Discussion 62
5 Conclusions 68
Acknowledgement 69
References 69

Intelligent Approaches to Safety Evaluation **74**
Derek J. Knight
1 Introduction 74
2 Hazardous Properties Assessment 75
3 Risk Assessment 75
4 Existing Studies and Literature Data 79
5 Surrogate Data: Calculation, Read Across and SAR/QSAR 80
6 Alternatives to Animal Toxicology Studies 81
7 Data Waivers 82
8 Expert Reports and Weight of Evidence 84
9 Preliminary Risk Assessments 85
10 EU Registration, Evaluation and Registration Scheme (REACH) 85
11 Data for REACH 87
12 Regulatory Planning & Liaison with the Regulators 90
13 Common Themes 92
Acknowledgements 93
References 93

Alternative Tests and the 7th Amendment to the Cosmetics Directive **95**
Carl Westmoreland
1 Introduction 95
2 Overview of Current Approaches to Risk Assessment for Cosmetics 95
3 7th Amendment to the EU Cosmetics Directive (76/768/EEC) 99
4 Status of Alternative Tests 100
5 Endpoints where Risk Assessment for Cosmetics can be Achieved without using Animal Data 103
6 Endpoints where Risk Assessment for Cosmetics cannot currently be Achieved without using Animal Data 106
7 New Approaches to Risk Assessment without Animals 114
Acknowledgements 115
Abbreviations 115
References 115

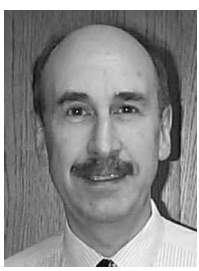
Subject Index **119**

Editors



Ronald E. Hester, BSc, DSc(London), PhD(Cornell), FRSC, CChem

Ronald E. Hester is now Emeritus Professor of Chemistry in the University of York. He was for short periods a research fellow in Cambridge and an assistant professor at Cornell before being appointed to a lectureship in chemistry in York in 1965. He was a full professor in York from 1983 to 2001. His more than 300 publications are mainly in the area of vibrational spectroscopy, latterly focusing on time-resolved studies of photoreaction intermediates and on biomolecular systems in solution. He is active in environmental chemistry and is a founder member and former chairman of the Environment Group of the Royal Society of Chemistry and editor of 'Industry and the Environment in Perspective' (RSC, 1983) and 'Understanding Our Environment' (RSC, 1986). As a member of the Council of the UK Science and Engineering Research Council and several of its sub-committees, panels and boards, he has been heavily involved in national science policy and administration. He was, from 1991 to 1993, a member of the UK Department of the Environment Advisory Committee on Hazardous Substances and from 1995 to 2000 was a member of the Publications and Information Board of the Royal Society of Chemistry.



Roy M. Harrison, BSc, PhD, DSc(Birmingham), FRSC, CChem, FRMetS, Hon MFPH, Hon FFOM

Roy M. Harrison is Queen Elizabeth II Birmingham Centenary Professor of Environmental Health in the University of Birmingham. He was previously Lecturer in Environmental Sciences at the University of Lancaster and Reader and Director of the Institute of Aerosol Science at the University of Essex. His more than 300 publications are mainly in the field of environmental chemistry, although his current work includes studies of human health impacts of atmospheric pollutants as well as research into the chemistry of pollution phenomena. He is a past Chairman of the Environment Group of the Royal Society of Chemistry for whom he has edited 'Pollution: Causes, Effects and Control' (RSC, 1983; Fourth Edition, 2001) and 'Understanding our Environment: An Introduction to Environmental Chemistry and Pollution' (RSC, Third Edition, 1999). He has a close interest in scientific and policy aspects of air pollution, having

been Chairman of the Department of Environment Quality of Urban Air Review Group and the DETR Atmospheric Particles Expert Group as well as a member of the Department of Health Committee on the Medical Effects of Air Pollutants. He is currently a member of the DEFRA Air Quality Expert Group, the DEFRA Advisory Committee on Hazardous Substances and the DEFRA Expert Panel on Air Quality Standards.

Contributors

Michael Balls, *Fund for the Replacement of Animals in Medical Experiments (FRAME), Russell & Burch House, 96-98 North Sherwood Street, Nottingham NG1 4EE, UK*

Robert D. Combes, *Fund for the Replacement of Animals in Medical Experiments (FRAME), Russell & Burch House, 96-98 North Sherwood Street, Nottingham NG1 4EE, UK*

Derek J. Knight, *Safepharm Laboratories Ltd., PO Box 45, Derby DE1 2BT, UK*

Paul Illing, *Centre for Occupational and Environmental Health, University of Manchester, Manchester, UK*

Carl Westmoreland, *Safety and Environmental Assurance Centre (SEAC), Unilever Colworth, Sharnbrook, Bedford MK44 1LQ, UK*

General Overview of the Safety Evaluation of Chemicals

PAUL ILLING

1 Introduction

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

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

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

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

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

Issues in Environmental Science and Technology, No. 23
Alternatives to Animal Testing
© The Royal Society of Chemistry, 2006

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

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

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

2 Legislation and Regulatory Requirements

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

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

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

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

strategies are given in Directive 92/32/EEC[4] and in a technical guidance document[5] and test protocols are given in Annex V of Directive 67/548/EEC.[1] The test protocols are those set up by the OECD.[6] Tests have to be conducted to these protocols and audited in accordance with the principles of good laboratory practice.[7] Animal welfare considerations, set out in Directive 86/609/EC,[8] should be taken into account when deciding on the need for a specific test.

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

The Registration, Evaluation, Authorization and restriction of CHemicals (REACH) proposals[9] currently before the EU Council and Parliament, intend to bring the regulation of both new and existing substances into a common scheme, and to require that specific information is available under specified conditions. For existing substances this implies those placing on the market the substance or preparations containing it may need to set up consortia to conduct testing to fill data gaps. The REACH proposals also have a testing strategy contained within them, and the test protocols are those set out in the OECD Guidelines. Although currently the classification is that for Directive 67/548/EEC, it is likely that the UNECE Globally Harmonized System (GHS)[10] will be substituted before implementation of the proposals. It is claimed that the REACH proposals should encourage the use of alternatives to animal testing.

The second is the area of 'cosmetics' – more properly personal care products. These are governed by Council Directive 76/768/EEC.[11] There are lists of acceptable ingredients and of banned or restricted ingredients. The regulator evaluates dossiers on certain ingredients, information on other ingredients and products are evaluated by the manufacturer/formulator placing the product on the market. The testing requirements have been set out in guidance from the Scientific Committee on Cosmetic Products and Non-Food Products (SCCNFP) intended for Consumers, the latest version being the fifth revision.[12] The tests protocols are based on those set out in Annex V of Directive 67/48/EEC, and are supplemented with SCCNFP opinions concerning studies on human volunteers[13] and mutagenicity/genotoxicity testing.[14] The seventh amendment to this Directive[15] requires that testing of cosmetic products and ingredients in live animals to meet the requirements of the Cosmetics Directive should cease. It gives the European Commission powers to set deadlines after which testing of cosmetics ingredients in animals will not be permitted if the intention is to meet the requirements of the Cosmetics Directive. The Directive sets maxima of a 6 year time limit for short-term testing and a 10 year time limit for repeated dose toxicity, reproductive toxicity and toxicokinetics, with provision that shorter deadlines can be introduced if suitable test methods become available earlier.

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

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

3 The Regulatory Paradigm

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

3.1 Hazard and Risk

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

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

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

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

3.2 Risk Assessment and Risk Management

The OECD/IPCS definitions for risk assessment and risk management[19] are given below.

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

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

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

and non-regulatory options and to select and implement [an] appropriate regulatory response to that hazard.
Risk management comprises three elements: risk evaluation; emission and exposure control; and risk monitoring'.

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

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

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

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

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

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

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

3.3 The Risk Evaluation Framework

The Royal Society study group[18] first put forward a risk evaluation framework. This framework (the 'tolerability of risk' concept) has been restated and slightly developed in 'Reducing risks, protecting people'.[21] It was originally developed to handle engineering risk, but it is equally applicable to risks to human health. Illing[22] and the Health and Safety Executive (HSE)[21] have discussed the application of this framework to the evaluation of health risks arising from exposure to chemicals.

Criteria for reaching decisions can be classified according to three 'pure' criteria.[21] These are:

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

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

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

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

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

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

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

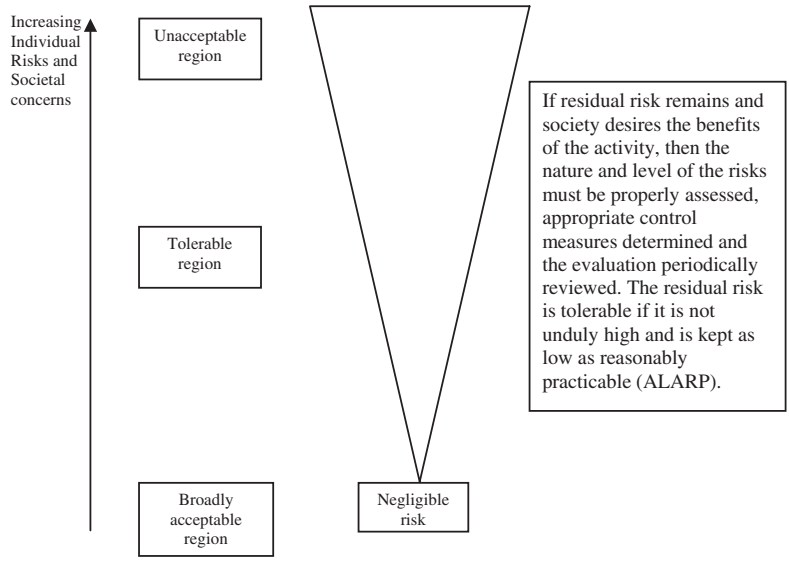


Figure 1 *Outline of the Royal Society approach to risk management/'Tolerability of risk' framework (based on Health and Safety Executive[21])*

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

3.4 Some Problems with the Engineering/Actuarial Approach to Risk

Until recently, the framework given above was seen as a framework for a technical or actuarial analysis of risk based on premature death as the end point. Engineers usually worked in terms of numbers derived from probabilities and consequences, and actuarial material in terms of causes of death. The framework can also be applied to the health risks associated with chemicals, especially when suitable historical data from human epidemiological studies are available. In 'Reducing risk, protecting people', the HSE[21] believes that the boundary between 'broadly acceptable' and 'tolerable' risk is an individual risk of death of one in a million per annum for both workers and the public. This is the original criterion given in the Royal Society's 1983 report. A less clear boundary of one in a thousand per annum for workers and one in ten thousand for the general public is considered as the 'just tolerable' risk for any substantial body of people. HSE first stated these numbers in 1992.[25] If this approach is pursued for death due to the toxicity caused by exposure to an agent, extremely sensitive epidemiological studies are required to give certainty to the numbers found. The critical numbers for much lesser forms of ill health are ill defined, but will be lower, and it might be possible to measure exposures at which these occur directly.

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

Social scientists have outlined a number of difficulties with this approach to risk.[26–29]

Jasanoff[26] suggests that scientists adopt three propositions in developing the technical or actuarial approach to determining the risk criteria representing 'unacceptable', 'tolerable' or 'broadly acceptable' risk, namely:

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

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

The chapter on risk perception in the 1992 report from the Royal Society Study Group[27] states that:

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

Slovic[29] emphasizes the subjective and value-laden nature of risk assessment. Even a relatively simple measure, such as the end point for expressing the risk, can be complex and judgmental (Table 1). Which is the correct measure?

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

Table 1 *Ten ways of expressing risk of death*

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

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

There are also other sociological aspects of risk that affect risk analysis. Anthropologists and cultural sociologists have suggested that social responses to risk are determined by prototypes of cultural belief patterns – clusters of related convictions and perceptions of reality. Renn[28] summarizes four such prototypes:

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

These cultural prototypes have different 'world-views'. Slovic[29] has identified five 'world-views':

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

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

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

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

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

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

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

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

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

4 The Interface between Toxicology and Risk Assessment

4.1 Introduction

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

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

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

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

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

The usual procedure for predictive health risk assessment has been based on animal testing, the 'No Adverse Observed (or Observable) Effects Level' (or surrogate) and 'Uncertainty Factors'. The aim is to develop a numerical value for the level of exposure or detriment judged to be trivial (*i.e*. judged to be a 'broadly acceptable' level of risk, or a safe level of exposure). This dose is normally referred to as the 'Reference dose' when limit setting, although the International Programme on Chemical Safety (IPCS) in its 1994 document[34] called it the 'tolerable intake'. The procedure, and the factor of 100 for use in extrapolating from the no-observed-adverse-effect-level (NOAEL) in animal tests to the reference dose, was first formally enunciated in the 1950s for use with food additives and contaminants.[35]

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

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

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

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

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

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

This approach mixes risk characterization with risk evaluation.

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

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

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

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

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

Within the European Community, the present Technical Guidance Document[5] for use with new and existing chemicals and biocides and the SCCNFP notes for guidance for use with cosmetics[12] use this 'margin of exposure' for risk assessment purposes, although the latter calls it a 'margin of safety'.

4.4 Margin of Safety (OECD Second Definition)

The OECD[19] defines two meanings for the term '*margin of safety*'.

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

Annex 1 of the REACH proposal[9] includes use of the 'DNEL', obtained from the no observed adverse effect level and uncertainty factors, and then sets exposure data against this DNEL. It therefore uses the 'margin of safety' (OECD second definition) approach and rolls risk characterization and risk evaluation together.

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

In the UK, health scientists have deemed certain effects, such as genotoxic carcinogenicity, to be of a type where the hazard is such that there is no 'safe' level of exposure. Exposure is eliminated or reduced 'as low as is reasonably practicable'.[37] Substitution of a different level of effect to the NOAEL, such as 'serious effect' into the uncertainty factors process described above results in a different maximum level of exposure, and hence level of risk. This is only possible when dealing with circumstances where utility/technology criteria are tolerated. If a sufficiently serious level of effect is used then the level of exposure would be that for the cut-off of the upper limit of risk, which should not be exceeded for any individual, and banning/prevention of exposure would be required. If the level of risk lies within the 'tolerable' region then the 'as low as is reasonably practicable' approach is feasible. In the US and elsewhere, extrapolation is undertaken using numerical approaches based on mathematical modelling, *i.e*. the approach so strongly criticized in the Royal Society study group[27] and by Jasanoff.[26] UK Regulators, whose views are based on a 1991 set of Guidelines,[38] do not generally accept such approaches. A more detailed discussion of this point is available.[36,39] This is an area where further developments are needed in order to develop a single, common approach.

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

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

4.6 Additional Comments

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

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

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

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

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

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

5 Interpreting Toxicology

5.1 Introduction

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

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

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

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

5.2 Types of Information

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

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

In vivo toxicity studies are studies carried out in animals or humans. Although the best species for studying effects in humans is humans, ethical considerations[16] limit the type and amount of human testing that can be undertaken. Thus animal studies are required, either as a preliminary to tests in humans or, particularly in the case of longer-term studies, because studies involving deliberate administration to humans would be unethical and/or impractical. Toxicity studies may cover some or all of the following end points:

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

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

5.3 The Key Pieces of Information

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

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

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

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

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

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

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

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

The aims behind obtaining these data are:

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

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

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

5.4 The 'Uncertainty Factors' (UFs) Approach

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

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

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

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

Uncertainty factors cover the following:

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

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

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

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

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

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

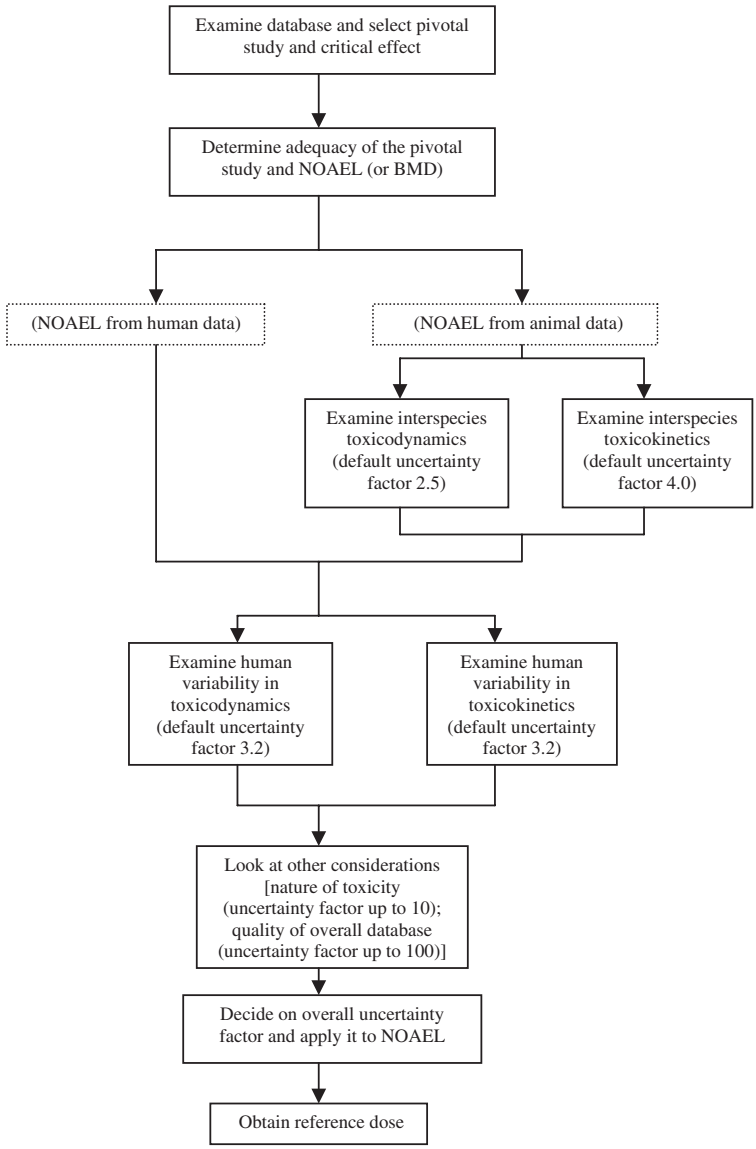


Figure 2 *Procedures for the derivation of uncertainty factors*[22,34]

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

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

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

5.5 The Margin of Exposure Approach

The margin of exposure (MoE) is:

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

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

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

5.6 The Margin of Safety (OECD second definition) Approach

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

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

or:

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

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

5.7 Medicines

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

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

5.8 Classification (and Labelling)

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

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

5.9 Non-Standard Information

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

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

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

In Annex IX to REACH[9] the case-by-case approach to use of structure-activity relationships and 'read across' is formalized. It also introduces the possibility of exposure-driven toxicity testing, *i.e.* the possibility that, if there is minimal or no exposure, some tests in animals can be omitted. Nevertheless, the ultimate arbiter concerning the acceptability of this material is the regulatory authority. The REACH proposals are intended to minimize animal testing by restricting the tests undertaken until the substance is placed on the market in substantial tonnage or has exhibited some alert as to potential toxicity.

The seventh amendment to the Cosmetics Directive[15] is aimed at avoiding animal testing entirely when developing personal care products. Possibly, society has also taken the view that the use of cosmetics and the choice of preparation used is relatively voluntary, and that animal welfare considerations should be given greater prominence in deciding on the need for animal testing. For cosmetics it is possible to obtain some information in humans. Thus the decision has been taken that there will be no need to conduct detailed tests in animals. A timetable is given for dates after which animal testing for various end points will no longer be permitted if the substance is for use as a cosmetic. Physicochemical parameters, structure–activity relationships, and *in vitro* testing will yield adequate information to permit short-term testing in humans and to eliminate enough of the substances likely to cause effects following prolonged dosing (carcinogenicity, mutagenicity and reproductive toxicity). This is an alternative approach not yet considered acceptable for substances generally.

All of these proposals lead away from the precision implied in the current animal test based models for toxic risk assessment/evaluation. Research is needed into whether the detail obtained by the regulatory test in animals is required in all circumstances and how alternatives to animal tests can be taken into the regulatory system. In addition, it is possible that less sophisticated risk assessment and evaluation processes may be all that is required for adequate risk management. For example, one approach to the control of substances hazardous to health (COSHH) in the workplace, set out in COSHH Essentials[45] has moved towards 'easier to understand' systems for risk management. These systems are based on the classifications into the categories of hazard used for classification of chemicals under Directive 67/548/EEC, coupled with assessments of exposure. This system is nominally more transparent than the previous system.

5.10 Comments

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

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

5.11 Some Problems with Animal Testing

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

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

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

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

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

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

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

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

6 Conclusions

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

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

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

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

References

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

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

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

International Validation and Barriers to the Validation of Alternative Tests

MICHAEL BALLS

1 Introduction

In 2001, the European Commission (EC) issued a White Paper entitled *Strategy for a Future Chemicals Policy*, which proposed a system for the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), to be applied to all chemicals manufactured and/or used in the European Union (EU), in order to ensure a high level of protection for human health and the environment.[1] This move was prompted by concern that as many as 100,000 "existing" chemicals were already in the EU market before the current legislation concerning "new" chemicals came into force in September 1981.

Concern about the safety of existing chemicals had already been voiced in the late 1980s, but little action followed until 1998, when the U.S. vice president, Al Gore, announced the introduction of the *Chemical Right-to-Know Program*, following the publication of a 1997 report entitled *Toxic Ignorance*, issued by the Environmental Defense group.[2] As a result, in 1999, the USA Environmental Protection Agency (EPA) negotiated an agreement with various stakeholders to establish the High Production Volume (HPV) challenge, whereby data consistent with the Organisation for Economic Cooperation and Development (OECD) Screening Information Data Set (SIDS) would be produced for 2800 chemicals produced or imported in the USA in amounts exceeding one million pounds per year.[2] The SIDS test battery includes about seven animal tests and two *in vitro* tests, and involves experiments on acute toxicity, repeat dose toxicity, developmental and reproductive toxicity, mutagenicity, ecotoxicity and environmental fate. The current situation with regard to the HPV challenge can be found on the EPA website.[3]

The aims of the REACH system proposals are far more ambitious than those of the HPV challenge, since they involve as many as 30,000 chemicals, divided into four groups according to annual marketing tonnage, namely, 1–10 tonnes (about 19,000 chemicals), 10–100 tonnes (about 5000 chemicals), 100–1000 (about 2500 chemicals) and >1000 tonnes (about 2700 chemicals), with more extensive testing, and more

Issues in Environmental Science and Technology, No. 23
Alternatives to Animal Testing
© The Royal Society of Chemistry, 2006

animal testing, required as the tonnage increases.[4] There will also be restrictions, irrespective of tonnage, and further testing where necessary, for CMR chemicals, *i.e.* chemicals known or suspected of being carcinogenic (C), mutagenic (M) or toxic to the reproductive (R) system, which are manifestations of toxicity of particular concern.

A report published by the EC's European Chemicals Bureau (ECB) reviewed the further testing that would be necessitated by the REACH system, and estimated that, over an implementation period of 11 years, 2.1–3.9 million animals would be required for REACH system compliance, and that the cost to the industry would be between €1143 million and €2274 million.[5]

The implications of these proposals have caused great concern not only in the chemical industry, but also among animal welfare organisations, and among scientists who are opposed to this checklist approach to hazard prediction based on tonnage, which fails to take sufficient account of the nature of the chemicals themselves or of likely human exposure.[6,7]

However, there is one great potential benefit which could accrue from the REACH system proposals, and also from the 7th amendment to the EU Cosmetics Directive[8] – huge human and economic resources are now being invested in the development of non-animal (alternative) test systems and testing strategies which could reduce, and in some cases replace, the need for animal experimentation. These non-animal approaches include computer modelling based on (Quantitative) Structure–Activity Relationships ([Q]SARs), expert systems and biokinetic modelling, and *in vitro* systems, involving animal or human cell, tissue and organotypic cultures, as well as strategies for their intelligent and integrated use.[6]

Nevertheless, if greater protection of human health and of the environment is to be achieved as a result of the identification of existing chemicals which represent currently unknown hazards, and if a vast increase in checklist animal testing is to be avoided, it is vital that the alternative (*i.e.* non-animal) tests and testing strategies are properly developed and proven to be a satisfactory basis for making predictions of toxic hazard for use in making risk assessments, *before* they are accepted into regulatory practice. How this can be achieved, and the pitfalls being encountered along the way, are the subjects of this chapter.

2 Background to the Validation Process

It is now 15 years since a set of principles for *validation*, the process by which the relevance and reliability of a test method are independently assessed for a particular purpose, were agreed at an international workshop held in 1990,[9] and more than 11 years since practical proposals for the conduct of the validation process, taking into account experience gained in an international study on alternatives to the Draize eye irritation test (DET),[10] were considered at a second workshop, held in 1994.[11]

A number of other significant developments took place during the mid-1990s. Bruner *et al.*[12] put forward the concept of the *prediction model* (PM), an algorithm to define how the results from a non-animal test should be used to predict an *in vivo* toxicity endpoint. Curren *et al.*[13] proposed that a three-stage *prevalidation* process, to be introduced between test development and a formal validation study, would greatly improve the efficiency and value of the validation process, by establishing satisfactory

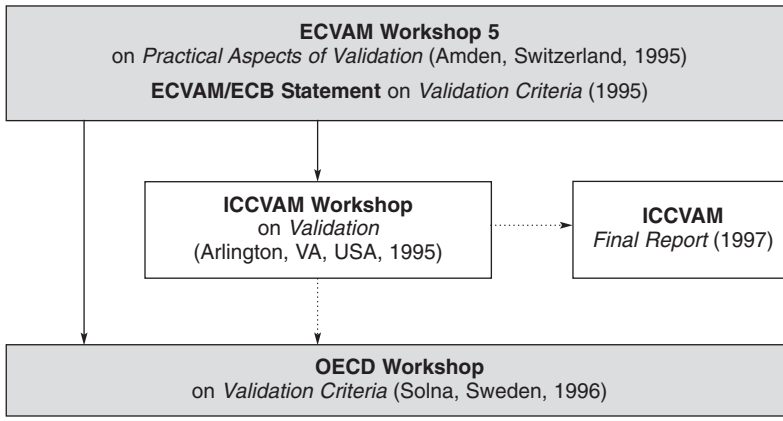


Figure 1 *Key events in the international development of the validation process*

levels of *protocol refinement*, *protocol transferability* and *protocol performance*. The heads of the European Centre for the Validation of Alternative Methods (ECVAM) and the ECB, the EC services responsible for technical aspects of validation and chemicals regulations, respectively, issued a joint policy statement, in which they spelt out criteria for judging the acceptability of test development and of validation studies, and urged that prevalidation and independent assessment should also be applied with equal force to all new or modified animal and non-animal Test Guidelines (TGs).[14]

Later in 1995, the publication of the ECVAM/ECB statement was followed by a workshop organised by the US Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM), which led to the subsequent publication of ICCVAM's position on the validation and regulatory acceptance of toxicological test methods,[15] and early in 1996, by an OECD workshop on the harmonisation of validation and acceptance criteria for alternative toxicological test methods[16] (Figure 1). As a result, the ECVAM/ICCVAM/OECD criteria and principles for validation and acceptance became the basis for practical validation studies, mainly conducted under the auspices of ECVAM, and for weight-of-evidence evaluations, mainly conducted under the auspices of ICCVAM. As a result, by 2002, the ECVAM Scientific Advisory Committee (ESAC) had endorsed the successful outcomes of validation studies on 10 non-animal toxicity tests for chemicals (Table 1),[17] and a number of other methods are currently progressing through the validation process (Table 2).[18]

This may all appear to be most encouraging, and to some extent it is, but it must also be admitted that a number of problems have arisen along the way, some of which now threaten the validation process itself.

3 The Validation Process

A test method consists of a test system and a PM (Figure 2). For example, in the 3T3 neutral red uptake (NRU) phototoxicity test,[19] the *test system* consists of cells of the Balb/c3 mouse embryo fibroblast 3T3 cell line, which are exposed to test chemicals

Table 1 *Validated replacement alternative tests (modified from Ref. 17)*

The 3T3 NRU test for phototoxic potential
The application of the 3T3 NRU phototoxicity test to UV filters
The EPISKIN™ skin corrosivity test
The EpiDerm™ skin corrosivity test
The rat transcutaneous electrical resistance (TER) skin corrosivity test
The CORROSITEX™ skin corrosivity test
The embryonic stem cell test for embryotoxicity
The whole-embryo culture test for embryotoxicity
The micromass test for embryotoxicity
The GM-CFU test for acute neutropenia

Table 2 *Examples of in vitro methods being validated or prepared for validation (modified from Ref. 18)*

The EPISKIN™ human skin model for dermal irritancy testing
The EpiDermTM human skin model for dermal irritancy testing
The mouse skin integrity function test (SIFT) for dermal irritancy testing
In vitro tests for cell transformation (Syrian hamster embryo cells, Balb/c 3T3 cells)
The *in vitro* micronucleus test (genotoxicity)
The use of dendritic cell models for predicting skin sensitisation potential
The NRU 3T3 and NHK basal cell cytotoxicity assays for acute lethal potency
The embryonic stem cell test with extracellular metabolism for embryotoxicity
The use of Leydig cells for reproductive toxicity testing
The use of Caco-2 cells for predicting trans-epithelial absorption
The use of perifusion systems to model chronic toxicity
The use of fish cells in ecotoxicity testing

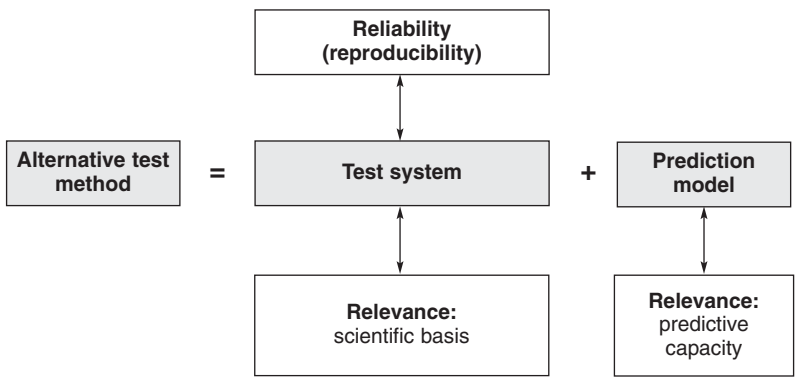


Figure 2 *The validation of replacement alternative test methods (modified from Ref. 20). Replacement test = any test which can be used to replace, or partially replace, the use of animals in the hazard prediction process*

with and without ultraviolet light (*the test regimen*), followed by a measurement of inhibition of cell proliferation after a further culture period (*the endpoint*), by measuring uptake of the vital dye, neutral red, which is dependent on the number of viable cells (*the endpoint assay*). The *result* is expressed as the EC_{50}, *i.e.* the concentration

of test chemical, which reduces NRU by 50%. The PM used to apply the result states that, if the $EC_{50\ without\ UV}$ value divided by the $EC_{50\ with\ UV}$ value is >5, the chemicals should be classed as having phototoxic potential.

The ECVAM validation process involves the following stages (Figure 3).[20,21] First, the *development* of a candidate test is assessed for its readiness for prevalidation. This requires a definition of the scientific purpose of the method, a statement concerning its proposed practical application, a description of the basis of the method, the case for its relevance, and in the case of a test method for the prediction of an *in vivo* pharmacotoxicological endpoint, this should refer to the mechanistic relevance of the test, and to any preliminary evidence supporting the predictive capacity of the test. There should also be an explanation of the need for the method in relation to existing *in vivo* (animal or human) methods (with reference to relevant TGs and legislation) and other non-animal methods, an optimised protocol, including: (a) any necessary standard operation procedures; (b) a specification of endpoints and endpoint measurements; (c) the method for deriving and expressing results; (d) the interpretation of the results in terms of one or more *in vivo* pharmacotoxicological endpoints by means of a PM; and (e) the use of adequate controls, together with a statement about the limitations of the test, and evidence of intra-laboratory reproducibility and, if available, inter-laboratory transferability.

Prevalidation is now widely accepted as an essential second step to optimise the test protocol and maximise its inter-laboratory transferability, before a formal validation study is undertaken. It also involves an evaluation of the performance of

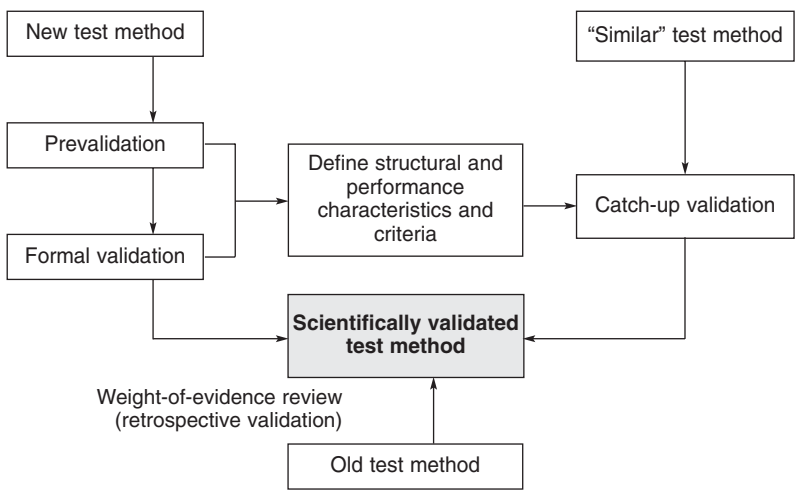


Figure 3 *The ECVAM validation process (from Ref. 20). In the ECVAM validation process, a scientifically validated method is one that has been endorsed by the ECVAM ESAC. If the method is appropriate for chemicals testing, a draft Annex V guideline, incorporating the method, will initially be submitted to the appropriate EC Services and the EU CAs for Directive 67/548/EEC, for consideration for regulatory acceptance and application*

the test, which, in certain circumstances, can be planned as a formal validation study.[13]

A *formal validation study* is characterised by management, test item selection, coding and distribution, data collection and analysis and reporting, which are independent of the laboratories conducting the test. Where such a practical study is not possible (*e.g.* because of a lack of test items backed by sufficient knowledge of sufficiently high quality), or is not considered necessary (*e.g.* because of long experience in the use of the test under evaluation), a weight-of-evidence approach may be appropriate.

An alternative method can only be judged to be valid if two conditions are met: (a) it is *reliable* (*i.e.* it provides results which are reproducible within and between laboratories and over time); and (b) it is *relevant* (*i.e.* it has scientific value and practical usefulness) in relation to its *stated purpose* (*i.e.* its intended application of the procedure in relation to the toxicity of concern). The PM should be defined in advance by the test developer, and in a practical validation study, which involves the blind testing of coded chemicals about which sufficient information is already available, the performance criteria to be met should be laid down in advance by the management team for the study. All laboratory work should comply with the principles of Good Laboratory Practice (GLP) and Good Cell Culture Practice (GCCP).[22]

The accepted indicators of test performance are: *sensitivity*, the proportion of toxic chemicals that are correctly identified; *specificity*, the proportion of non-toxic chemicals that are correctly identified; *positive predictivity*, the proportion of positive chemicals that are toxic; *negative predictivity*, the proportion of negative chemicals that are non-toxic; and *concordance*, the overall proportion of toxic and non-toxic chemicals that are correctly identified. Of course, the predictivity of a test performance is dependent on the relative numbers of toxic and non-toxic chemicals in the test set.

For a PM to be considered at an adequate stage of development, it should: (a) be associated with one or more specific protocols; (b) be associated with a clearly defined pharmacotoxicological endpoint; (c) have its limitations clearly defined; and (d) be associated with an indication of the accuracy of its predictions. For a PM to be considered valid, it should: (a) be assessed with independent data (*i.e.* the test set should be different from the training set) and (b) meet or exceed the criteria for predictive capacity defined in advance by the management team of a validation study.[23]

When a method is developed which is similar to a method which has been validated or is undergoing validation, and which can be shown to match its structural and performance characteristics,[24] the new method can proceed via a fast-track "catch-up validation" study. An example of this is the catch-up validation of the EpiDerm™ test for skin corrosivity, which was sufficiently similar to the equivalent and validated EPISKIN™ test.[25] In other cases, such as where a practical validation study is considered to be unnecessary, as in the case of *in vitro* methods for percutaneous absorption,[8] with which there was great industry experience and for which there was no established *in vivo* procedure, a weight-of-evidence evaluation can be conducted.

4 Problems being Encountered in Implementing the Validation Process

4.1 Unrealistic Expectations

Various pressure groups, and even some politicians, have been promoting demands for a risk-free environment, where there will be no hazardous chemicals or products, and others have promoted the belief that the development of non-animal tests and strategies is easy, or that alternative tests should be accepted as valid, merely because their developers say that they are. At the same time, some toxicologists and regulators claim that animal tests need not be validated, because their relevance can be assumed.

There is a desperate need for effective public education about the nature of hazard and risk, as well about the responsibilities of individuals, corporations and governments to define what levels of risk are unavoidable, acceptable and manageable, and how they should be managed. Human life, as we know, it would come to an end, or would not be worth living, if we lived in a risk-free world. All chemicals are hazardous at certain levels and in certain circumstances – this is often the essential basis of their usefulness. Thus, it is a question of how can risks best be evaluated, what levels of risk are acceptable when likely benefit is weighed against potential cost, and how risks can best be managed, so that they are avoided, or at least minimised.[26]

All new tests, and some established animal tests, should be rigorously and independently validated, because of the seriousness of the consequences which arise if they are used as a basis for making particular decisions, but are subsequently found not to be relevant or reliable for that purpose.

More emphasis should be placed on the fact that proper test development is the key to successful validation, acceptance and application, and that any unjustifiable bias in favour of a method later found to be wanting will rebound on the developers, as well as others along the way, who serve sectorial interests at the cost of honesty and transparency.

4.2 Toxicology versus Toxicity Testing

The science of toxicology is underdeveloped when compared with the science of pharmacology. Worse, toxicity testing has been correctly described as an art, rather than a science, since it has tended to involve subjective judgements based on a generalised checklist, black box approach, whereby animals are given large doses of chemicals, merely to see what happens. The problem is that, where protecting humans is the objective of the exercise, species differences inescapably limit the relevance of the animal data, and the high doses applied (because of fear that something might have happened at a higher dose than at a rational dose) provide data which have little meaning in terms of likely human exposure. As a result, "uncertainty factors" or "safety factors" have been introduced into the risk assessment and risk management process.

The way forward involves the use of human cells and tissues wherever possible, and predictions based on human experience, together with a concerted effort to increase our understanding of the mechanistic basis of toxicological processes. This will be facilitated by scientific developments in cell and molecular biology, and notably in genomics, proteomics and metabonomics, which are already showing

great promise.[18] However, it is also essential that ways are found to use hazard predictions obtained from (Q)SAR systems and *in vitro* tests in an appropriate risk assessment process, rather than trying to use them to provide data which merely match those currently provided by animal tests, which are themselves of doubtful relevance and reliability.

4.3 Test Development

Not enough candidate tests are coming forward for prevalidation, partly because this kind of applied science is unfashionable, and also because insufficient attention is paid to the nature of a test method (test system plus PM), the criteria for test development or the crucial importance of prevalidation and validation – *before* a method is promoted for regulatory application. One solution to this problem is to provide more funding specifically for test development, but experience has shown that scientists find fundamental research easier and more attractive, even when the funding granted to them was intended for test-development purposes.

In view of its pivotal position in relation to validation in the EU (Figure 4), ECVAM is responding to this problem by advising the EC's Directorate-General (DG) for Research on priorities for the development of alternative methods relevant to the chemicals policy,[27] and by establishing a dialogue with industry and academic institutions involved in test development. ECVAM has also coordinated a review of the prospects for alternative methods for cosmetics testing, in the context of the 7th

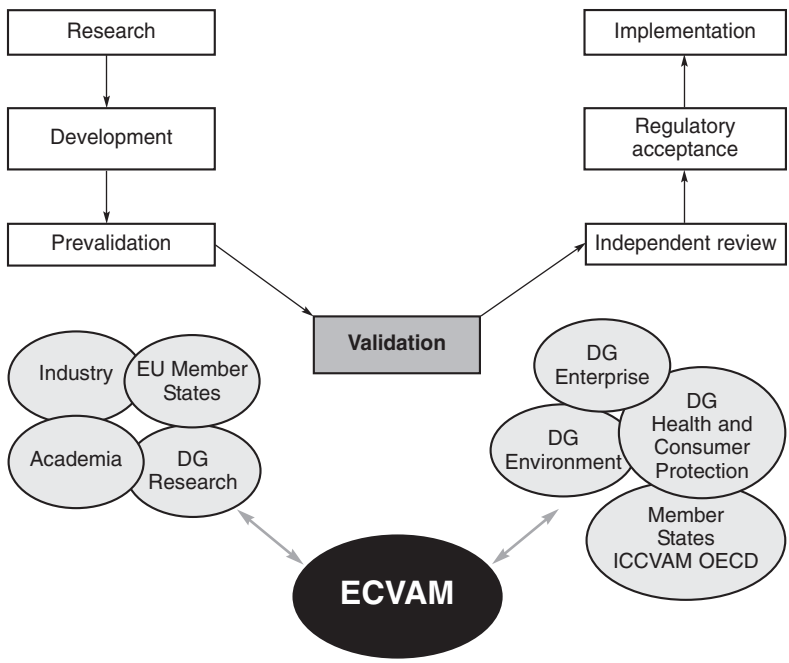


Figure 4 *ECVAM and the validation process in the EU (modified from Ref. 17)*

amendment to the EU Cosmetics Directive, and the establishment of a timetable for phasing out animal testing.[28]

4.4 Complexity and Cost

Problems have arisen because of the complexity of formal validation studies, which have involved too many tests, test chemicals and/or laboratories, leading to high costs and management difficulties. Such problems have often arisen because of political considerations, rather than scientific needs.

These problems can be reduced if the criteria for test selection (and especially for the novelty for a test and the need for it) and for prevalidation are rigorously applied. Test chemicals should be very carefully selected, and there is no reason for having more than three or four laboratories perform the same test in a multi-laboratory study, as was recommended in 1990[9] and in 1995.[11]

ECVAM is trying to make the validation process more readily managed, less expensive and time consuming, as well as more efficient, by introducing a modular approach.[21]

4.5 Lack of Cooperation

Some surprising episodes involving lack of cooperation have been encountered, including the assignment of staff not competent to perform the procedure, failure to adhere to an agreed protocol, sub-contracting work to a laboratory less familiar with the test under evaluation, failure to submit data in the specified format, the withdrawal of an essential test component by its manufacturers during a study, withdrawal from a study in progress because of altered priorities dictated by senior management, and refusal by companies to permit the outcomes of studies to be published (presumably because they were not in line with their expectations).

These problems can be avoided by having formal contracts between the sponsor of the study and all the participating laboratories, in which all the responsibilities are clearly spelt out, by having an appropriate training phase and evaluating the results of a preliminary phase, and by insisting that the results belong to the sponsors, with the freedom to publish them in the peer-reviewed literature. All these features were in place in the successful ECVAM-sponsored study on *in vitro* tests for skin corrosivity.[29]

4.6 Inadequacy of Standards for Comparison

Ideally, the evaluation of a new test method should include a comparison of the results it provides with "gold standard" results from an existing and established test. However, despite many decades of animal testing, it has been very difficult to find sufficient chemicals backed by sufficient *in vivo* data of sufficiently high quality for any of the main types of toxic effect. In addition, some companies have refused to provide *in vivo* data, either to protect their proprietary rights or for fear of reprisals from animal rights extremists. Hence, the selection of a test chemical set by independent experts, including an evaluation of all the animal and human

data, has become one of the most important aspects of any formal validation study.[30]

This is a very serious issue, since the literature (and therefore the minds of those involved) is full of references to "correct" and "incorrect" results and to "false-positive" and "false-negative" predictions, where "correctness" and "truth" are dependent on data from animal tests, which, time and again, have been shown to be unreliable as a basis for predicting effects in humans. A new method that gives precisely the same results as a conventional animal test could not be more reliable than that animal test. If the non-animal approaches are to contribute to improving the protection of human health, they must be better, *i.e.* more relevant and more reliable, than the animal tests they replace.

The direct search for alternatives to animal tests began in earnest in 1979, when animal welfare campaigners dressed as rabbits demonstrated outside the New York headquarters of Revlon in protest against the DET. It later transpired that this would be one of the most difficult topical tests to replace, because the DET score, on which the regulatory classification is based, is contrived and subjective, and there is wide variation among the responses of individual rabbit eyes. Nevertheless, one of the first formal validation studies involved an evaluation of the performance of nine *in vitro* tests in 37 laboratories with 50 chemicals backed by DET scores.[10] Not surprisingly, one would say with hindsight, none of the nine alternative methods could match the classification of the chemicals which resulted from the DET scores, which turned out to be a fool's gold standard.

More than 25 years later, the DET still has not been replaced, although a number of *in vitro* methods have been accepted as a basis for labelling chemicals as severe eye irritants, as a result of a weight-of-evidence evaluation, rather that their ability to match the DET scores or classifications.[31] The tests concerned are the Bovine Corneal Opacity and Permeability (BCOP) test, the Isolated Rabbit Eye (IRE) test, the Chicken Enucleated Eye Test (CEET), and the Hen's Egg Test on the Chorio-allantoic Membrane (HET-CAM). However, this leaves unresolved the problem of distinguishing between non-irritancy, mild irritancy and moderate irritancy, which is especially important for the cosmetic and household products industries.

One suggested way of dealing with this problem is the use of reference standards, *i.e.* carefully selected chemicals representative of a specific group of chemicals and about which sufficient is known, against which the results of tests on chemically similar substances can be measured and expressed. A feasibility study on the application of this approach to *in vitro* tests for eye irritation, gave promising results.[32] A different solution to the problem will be to conduct research aimed at providing a mechanistic understanding of the pivotal events in the development of eye irritation, as a basis for the development of new *in vitro* tests or *in silico* models, but this will take time.

4.7 Bias in the Conduct of Studies and Evaluations

As it has been explained, *independence* from vested interests is crucial at a number of points in the validation process, but this has often been difficult to achieve.[33]

Not surprisingly, *test developers* are biased in favour of the methods they have developed, and this can result in the selective submission of background information when a validation authority is asked to consider the suitability of a method for

prevalidation. Worse, however, is the problem that arises after a validation study, when test developers may not accept an outcome which does not favour their methods, and may even attempt to make direct contact with the authorities to promote their regulatory acceptance. Some candidate tests, which are unlikely to be independently judged to be relevant and reliable for regulatory use, appear to live on indefinitely.[33]

There is reluctance to accept the regulatory use of a method that is patented or that involves components, which are not freely available, and additional problems can arise when such methods are taken off the market by their *commercial suppliers*. This problem can be avoided by the acceptance of test methods which are defined according to structural and performance criteria,[24] rather than accepting specific patented or trade-marked procedures, as is the case, for example, with *in vitro* tests for skin corrosivity.[25,34]

Politicians, as representatives of the general public, can have unrealistic expectations about safety testing and the methods employed, including the prospects for developing valid replacement alternative methods, and this can be reflected in the demands they place on their civil servants. For example, the EC and the European Parliament have repeatedly set unrealistic deadlines for phasing out the animal testing of cosmetic ingredients, and the US Congress has placed an unreasonable burden on US government agencies in their demand for tests for endocrine disruption.

Regulations of many kinds can be used to protect national interests, and especially national commercial interests, and testing regulations and the procedures for the acceptance of new test procedures are no exception. This causes particular difficulties at the OECD level. For example, there are now three reduction/refinement *in vivo* alternatives to the LD50 test,[35] namely, the Acute Toxic Class method, the Up-and-Down procedure and the Fixed Dose procedure, developed and therefore favoured by Germany, the USA and the UK, respectively, whereas it should be possible to have one alternative test incorporating the best parts of all three of these innovations.

The need for independent *peer review* is an important aspect of procedures for acceptance of the outcomes of validation studies, and is central to the ICCVAM and OECD weight-of-evidence approaches. However, it raises the problems of who picks the peers and whether the peers have both the necessary experience and expertise and are sufficiently independent for a review to be acceptable. For example, can employees of regulatory agencies be considered sufficiently independent to take part in the peer review of a method, especially if there is political and administrative pressure within an agency for the acceptance of a particular procedure?[33]

There is only one way of avoiding these problems: all concerned, and especially ECVAM, ICCVAM and the OECD secretariat, backed by the EU member states, the US government and regulatory agencies and the OECD member countries, should insist that the validation principles are upheld and that the validation process is as transparent as possible.

4.8 Publication of the Outcomes of Validation Studies

It is not easy to secure the publication of reports on validation studies in the peer-reviewed literature, as editors are reluctant to publish such long and detailed

reports. Also, it is difficult to find referees with sufficient experience, but who lack bias and have sufficient time to consider such complex manuscripts – and who understand that they should be looking at the need for clarity, rather than suggesting better ways of conducting the validation study or even developing the tests being evaluated!

A small number of journals, including *ATLA* (*Alternatives to Laboratory Animals*) and *Toxicology in vitro*, have recognised this problem, and are willing to take on the burden of publishing reports on validation studies.[25,29]

4.9 The OECD

In 1996, the OECD secretariat worked closely with ECVAM and ICCVAM to agree an internationally acceptable set of validation criteria and principles,[16] then, in 1998, set out to produce a guidance document for its member countries. However, this appears to have been overtaken by events, when pressure from certain regulatory authorities induced the OECD secretariat to organise validation studies on animal tests for identifying potential endocrine disruptors, in ways, which conflicted with the previously agreed criteria and principles. Along the way, and apparently prompted by the endocrine disruptor situation, the OECD held a second workshop on validation,[36] at which attempts were made to avoid the discipline of having a PM, and it was even claimed that animal tests are inherently valid, a view which was challenged at the time and subsequently.[37]

The shortcomings of the OECD-sponsored study on the rat uterotrophic assay for oestrogens and anti-oestrogens have been discussed in detail elsewhere.[38–40] It failed to satisfy the criteria for test development, including the lack of an optimised test protocol and the lack of a clear purpose. It was characterised by poor management of the practical study, biased reporting, and unresolved conflicts in the peer review process, but, amazingly, it now appears that the OECD secretariat have allowed representatives of the OECD member states to agree a draft TG based on the outcome of the study, seemingly in an attempt to make the best of a bad job.

This assay should be considered invalid and should formally be declared as such.[33] Vast resources, including the lives of more than 7000 animals, have already been wasted, but it has not been satisfactorily and independently established that the method could reliably be used to afford greater protection either to human beings or to other animals in the general environment. The need for the test for its stated regulatory purposes should be reconsidered in the light of other test developments, and an optimised test protocol should be re-validated, should a need for it be reaffirmed, but with a different management team and with the minimum number of different participating laboratories.

The OECD secretariat should be advised to focus on updating its current TGs and improving its procedures for the acceptance of validated alternative methods, leaving the validation process to those with the experience and expertise required for doing it properly.[33] In the case of endocrine disruption, more effort should be invested in the development and validation of *in silico* and *in vitro* procedures,[41] as well as in improving the understanding of what "endocrine disruption" actually means in the real world.

4.10 Acceptance

The acceptance of validated test methods should primarily be a scientific matter, but, as it has already been indicated, it suffers from commercial and political implications and biases. For example, a problem not confined to this area is the fact that US scientists, institutions and authorities tend to favour home-grown products and to reject new developments from abroad on a Not Invented Here (NIH) basis. Many regulators feel "more comfortable" with animal test data and do not understand non-animal tests and testing strategies,[37] and many industry scientists and administrators seem to want to settle for a quiet life, seeing the registration of their products as more important that any wide ethical and social issues. Meanwhile, the commercial contract houses and animal and equipment suppliers do very well out of animal testing.

The OECD acceptance procedure is very slow and relies on consensus, which gives an individual member country an opportunity to impose and benefit from its national bias and interests by blocking progress. This particular problem is avoided in the EU, where a majority voting system operates. For a validated method to be considered for acceptance within the EU, ECVAM presents the outcome of a practical validation study or of a weight-of-evidence review to the ECVAM ESAC, on which are represented the other relevant services of the EC, industry, academic toxicology and animal welfare.[42] If the ESAC endorses the positive outcome of a study, this endorsement is published and reported to the other EC services with responsibilities for chemicals, pharmaceutical, cosmetics and other products, and for the protection of consumers and the environment, and to the Competent Authorities (CAs) in the EU member states, as well as to ICCVAM and the OECD (Figure 5). If the validated method is acceptable to the other EC services and to the CAs, a draft TG is produced and a guideline is eventually added to *Annex V* of Directive *67/548/EEC*,

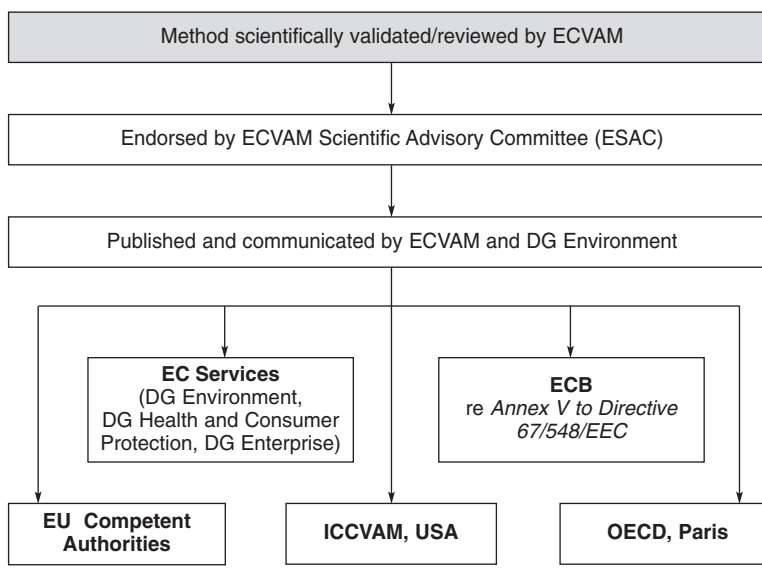


Figure 5 *Acceptance of replacement alternative test methods in the EU*

and the EC and/or one or more member states request that a TG is also drafted as part of the OECD TG programme.

The way forward with regard to acceptance is to encourage a mutual acceptance of satisfactorily validated methods approach, involving the EU, Japan, the USA and the other OECD member countries, and to require a public explanation of the reasons why a particular procedure has been found unacceptable for regulatory use in particular territories. In addition, legislation that says an animal procedure must not be performed if there is a satisfactory and valid non-animal alternative procedure should be rigorously enforced.

4.11 The Validation of In Silico Systems

Up to now, the main focus of this discussion has been on *in vitro* methods, but the proposed REACH system has also prompted much effort in devising computational (*in silico*) modelling systems and chemical-based approaches, including (Q)SAR systems and read-across methods. A recent report from the EC's Institute for Health & Consumer Protection has emphasised that such approaches are needed for scientific, logistical and cost reasons, and also for the sake of expediency, and inferred that they should be considered suitable for use immediately, or very soon, since they have been implemented elsewhere, particularly in the USA, in other strategies for chemicals testing.[43] It is also suggested that the validation process for (Q)SAR models and read across can be more straightforward, less costly and more rapid than that for *in vitro* methods, and that *In terms of time and cost needed for the application of the different elements of intelligent testing strategies, the following ranking can be done from fast and inexpensive to slow and costly: read-across > QSARs > in vitro tests > in vivo tests*.

However, it can also be argued that some caution is necessary, because of inherent problems in validating and accepting *in silico* methods as relevant and reliable for specific regulatory purposes in ways that are analogous to those that have been developed and applied to *in vitro* tests.[44]

It has become clear that how, and even whether, (Q)SAR models should be validated is still controversial, although there have been positive initiatives, discussions and workshops involving ECVAM, the ECB and the OECD.[45–47] Indeed, some individuals in the field take the view that such models are inherently reliable, and that it can be very difficult, if not impossible, to assess their relevance before they are used to predict the toxicities of novel substances, since all the relevant toxicological information was used in their generation. However, if this line of argument is accepted, it leads to what has been called "internal validation" or "self-validation", which runs contrary to the principles of independent and unbiased, *i.e.* "external", validation.[44]

This problem must be resolved, since progress in developing and benefiting from *in silico* approaches must not be impeded. Nevertheless, although it is unlikely that chemicals would be judged to be sufficiently safe for certain purposes on the basis of *in silico* methods alone, resources would be wasted and the benefits of using certain chemicals might not be fully realised, if their use were unnecessarily restricted or unnecessary precautions were taken, on the basis of predictions from methods which were later found to be unreliable or irrelevant.[44]

4.12 Re-Assessing the Currently-Accepted In Vivo Tests

It is often said that the animal tests that dominate the current OECD health effects TGs have not been validated according to the standards insisted on for *in vitro* methods, but the counterargument is that they are valid because of the weight of evidence built up over time. Be that as it may, the Fund for the Replacement of Animals in Medical Experiments (FRAME) was concerned when, in 2003, in its further explanation of the REACH system, the EC included an annex which reproduced the OECD TGs and implied that they will be the basis for satisfying the demands of the REACH legislation.[4] Therefore, as part of the FRAME's ongoing initiatives on the REACH system, we have assessed each TG with respect to its design and its scientific and animal welfare implications, the extent to which it makes use of modern techniques, and its suitability to be used in the REACH system for the testing of large numbers of new and existing chemicals.[48] In addition, the prospects and possibilities for applying the Three Rs (*reduction*, *refinement* and *replacement*) to each of the TGs were discussed. Our investigation showed that there is an urgent need to update the TGs to reflect modern techniques and methods, and to use current approaches for applying refinement strategies to improve the scientific and animal welfare aspects of the procedures used (see Table 3). Improvements can and should be made in all aspects of toxicity testing, from sample preparation, and animal housing, care and feeding, to dose formulation, test-material administration and the histopathological and clinical analysis of tissue samples.

Opportunities for streamlining individual assays are very limited, but testing could be made more efficient by: (a) only undertaking studies that provide relevant data; (b) making greater use of screens and preliminary testing; (c) applying some tests simultaneously to the same animals; (d) using one sex; and (e) eliminating redundant tests. It was concluded that, as they stand, the OECD health effects TGs are unsuitable for use in the EU REACH system, for which potentially very large numbers of laboratory animals will be needed for the testing of a very large number of chemicals.

The solution to this problem is for the representatives of the OECD member countries in the OECD TG programme, especially those from EU member states, to insist that the OECD TGs are updated, and in some cases eliminated, as a matter of urgency. In addition, and in parallel, the EU *Annex V* TGs would have to be reviewed and modified.

4.13 The Validation of Test Batteries and Integrated Testing Strategies

Although it did not seem particularly straightforward at the time, it must be recognised that the phase in which it could be hoped that animal tests would be directly replaced by non-animal tests is coming to an end. Replacement alternative tests for phototoxicity, skin penetration and skin corrosivity are already accepted into regulatory practice, the validation of alternative tests for skin irritation is under way, and what needs to be done about eye irritation and skin sensitisation has been identified, even if how to achieve it is not yet fully known.

We are now left with target-organ toxicity (such as hepatotoxicity and nephrotoxicity), target-system toxicity (*e.g.* to the nervous and immune systems) and multi-faceted

Table 3 *Examples of current OECD TGs, which should be reviewed with respect to their necessity, test design, and animal welfare implications (modified from Ref. 47)*

TG	*Test*
406	Guinea-pig skin sensitisation test – the necessity for this test can be questioned, in view of the availability of the mouse local lymph node assay (TG 429)
424	Neurotoxicity studies in rodents – the scientific justification for such testing is obscure
426	Rat developmental neurotoxicity test – the necessity for this test can be questioned, since the information it provides can be obtained from reproduction toxicity studies
427	Skin absorption, *in vivo* method – the necessity for this new test can be questioned, since it has not been validated and valid *in vitro* approaches are avail able (TG 428)
451	Rat and mouse carcinogenicity study – the need for studies in two rodent species has been questioned for nearly 30 years
478	Genetic toxicology: rodent dominant lethal test – the relevance of this test is questionable
483	Mammalian spermatogonial chromosome aberration test – the justification for this test is questionable, since micronucleus formation in the bone-marrow (TG 475) is a good indicator of germ-line effects
484	Genetic toxicology: mouse spot test – the need for this TG is highly questionable, especially if a TG 475 test is positive
485	Genetic toxicology: mouse heritable translocation assay – the need for this TG is highly questionable, especially as it provides little information and involves large numbers of animals
Draft	Acute dermal photoirritation screening test – the need for this new TG is highly questionable, since the method has not been validated, and the validated and accepted NRU *in vitro* phototoxicity test (TG 432) and the accepted *in vitro* skin absorption procedure (TG 428) are available

toxicity (such as developmental and reproductive toxicity, and carcinogenicity), all of which will require batteries of tests and stepwise testing for acute and chronic toxic effects, which will, in addition, require integrated and intelligent testing strategies.[4,44,49] This problem is being tackled in, for example, the DG Research-funded 35-partner AcuteTox[50] project and the 30-partner ReProTect project.[51]

We have proposed an overall scheme for predictive toxicity testing, whereby the non-animal methods identified and discussed in recent and comprehensive ECVAM reviews could be used in a tiered approach to provide a rapid and scientifically justified basis for the risk assessment of chemicals for their toxic effects in humans (Figure 6).[6] The scheme starts with a compound prioritisation phase and a preliminary risk-assessment process (involving all available information on hazard and exposure). This is followed by testing based on physicochemical properties, and the use of (Q)SAR approaches in conjunction with expert system and biokinetic modelling. The resulting information is then combined with production levels and patterns of use to assess potential human exposure, which, where possible, also includes an assessment of the potential for metabolism in humans and the metabolites, which might be produced. The nature and extent of any further testing should be based

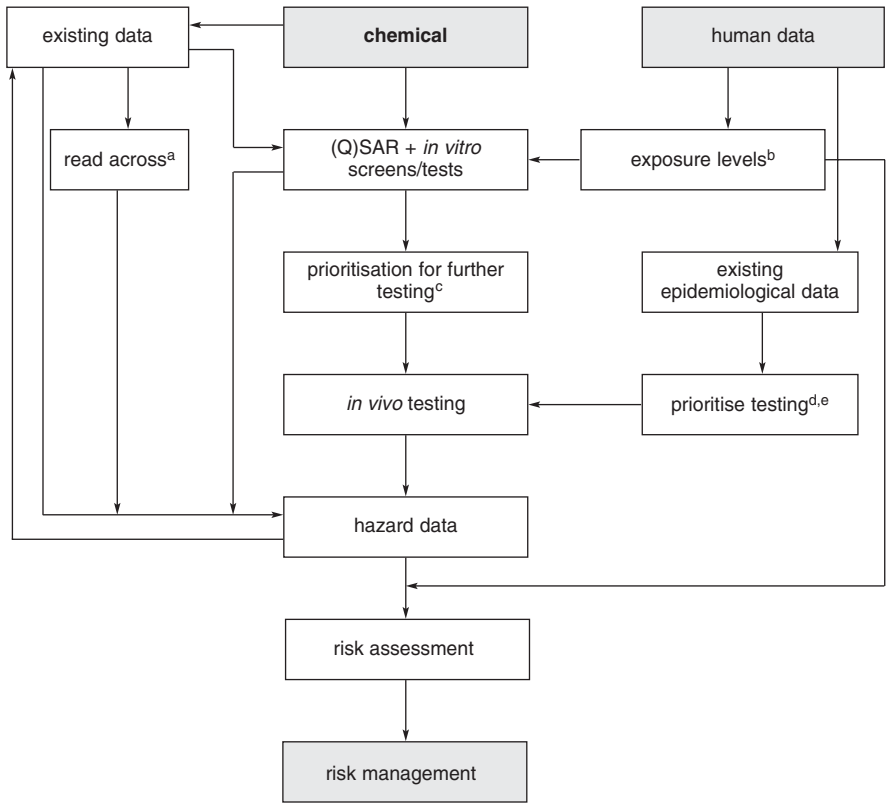


Figure 6 *The FRAME scheme for the integrated testing of chemicals (from Ref. 6) [a]As performed by the Health and Safety Executive; [b]if unavailable, generate before in vivo testing; [c]in a few cases, where in vitro tests have been accepted as replacements; [d]focus on subchronic/chronic, if acute human data are available; [e]prioritisation of testing: low priority and high priority are relative terms, meaning that chemicals of high priority need to be subjected to risk assessment before chemicals of low priority. Other information, including uses and benefits, availability of substitutes, relative toxic potencies and levels of anticipated human exposures, are all factors that should be taken into account when assigning further priorities within these two broad categories*

strictly on the need to fill essential information gaps in order to generate adequate risk assessments, and should rely on non-animal (*in silico* and *in vitro*) methods as far as possible. Animal testing should only be conducted as a last resort, and should require scientific justification on a case-by-case basis.

It must be recognised that there will be an abundance of non-animal tests, focused, for example, on a variety of mechanisms of toxicity, target organs and target systems and types of chemicals and products. Making intelligent and appropriate choices among them will be essential, as will their use in appropriate and case-by-case stepwise and decision-tree strategies – otherwise, what could be useful and effective will merely become impractical and unmanageable.

A read-across approach will also be vital to the success of strategies for implementing the REACH system, whereby inferences are made about the likely toxicity of a chemical from its similarity to another chemical with a known toxicity profile, with respect to physicochemical properties, manufacturing processes, uses and likely human exposure.[6] Specific *in silico* systems and *in vitro* tests could be used on a case-by-case basis, to provide the information necessary to justify a read-across approach to be applied, in order to extrapolate from chemicals about which sufficient data are available to chemicals that are structurally related, but about which much less is known.

These possibilities for intelligent assessments open up enormously challenging, but potentially very rewarding prospects, but it must be emphasised that not only every component of a test battery or integrated testing scheme, but also the reliability and relevance of the overall batteries and schemes themselves, must be subjected to independent evaluation. The manageability and dependability of such evaluations will very much depend on their careful planning, as well as on the experience and integrity of those involved.

4.14 The Need for a New Risk-Assessment Paradigm

The traditional risk-assessment strategy has been to conduct tests to characterise hazard, followed by hazard prediction, then to conduct an estimation of exposure, before proceeding to a risk assessment (Figure 7). However, past experience has shown that, in many cases, the margin of safety for the prospective use of a chemical (the safety factor, or the magnitude by which the no observed adverse-effect level exceeds the known or estimated exposure level), turns out to be so high that it becomes clear, with hindsight, that much of the testing that had already been undertaken was completely

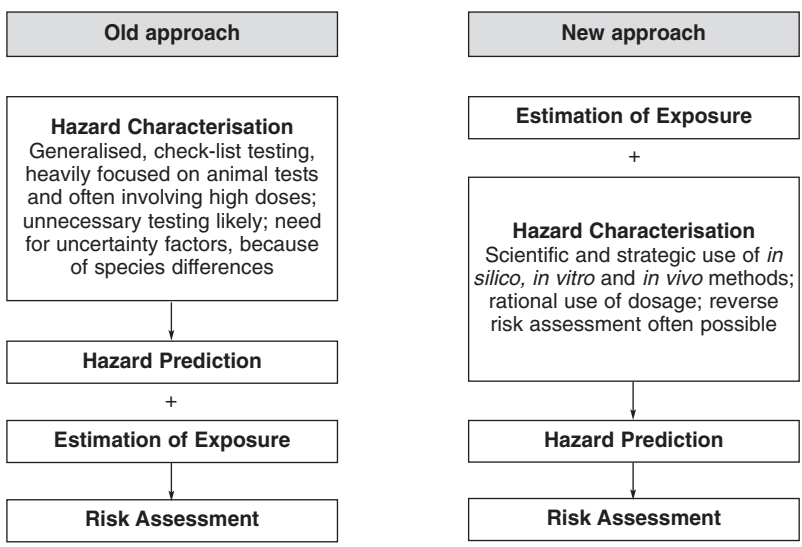


Figure 7 *Old and new approaches to risk assessment*

unnecessary, because it is inconceivable that the minimum toxic dose could ever be experienced in real life.

In some cases, it is possible to avoid such superfluous testing by applying a *reverse risk-assessment* process.[6] This entails selecting testing requirements specifically with regard to the likely expected exposure to the new chemical. In some cases, where the prospective exposure of the new chemical is extremely low, it might be deemed unnecessary for any animal tests to be performed at all.

A more realistic approach to risk assessment in general, which would require less testing and therefore fewer animals and less expenditure, would be to assess likely exposure levels as the first step (Figure 7). Only where potential exposure was scientifically established as a matter of concern, would hazard have to be predicted before assessing risk.

Conducting an estimation of likely exposure as the first step, combined, when appropriate, with read-across and reverse risk assessment, and followed by further tests to characterise hazard only when they were scientifically justified, would greatly reduce the cost of implementing the REACH system, as well as reducing the unnecessary use of animal tests or alternative testing (Figure 7).

Another problem to be faced is that, since the old approach to risk assessment was designed to provide hazard predictions based on dose–response studies in animals, which led, as noted previously, to the need for uncertainty factors because of species differences, there is little point in using animal data or predictions based on them as the gold standards to be met by *in silico* and/or *in vitro* tests, merely so that the same risk-assessment procedures could be applied. As Chamberlain (cited in Ref. 16) pointed out some time ago, ways must be found of using non-animal test data in different kinds of risk assessment. One way forward is to combine a rational estimation of likely exposure, combined with predictions of absorption, distribution, metabolism and excretion, ideally based on a sufficient mechanistic understanding of the likely basis of toxic effects induced by rationally estimated doses. How this could be achieved has been recently discussed by Combes[52] and is to be the subject of a workshop organised by FRAME and Unilever Research.

Solving this problem will be crucial to improving risk assessments and to providing greater protection of human health. To put it another way, questions concerning *relevance* and *reliability* must not be focused on whether an alternative method can produce the same classification for a chemical as that provided by tests in rodents, but on whether the outcome of a new test is credible and meaningful for the target species of concern. Similar remarks could be made about risk assessment and the protection of the environment, but that is another story – tests on fish, *Daphnia* and earthworms do not tell us much about ecosystems.

5 The Need for an Invalidation Process

Bearing all these points in mind, the director of FRAME and I have recently spelt out the need for an *invalidation* process, to parallel and protect the validation process, and to *render null and void* any test method that was judged to be unfit for its stated purpose.[33] An additional advantage would be that valuable resources would not be wasted in attempting to get inherently unfit tests validated or re-validated.

The invalidation process would be independent and transparent, and would be based on an assessment of whether the criteria for test development and validation, summarised above, had been satisfactorily met. The precise details of the process would, of course, depend on the particular circumstances, but a test method could be declared invalid, if the assessment revealed that:

1. satisfactory definitions of the scientific purpose of the method, and of its proposed practical application, were lacking;
2. there was no satisfactory description of the basis of the method;
3. the case for its relevance had not been made, including its mechanistic relevance to an *in vivo* pharmacotoxicological endpoint;
4. there was a lack of preliminary evidence supporting the predictive capacity of the test;
5. there was no satisfactory explanation of the need for the method in relation to existing *in vivo*, *in silico* or *in vitro* methods;
6. no optimised protocol was available, including: (a) the necessary standard operation procedures; (b) a specification of endpoints and endpoint measurements; (c) the method for deriving and expressing results; (d) the interpretation of the results in terms of one or more *in vivo* pharmacotoxicological endpoints, by means of a PM; and (e) the use of adequate controls;
7. there was no satisfactory statement about the limitations of the test; and/or
8. there was insufficient evidence of intra-laboratory reproducibility and inter-laboratory transferability.

Similarly, a test could be judged invalid, if an evaluation of the design, management, performance and outcome of a validation study showed that:

1. the clarity of the defined goals (in relation to the intended usage of the test method[s] and to the PM for converting the data obtained into a meaningful measure of toxicity) was not sufficient;
2. the overall design was not of sufficiently high quality;
3. the independence of the management of the study was not sufficiently well established;
4. the selection, coding and central distribution of chemicals in a multi-laboratory study was not sufficiently independent;
5. the data collection, assessment, statistical analysis and reporting of the outcome were not sufficiently independent;
6. the number and properties of the test materials studied (to cover different structural classes and with well-defined and different activities for the endpoint being predicted) were not sufficient;
7. the quality of the data interpretation (in relation to the PM defined at the start of the study) was not acceptable;
8. the performance of the methods in relation to the goals of the study (the extent to which the methods meet the performance criteria agreed at the start of the study) was not acceptable;
9. the outcome was not reported in the peer-reviewed literature and freely available;

10. all the raw data were not available; and/or
11. the independence of the assessment of the outcome was not satisfactory.

5.1 Types of Invalidation and Candidates for Invalidation

There would be three main types of invalidation: (a) invalidation because a test method was no longer needed, either because it had been superseded by an improved method, or because it produced results similar to those provided by another simpler, more humane method; (b) invalidation because a test method had not been developed according to the logistical requirements discussed above; or (c) invalidation because a test method had been shown not to be scientifically reliable and relevant, according to the validation criteria discussed above. Tests falling into categories (a) and (c) should be declared invalid permanently, and there should be no further proposal made for their potential regulatory use for the same stated purpose for which they have either been shown to be unsatisfactory in category (c) or for which they have been superseded in category (a). Further development would sometimes be possible for methods in category (b).

As it has already been pointed out, a number of existing and proposed animal tests would be candidates for invalidation. In addition, a number of non-animal tests have been proposed, and some are even being considered for validation or are undergoing validation at present, to which our suggested invalidation criteria would apply. Examples of these are to be found among the large number of candidate tests for predicting eye irritation, including the silicon microphysiometer, pollen tube germination and slug tests (see Ref. 53). If these test methods were assessed according to our proposed criteria for invalidation, it would be seen that they either rely on techniques that are costly and not widely applicable, or they involve biological systems and mechanisms that are insufficiently closely related to the process being modelled.

5.2 Implementing Invalidation in Practice

There are various ways in which the invalidation process could be implemented in practice, notably according to the stage reached by a test in terms of its development, validation and regulatory acceptance, as illustrated in Figure 8.

The first stage should be by the test developer as early as possible, and as soon as any of the suggested criteria for invalidation have been fulfilled. However, it is more likely that a method will not be declared invalid until it is formally submitted to a validating body, such as ECVAM, for consideration for prevalidation and validation. ECVAM has issued guidelines for the submission of a proposal for the validation of a new test method (available at ecvam.jrc.cec.eu.int). These have been specified in order to reduce the likelihood that validation studies are commenced with tests that are unsuitable or not ready for validation. ICCVAM has broadly similar requirements for a test being submitted for validation.[54] We suggest that bodies such as ECVAM and ICCVAM should consider ways of formally declaring a proposed regulatory test invalid for its stated purpose(s) if it fails to meet any of its submission criteria and any other relevant criteria that we have identified for invalidation. In

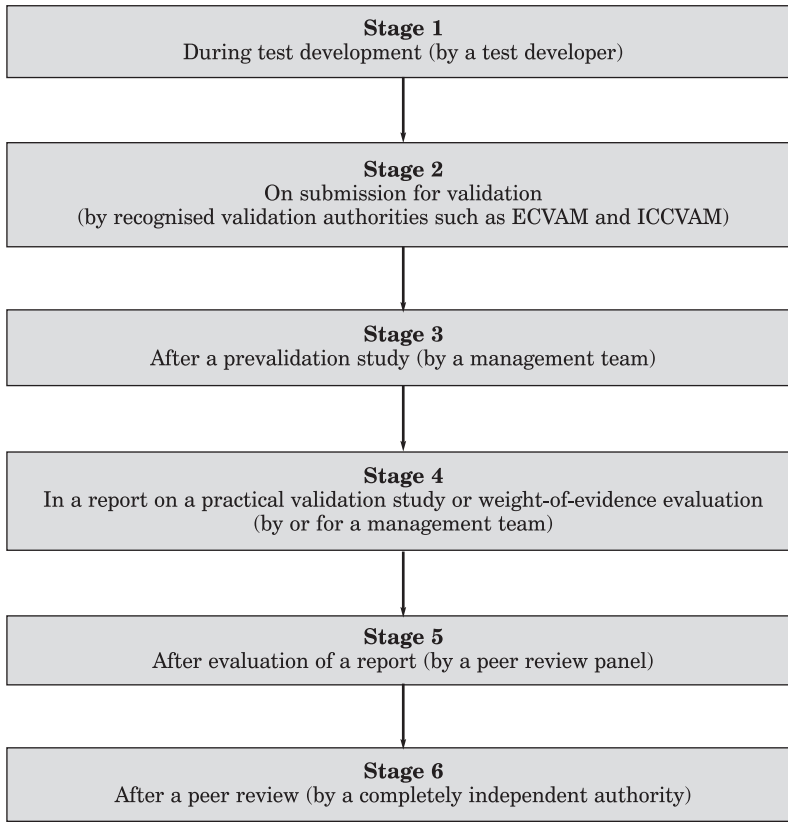


Figure 8 *Principal stages in the test development and validation process, at which a test method which does not satisfy the relevant criteria should be declared invalid (from Ref. 33)*

some cases, this would mean that test developers could re-submit a test method, while in others the test method might be declared to be inherently flawed, in which case it should be made clear that no re-submission of the test would be acceptable.

The third principal stage at which a test could be declared invalid is following its validation. For example, if a test is considered invalid for its stated purpose by a validation peer review process (either following prevalidation or full validation), then the management group and peer review panel concerned, respectively, should make a clear statement to this effect, adding the reasons for the decision, and what – if anything – could be done to allow the test method to be reconsidered for validation.

However, it is also possible that a test method could be considered invalid at a fourth stage, that is, following its formal peer review. For example, as discussed earlier, in the case of the OECD validation study of the uterotrophic assay, the study was not peer reviewed in an independent and transparent manner.[40] Peer review, to ensure that the results obtained show that the relevance and reliability of the test method have been properly established for its stated purpose, can only be achieved objectively if the peer review process is conducted by a panel that is completely independent

of the conduct of the validation study itself, although experts should be co-opted onto the review panel, as required. Where this has not been so, it is not feasible for a test method to be formally declared invalid by the peer review body itself, since this would require the peer review panel to judge itself. It is therefore necessary that the conduct of all peer reviews of validation studies is monitored by a body, which is completely independent of testing, test development, test validation, regulatory acceptance and regulatory usage. Such independent organisations exist, for example, those who provide quality assurance and standardisation services, and the use of these in the area of test development, validation and acceptance should be explored.

6 Concluding Remarks

Independent validation and invalidation procedures are necessary in relation to the development, evaluation and regulatory acceptance of test methods of various kinds, in order to ensure that they are truly and transparently established as scientifically relevant and reliable for their intended purposes, *before* they are used as a basis for making decisions which affect human health and the environment in general – as well as the legitimate commercial interests of chemical producers and product manufacturing companies. These procedures may appear to be time consuming and they may seem to be costly. However, while all possible steps should be taken to make them as efficient and cost-effective as possible, not to conduct them properly opens up the spectre of grave consequences for all concerned.

References

1. European Commission, *White paper on a Strategy for a Future Chemicals Policy, COM(2001)88 final*, European Commission, Brussels, 2001.
2. N. Cardello, *Analysis of the High Production Volume Challenge: Industry Violations and EPA Negligence*, Physician Committee for Responsible Medicine, Washington, DC, 2001.
3. Anon., *U.S. Environmental Protection Agency High Production Volume (HPV) Challenge Program*, EPA, Washington, DC, 2005. Available at: http://www.epa/gov/chemrtk/volchall.htm.
4. European Commission, *Proposal concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), COM(2003)644 final*, European Commission, Brussels, 2003. Available at: http://europa.eu.int/eur-lex/en/com/pdf/2003/com2003_0644en.html
5. K. Van der Jagt, S. Munn, J. Tørsløv and J. de Brujin, addendum to *Assessment of Additional Testing Needs under REACH. Effects of (Q)SARs, risk based testing and voluntary industry initiatives*, JRC Report EUR 21405 EN, Joint Research Centre, Ispra, Italy, 2004.
6. R. Combes, M. Barratt and M. Balls, *ATLA Altern. Lab. Anim.*, 2003, **31**, 7.
7. C. Grindon, N. Bhogal, R. Combes and M. Balls, *TOPRA Reg. Rapp.*, 2005, **2**, 16.
8. C. Westmoreland, in *Alternatives to Animal Testing*, R.E. Hester and R.M. Harrison (eds), RSC, Cambridge, UK, 2005, Chapter 5.

9. M. Balls, B. Blaauboer, D. Brusick, J. Frazier, D. Lamb, M. Pemberton, C. Reinhardt, M. Roberfroid, H. Rosenkranz, B. Schmid, H. Spilemann, A.-L. Stammati and E. Walum, *ATLA Altern. Lab. Anim.*, 1990, **18**, 313.
10. M. Balls, P.A. Botham, L.H. Bruner and H. Spielmann, *Toxicol. In Vitro*, 1995, **9**, 871.
11. M. Balls, B.J. Blaauboer, J.H. Fentem, L. Bruner, R.D. Combes, B. Ekwall, R.J. Fielder, A. Guillouzo, R.W. Lewis, D.P. Lovell, C.A. Reinhardt, G. Repetto, D. Sladowski, H. Spielmann and F. Zucco, *ATLA Altern. Lab. Anim.*, 1995, **23**, 129.
12. L.H. Bruner, G.J. Carr, M. Chamberlain and R.D. Curren, *Toxicol. In Vitro*, 1996, **10**, 479.
13. R.D. Curren, J.A. Southee, H. Spielmann, M. Liebsch, J.H. Fentem and M. Balls, *ATLA Altern. Lab. Anim.*, 1995, **23**, 211.
14. M. Balls and W. Karcher, *ATLA Altern. Lab. Anim.*, 1995, **23**, 884.
15. National Institutes of Health, *Validation and Regulatory Acceptance of Toxicological Test Methods: A Report of the ad hoc Interagency Coordinating Committee on the Validation of Alternative Methods*, NIH Publication No. 97-398, NIEHS, Research Triangle Park, NC, 1997.
16. OECD, *Final Report of the OECD Workshop on Harmonization of Validation and Acceptance Criteria for Alternative Toxicological Test Methods*, OECD, Paris, 1996.
17. Balls, M., *ATLA Altern. Lab. Anim.*, 2002, **30** (Suppl. 2), 5.
18. N. Bhogal, C. Grindon, R. Combes and M. Balls, *Trends Biotechnol.*, 2005, **23**, 299.
19. OECD, *OECD Guidelines for the Testing of Chemicals No. 432: In Vitro 3T3 NRU Phototoxicity Test*, OECD, Paris, 2002.
20. A.P. Worth and M. Balls, *ATLA Altern. Lab. Anim.*, 2002, **30** (Suppl. 2), 15.
21. T. Hartung, S. Bremer, S. Casati, S. Coecke, R. Corvi, S. Fortaner, L. Gribaldo, M. Halder, S. Hoffmann, A. Janusch Roi, P. Prieto, E. Sabbioni, L. Scott, A. Worth and V. Zuang, *ATLA Altern. Lab. Anim.*, 2004, **32**, 467.
22. S. Coecke, M. Balls, G. Bowe, J. Davis, G. Gstraunthaler, T. Hartung, R. Hay, O.-W. Merten, A. Price, L. Schechtman, G. Stacey and W. Stokes, *ATLA Altern. Lab. Anim.*, 2005, **33**, 261.
23. A.P. Worth and M. Balls, *ATLA Altern. Lab. Anim.*, 2001, **29**, 135.
24. M. Balls, *ATLA Altern. Lab. Anim.*, 1997, **25**, 483.
25. M. Liebsch, *ATLA Altern. Lab. Anim.*, 2000, **28**, 371.
26. P. Illing, in *Alternatives to Animal Testing*, R.E. Hester and R.M. Harrison(eds), RSC, Cambridge, UK, 2005, Chapter 1.
27. A.P. Worth and M. Balls, *ATLA Altern. Lab. Anim.*, 2002, **30** (Suppl. 1), 1.
28. C. Eskes and V. Zuang, *ATLA Altern. Lab. Anim.*, 2005, **33** (Suppl. 1), 1.
29. J. Fentem, G.E.B. Archer, M. Balls, P.A. Botham, R.D. Curren, L.K. Earl, D.J. Esdaile, H.G. Holzhütter and M. Liebsch, *Toxicol. In Vitro*, 1998, **12**, 483.
30. M. Barratt, P.G. Brantom, J.H. Fentem, I. Gerner, A.P. Walker and A.P. Worth, *Toxicol. In Vitro*, 1998, **12**, 471.
31. European Commission, *Manual of Decisions for Implementation of the 6th and 7th Amendments to Directive 67/548/EEC on Dangerous Substances*, Updated version of July 2004 (EUR 20519), European Chemicals Bureau, European Commission JRC, Ispra, Italy, 2004.

32. M. Balls, P.G. Brantom, S. Cassidy, D. Esdaile, J. Fentem, M. Liebsch, J. McPherson, U. Pfannenbecker and M. Prinsen, in *Proceedings of the Second International Scientific Conference Organised by the European Cosmetic Industry*, D Clark, S. Linsansky and R. Macmillan (eds), CPL Press, Newbury, 1999, 201.
33. M. Balls and R. Combes, *ATLA Altern. Lab. Anim.*, 2005, **33**, 299.
34. OECD, *OECD Guidelines for the Testing of Chemicals No. 431. In vitro Skin Corrosion: Human Skin Model Test*, OECD, Paris, 2004.
35 . L.Gribaldo, A. Gennari, K. Blackburn, C. Clemedson, A. Deguercy, A. Meneguz, W. Pfaller and I. Ruhdel, *ATLA Altern. Lab. Anim.*, 2005, **33** (Suppl 1), 27.
36. OECD, *OECD conference reaches agreement on principles and criteria for the international regulatory use of testing methods*, OECD, Paris, 2002.
37. M. Balls, *ATLA Altern. Lab. Anim.*, 2004, **32** (Suppl 1B), 755.
38. R.D. Combes, *ATLA Altern. Lab. Anim.*, 2000, **28**, 81.
39. R.D. Combes, *ATLA Altern. Lab. Anim.*, 2003, **31**, 489.
40. R.D. Combes, *ATLA Altern. Lab. Anim.*, 2004, **32**, 111.
41. R. Combes, in *Alternatives to Animal Testing*, R.E. Hester and R.M. Harrison (eds), RSC, Cambridge, UK, 2005, Chapter 3.
42. A.P. Worth and M. Balls, *Altern. Lab. Anim.*, 2001, **29**, 525.
43. Anon., *REACH and the Need for Intelligent Testing Strategies*, EUR 21554 EN, Institute for Health & Consumer Protection, EC JRC, Ispra, Italy, 2005.
44. R. Combes and M. Balls, *ATLA Altern. Lab. Anim.*, 2005, **33**, 289.
45. A.P. Worth and M.T.D. Cronin, *ATLA Altern. Lab. Anim.*, 2004, **32** (Suppl. 1B), 703.
46. A.P. Worth, C.J. van Leeuwen and T. Hartung, *SAR and QSAR in Env. Res.*, 2004, **15**, 331.
47. Anon., *The Report from the Expert Group on (Quantitative) Structure Activity Relationships ([Q]SARs) on the Principles for the Validation of (Q)SARs*, OECD Series on Testing and Assessment No. 49, OECD, Paris, 2004.
48. R.D. Combes, I. Gaunt and M. Balls, *ATLA Altern. Lab. Anim.*, 2004, **32**, 163.
49. D. Knight, in *Alternatives to Animal Testing*, R.E. Hester and R.M. Harrison (eds), RSC, Cambridge, UK, 2005, Chapter 4.
50. Anon., *ATLA Altern. Lab. Anim.*, 2005, **33**, 7.
51. S. Bremer, D. Balduzzi, R. Cortvrindt, G. Daston, B. Eletti, A. Galli, I. Huhtaniemi, S. Laws, G. Lazzard, U. Liminga, J. Smitz, M. Spano, A. Themmen, A. Tilloy and I. Waalkens-Behrends, *ATLA Altern. Lab. Anim.*, 2005, **33**, 391–416.
52. R.D. Combes, *Toxicol. In Vitro*, 2005, **19**, 921–924.
53. C. Eskes, S. Bessou, L. Bruner, R. Curren, J. Harbell, P. Jones, R. Kreiling, M. Liebsch, P. McNamee, W. Pape, M.K. Prinsen, T. Seidle, P. Vanparys, A. Worth and V. Zuang, *ATLA Altern. Lab. Anim.*, 2005, **33** (Suppl 1), 47.
54. W.S. Stokes, L.M. Schechtman and R.N. Hill, *ATLA Altern. Lab. Anim.*, 2005, **30** (Suppl. 2), 23.

In Vitro Testing for Endocrine Disruptors

ROBERT D. COMBES

1 Introduction

1.1 Background

Exposure to natural and synthetic chemicals in the environment, capable of mimicking or blocking the effects of endogenous sex steroid hormones, is generally considered to possess the potential to elicit a wide range of toxic effects in humans, especially reproductive and developmental toxicity and some types of cancer.[1,2] Also, there have been widespread reports of various effects on wildlife being attributed to chemicals.[3–8] Many different types of chemicals can exert deleterious effects on the endocrine system.[9] Much attention is now focused on exogenous substances called endocrine disruptors (EDs), which are thought to induce sex steroid-like effects.[10] It should, however, be noted that, contrary to the situation regarding damage to wildlife, the information supporting a direct link between exposure to chemicals and hormonal disruption in humans is scarce. Indeed, it would seem that such data are limited to the recent reports of breast cancer being due to the use of cosmetics containing certain ingredients, such as *p*-hyroxybenzoic acid esters.[11]

There is no universally accepted definition of an endocrine disruptor, but the following one will be used here: "*an exogenous substance or mixture that alters function(s) of the endocrine system and causes adverse health effects in an intact organism, or its progeny, or (sub) populations*", although other definitions have been used.[12]

1.2 Mechanisms of Action

EDs act by a diversity of mechanisms including receptor binding, altered post-receptor activation, steroidogenesis (modulation of hormone synthesis), perturbation of hormone storage, clearance and homeostasis (Figure 1). These processes can lead to agonistic or antagonistic effects, as well as to the suppression or induction of levels of endogenous hormones, and to other phenomena, such as the release of follicle-stimulating hormone

Issues in Environmental Science and Technology, No. 23
Alternatives to Animal Testing
© The Royal Society of Chemistry, 2006

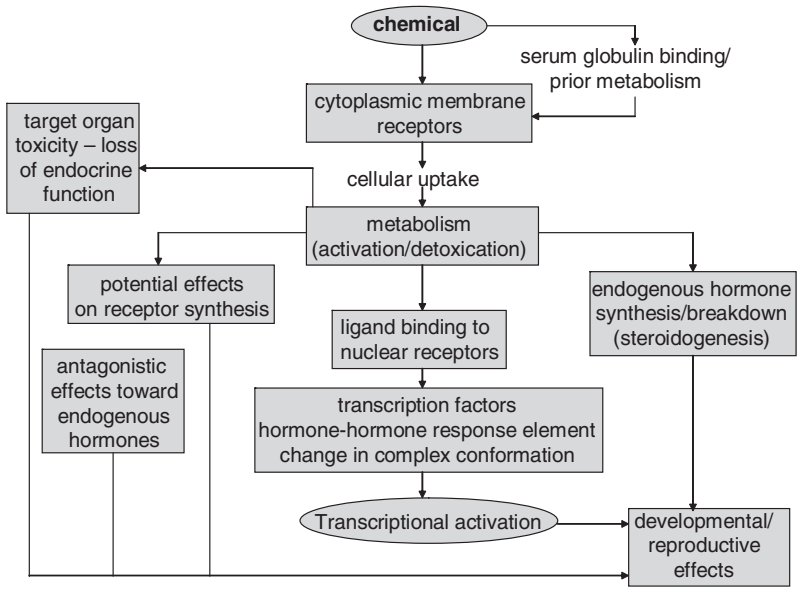


Figure 1 *Main pathways for endocrine disruption*

(FSH), luteinising hormone (LH) or prolactin. Binding occurs specifically to an extensive family of specific nuclear receptors (steroid hormone receptors (SHRs), including the oestrogen, progesterone, androgen, glucocorticoid and mineralocorticoid receptors).[13] These share several structural features including a central highly conserved DNA-binding domain that targets the receptor to specific DNA sequences (hormone response elements (HREs)). The hormone is able to interact directly with the terminal portion of the receptor as this possesses a ligand-binding domain with a hormone-dependent transcriptional activation region. Once binding of the steroid hormone to the receptor protein has occurred, a conformational change results in binding of the DNA to the HRE to form a complex that triggers or suppresses targeted gene transcription.[12,14,15] The conformation of the ligand–receptor complex is therefore crucial for binding to have a subsequent effect on inducing gene expression.[16]

There are at least two subtypes (isoforms) of the estrogen receptor (ER), each encoded by a separate gene, which differ in their tissue distribution and relative ligand-binding affinities for the same hormones.[17] Also, the androgen receptor has a number of isoforms that arise as a result of its differential phosphorylation following synthesis.[18] The arylhydrocarbon (Ah) and pregnane X receptors are also involved in mediating endocrine effects, as shown by the fact that some endogenous steroids induce cytochrome P450,[19,20] and chemicals with dioxin-like activity in pulp-mill effluents are positive in a mouse hepatoma cell line (Hepa1c1c7) possessing the Ah receptor, a dioxin response element and a luciferase reporter gene complex.[21] Indeed, it is clear that the interactions of chemicals with many other cytosolic and nuclear receptors are also potential mechanisms of action of EDs, due to the complex interplay between cellular control mechanisms.

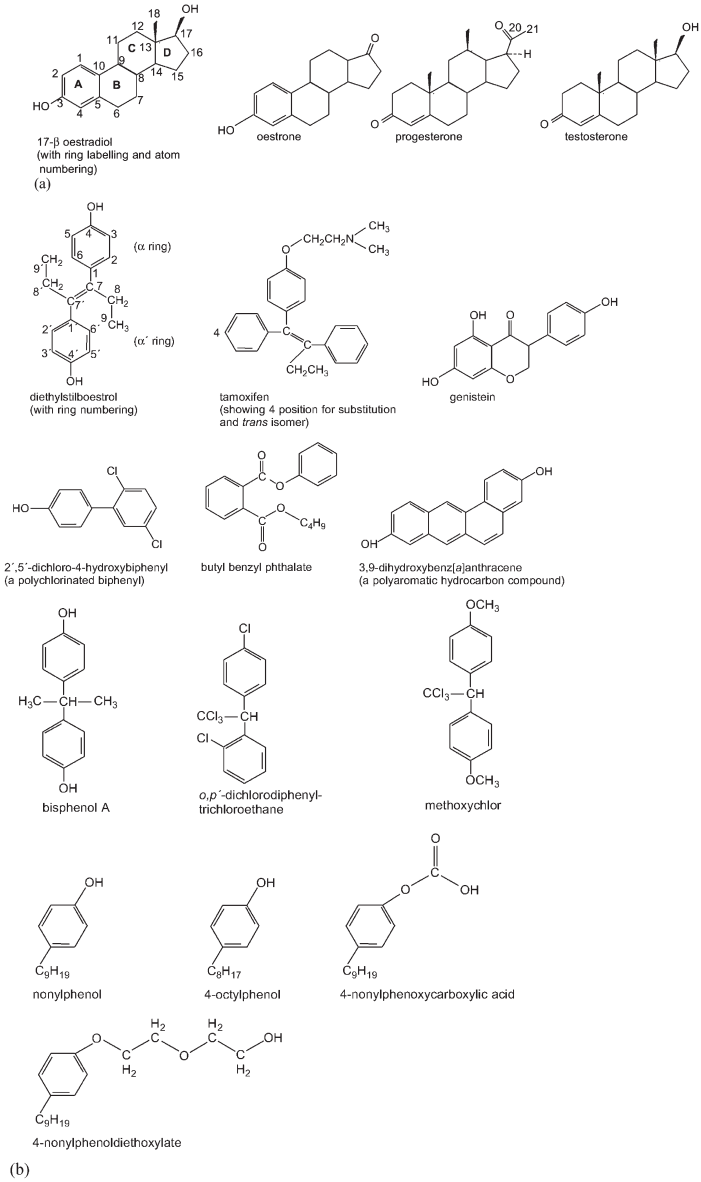


Figure 2 *Structures of (a) endogenous steroid hormones; and (b) various endocrine disruptors*

However, as has been noted elsewhere,[22] all components of the endocrine system are potential sites for hormonal disruption by xenobiotics. Those developing methods for detecting EDs have thus far largely neglected this fact. There is, therefore, a need to investigate the effects of target-organ toxicity that can cause loss of endocrine function. At present, the main way whereby this can be assessed is by using complex animal tests.

1.3 Structural Features of EDs

Natural and synthetic oestrogens exhibit great structural diversity (Figures 2a and b) but many share a phenolic OH group on a small lipophilic molecule of about 200–300 Da. However, the OH substituent is not essential for an ED activity.[23] Thus, endogenous sex steroid hormones, such as oestrogen, testosterone and progesterone are characterised by the presence of a four-ringed fused cyclopentenophenanthrene nucleus possessing a high lipid solubility. Other related substances include the corticosteroids, mineralocorticoids, vitamin D and retinoic acid.

1.4 Legislation for the Testing of EDs

The Food Quality Protection Act of 1996, Public Law 104–170 and Safe Drinking Water Act Amendments of 1996, Public Law 104–182,[24] in the USA requires that chemicals are assessed for their endocrine disrupting potential. Other regions of the world, especially in Japan and Europe, are also introducing legislation for ED testing,[25] and it is incorporated into the new registration, evaluation and authorisation of chemicals (REACH) legislation being developed in the EU.[26] The US Environmental Protection Agency (EPA), the Japanese Ministry of Heath and Welfare (MHW) and the Japanese Ministry of Economics, Trade and Industry (METI) and the OECD are all involved in developing and validating methods for screening of EDs, especially short-term *in vivo* tests.[25,27] For this purpose, the OECD has established a group to manage the validation of three new animal methods (the uterotrophic assay for

Table 1 *Summary of non-animal tests for endocrine disruptors*

Test name	*Endpoint*	*Mechanism of effect*
(Q)SAR	*In silico* modelling of receptor binding	Mathematical relationship between physicochemical properties of ligand and receptor
Receptor binding	Radioactive binding to hormone receptors	Specific binding of ligand to receptor
Cell proliferation	Induction of cell division in ED-sensitive cell lines	Specific binding to receptor, resulting in an induction of expression of genes controlling cell division
Reporter gene assays	Induction of transcriptional activation of regulatory element linked to a gene whose expression is easily monitored	Specific binding to receptor, resulting in an induction of gene expression
Steroidogenesis in Leydig cells (minced testis)	Measurement of testosterone	Susceptibility of Leydig cells to androgen-binding chemicals
Aromatase assay (26)	Aromatase inhibition	Inhibition of oestrogen biosynthesis by aromatase inhibitors

oestrogens and anti-oestrogens, the Hershberger test for androgens and anti-androgens and an enhanced 407 protocol for repeat-dose toxicity to detect agents acting via these and other mechanisms, such as steroidogenesis). The OECD has more recently established a validation management group (VMG) for non-animal assays for detecting EDs, and such a move is not before time.[27] An overview of non-animal testing methods is presented in Table 1.

2 *In Silico* Approaches

2.1 Molecular Modelling of Ligand–Receptor Complexes

Many oestrogens contain one or more phenolic OH groups on a small lipophilic molecule of about 200–300 Da.[28] Removal of the hydroxyl groups usually significantly reduces the binding efficiency. Computer-generated models of ligand–ER interactions have been developed, but such work is limited by the availability of crystal structures of the various SHRs.[29] Molecular models for some hormone receptors have been developed via derivation of crystal structures, for example, the human oestrogen receptor and the human retinoic acid receptor (hRXR).[30,31]

2.2 Identifying Structural Alerts

Several structural alerts (pharmacophores/biophores) and quantitative structure activity or structure activity ((Q)SAR) models are available for predicting the binding affinities of chemicals to the ER.[12,23] However, most commercially available expert systems lack extensive rule bases for detecting EDs.[32] The MULTICASE system identifies a 6 Å unit spacer biophore as being associated with an ED activity.[33]

2.3 (Q)SAR Modelling

Several (Q)SAR models for EDs have been published for ligand binding to the ER.[34–38] Also, (Q)SARs for receptor–ligand binding involving the analogous peroxisome proliferator-activated receptor (rPPAR) have been developed,[31] and these could be useful for modelling ligand interactions with hormone receptors.

The oestrogenicity of 42 structural analogues of 17β-oestradiol was investigated by Wiese *et al*.[39] by using the MCF-7 cell proliferation assay in conjunction with an assessment of ligand binding to the ER. The latter information was analysed by comparative molecular field analysis (CoMFA) to develop (Q)SARs for molecular binding. CoMFA is a method for obtaining (Q)SAR information, based on the separate contributions of shape and electrostatic molecular field lattices of a molecule to its overall biological activity. The approach is now used widely to predict the binding affinities of ligands to various receptor molecules. CoMFA has also been used by Waller *et al*.[36,40] to develop (Q)SAR models for predicting the binding affinities of both oestrogenic and androgenic chemicals. This involved correlating structural differences in molecules with their abilities to compete for binding to the ER. This technique was applied to a series of chemicals in several chemical subsets

for steroid hormones, phenols, phthalates, phytoestrogens, DDT congeners, PCBs, some chemically related pesticides and a series of chemicals related to diethylstilboestrol.

These preliminary studies were used to establish some 3D and physico-chemical characteristics of the ER pharmacophore, especially steric bulk and electrostatic potential, in order to use the model to eventually predict oestrogenicity. It was also noted that the strong ER-binding capacities of steroids containing C-17 hydroxyl groups were closely correlated with the slight positive electrostatic potentials of the H atoms in these substituents. In addition, Gantchev *et al*.[41] used CoMFA to generate (Q)SAR models for predicting oestrogenicity, based on a molecular analysis of a range of substituted oestradiol–receptor interactions. The model was tested with some 27 test molecules, and the predicted results correlated closely with the observed receptor-binding affinities.

However, several of these models have only limited applicability domains as they were developed with small databases for identifying training sets of chemicals. There have been later attempts to rectify this situation. For example, Hong *et al*.[42] developed a decision tree-based model to evaluate 58,000 chemicals for binding to the ER, by using physicochemical information and SARs derived from a training set of chemicals with known binding affinities for the rat ER. The model also involves rejection (exclusion) filters, the first of which is a molecular mass range of <94 or >1000. The second (inclusion) filter is the need for a ring structure and three structural alerts: (a) the steroid ring; (b) the phenolic ring; and (c) the double-ring structure of the diethylstilboestrol skeleton. Filtering is also based on log P (where P is the octanol–water partition coefficient), positive surface-charge area (incorporating molecular shape and electronic distribution) and the breadth of the molecule. This computational approach resulted in a low false–negative prediction rate, and predicted that 80% of the 58,000 chemicals (*i.e.* 46,000) would show negligible binding to the ER.

Tong *et al*.[43] have also used data from a standardised rat ER binding assay to obtain a training set of 232 chemicals with good quality binding data. The dataset contains chemicals covering a wide structural diversity of chemicals that have a wide spectrum of affinity for the ER. The information has been used to generate a series of (Q)SAR models for ER binding.

2.4 Applying (Q)SAR Models to Prioritise Chemicals

The information obtained by Tong *et al*.[43] was used to prioritise chemicals for further testing in a four-phased approach developed by the National Center for Toxicological Research (NCTR) in the USA. In phase I, chemicals with a molecular weight $>94 - >1000$, or those containing no ring structure are rejected. In phase II, three approaches (structural alerts, pharmacophores and classification methods) that included a total of 11 models, are used to make a qualitative prediction. In phase III, a 3D (Q)SAR/CoMFA model is used to make a more accurate quantitative activity prediction. Phase IV involves the use of an expert system to prioritise chemicals according to a set of rules. The different phases are hierarchical and involve complementary methods.[43]

The information generated in this multiphased approach can be applied in a decision-tree flowchart, based on a number of simple 'IF–THEN' rules for binding developed by the same authors as follows: (a) IF a chemical contains no ring structure, THEN it is unlikely to be an ER ligand; (b) IF a chemical has a non-aromatic ring structure, and IF it does not contain an O, S, N or other heteroatoms for H-bonding, THEN it is unlikely to be an ER ligand; otherwise, its binding potential is dependent on other key structural features; (c) IF a chemical has a non-OH aromatic structure, THEN its binding potential is dependent on other key structural features. (d) IF a chemical contains a phenolic ring, and IF it contains any other key structural features, THEN it tends to be an ER ligand; for the chemicals containing a phenolic ring that is separated from another benzene ring with the bridge atoms ranging from one to three, THEN it is likely to be an ER ligand.

However, a subsequent study by Cunningham *et al*.[33] has also suggested that the 6 Å biophore, which was associated with several carcinogens possessing oestrogenic activity, is not widespread among oestrogens. This conclusion was based on the inability of the CASE expert system to detect the biophore frequently in the chemical structures of a wide range of phytoestrogens, oestrogens and anti-oestrogens (frequency of occurrence = 16/79). Thus, although the biophore was found in oestradiol, it was absent from tamoxifen. The precise role that this biophore plays in receptor binding, and the relationship between binding, oestrogenicity and carcinogenicity, remains to be elucidated. Several chemicals containing the biophore, and therefore suspected of being EDs have, however, been identified (Rosenkranz, 1999, unpublished).

3 *In Vitro* Tests for EDs

3.1 Introduction

More than 50 different tests (including those based on subcellular, *in vitro* and animal systems) are documented,[44] and have been variously discussed in several key reviews.[45–50] The possibilities for investigating the reproductive effects of chemicals *in vitro* have been the subject of a workshop.[51] Evaluation of some of these tests is ongoing at the US EPA (http://www.epa.gov/scipoly/oscpendo/edsparchive/standards.htm), and as part of the work of the OECD in validating methods for endocrine disruptors.

The main *in vitro* tests comprise subcellular hormone receptor-ligand binding; the induction of proliferation (mitogenesis) in hormone-responsive mammalian cell lines and transactivation systems in yeast and mammalian cell lines (see Table 1 and also a report by ICCVAM).[52]

3.2 In Vitro Receptor Binding Assays

These assays involve assessing the molecular binding of the hormone to an isolated and purified receptor protein. They are designed to quantify binding to the ER and AR receptors especially, and to measure the competitive inhibition of binding of standard oestrogens.[53,54] Competitive ligand binding assays are used to detect chemicals that directly interact with an endogenous nuclear receptor.[55,56]

3.3 Assays based on Stimulation of Mitogenesis and Gene Expression

Cell-based assays are primarily based on induction of mitogenesis (cell proliferation) in oestrogen-responding cells (*e.g.* the MCF-7 human breast cancer cell line) and hormone-sensitive transcription of reporter genes (transactivation). Genetically engineered mammalian cells or strains of yeast (*Saccharomyces cerevisiae*) are used in which the cells have been transformed (or transfected either permanently or transiently) by introducing vectors bearing DNA sequences for human or mammalian ER nuclear receptors, in conjunction with the respective response elements linked to promoter regions for a reporter gene, together with the reporter gene itself. The transfected nuclear receptor functions as a steroid dependent, transcriptional activator whose induction is detected colorimetrically, by using a suitable substrate.

Mammalian transactivation systems involve the use of luciferase (*e.g.* chemically activated luciferase expression (CALUX®) or chloramphenicol transferase (CAT) reporter genes. Cell line systems for detecting chemicals binding to other receptors exist, for example, the androgen and glucocorticoid receptors.[57–59] The possibility of generating cell lines carrying more than one nuclear hormone receptor for the simultaneous detection of EDs with different activities has been reported by Sonneveld *et al*.[60] The cells are part of a panel of cell lines based on the CALUX® reporter system derived from the human U2-OS osteoblastoma cell line. They possess reporter gene constructs that have oestrogen, androgen, progesterone and glucocorticoid receptors, and exhibit specific sensitivities to extremely low concentrations of EDs that bind to these receptors (pM amounts, with up to 30 times induction of gene expression over respective controls).

Oestrogen-responsive reporter cell lines are being developed for screening EDs,[61] and as a means to identify biomarkers of endocrine disruption based on microarray and proteomics analysis (see below).

3.4 Cell Line Assays for Steroidogenesis and Adrenal Gland Function

Steroidogenesis (endogenous hormone synthesis) is crucial for the functioning of the gonads and the adrenal gland. However, few tests have been developed that take these potential mechanisms of hormonal disruption into account.[62] Cell-line assays are preferable to the minced (sliced) testis assay, as animals are not required.

There are three main types of cell-line assays for steroidogenesis:[63] (a) the mouse Leydig cell tumour cell line (MA-10); (b) the rat Leydig cell tumour cell line (R2C); and (c) the human adrenocortical carcinoma cell line (H295R). This endpoint can also be measured in mouse Leydig cells.[64,65] The H295R assay is receiving the most attention at the moment with respect to further development and validation by the US EPA and the OECD.[62] This is because it possesses all the enzymes in the steroid synthesis pathway (from cholesterol recruitment to the generation of endogenous oestrogens and androgens). The H295R cell line also expresses the *CYP11A*, *CYP17*, *CYP21* and *CYP11B1* isozymes of cytochrome P450 and responds to some known steroidogenesis inhibitors, and also is suitable for assaying cortisol production.[66] Effects on steroidogenesis are also studied by measuring effects of chemicals on

aromatase activity in placental cells, and also by using other cell types in culture, including yeast.[67,68]

3.5 In Vitro Testing for Thyroid Effects

The hypothalamic-pituitary-thyroid (HPT) axis plays a crucial role in the endocrine control of normal development, and as such is therefore an important target for EDs. Such effects can occur in a variety of ways, including interactions of chemicals with the thyroid hormone receptor (TR),[69,70] and there is evidence for the ability of several chemicals, including bisphenol A to act via this mechanism.[71] Hohenwarter *et al.*[72] have developed an *in vitro* test, which detects chemicals by their binding to the TR by inhibition of the proliferation of a thyroid hormone-dependent rat pituitary somatotroph cell line.

Also chemicals can directly affect the thyroid gland itself, and this can be investigated by using *in vitro* methods.[73,74] Lastly, the hormone binding globulin function of the blood can be modulated, thereby affecting the transport of thyroid hormone.

3.6 Specific Tests for Environmental Effects of EDs

A chemical analysis of environmental samples for EDs is beyond the scope of this chapter, but the reader is referred to Scrimshaw and Lester[75] for further information. No fully validated test guidelines for environmental endocrine disruptor testing[76] currently exist, although the fish screening assay and the amphibian metamorphosis assay have both been extensively reviewed[77,78] and are being evaluated by the US EPA (www.epa.gov; see also Refs. 79 and 80).

There are several studies aimed at using non-invasive techniques or non-animal methods. Thus, some *in vivo* biomarker endpoints (*e.g.* plasma steroid concentrations) have been developed for detecting the oestrogenicity of xenobiotics in water samples.[81] Other indicators used include binding to, and activation of, the oestrogen receptor,[82] and the induction of vitellogenin (the major egg yolk precursor protein) production in male and female fish.[83–85] Vitellogenin is oestrogen dependent, and can be assessed *in vitro* by using either primary fish hepatocytes or permanent fish cell lines with hepatoma cells.[86,87]

A reporter gene system in the Rainbow trout cell line RTG-2 is based on the transfection of a reporter gene plasmid consisting of an oestrogen-responsive element fused to a firefly luciferase gene, which induces luciferase expression after binding of an oestrogenic agonist to the oestrogen receptor, and transcriptional activation.[88] In addition, cell lines should be developed for testing from other species, such as carp, medaka, stickleback and zebrafish, to take account of any species specificity in response. Primary hepatocytes from the carp have been used to detect chemicals binding to the oestrogen receptor and inducing vitellogenin.[89] It should also be noted that mammalian cell lines can be generated that have nuclear receptors linked to fish-specific gene expression systems, such as vitellogenin synthesis.[90]

Further studies are required to assess the importance of tests with fish cell lines for screening endocrine disruptors, and in the meantime it is likely that the recombinant yeast assay will remain one of the most widely used ways for screening environmental

samples, as it responds to a wide range of endogenous and foreign oestrogens, and is simple to use.[91–93]

3.7 Using Microarrays in a Toxicogenomic Approach

This approach is based on the premise that physiological, pharmacological and toxicological events ultimately modulate protein composition, and therefore the structure and function of cells and tissues. Toxicogenomics involves comparing gene expression profiles of the entire genome, or a set of selected relevant genes, at either the transcriptional (genomic analysis, by using mircoarrays of oligonucleotides corresponding to gene sequences) or translational (proteomic analysis, by 2D electrophoretic patterns of total cellular protein) levels after exposure to a test chemical with those obtained with expression profiles specific to exposure to known toxicants.[94]

There have been several reports on the application of toxicogenomic and proteomic analyses to the detection of EDs, and the expression levels of a number of genes have been shown, or are expected, to be modulated in specific ways by a number of EDs[95] (see also http://dbZach.fst.msu.edu). Terasaka *et al*.[96] have developed a DNA microarray comprising 172 genes that were shown to be oestrogen responsive in previous studies with MCF-7 cells. These genes were classified into four groups according to their associations with cellular function: (a) tumour genes; (b) growth genes; (c) ion binding and transporter genes; and (d) other genes (for certain structural and neuronal proteins). The microarray was then used to analyse the effects of a number of naturally occurring and synthetic chemicals with results that were highly predictive of effects seen in other assays for EDs. Importantly, the microarray proved to be more sensitive than conventional tests to the effects of some EDs.

4 Discussion

4.1 Advantages and Limitations of In Vitro Methods

These *in vitro* tests are extremely useful as screens to prioritise chemicals for further testing, but like all *in vitro* methods they have certain advantages and limitations.[97] They are generally very sensitive to a wide range of natural and synthetic EDs; for example, there have been reports that nanomolar concentrations of oestradiol are detectable in the MCF-7 cell proliferation assay and picomolar concentrations in several other test systems. The methods can be used to rank chemicals according to their relative potencies, and both agonistic and antagonistic effects can be detected and distinguished relatively easily. In principle, it should be more straightforward to investigate the possibility of synergistic or antagonistic activity between different chemicals in mixtures than by using *in vivo* methods, although there have been some problems reported.[12] Strains of indicator cells have been isolated that express all the known relevant hormone receptors and also both subtypes of the ER.

Lastly, most of the strains of indicator cells are easy to culture and to genetically engineer and normally have no hormone receptor background. The generation of dose–response curves is also usually easier *in vitro* than *in vivo*, although with EDs

this has been studied more often with *in vivo* tests to investigate the controversial so-called 'low-dose' effect where potency can decrease with the increasing dose.[98]

However, *in vitro* methods have a number of disadvantages. Important intra- and inter-laboratory differences in data have been obtained, for example, when using the MCF-7 cell proliferation assay[12] with the results depending on subclone, protocol, culture conditions (*e.g.* the absence of hormone-binding substances such as serum) and also time-dependent specific effects have been obtained. Moreover, responses *in vitro* can be non-specific in the sense that some cell lines become oestrogen independent with repeated subculturing. In addition, while several of the indicator cells used, retain some residual metabolising capacity, this is variable, often low and there can be problems adding exogenous metabolising fractions (discussed in more detail below).

False negative data can arise due to the non-specific interaction of potential EDs with tissue culture medium components such as serum, and/or lack of cell permeability (*e.g.* yeast). In common with other assays involving genetically modified cell lines, the results from using the transcriptional gene activation systems will be affected by gene copy numbers and plasmid stability. Recently, Wilson *et al*.[99] developed a T47D human breast cancer cell line that stably expresses a luciferase reporter gene system and both ER subtypes, and which was responsive to oestrogen agonists and antagonists. The cell line was shown to detect subnanomol concentrations of various oestrogens and anti-oestrogens, but was insensitive to the potent glucocorticoid receptor agonist, dexamethasone (used as a control to show specificity of the assay for oestrogens). The authors consider the assay adaptable for high-throughput screening.

The mechanistic basis of some of the assays is obscure and the relevance of gene activation, resulting from receptor binding, to phenotypic changes (*e.g.* the induction of mitogenesis) needs further work before prediction models can be developed for their formal validation. Validation will also require the development of standardised protocols and standard operating procedures (SOPs) for the tests, many of which are so far lacking. However, the relative lack of interest in validating these *in vitro* methods is more likely to be due to the fact that the above-mentioned limitations are considered to relegate the use of the tests to one of providing indicative, rather than definitive data, *in vivo* assays being considered necessary for the latter. In this regard, it might be possible to detect the potential non-genotoxic carcinogenicity of some EDs *in vitro* by using cell transformation assays.[100]

4.2 Incorporating Metabolism into In Vitro Assays

The omission of metabolism from *in vitro* test methods for endocrine disruptors is a serious problem and has been reviewed elsewhere.[101] Morohoshi and co-workers[102] have recently shown that the addition of a post-mitochondrial supernatant (S9 fraction) to several *in vitro* assays for EDs can modulate their activity substantially. There are compelling reasons why metabolism should be incorporated into the test systems: (a) hormonally active chemicals, including some naturally occurring EDs, are susceptible to metabolism *in vivo*, and some are activated; (b) many chemicals, including known or potential EDs, require metabolic activation to exert other toxicities; and (c) metabolising systems have been shown to modulate ED activity substantially.

However, the use of exogenous metabolism in *in vitro* assays for EDs can be problematical, due to the fact that several potential EDs are known to non-specifically bind to protein in the enzyme fractions. Nevertheless, such problems could be overcome by using cells that have been rendered metabolically competent, instead of adding subcellular enzyme fractions to the test system. Moreover, a cellular metabolising system would be more likely to express both phase 1 and phase 2 metabolism, instead of just the former, as happens with commonly used exogenous systems, to reduce the chances of preferential activation occurring *in vitro.* This is particularly important as, although phase 1 ring-hydroxylated products would be expected to be hormonally active (based on (Q)SAR studies discussed earlier), they would also act as substrates for conjugation. The ideal way of taking account of the need for metabolism in *in vitro* assays for EDs would be to generate multiply expressing genetically engineered mammalian cell lines containing steroid hormone nuclear receptors, their response elements and reporter genes, together with genes expressing specific metabolising enzymes all within the same cellular environment. The possibility of developing such strains should seriously be investigated. Instead of taking account of the need for metabolism in the above ways, a further possibility is to identify, synthesise and test the principal suspected or known metabolites of a parent test chemical. This approach was adopted by Freyberger and Sholz,[54] who showed that metabolites of testosterone and vinclozolin had different activities in a cytosolic AR-binding assay, compared with their respective parent molecules. However, the possibilities for testing metabolites are severely limiting as they depend on knowing what metabolites to test, and on the availability of suitable samples. However, it is a way of avoiding the potential technical problems mentioned above.

4.3 Published Testing Strategies for EDs

The US EPA's Endocrine Disruptor Screening and Testing Advisory Committee's (EDSTAC) tier I testing scheme (Figure 3) and the OECD testing strategy (Figure 4)[103] involves the use of both *in vitro* and *in vivo* screens. While much effort is being expended in developing and validating the methods for these schemes, it is most important that non-animal test methods are developed and validated as rapidly as possible. This is so that they can be part of an integrated testing battery, and used first in the test battery for prioritising chemicals for further testing, and also to facilitate the interpretation of data from animal tests.[104]

Very large numbers of animals would be required for endocrine disruptor testing, should the current EPA and EU testing proposals go ahead. The actual numbers required would depend on the sensitivity (ability to correctly identify hazardous chemicals) and specificity (ability to correctly identify non-hazardous chemicals) of the testing methods that will be used.[105]

4.4 The Potential Role of Non-animal Tests for EDs

FRAME has produced a decision-tree testing scheme for endocrine disruptors (Table 2). This scheme makes the maximum use of non-animal (*in silico* and *in vitro*) approaches in a tiered approach, in which positive results in these assays at

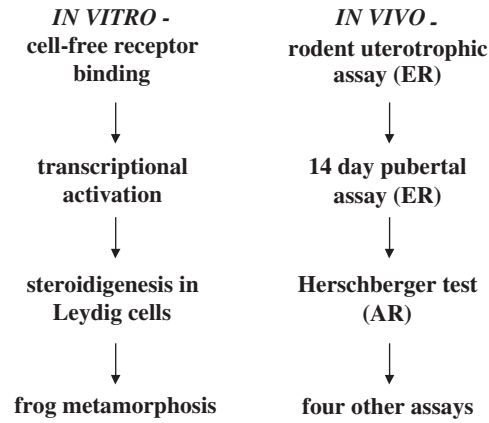


Figure 3 *The EDSTAC tier I testing strategy*

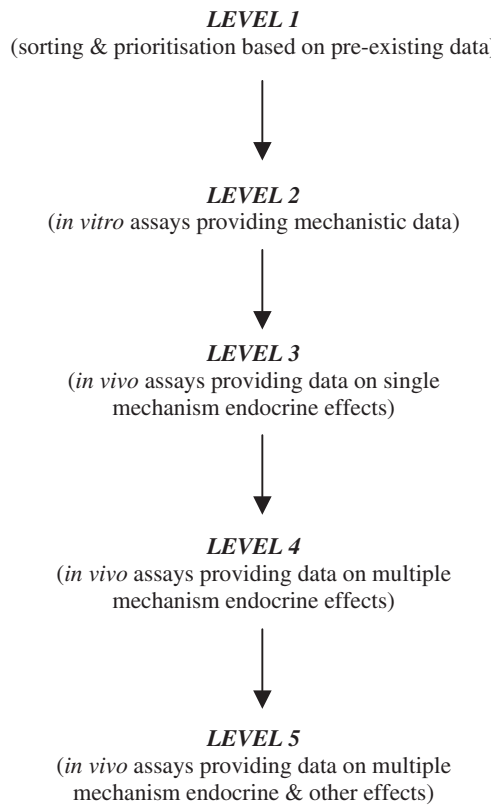


Figure 4 *The OECD tiered testing scheme for endocrine disruptors*

Table 2 *Suggested decision tree integrated testing strategy for detecting endocrine disruptors affecting human health*

Step	*Questions and commentary*	*Command*
1	Are there existing human and/or animal data to suggest that the substance is an endocrine disruptor? (These data include results from previous developmental and reproductive studies including *in vitro* embryotoxicity assays, acute toxicity studies, repeat dose/chronic studies, epidemiology studies, occupational exposure, *etc*.)	
	If data quality and quantity are adequate for C&L and/or RA	No further testing
	If not	Step 2
2	Can predictions of likely metabolites and their toxicities be made using existing *in silico* models and data from *in vitro* methods (non-validated*)? [Prediction of metabolites and their toxicity can be used in PBPK studies in step 7]	
	If yes	Step 3
	If not, undertake metabolism screening using *in vitro* methods	Step 3
3	Can predictions of receptor binding be made via *in silico* methods? (*e.g*., nuclear hormone receptors such as ER and AR; cell-surface receptors such as thyroid and glucagon)	
	If yes, predictions of nuclear receptor binding can be used in a weight of evidence evaluation	Step 8
	If not, or more testing is required	Step 4
4	Perform *in vitro* subcellular competition in ligand-binding studies (being validated) – see review on alternatives	Step 5
	[Results from competitive ligand-binding studies can be used in a weight of evidence evaluation.]	In step 8
5	Perform *in vitro* cell proliferation assays in oestrogen/androgen responsive cell lines (non-validated)	Step 6
	[Results from cell proliferation assays can be used in a weight of evidence evaluation.]	In step 8
6	Perform further *in vitro* tests such as: transactivation/reporter gene assays containing cloned nuclear receptors (being validated); thyroid receptor binding study (in-house test); and tests for steroidogenesis, such as the aromatase assay and H295R assay (both non-validated) – see review on alternatives	Step 7
	[Results from further *in vitro* tests can be used in a weight of evidence evaluation. In step 8]	
7	Perform PBPK and biokinetic modelling (non-validated) Tissue concentrations from PBPK modelling could be used to re-test positive results in the most sensitive of the above *in vitro* tests to give a more relevant result	Step 8
	[Results from PBPK modelling can be used in a weight of evidence evaluation. In step 8]	
8	Perform a weight of evidence evaluation on all data so far (including any data from step 1, which was not adequate enough to require waiving of further testing)	

(*Continued*)

Table 2 (*Continued*)

Step	*Questions and commentary*	*Command*
	If weight of evidence evaluation can adequately show the substance is an endocrine disruptor or not, perform C&L and/or RA	No further testing
	If not	Step 9
9	Perform short-term *in vivo* studies, *e.g*. uterotrophic assay, Hershberger assay, enhanced 28d repeat dose study (all in the various stages of validation)	
	If results from short-term assays are positive, perform C&L and/or RA	No further testing
	If not	Step 10
10	Perform a new weight of evidence evaluation on all data so far (including data from ADME and PBPK modelling to interpret negative results)	
	If weight of evidence evaluation can adequately show the substance is an endocrine disruptor or not, perform C&L and/or RA	No further testing
	If not	Step 11
11	Perform a multi-generation reproductive toxicity study (OECD TG 416), but only as a last resort if accurate quantitative data are required for RA. Use results to perform C&L and/or RA	End

**In silico* models for metabolism prediction are listed in the review of alternatives and include expert systems such as METEOR, COMPACT and TIMES.[93] *In vitro* methods for metabolism include batteries of engineered cell lines, the use of specific inhibitors, inducers or anti-bodies for metabolism and the addition of specific co-factors to exogenous metabolising systems.

each stage trigger a decision to either undertake classification and labelling and/or some form of risk assessment. If this is impossible to do, then it is recommended that further testing is conducted. The scheme comprises tests that are either in the process of being validated or which are under development, and thus its usage for regulatory testing is currently not possible. However, it highlights the urgent need for the OECD and the US EPA to accelerate their efforts to validate these non-animal methods.

A major limiting factor when developing and proposing non-animal approaches for the detection of EDs is that many cytosolic and nuclear receptors are known to be involved in mediated effects that can perturb the hormone system, and it is impossible to be sure that all of these potential mechanisms have been covered even when using a system that involves detection of binding to a panel of these receptors. Moreover, the scheme (Table 2) takes account of the need to assess disrupting potential by steroidogenesis. However, at this stage of knowledge concerning potential mechanisms of action of EDs, it is impossible to devise a non-animal testing strategy that would allow confidence in negative data obtained at any stage of testing. However, this is the same with the other *in vivo* assays under development and validation, although at least the enhanced 28d test should detect chemicals that exert their effects by mechanisms other than receptor binding.

4.5 Concerns about Using Animal Screens and Potential Ways to Avoid them

The proposed animal tests are inherently unsuitable for screening purposes in any case, since they involve the use of relatively large numbers of animals and severe procedures, including surgery (ovariectomisation and castration), making their use unacceptable from an animal welfare perspective.

The uterotrophic and Hershberger assays effectively assess whether a substance is capable of binding to the oestrogen and androgen receptors, respectively, *in vivo* in the target tissue of interest. The scheme in Table 2 incorporates a step involving PBPK modelling to be applied in conjunction with the *in vitro* assays proposed. This is so that target tissue concentrations can be predicted from doses usually applied during *in vivo* testing, so that these can then be used as test concentrations in further *in vitro* assays in which positive data were previously obtained.[106] In this way, it is hoped that both the uterotrophic and Hershberger assays could be avoided.

4.6 The Validation of In Vitro Tests for EDs

Validation involves the transparent and independent establishment of the reliability (reproducibility) and relevance (predictivity) of a test for a specific purpose, and is conducted according to international criteria.[26]

In view of the complexity of many of the assays for EDs, and also the wide range of possible mechanisms of action of such chemicals, the developing and validation of these tests needs very careful thought.[107] However, the following non-animal tests are being considered for validation, or are in the process of being validated variously by the US EPA, METI (Japan), the OECD and ECVAM: (a) (Q)SAR models for ER binding; (b) transcriptional activation assays; (c) ER binding assays; (d) the use of fish cell lines; and (e) steroidogenesis and aromatase inhibition assays (involving the use of Leydig cells and the H259 cell line. Most of these activities are now being coordinated by the OECD VMG for non-animal tests. However, it is unfortunate that there is a lack of planning, and studies are not being undertaken completely according to ECVAM/ICCVAM international criteria, especially with respect to validating (Q)SAR methods.[108]

5 Conclusions

While there is widespread agreement that chemicals in the environment can elicit adverse effects on the hormonal systems of wildlife, the evidence for the potential effects of such chemicals on humans is less well established. In fact, there is a paucity of information on human exposure levels to potential EDs and on the causal links between such exposures and adverse human health effects. Notwithstanding this, there has been a recent call for urgent action by many chemists and toxicologists (the so-called Prague Declaration) for the testing of chemicals for their endocrine disrupting potential. However, interestingly, a recent meeting on risk assessment and EDs concluded that more data concerning the bioavailability of chemicals that could be EDs in animals and humans are required to evaluate any risks that such agents could pose.[109]

In addition, due to the wide range of mechanisms whereby EDs can potentially act, there are difficulties in modelling their effects *in vitro* on the hormonal systems of animals, and particularly of humans. There are consequently uncertainties over the relevance of many of the *in vivo* and *in vitro* tests thus far developed and the hazard data they produce. There is, however, a range of non-animal methods that can be used for screening, including (Q)SAR modelling, subcellular ligand-binding assays and a variety of tissue culture methods that are responsive to chemicals that act via several of the known mechanisms of action of EDs. It might also be possible to apply these methods to provide useful data for risk assessment if they were to be used in conjunction with PBPK modelling to allow the testing of realistic target organ doses of chemicals, instead of using short-term *in vivo* screens for receptor binding and steroidogenesis.

The regulatory testing of EDs has important implications for animal welfare as many tests will be required, involving large group sizes and often complicated and stressful procedures. Therefore, the further development and validation of these non-animal approaches is urgently required so that they can be used to optimum effect as part of an integrated testing strategy, according to the decision-tree system proposed in this article. Lastly, much better co-ordination of the ongoing validation studies is required, and they need to involve the use of methods that take account of the need for metabolism.

Acknowledgement

Grateful thanks are due to Dr. C. Grindon for help in the preparation of this manuscript.

References

1. D. Neubert, *Regul. Toxicol. Pharmacol.*, 1997, **26**, 9.
2. F. Eertmans, W. Dhooge, S. Stuyvaert and C. Comhaire, *Toxicol. In Vitro*, 2003, **17**, 515.
3. S. Gimeno, H. Komen, S. Jobling, J. Sumpter and T. Bowmer, *Aquatic Toxicity*, 1998, **43**, 93.
4. S. Jobling, M. Nolan, C.R. Tyler, G. Brighty and P. Sumpter, *Environ. Sci. Technol.*, 1988, **32**, 2498.
5. S. Jobling and J.P. Sumpter, *Aquatic Toxicity*, 1993, **27**, 361.
6. D.G.J. Larsson, E.R. Adolfsson, J. Parkkonen, M. Pettersson, A.H. Berg, P.E. Olsson and L. Forlin, *Aquatic Toxicity*, 1999, **45**, 91.
7. D.G.J. Larsson, H. Hallman and L. Forlin, *Environ. Toxicol. Chem.*, 2000, **19**, 2911.
8. C.E. Purdom, P.A. Hardiman, V.J. Bye, N.C. Eno, C.R. Tyler and J.P. Sumpter, *Chem. Ecol.*, 1994, **8**, 275.
9. C. Atterwill and J.D. Flack, *Endocrine Toxicology*, Cambridge University Press, Cambridge, UK, 1992.
10. P.W. Harvey, K.C. Rush and A. Cockburn, *Endocrine and Hormonal Toxicology*, Wiley, Chichester, UK, 1999.

11. P.W. Harvey and P. Darbre, *J. Appl. Toxicol*., 2004, **24**, 167.
12. R.D. Combes, *ATLA*, 2000, **28**, 81.
13. M. Warner, S. Nilsson and J.-Å. Gustafsson, *Curr. Opin. Obstet. Gynecol.*, 1999, **11**, 249.
14. N. Weigel, *Biochem. J.*, 1996, **319**, 657.
15. M. Jacobs, *Pesticide Action Briefing Paper 4*, 2001, January. Website: http://www.pan-uk.org (accessed 14.5.04).
16. S.P. Bradbury, O.G. Mekenyan and G.T. Ankley, *Environ. Toxicol. Chem.*, 1998, **17**, 1552.
17. K.C. Fertuck, S. Kumar, H.C. Sikka, J.B. Matthews and T.R. Zacharewski, *Toxicol. Lett.*, 2001, **121**, 167.
18. H.Y. Wong, J.A. Burghoorn, M. van Leeuwen, P.E. de Ruiter, E. Schippers, L.J. Blok, K.W. Li, H.L. Dekker, L. de Jong, J. Trapman, J.A. Grootegoed and A.O. Brinkmann, *Biochem. J.*, 2004, **383**, 267.
19. T. Steimer, *Steroid Hormone Metabolism*, Geneva Foundation for Medical Education and Research, Geneva, Switzerland, 2003. Website: http://www.gfmer.ch/ (accessed 14.5.04).
20. E. Mikamo, S. Harada, J.-L. Nishikawa and T. Nishihara, *Toxicol. Appl. Pharmacol*., 2003, **193**, 66.
21. T.R. Zacharewski, K. Berhane, B.E. Gillesby and B.K. Burnison, *Environ. Sci. Technol.*, 1995, **29**, 2140.
22. P.W. Harvey and I. Johnson, *J. Appl. Toxicol.*, 2002, **22**, 241.
23. R.D. Combes, *Toxicol. In Vitro*, 2000, **14**, 387.
24. Anon., *Endocrine Disruptor Screening Program*, US Environmental Protection Agency, Washington, DC, USA, 2001. Website: http://www.epa.gov/scipoly/oscpendo/index.html (accessed 14.5.04).
25. R.D. Combes, *ATLA*, 2003, **31**, 489.
26. M. Balls, *This Volume*, 2006.
27. T. Seidle, *ATLA*, 2004, **32S1**, 669.
28. R.M. Blair, H. Fang, W.S. Branham, B.S. Hass, S.L. Dial, C.L. Moland, W. Tong, L. Shi, R. Perkins and D.M. Sheehan, *Toxicol. Sci.*, 2000, **54**, 138.
29. M.N. Jacobs and D.V.F. Lewis, *Proc. Nutr. Soc.*, 2002, **61**, 105.
30. W. Bourguet, M. Ruff, P. Chambon, H. Gronemeyer and D. Moras, *Nature*, *London*, 1995, **375**, 377.
31. D.F.V. Lewis and B.G. Lake, *Toxicol. In Vitro*, 1998, **12**, 619.
32. R.D. Combes and R.A. Rodford, in *Predicting Chemical Toxicity and Fate*, M.T.D. Cronin and D.J. Livingstone (eds), CRC Press, Boca Raton, FL, USA, 2004, 193.
33. A.R. Cunningham, H.S. Rosenkranz and G. Klopman, *Proc. Soc. Exp. Biol. Med.*, 1998, **217**, 288.
34. W.D. Tong, R. Perkins, R. Strelitz, E.R. Collantes, S. Keenan, W.J. Welsh, W.S. Branham and D.M. Sheehan, *Environ. Health Perspect.*, 1997, **105**, 1116.
35. W.D. Tong, R. Perkins, L. Xing, W.J. Welsh and D.M. Sheehan, *Endocrinology*, 1997, **138**, 4022.
36. C.L. Waller, T.I. Oprea, K. Chae, H.K. Park, K.S. Korach, S.C. Laws, T.E. Wiese, W.R. Kelce and L.E. Gray Jr., *Chem. Res. Toxicol*., 1996, **9**, 1240.

37. T. Ghafourian and M.T.D. Cronin, *SAR (Q)SAR Environ. Res.*, 2005, **16**, 171.
38. P.A. Schmeider, G. Ankley, O. Mekenyan, J.D. Walker and S. Bradbury, *Environ. Toxicol. Chem.*, 2005, **22**, 1844.
39. E. Wiese, L.A. Polin, E. Palomino and S.C. Brooks, *J. Med. Chem.*, 1997, **40**, 3659.
40. C.L. Waller, B.W. Juma, L.E. Gray Jr. and W.R. Kelce, *Toxicol. Appl. Pharmacol.*, 1996, **137**, 219.
41. T.G. Gantchev, H. Ali and J.E. van Lier, *J. Med. Chem.*, 1994, **37**, 4164.
42. H. Hong, W. Tong, H. Fang, L. Shi, Q. Xi, R. Perkins, J.D. Walker, W. Branham and D.M. Sheehan, *Environ. Health Perspect.*, 2002, **110**, 29.
43. W. Tong, H. Fang, H. Hong, Q. Xie, R. Perkins and D.M. Sheehan, in *Predicting Chemical Toxicity and Fate*, M.T.D. Cronin and D.J. Livingstone (eds), CRC Press, Boca Raton, FL, USA, 2004, 285.
44. ECETOC, ECETOC Document No. 33, *Environmental Oestrogens: A Compendium of Test Methods*, European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels, Belgium, 1996, 27.
45. V. Baker, *Toxicology in Vitro*, 2001, **14**, 413.
46. J.R. Reel, J.C. Lamb and B.H. Neal, *Fundam. Appl. Toxicol.*, 1996, **34**, 288.
47. P. Holmes, C. Humfrey and M. Scullion, *Appraisal of Test Methods for Sex-Hormone Disrupting Chemicals Capable of Affecting the Reproductive Process*, Document prepared by the Medical Research Council's Institute for Environment and Health, OECD, Paris, France, 1998. Website: http://www.oecd.org/ (accessed 14.5.04).
48. G. Langley, *Endocrine Disrupting Chemicals: A Non-Animal Testing Approach*, BUAV, London, 2004.
49. G.D. Charles, *ILAR J.*, 2004, **45**, 494.
50. A. Worth and M. Balls (eds), ATLA, 2002, **30**(Suppl. 1), 1–125.
51. N.A. Brown, H. Spielmann, R. Bechter, O.P. Flint, S.J. Freeman, R.J. Jelinek, E. Koch, H. Nau, D.R. Newall, A.K. Palmer, J.-Y. Renault, M.F. Repetto, R. Vogel and R. Wiger, *ATLA*, 1995, **23**, 868.
52. ICCVAM, *Development of Stably Transfected Cell Lines to Screen Endocrine Disrupters*, 2002. Website: http://iccvam.niehs.nih.gov/methods/endodocs/final/erta_brd/ertaappx/erta_b4.pdf, accessed 06/04/2006.
53. J. Häggblad, B. Carlsson, P. Kivelä and H. Siitari, *Biotechniques*, 1995, **18**, 146.
54. A. Freyberger and G. Scholz, *ALTEX*, 2004, **21**, 20.
55. L.E. Gray Jr., W.R. Kelce, T. Wiese, R. Tyl, K. Gaido, J. Cook, G. Klinefelter, D. Desaulniers, E. Wilson, T. Zacharewski, C. Waller, P. Foster, J. Laskey, J. Reel, J. Giesy, S. Laws, J. McLachlan, W. Breslin, R. Cooper, R. Di Giulio, R. Johnson, R. Purdy, E. Mihaich, S. Safe, C. Sonnenschein, W. Welshons, R. Miller, S. McMaster and T. Colborn, *Reprod. Toxicol.*, 1997, **11**, 719.
56. S. Stoessel and G. Leclercq, *J. Steroid Biochem.*, 1986, **25**, 677.
57. P.C. Hartig, K.L. Bobseine, B.H. Britt, M.C. Cardon, C.R. Lambright, V.S. Wilson and L.E. Gray Jr., *Toxicol. Sci.*, 2002, **66**, 82.
58. J.A Kemppainen, E. Langley, C.I. Wong, K. Bobseine, W.R. Kelce and E.M. Wilson, *Mol. Endocr inol.*, 1999, **13**, 440.
59. V.S. Wilson, K. Bobseine, C.R. Lambright and L.E. Gray Jr., *Toxicol. Sci.*, 2002, **66**, 69.

60. E. Sonneveld, H.J. Jansen, J.A.C. Riteco, A. Brouwer and B. van der Burg, *Toxicol. Sci*., 2005, **83**, 136.
61. W.D. Pennie, T.C. Aldridge and A.N. Brooks, *J. Endocrinol.*, 1998, **158**, R11.
62. P.W. Harvey and D.J. Everett, *J. Appl. Toxicol*., 2003, **23**, 81.
63. G. Timm, *The Story of the Steroidogenesis Assay in EPA's Endocrine Disruptor Screening Program*, US Environmental Protection Agency, 2005. www.epa.gove?scipoly/oscpendo/docs/edmvac/ steroidogenesis_story.htm.
64. L.R. Chaudhary and D.M. Stocco, *Biochem. Int.*, 1989, **18**, 251.
65. H. Nikula, T. Talonpoika, M. Kaleva and J. Toppari, *Toxicol. Appl. Pharmacol.*, 1999, **157**, 166.
66. M.K. Johansson, J.T. Sanderson and B.-O. Lund, *Toxicol. In Vitro*, 2002, **16**, 113.
67. P. Mak, F.D. Cruz and S. Chen, *Environ. Health Perspect.*, 1999, **107**, 855.
68. R.W. Bueggemeier and N.E. Katlic, *Cancer Res.*, 1990, **50**, 3652.
69. Y. Wu, B. Xu and R.J. Koenig, *J. Biol. Chem.*, 2001, **276**, 3929.
70. A.O. Cheek, K. Kow, J. Chen and J.A. McLachlan, *Environ. Health Perspect.*, 1999, **107**, 273.
71. S. Kitamura, N. Jinno, S. Ohta, H. Kuroki and N. Fujimoto, *Biochem. Biophys. Res. Comm.*, 2002, **293**, 554.
72. O. Hohenwarter, A. Waltenberger and H. Katinger, *Anal. Biochem.*, 1996, **234**, 56.
73. H. Kuzuya and K. Naka, *J. Clin. Endocrinol. Metab.*, 2004, **87**, 5185.
74. F.M.A. McNabb, D.A. Jang and C.T. Larsen, *Toxicol. Sci*., 2004, **82**, 106.
75. M.D. Scrimshaw and J.N. Lester, *Anal. Bioanal. Chem.*, 2004, **378**, 576.
76. T. Zacharewski, *Environ. Sci. Technol.*, 1997, **31**, 613.
77. OECD, Detailed Review Paper on Fish Screening Assays for the Detection of Endocrine Active Substances, OECD Series on Testing and Assessment, Number 47, OECD, Paris, 2004.
78. OECD, Detailed Review Paper on Amphibian Metamorphosis Assay for the Detection of Thyroid Active Substances, OECD Series on Testing, Number 46, OECD, Paris, 2004.
79. G.T. Ankley and R.D. Johnson, *ILAR J.*, 2004, **45**, 469.
80. K.L. Thorpe, T.H. Hutchinson, M.J. Hetheridge, J.P. Sumpter and C.R. Tyler, *Environ. Toxicol. Chem.*, 2000, **19**, 2812.
81. V.J. Kramer, S. Miles-Richardson, S.L. Pierens and J.P. Giesy, *Aquatic Toxicol.*, 1998, **40**, 335.
82. J.P. Sumpter and S. Jobling, *Environ. Health Perspect.*, 1995, **103**, 173.
83. C. Pelissero, G. Flouriot, J.L. Foucher, B. Bennetau, J. Dunogues, F. Gac and J.P.Sumpter, *J. Steroid Biochem. Biol.*, 1993, **44**, 263.
84. J.M.W. Smeets, T. R. Rankouhi, K.M. Nichols, H. Komen, N.E. Kaminski, J.P. Giesy and M. van den Berg, *Toxicol. Appl. Pharmacol.*, 1999, **157**, 68.
85. F. Gagné and C. Blaise, *Bull. Environ. Contam. Toxicol.*, 2000, **65**, 494.
86. J.L. Maitre, Y. Valotaire and C. Guguen-Guillouzo, *In Vitro Cell. Dev. Biol.*, 1986, **22**, 337.
87. A. Castaño, N. Bols, T. Braunbeck, P. Dierickx, M. Halder, B. Isomaa, K. Kawahara, L.E.J. Lee, C. Mothersill, P. Pärt, G. Repetto, J.R. Sintes, H. Rufli, R. Smith, C. Wood and H. Segner, *ATLA*, 2003, **31**, 317.
88. K. Fent, *Toxicol. In Vitro*, 2001, **15**, 477.

89. J.T. Sanderson, R.J. Letcher, M. Heneweer, J.P. Giesy and M. van den Berg, *Environ. Health Perspect.*, 2001, **109**, 1027.
90. C.J. Mao and D.J. Shapiro, *Endocrinology*, 2000, **141**, 2361.
91. N. Garcia-Reyero, E. Grau, M. Castillo, M.L. de Alda, D. Barcelo and B. Pina, *Environ. Toxicol. Chem.*, 2001, **20**, 1152.
92. A.J. Murk, J. Legler, M.M.H. van Lipzip, J.H.N. Meerman, A.C. Belfroid, A. Spenkelink, B. van der Burg, G.B.J. Rijs and R. Vethaak, *Environ. Toxicol. Chem*., 2002, **21**, 16.
93. S. Pawlowski, T.A. Ternes, M. Bonerz, A.C. Rastall, L. Erdinger and T. Braunbeck, *Toxicol. In Vitro*, 2004, **18**, 129.
94. N. Bhogal, C. Grindon, R. Combes and M. Balls, *Trends Biotechnol.*, 2005, **23**, 299.
95. M. Tanji and R. Kiyama, *Curr. Pharmacogenet.*, 2004, **2**, 255.
96. S. Terasaka, Y. Aita, A. Inoue, S. Hayashi, M. Nishigaki, K. Aoyagi, H. Sasaki, Y. Wada-Kiyama, Y. Sakuma, S. Akaba, J. Tanaka, H. Sone, J. Yonemoto, M. Tanji and R. Kiyamal, *Environ. Health Perspect.*, 2004, **112**, 773.
97. J.T. MacGregor, J.M. Collins, Y. Sugiyama, C.A. Tyson, J. Dean, L. Smith, M. Andersen, R.D. Curren, J.B. Houston, F.F. Kadlubar, G.L. Kedderis, K. Krishnan, A.P. Li, R.E. Parchment, K. Thummel, J.E. Tomaszewski, R. Ulrich, A.E.M. Vickers and S.A. Wrighton, *Toxicol. Sci.,* 2001, **59**, 17.
98. F.S. vom Saal, C.A. Richter, R.R. Ruhlen, S.C. Nagel, B.G. Timms and W.V. Welshons, *Birth Defects Res. Part A: Clin. Mol. Teratol.*, 2005, **73**, 140.
99. V.S. Wilson, K. Bobseine and E. Gray Jr., *Toxicol. Sci*., 2004, **81**, 69.
100. R. Combes, M. Balls, R. Curren, M. Fischbach, N. Fusenig, D. Kirkland, C. Lasne, J. Landolph, R. LeBoeuf, H. Marquardt, J. McCormick, L. Muller, E. Rivedal, E. Sabbioni, N. Tanaka, P. Vasseur and H. Yamasaki, *ATLA*, 1999, **27**, 745.
101. R.D. Combes, *ATLA*, 2004, **32**, 121.
102. K. Morohoshi, H. Yamamato, R. Kamata, F. Shiraishi, T. Koda and M. Morita, *Toxicol. In Vitro*, 2005, **19**, 457.
103. H.P. Gelbke, M. Kayser and A. Poole, *Toxicology*, 2004, **205**, 17.
104. J.W. Bridges and O. Bridges, *Toxicology*, 2004, **205**, 11.
105. I.F. Purchase, *Toxicol. Sci.*, 1999, **52**, 141.
106. B.J. Blaauboer, *Toxicol. Lett.*, 2003, **138**, 161.
107. J. Ashby, *Toxicol. Pathol.*, 2000, **28**, 432.
108. R.D. Combes and M. Balls, *ATLA*, 2005, **33**, 289.
109. W. Wutke, *Toxicology*, 2005, **205**, 1.

Intelligent Approaches to Safety Evaluation

DEREK J. KNIGHT

1 Introduction

Chemicals play a vital role in modern industrial society, and with an estimated 100,000 synthetic chemicals in commerce today, the monitoring and regulation of these materials with respect to their safety has become a significant task involving many scientific disciplines.

There is a complex path from development to eventual successful marketing of new industrial and household chemicals and other regulated products such as biocides, plant protection products, cosmetics and toiletries and food packaging. The process is costly in terms of money, time and human resources, and there are many potential pitfalls. A key element is regulatory planning to design a testing programme to meet the appropriate regulatory requirements. There is often a balance between reducing the cost of testing at the expense of delay and uncertainty while the regulators are consulted. It is therefore essential for regulatory advisors to establish the priority and make the best use of the various regulatory options. For example, it may be possible to get a product to market more easily if its use is restricted so it is defined as a food or cosmetic, than if it falls into the scope of a more demanding approval scheme such as those covering medicines.

The obligation to establish that products are safe has been brought into conflict with strong public opinion, especially in Europe, against animal testing. Hence industry, academia and regulators have been working in partnership to find other ways of evaluating the hazardous properties of chemicals.

Contract research laboratories and registration consultants are employed by industry to get new products through the increasingly complex web of legal and regulatory requirements that have to be complied with before marketing. They help their clients develop safe products as quickly and cheaply as possible without compromising safety. There is often a balance between reducing the cost of testing at the expense of delay and uncertainty while the regulators are consulted. It is therefore essential to find out the clients priority and keep them informed of the

Issues in Environmental Science and Technology, No. 23
Alternatives to Animal Testing
© The Royal Society of Chemistry, 2006

regulatory options: it may be possible to get a product to market more easily if its use is restricted so that it falls outside the scope of a demanding approval scheme.

It is possible to adopt an intelligent approach to safety evaluation and testing based on a tiered approach:

- Make maximum use of existing studies and literature data.
- Predict properties, using calculations and estimations or by analogy to tested substances with similar chemical structures, *i.e.* "read-across" and chemical "categories".
- Alternatives to animal testing should be used if available for the missing end-points.
- Use data waivers to justify omitting studies on the grounds of low exposure.
- Make a preliminary risk assessment.
- Make use of expert reports.
- Adopt a "weight of evidence" approach.
- Discuss the testing programme with the regulators.

2 Hazardous Properties Assessment

Toxicological, ecotoxicological, environmental fate and physico chemical properties have to be determined for safety assessment and regulatory approval. Traditionally such hazardous properties are determined by testing using standardised laboratory tests with the studies conducted in compliance with Good Laboratory Practice (GLP). These laboratory studies though are only in effect modelling the impact that the chemical may have on the systems really of concern, namely human health and the environment. The data are used to identify hazardous properties, which are described by various classification and labelling schemes using stylised interpretation criteria. The magnitude of hazardous properties from many of the laboratory tests can then be used to predict the adverse effects on humans or the environment.

Hazards from chemicals can be categorised into physico-chemical, toxicological and environmental. Physico chemical hazards, such as explosivity, oxidising properties and flammability, are caused by the intrinsic physical or chemical properties of the substance. Toxicological hazards arise from a chemical causing harmful effects to living organisms, which in practice normally means death, injury or adverse effects when ingested, inhaled or absorbed through the skin. Toxic effects may be acute or chronic, local or systemic and reversible or irreversible. They include corrosivity and irritancy to skin, eyes and the respiratory tract. Specific toxic effects are skin and respiratory sensitisation, carcinogenicity, mutagenicity and reproduction effects. Environmental hazards relate to air, soil or water, including groundwater. Finally, there may be hazards to human health by exposure to the chemical *via* the environment. Safety studies are summarised by Knight and Thomas.[1]

3 Risk Assessment

The exercise of risk assessment is an essential element of most regulatory processes. It can also be used as an aid to intelligent approach to designing testing programmes

Table 1 *Steps for risk assessment*

Assessment of effects
• Hazard identification (intrinsic hazardous properties of the substance) • Elucidation of the dose (concentration) versus response (effects) characteristics
Exposure assessment
• For the human population, workers, consumers and indirect exposure *via* the environment • For the different environment compartments likely to be exposed to the substances; water, sediment, soil, and air
Risk characterisation
Comparison of information on hazardous properties and effective dose levels/concentrations with exposure levels in order to characterise the degree of risk posed by the substance to human health or to the environment

(see Section 9), by performing a preliminary risk assessment using the available hazardous properties, information and simplified or approximate exposure assessments.

The risk from a substance is determined from its intrinsic hazardous properties and the likely exposures of humans and the environment from use throughout the life cycle.[2–4] Table 1 shows the steps for risk assessments used in the EU regulatory environment (see Chapter 1 in this book), to illustrate the scientific principles.

The risk assessment is one of the main factors in deciding whether the product can be approved in schemes that require marketing authorisation, such as biocides and plant protection products. It is very important to appreciate that risk assessment and regulatory approval is usually an iterative process, *i.e.* a cycle of successive improvements, where the assessment is revised several times (Figure 1). If the initial outcome is unfavourable, the risk assessment can be made more realistic and less precautionary either by improving the exposure assessment, for example by revising use conditions to lower exposure, or by generating more safety data to improve the knowledge of the hazardous properties. The outcome of risk assessment may be risk management, if necessary, to prevent harm to humans or the environment.

The EU risk assessment is required for a particular hazardous property if the substance is classified as "dangerous" according to Annex VI of Council Directive 67/548/EEC,[5] or if there are other "reasonable grounds for concern" on the basis of exposure or properties. If there is a consumer exposure, this is exposure-based reasonable grounds for concern, and a risk assessment is needed, including consumer exposure assessment, whether the substance is classified as dangerous or not, and whether it is neat or in a preparation. If a substance is in a preparation for certain stages in the EU life cycle, a risk assessment for the hazardous property is generally only necessary if the preparation is also classified as dangerous. The calculation methods of the Dangerous Preparations Directive[6] are generally applied, although if adequate test data on the preparation are available for general toxic properties or ecotoxicity then these take precedence for classification of the preparation. Impurities in the technical grade of a substance that have specific toxic effects (*i.e.* carcinogenicity, mutagenicity or toxicity for reproduction) will contribute to the dangerous

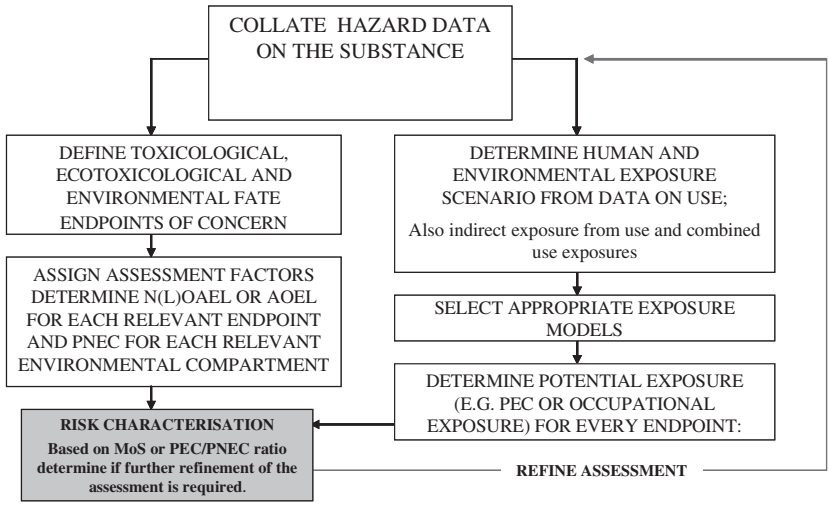


Figure 1 *Risk assessment*

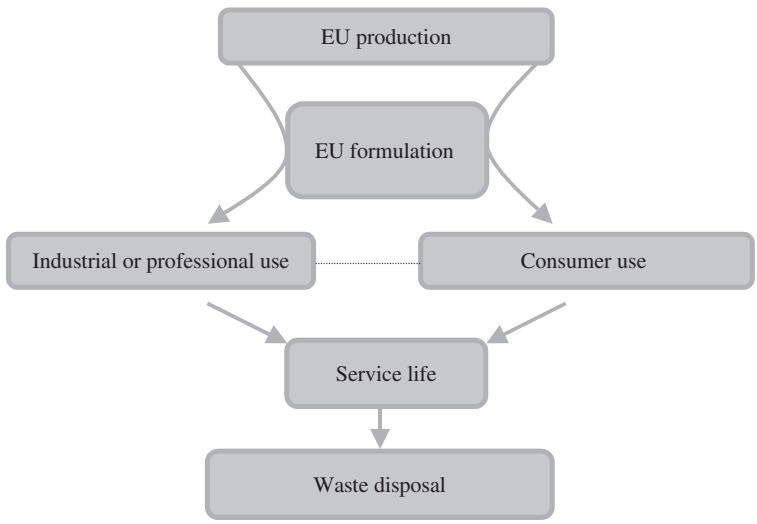


Figure 2 *Schematic representation of the lifecycle stages of a substance*

classification of the substance using the calculation methods of the Dangerous Preparations Directive:[6] hence, a risk assessment will be required for these.

Exposure to a substance can potentially occur during each stage of the life cycle of a substance, from production to ultimate disposal or recovery. At each of the stages of the lifecycle, the exposure to the various environmental compartments and human populations is assessed. The main stages in the life cycle that may be discussed in the risk assessment are shown in Figure 2.

The "dangerous" properties of explosivity, flammability and oxidisibility are evaluated in terms of potential adverse effects to workers and consumers, based on the known or reasonably foreseeable conditions of use. Risk assessment entails evaluation of the likelihood of an adverse effect based on how the hazard is dealt with.

The "dangerous" properties of acute toxicity, skin and eye irritation, corrosivity, sensitisation, repeated-dose toxicity, mutagenicity, carcinogenicity and toxicity for reproduction are evaluated in terms of their potential toxic effects to workers, consumers and man exposed indirectly *via* the environment, based on the known or reasonably foreseeable conditions of use for each stage in the life cycle of the substance from which exposure can occur. Risk assessment is also required if there are reasonable grounds for concern for properties, such as positive *in vitro* mutagenicity tests or structural alerts. Comparison of occupational and consumer exposure levels with the exposure levels at which no adverse effects are expected to be seen will allow an assessment of risk to be made, based on the ratio between the two (as discussed in detail in Chapter 1). In practice this will often be a qualitative evaluation, especially for acute toxicity and irritation or corrosivity.

Environmental risk assessment is required for substances that are classified as "dangerous for the environment", or if there are the following other reasonable grounds for concern in relation to environmental effects:

- Indications of bioaccumulation potential.
- The shape of the toxicity/time curve in ecotoxicity testing.
- Indications of adverse health effects from long-term exposure on the basis of toxicity studies.
- Structural alerts from ecotoxic analogous substances.

Environmental risk assessment involves examination of the potential exposure to ecosystems of the aquatic, terrestrial and air components. Initial assessment normally focuses on the aquatic compartment, including micro-organisms in sewage treatment systems. This first tier risk assessment can be extended to cover the sediment part of the aquatic compartment and the soil compartment. At higher tonnage levels, effects relevant to the food chain (secondary poisoning) are evaluated. Diderich in Chapter 8 of Ref. 1 discusses the principles of environmental risk assessment as applied in the EU. The objective is to predict the concentration of the substance below which adverse effects in the environmental compartment of concern are not expected to occur, *i.e.* the predicted no effect concentration (PNEC). However, in some cases, it may not be possible to establish a PNEC, and a qualitative estimation of the concentration, effect has to be made instead. Applying an assessment factor to the numerical results of acute and long-term ecotoxicity studies allows the PNEC to be estimated. The assessment factor is an expression of the degree of uncertainty in extrapolation from test data on a limited number of species to the real environment, and thus in general, the more extensive the data and the longer the duration of the tests, the smaller is the degree of uncertainty and the size of the assessment factor. The aim of the exposure assessment is to predict the concentration of the substance that is likely to be found in the environment, *i.e.* the predicted environmental concentration (PEC). Again it may be not be possible to

establish a PEC and a qualitative estimation of exposure has to suffice. For any given environmental compartment, the risk assessment entails comparison of the PEC with the PNEC, in the PEC/PNEC ratio. If this ratio is below 1, there is no immediate concern. If the ratio is above 1, the assessor decides on the basis of its value, and the factors giving reasonable grounds for concern described above, what conclusions and recommendations apply for further testing or risk management. If it has not been possible to derive a PEC/PNEC ratio, the risk assessment is a qualitative evaluation of the likelihood that an adverse effect will occur. For assessing potential adverse effects in the air compartment, *e.g.* the ozone-depleting potential, for which the above steps are impracticable, risk assessment is conducted on a case-by-case basis.

4 Existing Studies and Literature Data

The ideal is to have modern laboratory studies conducted to standard EU or international methods in compliance with the quality standard of GLP. In practice, however, there will often be old studies or published papers which can provide equivalent information to assess the properties of the substance.

The first step in obtaining existing studies is normally to check for in-house data, or studies generated by competitor companies who may agree to share them. The next stage is to conduct appropriate literature searches, which can sometimes be problematic without experience in using the relevant databases. In practice, however, it is often possible to collect much of the published safety information, at least in summary form, using one or two well chosen data bases and by searching the internet: many suppliers of chemicals put safety data sheets on their websites. Information may also be available on regulators' websites. Information may be published in standard chemistry and toxicology textbooks and manuals for common chemicals or for particular types of chemistries, *e.g.* particular metal ions.

The quality of existing studies can vary greatly, and it is important to assess their reliability before using them to make scientific regulatory decisions on whether a substance is safe and can be approved or commercial decisions on what new studies are needed and whether they are economically viable. Guidance on evaluating data for use under the EU chemical schemes is given in the Risk Assessment Technical Guidance Document,[3] and insight can also be gained into this matter from the US Environmental Protection Agency (EPA) guidance on searching for literature data.[7] It may be that a professional judgment can be made to get an overall impression of the hazardous properties, and this may be all that is needed for simple preliminary decision-making. Nevertheless, for more formal decisions and in preparing registration dossiers it is essential to evaluate existing studies more systematically, with regard to their adequacy and completeness. The adequacy of each test can be defined by two basic elements:

- *Reliability*: the inherent quality of a study relating to the test method, performance of the study and the reporting (Table 2).
- *Relevance*: the extent to which a test is appropriate for a particular hazard or risk assessment covering, for example, whether the appropriate species and route of exposure studied has been used in a toxicology study, and whether the endpoints are studied under relevant conditions in ecotoxicology studies.

Table 2 *Factors to consider when evaluating the reliability of an existing study*

- Purity/impurities and origin of the test substance
- Complete test report is available or published in sufficient detail
- Even if the reliability cannot be established or the study does not meet current standards, the data might still be used, taking into account
  - Other studies or estimated results are consistent with the study results
  - Other studies on close chemical analogues give similar results
  - A approximate value is sufficient for use in the risk assessment
  - In the cases where differing results from similar studies were obtained or an extensive data set is available for an individual species or a taxonomic group, it may be possible to use the distribution of these data to draw general conclusions on the toxicity

Table 3 *Klimisch scoring scheme for reliability studies*

1 – reliable without restrictions studies or data … generated according to generally valid and/or internationally accepted testing guidelines (preferably performed according to GLP) or in which the test parameters documented are based on a specific (national) testing guideline … or in which all parameters described are closely related/comparable to a guideline method

2 – reliable with restrictions studies or data … (mostly not performed according to GLP), in which the test parameters documented do not totally comply with the specific testing guideline, but are sufficient to accept the data or in which investigations are described which cannot be subsumed under a testing guideline, but which are nevertheless well documented and scientifically acceptable

3 – not reliable studies or data … in which there were interferences between the measuring system and the test substance or in which organisms/test systems were used which are not relevant in relation to the exposure (*e.g.* unphysiologic pathways of application) or which were carried out or generated accordi ng to a method which is not acceptable, the documentation of which is not sufficient for assessment and which is not convincing for an expert judgment

4 – not assignable studies or data … which do not give sufficient experimental details and which are only listed in short abstracts or secondary literature (books, reviews, *etc*.)

The reliability of the data is a key-initial consideration and can be checked quickly to filter out unreliable studies, so focusing on studies is considered most reliable. Klimisch *et al.*[8] have developed a scoring system for checking reliability (Table 3).

5 Surrogate Data: Calculation, Read Across and SAR/QSAR

There are many methods to estimate physico chemical properties,[9–11] and calculate these using thermodynamic and empirical relationships.[12] For example, the partition co-efficient between n-octanol and water (Pow) can often be fairly reliably estimated from the chemical structure based on the concept that fragments contribute to hydrophilicity or hydrophobicity, and from this the water solubility can be calculated. Another example is the calculation of boiling point from measured vapour pressure. Such predictions and calculations may be used in deciding, which of the experimental methods to use or for providing an estimate or limit value where the experimental method cannot be applied for technical reasons. It is particularly

important where estimation methods have been used, to ensure that the results derived are consistent with one another and are reasonable based on chemical structure. This is true for endpoints measured directly, as an unexpected result for the particular structure can lead to inconsistency and deficient data.

If there is no, or only limited, data available Structure–Activity Relationships (SARs) may be considered, including quantitative SARs (QSARs). These can indicate a potential hazard, toxicokinetic properties or the need for further testing. There are validated QSARs for some physico chemical properties and for the ecotoxicity of certain chemical classes. SARs are not well developed in relation to mammalian toxicity, and "expert judgment" (perhaps based on "structural alerts") is usually required.

The properties of substances can sometimes be assessed on a case-by-case basis using the "read-across" approach. The properties of a substance are predicted from data on close analogues with similar physico chemical properties and impurity profiles. Similar biological properties are anticipated, since toxicokinetics (especially absorption and metabolism) would be comparable. As many matching toxicological properties as possible are required to provide evidence in support of read-across for missing endpoints.

The use of SAR and "read-across" to form "categories" of chemicals for evaluation as a group is permitted in the International Council of Chemical Associations (ICCA) Organisation for Economic Co-operation and Development (OECD) high production volume (HPV) schemes and encouraged for US HPV chemicals. There is guidance on SAR and categories in the OECD manual[13] and by the EPA.[14,15]

A key use for calculation in combination with QSAR and read-across is for justification of omission of experimental measurements, *i.e.* for a data waiver.

6 Alternatives to Animal Toxicology Studies

For many substances there will be some essential endpoints for which there are no existing data and for which there are no satisfactory (Q)SARs or other surrogate data. In such cases it may nevertheless not be necessary to conduct animal studies, because some properties can be adequately assessed using *in vitro* or *ex vivo* tests.

In their 1959 book, *The Principles of Humane Experimental Technique*,[16] Russell and Burch defined a strategy for minimising animal use without compromising the quality of scientific work. This was to be achieved through the use of three different categories of alternative testing, which they termed the "Three Rs": reduction, refinement and replacement. Replacement methods are the ultimate goals, in that they do not involve conducting experiments on whole living animals. Some of these approaches are only relative replacements, because they entail the humane killing of an animal to obtain cells, tissues or organs for *in vitro* studies. Others are absolute replacements that do not require any biological material from a fully developed, vertebrate animal.

Progress has been made in reducing the animal studies used for hazard evaluation by using *in vitro* tests and property predictions based on chemical structure.[17,1] Alternative tests can be used for eliminating substances from further development on the basis of their potential toxicity or for regulatory safety evaluation. Improved scientific understanding of the toxicological effect can better predict the risk to humans.

Table 4 *ECVAM validated tests for use in the EU*

- EU Method B40 for evaluating skin corrosivity (the transcutaneous electrical resistance (TER) assay and the human skin model)
- EU Method B41 for phototoxicity (3T3 neutral red uptake photoxicity test (3T3 NRU PT))
- EU Method B42 the Local Lymph Node Assay

There are also commercial benefits in reducing the cost and increasing the speed of product development. The key consideration, though, is animal welfare, and animal studies should only be undertaken if scientifically necessary and there is no other method of obtaining the information.

Efforts to develop *in vitro* tests capable of evaluating potential systemic toxicants have been hindered by the complexity of the systems involved; however, more progress has been made with assays that evaluate local effects. A group of different *in vitro* assays (the "test battery" approach) will provide the most accurate and comprehensive prediction of the toxic potential of a test material in man.

All alternative tests have to be validated scientifically and established as acceptable by industry and the regulators before being used in safety assessment. The key international bodies that undertake this validation process are the European Centre for the Validation of Alternative Methods (ECVAM) and the US Interagency Coordinating Committee on the Validation if Alternative Methods (ICCVAM).

ECVAM-validated tests become official EU test methods (Table 4), and alternative tests not incorporated into EU legislation may still be used for regulatory purposes on a case-by-case basis if there is sufficient information to define the substance as "dangerous" and the result is adequate for risk assessment. In practice this currently usually requires a positive result to trigger classification as dangerous, whereas a negative result in an *in vitro* study not recognised by the EU (*e.g.* skin irritation tests) normally means that the standard animal study has to be conducted to confirm the absence of the hazardous property. However, increasingly a "weight of evidence" approach is taken (*i.e.* considering the chemical structure and any data on analogous substances), and this will certainly be the case under the proposed EU chemicals scheme (see Section 10).

Alternative tests that are not yet validated can be very useful in tiered testing strategies, to screen out clearly positive chemicals and avoid unnecessary animal testing. For example, before rabbit skin or eye irritation tests are conducted, it is necessary first to collect the existing data, consider the acidity and alkalinity and then *in vitro* tests are often conducted.

7 Data Waivers

The standard set of toxicological studies for various regulatory schemes can be very extensive and use many experimental animals. Hence, before conducting new studies it is important to make maximum use of "data waivers". These are justifications to omit standard studies because either the study is impossible to conduct or because it is scientifically unnecessary on the grounds of low exposure.

Insight into the criteria for low-exposure data waivers can be gained from the EU biocidal products scheme.[18] The technical guidance document for data requirements[19] makes provision for certain long-term toxicology studies (Table 5) to be waived on a case-by-case basis based on a consideration of the level, frequency and duration of exposure (Table 6). For example, it may be possible to demonstrate that, taking into account all life-cycle stages (*e.g.* processing, formulation, use and disposal) exposure will be limited, so the idea is that there will be low risk from low exposure (Figure 3). This type of approach is likely to be useful in REACH registrations, perhaps using established technical guidance from other EU regulatory schemes, such as biocides.

A case can often be made to omit studies as scientifically unnecessary, *i.e.* it is possible to conduct an adequate risk assessment without them. This is most often the case if the substance decomposes to substances that are already established as safe. For example, the substance may hydrolyse rapidly to non-toxic products and the key issue here is to establish that this happens rapidly in the stomach before the parent substance can be absorbed. There are then excellent grounds for omitting the expensive long-term animal studies. It would also be necessary to establish that there is no absorption from the other exposure routes, *i.e.* dermal and inhalation. A useful study is therefore the *in vitro* skin penetration study, although in many cases radiolabelled test material is needed, which is expensive and time-consuming to prepare. Before conducting such a study it may be worth estimating skin penetration using a QSAR approach. In a similar way it is often justified to omit ecotoxicity studies on a substance, which hydrolyses or otherwise decomposes in the aquatic environment to stable products that have already been tested. In this case, it is acceptable for the rate of hydrolysis to safe degradants under environmental conditions to be somewhat

Table 5 *Possible toxicology studies for data waiving for biocides in the EU*

- 90-day repeat-dose toxicity study in the second species (non-rodents)
- Chronic toxicity studies in rodents and non-rodents
- Carcinogenicity studies in both species (normally rat and mouse)
- Developmental toxicity in the second species (normally the rat)
- 2-generation fertility study in the rat

Table 6 *Thresholds of concern for human exposure for biocides in the EU*

Level of exposure (How much?)
Threshold for lower concern: there must be a provisional margin of safety of at least 1000 between the exposure and the overall NOAEL established from the core toxicological studies. There is negligible exposure if specific exposure or use conditions apply, *e.g.* automated application of a ready-to-use product in a completely closed industrial system
Frequency of exposure (How often?)
Threshold for lower concern: up to once per month (may be average per year, *i.e.* up to 12 exposures per year).
Duration of exposure (How long?)
Threshold for lower concern: up to 3 months per year

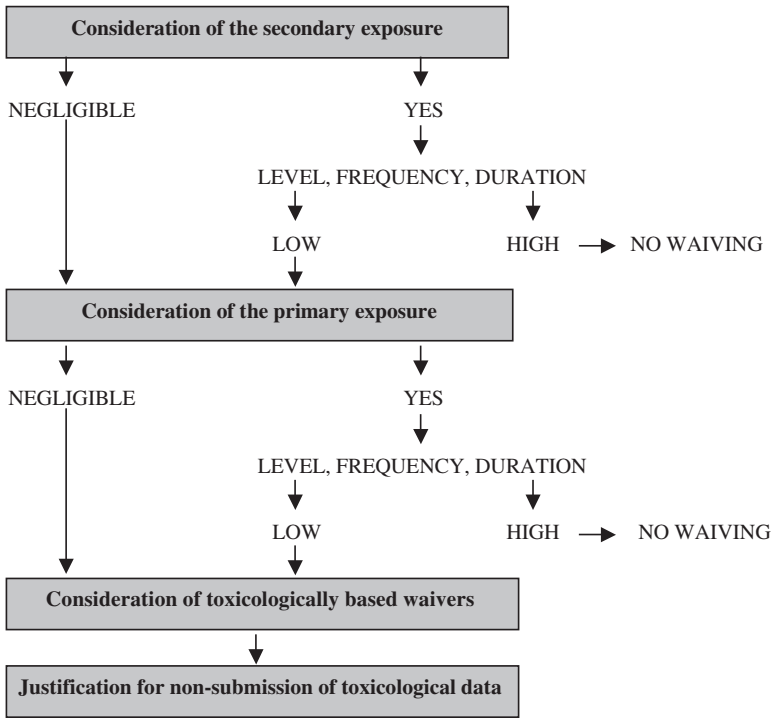


Figure 3 *Human exposure assessment for possible data waiving of biocides in the EU*

slower, perhaps even up to several hours, to justify omitting ecotoxicology studies. However, to justify omitting toxicology studies, degradation under physiologically relevant conditions must be almost instantaneous, ideally in seconds rather than minutes. It can take some hours for chemicals discharged to drain to reach the wastewater treatment plant whereas, in principle, absorption from the stomach could be immediate.

8 Expert Reports and Weight of Evidence

From experience with HPV chemicals and existing biocide active substances, it can be invaluable to have an expert report to collect, summarise and evaluate all the available literature data and in-house studies to get an overall view of the toxicity, environmental fate or ecotoxicity of the substance. Such an expert report can take into account the chemical and physical properties and draw upon information on chemical analogues. It can also make predictions based on biological processes, *i.e.* human biology or ecology, often considering possible degradation by chemical, metabolic or environmental processes.

A simple expert report would be on a single hazardous property, such as toxicokinetics or long-term general toxicity, and in effect would be a review article. This review can then be regarded administratively as a single data element for that property (arguably of reliability 2 and hence acceptable), whereas each individual reference in the

expert report would be inadequate alone (*i.e.* of reliability 3 or 4). At its most basic, it can often be adequate simply to list the published set of numerical results for a simple endpoint (*e.g.* acute oral toxicity or acute fish or Daphnia toxicity) and point out that they are comparable, so that although each result on its own is inadequate, overall the particular hazardous property is adequately assessed without new standard testing being needed.

More complex expert reports, covering a wider range of endpoints (*e.g.* all long-term toxic effects or effects on organisms in a particular environmental compartment) and/or using other arguments (such as "read across", *etc*.) may in effect be regarded as a data waiver argument to justify omitting studies as scientifically unnecessary. In addition, if there is low but not insignificant exposure to humans or a particular environmental compartment, so that a data waiver purely on low-exposure grounds is not justified, this argument can be combined with the scientific argument from an expert report in a weight of evidence approach.

9 Preliminary Risk Assessments

Before the testing programme is finalised it may be useful to conduct a preliminary risk assessment (Table 7). This helps to forecast whether the substance is safe when used as intended, if restrictions are likely to apply, what extra data are likely to be needed, and to indicate potential problem areas where new testing could initially be focused to get a rapid and more economical assessment of the final testing programme. This allows industry to decide whether the substance can still be supported or if it may not be commercially viable. The preliminary risk assessment is based on the available safety data, including from a literature search, in combination with exposure assessment using EU models or adaptations of other models if the EU models are not yet developed. Likely toxicology data waivers and possible additional tiered testing for ecotoxicology or environmental properties may be forecast.

10 EU Registration, Evaluation and Registration Scheme (REACH)

It is useful to illustrate the principles of intelligent safety assessment within a specific regulatory context. The example chosen is the forthcoming new EU scheme to control chemicals, because this is very much in the forefront of regulatory thinking

Table 7 *Preliminary risk assessment as an aid to decision-making*

- Forecast whether the substance is approvable (and any restrictions)
- Predict data needed, with potential problem areas (to focus initial work)
- Use available safety data
- Use EU exposure models, or adapt EU or non-EU models
- Suggest toxicology data waivers
- Alert ecotoxicology/environmental fate higher-tier testing
- Iterative process: refine as new data are generated
- Use to decide whether to support the active substance
- Use to negotiate with the regulator
- Finalise to form part of the dossier for submission

on safety testing with the focus on risk assessment, and it is likely to set the scene for other regulatory developments, in the EU and elsewhere.

On 13 February 2001 the European Commission adopted the much-discussed White Paper "Strategy for a Future Chemicals Policy".[20] This proposed a wide-ranging fundamental overhaul of EU chemical control legislation, *i.e.* the Dangerous Substances Directive,[5] including the notification scheme for new substances,[21] the Dangerous Preparations Directive,[6] the Existing Substances Regulation[22] and the Marketing and Use Directive.[23] In essence, the legislation for new and existing substances would be merged. The current EU chemical control measures were considered to result in too great a disparity between new and existing substances, with the high cost of new substance notification stifling innovation. Furthermore, although existing substances account for >99% by volume of chemicals in commerce, they were considered to be relatively poorly assessed and controlled in comparison with new substances.

The main objective of the new EU Chemical Strategy is to ensure a high level of protection for human health and the environment, while maintaining an efficient internal market within the EU and stimulating innovation and competitiveness in the chemical industry. The White Paper also takes account of the need to increase transparency to decision-making and improve public access to information on chemicals. There will be integration with international initiatives. The Precautionary Principle will be applied, and substitution of substances with less dangerous alternatives will be encouraged.

The European Commission published the first draft of legislation[24] intended to implement the White Paper on 7 May 2003. This is the Registration, Evaluation and Authorisation (REACH) scheme for chemicals. There was an 8-week Internet consultation on the "workability" of this legislation from 16 May to 10 July 2003 and about 6400 contributions were received. A revised version emerged in September 2003, jointly from the Enterprise and Environment Directorates of the European Commission and after consultation within the commission; the final proposed Regulation was presented on 29 October 2003. These formal legislative proposals have to be discussed by the Council of Ministers and the European Parliament, under the Co-decision Procedure. After some debate it was agreed on 5 February 2004 that the Environment and Public Health Committee of the European Parliament would handle the dossier.

REACH will place a duty on companies that produce, import and use chemicals to assess the risks arising from their use (with new studies conducted in justified cases), and to take the necessary measures to manage any risks identified. Hence, the burden of proof will be transferred from the regulators to industry for putting safe chemicals on the market. Testing results have to be shared to reduce any animal testing, and registration of information on the properties; uses and safe use of chemical substances will be an integral part of the new system. These registration requirements will vary depending on the volume in which a substance is produced, and on the likelihood of exposure to humans or the environment. A phase-in system lasting up to 11 years is planned. Higher tonnage substances would require the most data, and would have to be registered first; lower tonnage substances would require less data and be registered later. It is anticipated that ca. 30,000 substances will be registered, and that around 80% will not proceed to the next stage of evaluation.

The information included in the registration of the ca. 5000 substances exceeding a production or import volume of 100 tonnes per annum will have to be evaluated. National "rapporteur" Competent Authorities will be allocated substances to evaluate on behalf of the EU. When the volume reaches 100 tonnes per annum the manufacturer or importer has to submit the available information and a proposal for a testing programme (according to Annex VII of the Regulation). The rapporteur Competent Authority, in consultation with other Competent Authorities and the European Chemicals Agency (ECA), agrees the final testing programme, and evaluates the studies when they are submitted. At 1000 tonnes per annum an equivalent procedure is followed for Annex VIII testing. For phase-in substances already exceeding the evaluation trigger values, a tiered approach is proposed following initial registration.

Tighter controls will be introduced for the chemicals of highest concern, *i.e.* carcinogens, mutagens and reproductive toxicants (CMRs), persistent, bioaccumulative and toxic substances (PBTs) and very persistent and very bioaccumulative substances (vPvBs) will be subjected to an authorisation regime, and hence will be registered early. Other substances of concern, such as endocrine disrupters, will be included on a case-by-case basis within the authorisation system. Substances subject to authorisation will have to be approved for a specific use, with decisions based on a risk assessment and consideration of socio-economic factors.

11 Data for REACH

The information on hazardous properties is linked to the manufacture/import level, on the grounds that there is a potential for more exposure as more substance is in the EU. Annexes in the proposed REACH Regulation specify the standard data requirements and give rules on the circumstances in which data may be omitted, essentially corresponding to the current data waiver situations for notification of new substances, and when extra data are triggered, either as an outcome of risk assessment or to clarify ambiguous study results. Clearly any studies that are technically impossible can be omitted. The general technical, commercial and administrative information needed for all registrations for the technical dossier is specified in Annex IV of REACH (Table 8) and Annexes IV–VIII specify the supply-level triggered standard hazardous properties information needed for registration (Tables 9–12). Note that there are likely to be changes to these standard data requirements in the final version of REACH eventually adopted.

In addition to these specific rules, the registrant can adapt the standard data set and provide adequate information on the properties of the substance by alternative means: *i.e.* with non-standard or non-GLP tests, historical human data, a weight of evidence, SAR or "read-across" to tested analogues (Table 13). There is also a provision for "substance-tailored exposure-driven testing" to allow data waivers at higher supply levels for low-exposure substances.

The overall initial cost of testing substances for registration under REACH will depend on the EU tonnage, with data to be provided in advance of supply at $\geqslant$1 and $\geqslant$10 tonnes per annum. Compared with the current EU notification scheme,[21] the REACH proposed Annex V data set (Table 9) is effectively an extended "reduced notification" (*i.e.* including some "base set" tests) and Annex VI (Table 10) is an extended

Table 8 *General Annex IV information for registration under REACH*

- Technical dossier (in a specified electronic format):
  - Annex IV technical data on the registrant, identification of the substance, manufacture and use and guidance on safe use
  - Robust summaries of safety data
  - Proposed classification and labelling
  - Statement whether animal testing was conducted
  - Proposal for any further testing
- Chemical Safety Report (using ECA developed software), only at $>$ 10 tonne per annum under the revised proposals. This is a risk assessment including PBT and vPvB assessment (but for substances used in cosmetics and food-packaging materials only the environment is covered)

Table 9 *Proposed REACH Annex V data for substances at ⩾1 tonne per annum in the EU*

Melting/freezing point
Boiling point
Relative density
Vapour pressure
Surface tension
Water solubility
n-Octanol-water partition coefficient
Flash point or flammability
Explosivity
Auto-flammability
Oxidising properties
Granulometry
Skin irritation or corrosivity evaluation or *in vitro* tests
Eye irritation evaluation or *in vitro* test
Skin sensitisation evaluation or local lymph node assay
Ames test
Acute *Daphnia* toxicity
Possible additional studies:
Further mutagenicity tests
21-day *Daphnia* reproduction study

"base set" (*i.e.* with some studies from Level 1 of the notification scheme). For supply at ⩾100 and ⩾1000 tonnes per annum the testing programme is negotiated with the Competent Authority, and hence the costs will be highly variable. These REACH Annex VII and VIII data (Tables 11 and 12) essentially comprise the notification Level 1 and 2 studies respectively, with various additions and modifications. In practice there is likely to be ample opportunity to reduce testing costs by sharing data and making use of "surrogate data" and "data waivers" for many chemicals. Nevertheless, in contrast, there will be some cases where extra testing is triggered from the results of the standard tests or as an outcome of risk assessment. Therefore it is difficult to compare the cost of registration with the current notification scheme, but Table 14 describes the standard notification testing to put the proposed REACH data into context.

Table 10 *Proposed REACH Annex VI data for substances at ⩾10 tonnes per annum in the EU*

In addition to the Annex V data:
In vivo skin irritation (unless classified from Annex V data)
In vivo eye irritation (unless classified from Annex V data)
In vitro chromosome aberration test
In vitro gene mutation assay
Acute oral toxicity
Acute inhalation or dermal toxicity
28-day (or 90-day) repeat-dose study in the rat (normally oral exposure)
Developmental toxicity screening study (OECD 421)
Developmental toxicity study (OECD 414)
Toxicokinetics assessment (a prediction based on the available data)
Acute fish toxicity
Algal growth test
Ready biodegradation
Activated sludge respiration inhibition test
Hydrolysis test
Adsorption/desorption screening test
Possible additional studies:
In vivo mutagenicity studies
Further repeat-dose study in the rat
Second developmental toxicity study
Two-generation fertility study in the rat
Chronic fish toxicity study
Biodegradation simulation studies

Table 11 *Proposed REACH Annex VII data for substances at ⩾100 tonnes per annum in the EU*

The registrant makes a testing programme proposal covering
Stability in organic solvents and identification of degradants
Dissociation constant
Viscosity
Reactivity to container material
In vivo mutagenicity studies
28-day or 90-day repeat-dose study in the rat (if not part of the Annex VI data)
Developmental toxicity studies in two species (if not part of the Annex VI data)
Two-generation fertility study in the rat (if there are adverse findings from the 28-day or 90-day studies)
21-day *Daphnia* reproduction study
Chronic fish toxicity study
Simulation test on the ultimate degradation in surface water
Soil simulation test
Sediment simulation test
Fish bioaccumulation study (unless there is a low predicted bioaccumulation potential)
Further adsorption/desorption study
14-day earthworm toxicity
Study of the effects on soil micro-organisms
Short-term toxicity to plants

Table 12 *Proposed REACH Annex VIII data for substances at ⩾1000 tonnes per annum in the EU*

The registrant makes a testing programme covering, if appropriate
Further mutagenicity studies
Long-term repeat-dose (⩾12 months) study in the rat
Further toxicity study to investigate specific concerns
Two-generation fertility study in the rat (if not part of the Annex VII data)
Carcinogenicity study (often combined with a 2-year chronic toxicity study, usually in the rat)
Further biodegradation in water, sediment and soil – covering degradation rate and identification of relevant degradants
Further environmental fate and behaviour studies
Long-term earthworm toxicity
Long-term toxicity to other soil invertebrates
Long-term plant toxicity
Long-term toxicity to sediment organisms
Long-term or reproductive bird toxicity

Table 13 *Options to adapt the standard REACH registration safety data*

- Annexes V–VIII give standard data requirements (in column 1) and rules for omitting tests or additional studies (in column 2)
- Annex IX covers adapting the standard data requirements
  - existing non-standard and/or non-GLP data
  - historical human data
  - weight of evidence
  - SAR
  - grouping and "read across"
  - data waivers, *i.e.* the study is technically impossible
  - substance-tailored exposure driven testing (data waivers for high tonnage substances with low exposure)

12 Regulatory Planning & Liaison with the Regulators

Most industrialised countries have notification schemes for new industrial and household chemicals.[1] Therefore testing programmes and registration strategies have to be designed for worldwide marketing. Notification of polymers is especially complex, with the low-concern criteria differing between schemes. For non-polymers, it is normal practice to present a standard EU base set, with data waivers to eliminate the extra national studies. Using "read-across" is becoming even more popular: since the EU competent authorities have been the most conservative in using such surrogate data, it is normal to negotiate the EU testing first, and then present this to other regulators. The US EPA has a long history of using surrogate data, including categories as consolidated pre-manufacturing notices, and (Q)SAR predictions of properties from chemical structure or measured physio-chemical properties.

Intelligent safety evaluation will be especially important for the new EU REACH scheme. Decisions have to be made on whether to use literature data and/or "surrogate data" and if "data waivers" are appropriate. With other registration schemes (*e.g.* notification of new substance and the EU review of active substances for biocidal

Table 14 *The current EU notification studies*

Reduced notification for substances at ⩾0.1 tonne per annum (or ⩾0.5 tonnes cumulative): Annex VII B of the directive [21]

Melting/freezing point	Acute oral toxicity
Boiling point	Skin irritation
Water solubility	Eye irritation
n-Octanol-water partition coefficient	Skin sensitisation
Flash point or flammability	Ames test
	Ready biodegradation

Possible additional studies:
Vapour pressure
Acute *Daphnia* toxicity

Base Set for substances at ⩾1 tonne per annum (or ⩾5 tonnes cumulative): Annex VII A of the directive [21]

Melting/freezing point	Eye irritation
Boiling point	Skin sensitisation
Relative density	28-day repeat dose study in the rat
Vapour pressure	Ames test
Surface tension	*in vitro* gene mutation or chromosome aberration
Water solubility	test
n-Octanol-water partition coefficient	Toxicokinetics assessment (a prediction based on
Flash point or flammability	the available data)
Explosivity	Acute fish toxicity
Auto-flammability	Acute *daphnia* toxicity
Oxidising properties	Algal growth test
Granulometry	Ready biodegradation
Acute oral toxicity	Activated sludge respiration inhibition test
Acute dermal and/or inhalation toxicity	Hydrolysis test
Skin irritation	Adsorption/desorption screening test

Level 1 for substances at 100 tonnes per annum (or 500 tonnes cumulative), possibly with some studies before: Annex VIII of the directive[21]

One- or two-generation fertility study	Fish bioaccumulation study
in the rat	(or prediction)
Developmental toxicity study	14-day earthworm toxicity study
90-day repeat-dose study in the rat	Test on higher plants
Further mutagenicity study	Further biodegradation test
Basic toxicokinetic information	Further adsorption/desorption study
21-day *Daphnia* reproduction study	Sediment dwelling organism toxicity study
Chronic fish toxicity study	

Level 2 for substances at ⩾ 1000 tonnes per annum (or ⩾ 5000 tonnes cumulative): Annex VIII of the directive[21]

It is not possible to define exactly what testing is required at Level 2, because this depends upon the Level 1 results and Risk Assessment at Base Set and Level 1.

products[18] and plant protection products[25]), there is the opportunity to consult the particular competent authority evaluating the substance before the registration dossier is submitted. In contract with REACH registrations the dossier is submitted to the ECA, and initially there is only to be an administrative check. This situation is rather like the US HPV scheme, where companies propose a test plan, in which they can make proposals for categories, using literature or surrogate data and data waivers. Therefore it will probably be left to the registrant to decide on the final testing programme for new studies under REACH. Nevertheless, certain substances will be evaluated under the REACH scheme in detail by a rapporteur competent authority on behalf of the EU with a view to further testing. In these circumstances it will be useful to base these discussions on the risk assessment included in the original registration in the Chemical Safety Report, supported by appropriate expert reports if necessary to improve the scientific arguments of the case.

13 Common Themes

The first theme is animal welfare. Many companies are reluctant to have animal testing conducted because of pressure from customers. Many regulatory changes are driven by animal welfare and the desire to reduce animal testing. This in turn follows from the scientific development of *in vitro* alternatives to animal tests,[17] and their validation, by ECVAM for the EU and ICCVAM in the USA. For example, the EU has already accepted the results of certain validated positive alternative eye irritation tests for notification without the need for a confirmatory animal study. Nevertheless there has been considerable pressure, especially on UK contract laboratories, well in advance of official EU reforms. For example, in the UK the regulator responsible for licensing laboratories using experimental animals, the Home Office, brokered an agreement in 2000 with UK laboratories to stop all animal testing on cosmetics, yet the associated EU restrictions from the Cosmetics Directive[26] must come into force as late as March 2013. Also, the Home Office virtually banned the traditional guinea pig skin sensitisation tests, in favour of the milder mouse local-lymph node assay (LLNA), but before this new method has gained full international regulatory acceptance.

A related theme is the rapidly increasing use of surrogate data, by predicting properties from chemical structure, other measured properties or from tested analogues. This process can be linked with making use of data waivers justified on the grounds of low exposure. Together with making maximum use of existing published data, "surrogate data" and "data waivers" allow "intelligent toxicity testing", to minimise the cost and animal usage, yet still to have adequate information to assess the hazardous properties and conduct a risk assessment. The consequence is that testing is by no means a routine "box checking" experience, and increasingly more time and effort by even more highly qualified experts are required.

The third and final thread, which again is interlinked with the others, is that risk assessments are increasingly needed and are becoming more and more complex and sophisticated. Such assessments are now needed for EU notifications, and the standard required has gone up, and they are part of EU biocide and plant protection product dossiers. Risk assessments form the basis of EU regulatory policy, and they

are used in some form in all the major chemical or pesticide schemes globally. Hence, preliminary risk assessments are often key in deciding whether to proceed with a product and to refine the testing programme. The consequence for industry and consultants is that expert regulatory affairs professionals experienced in risk assessment are essential. This will become more pronounced with REACH, leading to the suggestion that in time there will be a shortage of (eco) toxicologists and regulatory affairs professionals.

Acknowledgements

This article is based on a paper given at the "Alternatives to Animal Testing are not just *in vitro* Testing" conference organised by the Royal Society of Chemistry Occupational and Environmental Toxicology Group in London on 23 November 2004. Thanks are also due to Karmel Biring of SafePharm Laboratories Ltd for detailed comments and suggestions to improve this chapter.

References

1. D.J. Knight and M. B. Thomas (ed), *Practical Guide to Chemical Safety Testing – Regulatory Consequences: Chemicals, Food Packaging and Medical Devices,* RAPRA, Shrewsbury, UK, 2003.
2. H.P.A. Illing, *Toxicity and Risk, Context, Principles and Practice*, Taylor & Francis Group, London, UK, 2001.
3. *Technical Guidance Document in Support of Commission Directive 93/67/EEC on Risk Assessment for New Notified Substances and Commission Regulation (EC) No, 1488/94 on Risk Assessment for Existing Substances*, European Commission, Brussels, Belgium, 1996 and 2003 (partial update and revision).
4. *EUSES the European Union System for the Evaluation of Substances*, European Commission European Chemicals Bureau Joint Research Centre, Ispra, Italy, 1997.
5. *Classification, packaging and labelling of dangerous substances in the European Union*, European Commission, Luxembourg, 1997, as updated, in particular by the consolidated classification and labelling guide of Commission Directive 2001/59/EC of 6 August 2001, *Off. J. Eur. Communities*, 2001, **L225**, 1.
6. Council Directive 1999/45/EC of 31 May 1999, *Off. J. Eur. Communities*, 1999, **L200**, 1 (as amended).
7. *Guidance on Searching for Chemical Information and Data*, Environmental Protection Agency, Washington, DC, USA (latest version).
8. H.J. Klimisch, M. Andrene and U. Tillmann, A Systematic Approach for Evaluating the Quality of Experimental Toxicological and Ecotoxicological Data, *Regul. Toxicol. Pharm.*, 1997, **25**, 1–5.
9. P. Fisk, Using Structure – Property Relationships for Chemicals in Regulatory Submissions, *Inveresk Research Regulatory Affairs Bulletin*, 2000, **81**, 27–35.
10. P.H. Howard and W.M. Meylan, Prediction of Physical Properties, Transport and Degradation of Environmental Fate and Exposure Assessments, *Quantitative Structure-Activity Relationships in Environmental Sciences*, VII, Chapter 13, SETAC Press, USA, 1997.

11. *QSARs in the Assessment of the Environmental Fate and Effects of Chemicals*, ECETOC Technical Report No. 74, Brussels, Belgium, 1998.
12. W.J. Lyman, W.F. Reehl and D.H. Rosenblatt, *Handbook of Chemical Property Estimation Methods*, McGraw-Hill, 1982.
13. *Manual for Investigation of HPV Chemicals*, Organisation for Economic Co-operation and Development, Paris, France (latest version).
14. *The use of Structure-Activity Relationships in the High Production Volume Chemicals Challenge Program*, Environmental Protection Agency, Washington DC, USA (latest version).
15. *Development of Chemical Categories in the HPV Challenge Program*, Environmental Protection Agency, Washington, DC, USA (latest version).
16. W.M.S. Russell and R.L. Burch, *The Principles of Humane Experimental Technique*, Methuen Publishers, London, UK, 1959.
17. D.J. Knight and D. Breheny, Alternatives to Animal Testing in the Safety Evaluation of Products, *Altern. Lab. Anim.*, 2002, **30**, 7–22.
18. Directive of the European Parliament and the Council of 16 February 1998 concerning the placing of biocidal products on the market (98/8/EC), *Off. J. Eur. Communities*, 1998, **L123**, 1.
19. *Technical Guidance Document in Support of the Directive 98/8/EEC Concerning the Placing of Biocidal Products on the Market Guidance on Data Requirements for Active Substances and Biocidal Products*, Brussels, Belgium, October 2000 (final draft).
20. *White Paper Strategy for a Future Chemicals Policy*, European Commission, Brussels, Belgium, reference COM (2001) 88 final, February 2001.
21. Council Directive 92/32/E EC of 30 April 1992, *Off. J. Eur. Communities*, 1992, **L154**, 1 (as amended).
22. Council Regulation (EEC) No. 793/93 of 23 March 1993, *Off. J. Eur. Communities*, 1993, **L84**, 1.
23. Council Directive 76/769/EEC of 27 July 1976, *Off. J. Eur. Communities*, 1976, **L262**, 201 (as amended and adapted to technical progress).
24. *Consultation Document Concerning the Registration, Evaluation and Authorisation of Chemicals (REACH)*, European Commission, Brussels, Belgium, May 2003.
25. Directive, 91/414/EEC of 15 July 1991, *Off. J. Eur. Communities*, 1991, **L230**, 1 (as amended and adapted to technical progress).
26. Directive 2003/15/EC of 27 February 2003*, Off. J. Eur. Communities*, 2003**,** **L66**, 26–35.

Alternative Tests and the 7th Amendment to the Cosmetics Directive

CARL WESTMORELAND

1 Introduction

Consumers use a wide variety of cosmetic and personal care products in their everyday lives. It has been reported that the average European will use at least six different products in a normal day including the use of items such as soaps, shower gels, mouthwashes, deodorants, shaving products, baby care products, perfumes, make-up and hair care products. In other parts of the world such as Korea, this number can reach 22 products per day. The safety of consumers using cosmetic products is at the core of the European Union (EU) Cosmetics Directive (76/768/EEC[1]).

2 Overview of Current Approaches to Risk Assessment for Cosmetics

The basic requirement of the Cosmetics Directive is that cosmetic products must be safe under normal or reasonably foreseeable conditions of use. An assessment of the safety for human health is therefore performed for each product prior to marketing, taking into account the general toxicological profile of the ingredients, their chemical structure and level of exposure. The toxicological endpoints that are generally assessed for human health are listed in Table 1.

For each of these toxicological endpoints, the risk assessment will comprise a number of pieces of information. The risk to human health is generally regarded as a function of both the hazard represented by the chemicals present in the product and the levels of exposure that the user will have to the product. This is often expressed simply as 'Risk = Hazard × Exposure'.

When performing a risk assessment for a cosmetic product the process involves four distinct phases:

1. *Hazard identification*: This describes the intrinsic properties of the chemical under consideration. Does the chemical have the potential to damage human health?

Issues in Environmental Science and Technology, No. 23
Alternatives to Animal Testing
© The Royal Society of Chemistry, 2006

Table 1 *Toxicological endpoints used in safety assessment of cosmetics products**

1. Acute toxicity
2. Skin irritation and corrosivity
3. Eye irritation and corrosivity
4. Skin sensitisation
5. Dermal/percutaneous absorption
6. Mutagenicity/genotoxicity
7. Carcinogenicity
8. Repeated dose toxicity
9. Reproductive toxicity
10. Toxicokinetics
11. Photo-induced toxicity

Note: *Based on the European Commission's Scientific Committee on Consumer Products (SCCP) guidance document for the testing of cosmetic ingredients and their safety evaluation.

2. *Hazard characterisation*: This describes the relationship between the toxic response and the levels of exposure to the chemical. In many cases, a No Observed Adverse Effect Level (NOAEL) is determined, *i.e.* the highest dose of the chemical that does not cause toxicity.
3. *Exposure assessment*: The amount and frequency of human exposure to the chemical under normal and foreseeable misuse conditions (including consideration of specific groups at risk, *e.g.* children, pregnant women, *etc.*).
4. *Risk characterisation/Risk management*: The probability that the chemical in question in the proposed use scenario will cause damage to human health and to what extent.

In practice, these phases of a risk assessment are not necessarily sequential in occurrence. Prior to generation of any new experimental data on a chemical, the necessary components of the risk assessment will be considered in detail. For example, the nature of the product itself will necessitate consideration of routes of human exposure to the material, *e.g.* is skin contact likely? Is it likely that the product could be inhaled? Is there the possibility of accidental eye contact? Could children accidentally ingest the product? To illustrate this point, when considering relevant endpoints for the safety assessment of deodorant products, the inhaled route of particulate exposure is very relevant to subsequent risk assessment of aerosolised products, but not to solid (*e.g.* roll-on) formulations. This demonstrates that, in performing risk assessments for cosmetic products, relevant exposure data may indicate that hazard data are not always actually required for all of the toxicological endpoints listed in Table 1.

It is also very important before performing a risk assessment to understand the quality of the material that is under consideration. Is it suitable for use in cosmetics? What are the manufacturing and processing procedures used to produce the material? What impurities are present within the material?

2.1 Data Generation to Support Exposure Assessment

If exposure data for a chemical in a particular use scenario indicates that exposure will be low, this will guide what (if any) hazard data will be needed to conduct a robust risk assessment for the particular toxicological endpoint in question. In some cases, additional data can be generated with the test chemical to give a better estimate of likely human exposure. Two examples of such additional data include the use of skin penetration data for chemicals where there is dermal exposure, and respirable dose information for chemicals where exposure via inhalation is a possibility (*e.g.* aerosols, pumps, trigger sprays).

2.1.1 Skin Penetration. Percutaneous absorption can be a major route of entry into the body. If the skin is exposed to a chemical, it is essential to assess the systemic bioavailability of the chemical following exposure to allow accurate risk assessment for its use. The simplest risk assessment can be made by assuming that 100% of a topically applied chemical can be absorbed. For a non-toxic chemical this assumption can result in an adequate margin of safety. In some cases, where the product may be applied to abraded skin (*e.g.* the underarm) it may be necessary to consider risk assessments assuming 100% penetration. Where a more accurate measure of the percentage of a test ingredient that is likely to penetrate the skin is needed, *in vitro* test methods to assess dermal absorption and percutaneous penetration can be used.[2] The use of *in vitro* methods for measurement of skin absorption of topically applied chemicals is included in a recent OECD guideline.[3]

A variety of *in vitro* techniques can be used that measure the amount of material (per cm^2) absorbed into treated, excised skin plus the total amount that penetrates through the skin into a chosen receptor fluid over a defined period of time. These techniques generally require the availability of radiolabelled test material (^{14}C or ^{3}H) to enable detection of the chemical in the skin and receptor fluid. For this reason these techniques are limited to well-defined chemical substances and are of limited use for complex mixtures (*e.g.* plant extracts). The choice of skin used in these experiments can be human skin from cadaver or cosmetic surgery, excised pig skin or rodent skin, depending upon the nature and requirements of the study. If the penetration processes through the skin for a chemical are thought to be passive, and if metabolism by the skin is thought to be unlikely, it is also possible to use adequately frozen human skin for this assay. *In vitro* skin penetration studies to support risk assessment of cosmetic ingredients are designed to provide data that is relevant to the potential human exposure in terms of both test material concentration and the formulation in which the test material is applied to the skin. The absorption of a chemical through the skin can be greatly influenced by the vehicle in which it is delivered. For example, the formulation vehicle in which a sunscreen is presented to the skin can have a significant effect on absorption into and through the skin. Alcohol-based formulations appear to increase sunscreen absorption. In addition, some sunscreen chemicals may enhance the skin absorption of other sunscreens when applied in combination.[4]

Since the barrier properties of skin originate from the low permeability of the stratum corneum, many attempts have been made to predict the ability of a molecule to

penetrate the skin based on its physicochemical properties. *In silico* modelling and quantitative structure activity relationships (QSAR) can be used to give an estimate of a molecule's ability to penetrate the skin.[5,6] Such models are not currently acceptable alternatives to the *in vitro* systems described for use in risk assessments, but may provide useful information in screening strategies. As mentioned above, the formulation in which a chemical is delivered to the skin can have a great influence on its dermal penetration. To date, this parameter is not well modelled in *in silico* predictions.

2.1.2 Respirable Dose. Particles of various sizes are discharged during the use of pressurised aerosols and pump spray products. Some of the airborne particles may be inhaled by the person using the product and by other persons in the vicinity. It is important therefore prior to conducting a risk assessment, to gain an understanding of the exposure that the respiratory tract may have during the use of such products. The respirable discharge from aerosols and pumps is the mass fraction of the particles discharged from the product that are capable of being deposited on the alveolar surface of the human lung. This parameter can be measured analytically from each new product type.

In order to obtain an estimate of the likely exposure that the respiratory tract may have during use of these products, additional information is also required relating to the use of the product itself. The respirable dose is an estimation of the weight of non-volatile material that may be deposited on the alveolar surface of the human lung under simulated 'in use' conditions. By simulating the normal use of specific products (*e.g.* antiperspirants, hairsprays, bodysprays) in a reproducible manner and sampling air from the breathing zone of a mannequin using an aerodynamic particle sizer, it is possible to determine the number and mass of particles that are inhalable (0.2–20 μm) and respirable (0.5–7 μm) in the breathing zone. The estimation of respirable dose is, in part, dependent on reliable information regarding consumer/professional use habits of particular products, which can differ between different populations.

This detailed information on the respirable discharge from the product and the estimated respirable dose for the proposed use scenario can then be used to define further the hazard data that will be needed to support a robust risk assessment of the product. For example, it may be demonstrated that a pump formulation for a particular product produces minimal particles of a novel ingredient that are small enough to be inhaled and therefore this would influence significantly the additional practical work needed to support the subsequent risk assessment for local effects in the lung.

2.2 *Sources of Hazard Data*

When the necessary components of a risk assessment have been determined, ingredient-specific information is then needed on hazard identification and characterisation for several of the toxicity endpoints mentioned in Table 1. Often the manufacturers of the ingredient may have hazard data that can be used in a risk assessment. This data may have been produced for other purposes (*e.g.* for transport or occupational safety of those

producing the chemical) or for the Material Safety Data Sheet supplied with the chemical. Likewise, for registration of new chemicals, toxicity information is required, which may be applicable to use for a risk assessment of that chemical in a cosmetic product.

If little, or no, data are available for a toxicity endpoint of interest, 'expert judgement', 'read-across' and/or (Q)SAR can be used, based on knowledge of the effects of other compounds with comparable structure are used mainly to verify the plausibility of the test results and assess the need for further testing for endpoints. At present these techniques are not normally used to replace experimental data, for which there are formal requirements in the legislation. To use these techniques to provide information with the same confidence as experimental data, a number of considerations need to be addressed. These include the way the surrogate data are obtained (use of QSARs, test batteries, *etc*.), how the data are interpreted in relation to criteria, and the confidence placed on the models used (validation).

2.3 Generation of Hazard Data

In some cases, where hazard data for a specific toxicological endpoint are unavailable and these data are required to perform an adequate risk assessment, new data will be generated on the chemical. The currently accepted methods for safety evaluation studies on chemicals within the EU are published in Annex V of the Dangerous Substances Directive.[7] Toxicology data derived from studies using live experimental animals (*in vivo* studies) form a large part of the dataset used in the hazard identification and risk assessments for the safe use of chemicals and products.

Russell and Burch[8] first put forward the ideas of refinement, reduction and replacement as potential ways to minimise the numbers of animals used in scientific experiments and to prevent unnecessary suffering to animals that are used. The use of alternative methods based on these 'three Rs' principles for the generation of toxicology data has a long history and some impact on animal use. For example, the total annual number of animals (mice, rats, rabbits, hamsters, guinea pigs and fish) used for testing cosmetics in the old Member States of the EU decreased significantly from about 4200 to 1600 from 1998 to 2003. Nevertheless, the 7th amendment to the Cosmetics Directive[9] made it clear that a new approach is needed to achieve the total replacement of animal tests for evaluating the safety of personal care and cosmetic products.

3 7th Amendment to the EU Cosmetics Directive (76/768/EEC)

The 7th amendment to the EU Cosmetics Directive (76/768/EEC) was published in March 2003.[9] This amendment sets out several points relating to the use of animals in the safety testing of cosmetic products and ingredients. An immediate ban has been imposed on animal testing for finished cosmetic products (effective from September 2004). This is because it is recognised that the safety of finished products can already be assessed from knowledge about the safety of ingredients that they contain, and by methods that do not involve the use of animals. An immediate ban was also imposed on the marketing of new cosmetics (finished products and ingredients) tested on animals where validated alternative (non-animal) test methods exist.

Table 2 *Timetable for establishing the deadlines for phasing-out animal testing of cosmetic ingredients within the EU*

Toxicological endpoints	*Cut-off dates (7th amendment)*	*Expert opinion (stakeholders)*	*SCCNFP opinion (1 July 2004)*
1. Acute toxicity	11 March 2009	>2014	>2014
2. Skin irritation	11 March 2009		
For hazard identification		2008–2009	>2014
For risk assessment		>2010	
3. Eye irritation	11 March 2009	2010	>2010
4. Skin sensitisation	11 March 2013	2016–2018	>2019
5. Skin penetration	11 March 2009	2005	>2006
6. Sub-acute/chronic toxicity	11 March 2013	Not estimated	Not foreseeable
7. Genotoxicity	11 March 2009	>2016	>2016
8. UV-induced toxicity	11 March 2009	2009	Not foreseeable
Photogenotoxicity	11 March 2013	>2019	
Photoallergy			
9. TK and metabolism	11 March 2013		Not foreseeable
Without excretion		2014	
Excretion included		2016	
10. Carcinogenicity	11 March 2013	Not estimated	Not foreseeable
11. Reproductive and developmental toxicity	11 March 2013	Not estimated	Not foreseeable

A testing and marketing ban will also be imposed on cosmetics that contain ingredients that have been tested on animals up to 6 years after entry into force of the Directive. The marketing ban is extended to 10 years in the case of three particular types of toxicity tests (repeated-dose toxicity, reproductive toxicity and toxicokinetics) as it is recognised that there are no alternative tests currently under consideration for these endpoints. The exact timetable for establishing the deadlines for phasing-out animal testing was published by the European Commission in September 2004 (Table 2) following a report on the field drafted by an *ad hoc* group of experts[10] and comment from the Scientific Committee on Cosmetic Products and Non-Food Products (SCCNFP).

4 Status of Alternative Tests

As mentioned previously, the majority of accepted methods for generation of toxicological data within the EU currently rely on the use of experimental animals. The development, validation and regulatory acceptance of alternative methods that can be used to reduce, refine or replace the use of animals in testing is an area which has received much attention. Five main stages have been identified in the evolution of alternative test methods: test development, prevalidation, validation, independent assessment and progression towards regulatory acceptance.[11] Within the EU, the main role of European Centre for the Validation of Alternative Methods (ECVAMs) is to coordinate the prevalidation, validation and independent assessment of alternative methods. Importantly, with respect

Table 3 *Current status of alternative methods* for chemicals testing*

Type of test/endpoint	*Status*
Acute toxicity	Reduction and refinement alternatives in use R&D† on-going for replacement alternatives ICCVAM–ECVAM validation study in progress
Skin irritation	R&D on-going for replacement alternatives ECVAM validation study to be completed by end of 2005
Skin corrosion	Validated replacement alternatives in use (OECD TGs 430, 431)
Eye irritation	Refinement alternative (rabbit LVET‡) available but not widely used Replacement *in vitro* alternatives used as in-house screens (partial replacement) and for identification of severe eye irritants R&D (existing methods not formally validated)
Skin sensitisation	Reduction and refinement alternative (mouse LLNA§) validated R&D on-going for replacement alternatives
Phototoxicity	Validated *in vitro* method in use (OECD TG 432)
Skin absorption	*In vitro* methods considered to be valid (no formal validation study conducted), but used alongside *in vivo* studies
Genotoxicity	*In vitro* methods used routinely, but not specifically as alternatives to animal testing (part of accepted testing strategy). Used in combination with *in vivo* studies Validity of *in vitro* micronucleus assay under review
Systemic toxicity and toxicokinetics (repeated dose)	R&D (Basic research on target organ toxicity and repeat dosing *in vitro*) (Computer-based modelling approaches)
Reproductive toxicity	R&D (basic research on *in vitro* systems) *In vitro* screens (partial replacement) for embryotoxicity have been validated
Carcinogenicity	R&D (basic research on *in vitro* systems); validity of *in vitro* cell transformation assay under review, but would not replace rodent bioassays

Notes: *Alternative methods are defined as those that reduce, refine or replace the use of animals in testing; †R&D = research and development; ‡LVET = low volume eye test, §LLNA = local lymph node assay.

to the 7th amendment of the Cosmetics Directive, it is the eventual regulatory acceptance of alternative tests that will allow their use in place of the currently accepted *in vivo* tests. The rest of this chapter will summarise the status of the development of alternative methods for different types of toxicological endpoints. Table 3 contains a summary of the current status of alternative methods for chemicals testing.

4.1 Alternative Tests with Current Regulatory Acceptance

4.1.1 Skin Corrosion. Validated *in vitro* tests already exist for the assessment of skin corrosion (severe and irreversible tissue destruction). These tests are the rat skin transcutaneous electrical resistance (TER) method[12] and tests which employ three-dimensional models of human skin.[13] These *in vitro* tests were evaluated as alternatives

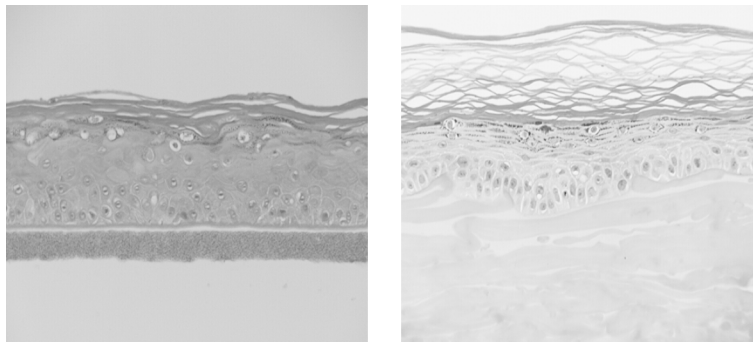


Figure 1 *Histological sections through commercially available three-dimensional cultures of human skin (EpiDerm™ and EPISKIN™)*

to the conventional *in vivo* rabbit skin corrosivity test by ECVAM in a formal validation study[14] where predicitivity of the assays were assessed as well as intra- and inter-laboratory reliability of the methodology.

The three-dimensional models of human skin mentioned above consist of normal, human-derived epidermal keratinocytes that have been cultured to form a multilayered, highly differentiated model of the human epidermis and are commercially available (*e.g.* EPISKIN™ and EpiDerm™). These cells, which are cultured on specially prepared cell culture inserts using serum-free medium, attain levels of differentiation and ultrastructure that closely parallel human skin (see Figure 1). To test a material for its corrosive potential using these three-dimensional models the material is topically applied to the epidermis unit (as either a solid or a liquid) and cell viability is subsequently assessed using the MTT (a tetrazolium salt) assay.

An additional *in vitro* test (CORROSITEX) is also considered valid for testing specific classes of chemicals, such as organic bases and inorganic acids.[15] The use of these validated *in vitro* tests for skin corrosion is now mandatory in the EU for regulatory toxicity testing.

4.1.2 Phototoxicity. If the skin is exposed to chemicals that absorb energy from sunlight, it must be ensured that this chemical excitation does not lead to adverse skin reactions. Many types of chemicals induce phototoxic effects and their common feature is the ability to absorb light energy within the sunlight region. Therefore, before any biological testing is considered, a UV/visible absorption spectrum of the test chemical must be determined. If the molar extinction/absorption coefficient is less than 10 L mol^{-1} cm^{-1}, the chemical has no photoreactive potential and does not need to be tested for phototoxicity. If, however, a test is needed, a validated *in vitro* test already exists for the assessment of phototoxicity. This test is the Neutral Red Uptake Phototoxicity Test using 3T3 Mouse Fibroblasts.[16] Cytotoxicity in this test is expressed as a concentration-dependent reduction in the uptake of the vital dye, Neutral Red, 24 h after treatment with the test chemical. Treatment occurs both with and without exposure to a non-cytotoxic UVA/vis light dose of 5 J cm^{-2} UVA. To predict the phototoxic potential of the test chemical, the cyotoxicity responses obtained

in the presence and absence of irradiation are compared. This standardised test protocol is internationally accepted and mandatory for testing chemicals in the EU.

It has been suggested that a tiered testing strategy involving the methods described above could be used for the assessment of phototoxicity when equivocal results are obtained in the 3T3 cell Neutral Red uptake phototoxicity test. To evaluate further the phototoxic hazard in skin, a phototoxicity assay using a three-dimensional human skin model could be considered.[17]

There are no standard guidelines for testing other potential consequences of photoactivation (*e.g.* photoallergy, photogenotoxicity or photocarcinogenicity), though *in vitro* assays for photogenotoxicity are considered to be valid and used when appropriate.

4.1.3 Skin Penetration. While an OECD guideline exists for *in vivo* animal testing to determine skin penetration, as mentioned earlier, the use of skin penetration data derived from *in vitro* systems is widely used in risk assessments for cosmetic products and also has regulatory acceptance.[3]

The OECD test guideline recommends the use of excised human or pig skin for the *in vitro* investigation of dermal absorption and penetration, and excised rat skin can also be used. The use of *ex vivo* tissues for routine testing is not ideal from an animal use perspective, but a reliable source of human tissue for routine testing is not always possible. The three-dimensional models of human skin mentioned above have recently been examined as a potential alternative source of tissue for use in skin penetration studies.[18] Such reconstructed skin models could be acceptable if their suitability can be shown by appropriate correlation studies. However, at present, the barrier function of such models is not sufficiently well developed for use in penetration assays. In particular, chemicals with lower octanol–water partition coefficients have distinctly higher penetration rates through the three-dimensional human skin models than they do through excised pig skin. For example sodium dodecyl sulphate has approximately 100 times higher permeability with EpiDerm™ in comparison to split thickness pig skin.

5 Endpoints where Risk Assessment for Cosmetics can be Achieved without using Animal Data

5.1 Skin Irritation

Animal data on acute skin irritation typically form a part of the mandatory testing package required for notification of new chemical substances (Draize rabbit skin irritation test.[19,20] In the standard animal test, 0.5 mL or 0.5 g of the chemical is tested on the skin of rabbits for four hours and then the site examined for irritation at several timepoints up to 72 h later.

For many cosmetic and personal care products, the skin is often exposed, either intentionally or unintentionally, and therefore a risk assessment regarding the potential to cause irritation of the skin is needed prior to marketing. For such a risk assessment of human skin irritation potential of cosmetic and personal care products, a strategy is generally used that avoids the use of animals and at the same time delivers

a more relevant assessment of skin irritation potential than is possible using data generated in rabbits. Any data that are generated in the species of interest for the risk assessment (*i.e*. humans) will avoid the need for inter-species extrapolation. It has been shown that rabbit skin can both underestimate and overestimate the irritant potential for chemicals in human skin.[21] The approach is therefore to identify corrosive materials (that have the potential to cause severe irreversible damage) *in vitro* and then to proceed to human volunteer testing for skin irritation potential. As described earlier, to conduct a robust risk assessment, data on both hazard identification and characterisation are necessary. Skin irritation studies in human volunteers can be designed to provide information on the primary skin irritation potential of test and control products and to compare results with a similar formulation, *e.g*. a relevant market leader.[22] The exposure used in these human studies is often an exaggeration of in-use exposure to enable a conservative risk assessment to be performed. Typical exposure scenarios used in human skin irritation testing include 48 and 96 h patch tests, arm immersion tests and 21 day repeat insult patch tests for cumulative skin irritation. For particular 'in-use' scenarios of some products human studies can be tailored to provide data relevant to that product type (*e.g*. studies in the axilla for risk assessment of underarm deodorants).

As with all work conducted in humans, volunteer studies for risk assessment of skin irritation must meet any local legal and ethical requirements and conform fully to the principles of the Declaration of Helsinki.[23] Substances should not be tested in humans when they have been shown (or predicted) to be corrosive, or if they present any likelihood of skin sensitisation, any acutely toxic hazard and/or any genotoxic hazard.

Although the use of human data in risk assessment for skin irritation potential of cosmetic products is well established, the use of the *in vivo* regulatory assay remains the standard way to generate data for hazard identification of chemicals. It has been suggested that the use of a 4 h patch test in humans may also be suitable for such hazard identification purposes.[24]

Despite the widespread use of *in vitro* models for studying acute skin irritation for internal decision-making within companies, to date, there are no *in vitro* tests that have been formally validated to allow their use in regulatory submissions.[25] Three *in vitro* tests (EpiDerm™, EPISKIN™ and the mouse skin integrity function test) are currently under evaluation in a formal validation study funded by ECVAM during 2003/2005 following successful prevalidation studies with these models.[15]

It is recognised that new, more predictive, *in vitro* tests for skin irritation also need to be developed,[25] based on better understanding and modelling of the mechanisms of human skin irritation. Such mechanistically based tests could provide better information when carrying out new risk assessments. Currently, the three-dimensional human skin models appear to be the most promising test systems for assessing skin irritation *in vitro*. Ultimately, it is hoped that the use of a battery of markers for inflammation in these models may deliver more relevant models of skin irritation that can also be used to address specific problems (*e.g*. reversibility of effects, effects of repeat dosing). The application of genomics and proteomics technologies may aid in the identification of specific biomarkers. This promising approach is currently

being progressed in an on-going research programme supported by the European Cosmetics, Toiletries and Perfumery Association (Colipa).[26]

5.2 Eye Irritation

As for the skin, the eye can also be exposed to cosmetics and personal care products either intentionally or unintentionally. It is assumed that the validated *in vitro* tests for skin corrosion would also identify chemicals that would be corrosive to the eye.[27] However, there are no validated alternatives to the acute eye irritation test which typically forms part of the mandatory testing package required for notification of new chemical substances (Draize eye irritation test[28]). A refinement alternative to the traditional Draize eye irritation test (the Low Volume Eye Test[29]) has been developed, but is not widely used despite it providing some benefits from an animal welfare viewpoint.

A large number of *in vitro* models exist for studying eye irritation. These assays can be grouped into *ex vivo* target organ/tissue assays (*e.g*. Bovine Corneal Opacity and Permeability Test (BCOP), Isolated Rabbit Eye Test, Chicken Enucleated Eye Test), organotypic models (*e.g*. the Hen Egg Test-Chorioallantoic Membrane (HET-CAM) assay), cytotoxicity assays (*e.g*. Neutral Red Assays, Red Blood Cell Lysis Assay) and chemical reaction assays (*e.g*. EYTEX).[30] To date, none of these methods has met all the formal validation requirements for replacing the currently accepted animal test in a regulatory context. Isolated eye tests and the BCOP are accepted by several European regulatory authorities (as well as the HET-CAM assay in Germany) on a case-by-case basis for the identification of severe eye irritants for the purposes of classification and labelling within EU directives on raw materials and products.[27]

The in-house use of some of these *in vitro* models (primarily the isolated eye tests and the BCOP) for prediction of eye irritation potential within the cosmetics industry is well established. As for skin irritation testing, risk assessment for eye irritation potential is often carried out using these *in vitro* models using a benchmarking/comparative risk assessment approach where the response of a new chemical/formulation is compared to reference samples with a similar in-use pattern in humans of known irritant potential. The limitations of these *in vitro* assays (*e.g*. lack of blink reflex, no ability to study recovery, *etc*.) are well documented and must be taken into account when using data from these studies.

As mentioned above, none of the methods currently available for studying eye irritation *in vitro* has received regulatory acceptance as a valid alternative to the Draize eye irritation test. This is due, in part, to the lack of understanding of the underlying physiological mechanisms of eye irritation. Further research in this area should enable the identification of new *in vitro* endpoints that are more predictive of the *in vivo* human response to chemical injury. Questions being addressed by on-going research funded via Colipa include: Can investigation into the kinetics/patterns of change in physiological function, and signals of injury released from the cornea *in vitro*, be used to predict a chemical's potential to damage the eye and give information on the ability of the eye to recover from this damage? Can the use of human immortalised cells and three-dimensional human conjunctival and corneal constructs provide better *in vitro* models for studying eye irritation?[31]

6 Endpoints where Risk Assessment for Cosmetics cannot currently be Achieved without using Animal Data

6.1 Skin Sensitisation

Skin sensitisation, resulting in allergic contact dermatitis is a relatively common adverse skin reaction. It arises as a result of repeated skin contact with a wide range of chemicals encountered in the environment and is an important occupational and consumer health problem. Currently, the only methods for detecting skin sensitisers involve the use of animals.

Traditionally, guinea pigs have been used to detect a chemical's ability to cause skin sensitisation. Two types of tests using guinea pigs have been developed: the adjuvant-type test, in which an allergic state is potentiated by dissolving the test chemical in Freund's complete adjuvant (*e.g*. the guinea pig maximisation test), and non-adjuvant tests (*e.g*. the Beuhler test). In both cases, the test animals are exposed to the chemical by intradermal injection and/or epidermal application. Following a rest period of 10–14 days, during which an immune response may develop, the animals are exposed to a challenge dose. The extent and degree of the skin reaction to the challenge dose is then compared to that of control animals.

A recent addition to the regulatory guidelines for detection of skin sensitisers is the murine local lymph node assay (LLNA).[32] This method provides a reduction and refinement alternative to the conventional tests used to detect skin sensitisers that use the guinea pig. The advantages which the LLNA has over these traditional guinea pig tests are that it uses approximately half the number of animals, and does not require the use of Freund's complete adjuvant, intradermal injection, fur removal, occlusive dressings or the elicitation of an allergic skin reaction.[33] Results from chemicals tested in the LLNA compare favourably with those from guinea pig tests and show that the LLNA is at least as good as the guinea pig in predicting human responsiveness.

An additional advantage of the LLNA over the guinea pig tests is that a quantitative test outcome is obtained that can be used to provide more refined risk assessments for skin sensitisation. Chemicals that are skin sensitisers are capable of causing allergic contact dermatitis in humans. However, this is not an all-or-nothing phenomenon, clear dose-response relationships occur and thresholds can be identified for the induction of skin sensitisation.[34] The parameter used in the LLNA to identify the sensitisation potency of a test chemical is the concentration required to produce a 3-fold increase in draining lymph-node cell proliferative activity in the treated mice (EC3 value). The derivation of the EC3 value for a chemical provides an objective and quantitative estimate of potency (hazard characterisation information) that can then be used with subsequent risk assessment calculations for cosmetic products, together with the information on the concentration of the test chemical in the proposed product and details of the human exposure scenarios.

At present there are no non-animal-based assays that can be used to identify the skin sensitising potential of a novel chemical. The search for alternative methods to detect skin sensitisers has focussed on the use of computer systems (*e.g*. QSAR) and *in vitro* assays.

For a chemical to be a skin sensitiser, there is a recognition that the chemical (or a metabolite of the chemical formed within the epidermis) should possess reactive

groups capable of modifying proteins and thus causing an immune response.[35] The physicochemical properties of a molecule are key to eliciting the relationship between chemical structure and the ability to cause sensitisation. This provides evidence of the value of structure–activity relationships and QSARs could have in identifying mechanisms of sensitisation without the need for extensive animal testing. At present, several systems exist for predicting sensitisation potential from chemical structure, but none is currently validated for use as an alternative to animal testing.[36] One example of such a system is Deductive Estimation of Risk from Existing Knowledge (DEREK).

There are several potential targets for research into *in vitro* assays for skin sensitisers based on the pathway by which chemicals cause skin sensitisation (see Figure 2). For a chemical to be a skin sensitiser, it must be able to penetrate into the viable epidermis, to react with skin proteins in the epidermis, to cause cytokine release stimulating Langerhans cell migration/maturation and final recognition by T-lymphocyte receptors.[37] Since a chemical must react with protein to become a skin sensitiser, the modelling of direct protein reactivity has recently become a focus of research effort which may eventually lead to tests which could be used in a testing strategy for skin sensitisers. For example, Gerberick *et al*.[38] have tested 38 chemicals representing allergens of different potencies (weak to extreme) for their ability to react chemically with glutathione or three synthetic peptides containing either cysteine, lysine or histidine. This initial work has indicated that measuring peptide reactivity could be useful for screening chemicals for their skin sensitisation potency and thus potentially reducing the reliance on animal-based test methods.

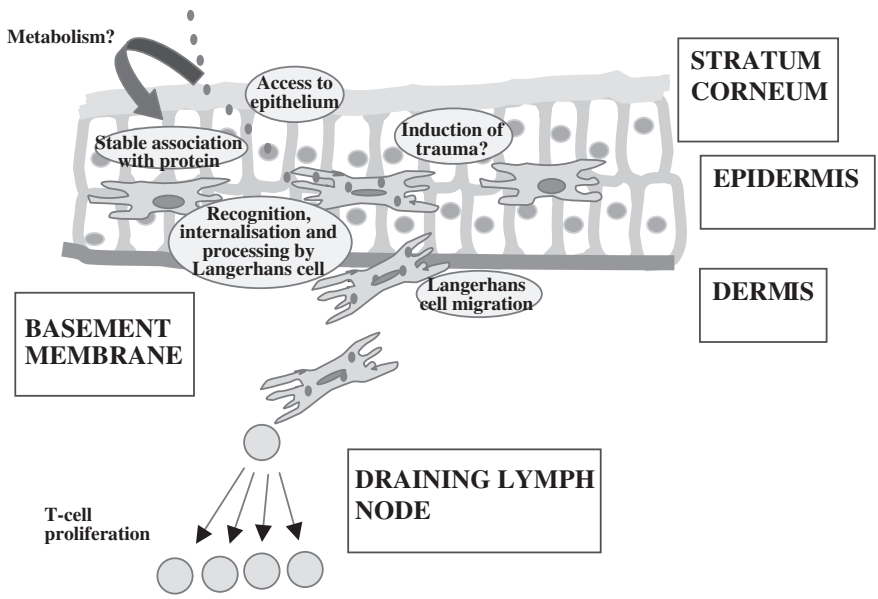


Figure 2 *Diagrammatic representation of the molecular events thought to be involved in the process of skin sensitisation*

The main efforts to develop cell culture-based *in vitro* alternative for skin sensitisation testing have concentrated on the Langerhans dendritic cells since these are the key mediators of contact allergenicity in the skin.[27] This research has not resulted in any specific tests that could be used to replace the use of animals, but has given several new leads into areas for future research that may ultimately provide the mechanistic data required to design future testing strategies. Several human monocytic cell lines (*e.g.* THP-1, MUTZ-3 and U-937) that are thought to possess some characteristics of human Langerhans cells are currently being evaluated for their ability to predict the sensitisation potential of chemicals *in vitro*. Measurement of expression of specific cell surface markers by these cells (*e.g.* CD86) following treatment with sensitising chemicals, but not following treatment with non-sensitisers, is currently showing promising results.[39,40]

6.2 Genetic Toxicology

The assessment of the genotoxic potential of a chemical determines the ability of the chemical to induce changes in a cell's DNA as a consequence of gene mutation, chromosome damage or alteration in chromosome number. Genotoxic compounds may be carcinogenic as it is assumed that mutation is a key event in carcinogenesis. In addition, genotoxic chemicals may cause genetic damage in germ cells and therefore have the potential to increase the incidence of genetic diseases. A number of computer-based expert systems exist to predict genotoxicity based on chemical structure (*e.g.* Computer Automated Structure Evaluation and DEREK[27]). These provide valuable tools in the tiered testing strategy for identification of genotoxins, but (as with QSARs for skin sensitisation) are not currently validated for regulatory purposes.

For testing of chemicals for use in cosmetics and personal care products, the genotoxicity-testing strategy usually applied uses at least two *in vitro* tests to detect gene mutations and chromosome damage (clastogenicity).[41] These *in vitro* tests include one bacterial test for microbial mutagenicity (*e.g.* the Ames test using *Salmonella typhimurium*) and one test in mammalian cells. There are three main tests in mammalian cells that are currently used: (1) metaphase analysis in cultured mammalian cells (*e.g.* human peripheral lymphocytes or Chinese hamster ovary cells) to detect chemicals that damage chromosome structure, (2) gene mutation tests in cultured mammalian cells (*e.g.* mouse lymphoma assay in L5178Y tk^+/tk^-cells), and (3) *in vitro* micronucleus test in cultured mammalian cells (*e.g.* human peripheral lymphocytes or Chinese hamster ovary cells) to detect chemicals that damage chromosome structure and cause alterations in chromosome number. As many carcinogenic chemicals are thought to be activated to reactive molecules that interact with DNA following metabolism in the body, all of the *in vitro* tests described above are carried out in both the presence and absence of an exogenous metabolising system (usually rat liver S9).

The performance of these *in vitro* genotoxicity tests has recently been evaluated for their ability to discriminate between rodent carcinogens and rodent non-carcinogens.[42] Generally, the sensitivity of these tests (*i.e.* ability to detect carcinogens) is high. Combinations of 2 or 3 of the test systems have a greater sensitivity (approximately 90%) than individual tests. However, the specificity of these tests (*i.e.* ability to accurately predict non-carcinogens) is poor. When 3 tests are performed 75–95% of non-carcinogens gave positive results (*i.e.* false positives) in at least one test in the battery.

For chemicals where negative results are obtained in *in vitro* genotoxicity tests, no further testing is required and therefore no animals are used in the genotoxicity testing battery. However, for those chemicals where positive *in vitro* genotoxicity results are obtained, at present the next stage in most genotoxicity test strategies is to determine whether the genotoxicity that is observed *in vitro* is realised in the intact animal, *i.e*. are the positive results obtained *in vitro* confirmed as positive *in vivo*? *In vivo* genotoxicity tests usually detect cytogenetic damage in the rodent bone marrow following oral or intraperitoneal dosing (*e.g*. mouse/rat micronucleus test), but other tests do exist for studying additional organs (*e.g*. rat liver unscheduled DNA synthesis test). A negative result in an *in vivo* genotoxicity assay together with an understanding of the reason why a positive result was obtained with that chemical in an *in vitro* assay may provide sufficient evidence to conclude that the chemical is unlikely to pose a genotoxic hazard if included in a cosmetic or personal care product.

There are many reasons why chemicals may give positive results in *in vitro* genotoxicity assays that do not then translate to the intact animal. For example, the metabolising system used in *in vitro* genotoxicity assays does not mimic completely that of the intact animal; phase II (detoxifiying) enzymes, for example, are poorly represented in rat liver S9. Chemically-induced cytotoxicity and alterations in pH/osmolality of culture media are also known to affect the results of *in vitro* genotoxicity assays. In addition, the DNA repair mechanisms and the anti-oxidant capacity of the cells used in the *in vitro* assays are not well understood in relation to the *in vivo* situation.

In vivo genotoxicity assays are used to determine whether the *in vitro* assays have given a false positive response (*e.g*. due to lack of appropriate metabolism in the *in vitro* test). As such, research in the area of replacing *in vivo* genotoxicity assays in line with the 7th amendment to the Cosmetics Directive is concentrating on improving the predictivity of current *in vitro* tests and the development of novel *in vitro* approaches to provide a greater understanding of the mechanisms of genotoxicity and their relevance to cancer in humans.

6.3 Respiratory Toxicity

Many cosmetic and personal care products such as pumps, sprays and aerosols release particles and droplets into the atmosphere, some of which have the potential to be inhaled into the respiratory tract of the user or other individuals. It is therefore necessary to conduct a risk assessment for respiratory toxicity for each of these products. As mentioned earlier, the first step of such risk assessments is to measure the respirable discharge from the product and to conduct an assessment of the respirable dose in the proposed use scenario. If significant exposure to the respiratory tract is predicted from these calculations, hazard data concerning effects on the respiratory system are required to conduct a robust risk assessment.

The safety assessment of chemicals for inhaled effects currently involves determining the type of lung pathology that can be caused in animal inhalation exposure studies, and establishing the NOAEL for these pathologies.[43] The most commonly used animal study for such investigations is a 3-month inhalation study in the rat. An important part of this process is to recognise the intrinsic differences between

rodents and humans in terms of the respiratory doses that each species experiences during inhalation exposures. Interspecies scaling factors become necessary when comparing the exposure doses experienced by rats, compared to humans, because of basic differences between species in lung clearance rates and alveolar area in the lungs. Of particular relevance to the design of safety studies to support products that contain particulate matter is the fact that chronic pulmonary inflammation, pulmonary fibrosis and even lung tumours can develop in inhalation studies in the rat with highly insoluble non-fibrous particles of low cytotoxicity.[44] These responses are due to excessive particulate lung burdens, and the term 'particle overload' has been used to describe this condition which is characterised by the impairment of alveolar macrophage-mediated lung clearance. Significant species differences exist with respect to the induction of adverse chronic effects in response to lung overload.

The assessment of inhalation toxicity is currently not possible without animal testing. Most research in this area has concentrated on the target cells from the conducting airways and relevant toxicity endpoints that could be measured in such cells *in vitro*. An ECVAM workshop on the subject of *in vitro* alternatives for inhalation toxicity testing[45] envisaged a tiered strategy that would consist of literature searching and measurement of several physicochemical parameters prior to a 2-stage *in vitro* testing strategy. The first stage would consist of identification of likely target cells by the use of culture systems such as tracheal rings, lung slices and cultured alveolar macrophages. The second stage would then consist of a more detailed battery of tests selected from the results of the first stage and potentially including cultures of nasal olfactory cells, airway epithelial ciliated or non-ciliated cells, Clara cells, type II cells, alveolar macrophages, vascular endothelial cells, fibroblasts and mesothelial cells.

There are still many technical challenges associated with studying toxicity in the respiratory tract *in vitro*. Representative cultures of some of the cell types from the respiratory tract are not easily available for routine testing. Advances in tissue and cell engineering may provide some useful tools in this area for the future. For example, Type II alveolar cells (responsible for the synthesis, secretion and re-uptake of surfactant and essential for normal lung function) are difficult to culture *in vitro*, and cell lines that are available to model Type II cells (*e.g.* the tumour cell line A549) are, in many cases, phenotypically very different from the target cell *in vivo*. Recent advances in coculture of stem/progenitor cells have shown some initial promise in maintaining the differentiation of Type II cells, as shown by the presence of surfactant protein C.[46] Another advance in this area has been the commercial availability of three-dimensional models of human respiratory tract tissue. These models (*e.g.* EpiAirway™) have many of the features of *in vivo* airways, including ciliated epithelium, mucin-producing goblet cells and the presence of tight junctions.

As well as issues surrounding test systems for *in vitro* investigations of inhalation toxicity, delivery of aerosols/sprays to *in vitro* systems requires considerable thought to allow adequate information on dosimetry for future use in risk assessment decisions. Likewise, the selection of appropriate markers of toxicity within the cultures and their subsequent interpretation in the context of a risk assessment for cosmetics and personal care ingredients represents a major challenge for the future. Studies are currently ongoing with model respiratory toxins using transcriptomic and proteomic

technologies alongside conventional histological analysis that hope to identify early markers of change at the gene/protein level that could be used in subsequent *in vitro* studies.

In addition to causing local effects in the respiratory tract, inhaled materials may also be absorbed through the lining of the lung into the systemic circulation. At present, using *in vivo* studies, the systemic effects of inhaled materials can be investigated in the same animals used for assessment of local effects in the lung. In the absence of animal tests however, an assessment must first be made of how much of the test material is likely to penetrate through the lungs into the systemic circulation. *In vitro* models using airway cell lines (16HBE14o- and Calu-3) and primary cultured human alveolar type I like cells are currently being investigated as absorption models for this purpose. Progress is being made towards *in vitro–in vivo* correlation for pulmonary absorption and in the use of cultured respiratory cells to evaluate metabolism, toxicity and drug targeting strategies.[47]

6.4 Systemic Toxicity/ Toxicokinetics

As mentioned earlier, skin absorption and lung absorption of ingredients from dermally applied or inhaled cosmetic and personal care products can result in systemic exposure to those ingredients. The same is true of products where exposure is via the oral route (*e.g*. toothpaste, mouthwash). If significant systemic exposure to a chemical is expected, a risk assessment for potential systemic effects must be carried out to assure safety.

The identification of target organ toxicity following repeat dosing (14–28 and 90 days) of a chemical to rats forms a key part of the data used to ensure chemical and product safety. A large number of clinical, biochemical, haematological and histological measurements are made during these studies. An NOAEL is often generated in these studies to enable calculations of Margins of Safety when conducting further risk assessments. The association between changes in target organs, duration of dosing, administered dose and the absorption, distribution, metabolism and excretion (ADME) of the test compound cannot be considered in isolation as all of these parameters are inter-dependent. The development of alternative methods to determine the systemic toxicity of chemicals as well as to adequately predict their toxicokinetics, presents a considerable scientific and technical challenge. It is likely that an integrated approach to chronic toxicity testing based on the use of alternative methods with complementary endpoints will need to be developed.

A number of *in vitro* models exist to model various aspects of the ADME profile of a chemical. Many liver-based models allow the study of the metabolic fate of chemicals, *e.g*. microsomal preparations, isolated hepatocytes, liver slices, complex long-term culture of hepatocytes and genetically engineered cell lines.[48] Such systems can provide information on the intrinsic clearance of molecules, identify routes of potential metabolism and may be useful in the prediction of enzyme induction/inhibition. There are also several computer models that can predict possible routes of metabolism.[49] As mentioned earlier, *in vitro* methods can be used for assessment of skin penetration. Work is also currently on-going to define similar models to enable prediction of gastrointestinal absorption and the ability of molecules to penetrate the blood/brain barrier.[27]

A vast number of *in vitro* models exist to study different types of target organ toxicity, *e.g*. hepatotoxicity,[50] nephrotoxicity,[51] haematotoxicity[52] and neurotoxicity.[53] There are many examples where such systems have been useful in screening situations to rank chemicals within a series to identify their ability to induce specific types of toxicity. Likewise, such models can be very useful to identify the mechanistic basis of known toxicities and to establish the relevance to man of toxicity data derived in animals. The use of *in vitro* models in these ways can be valuable in toxicity screening programmes and will significantly reduce the number of animals used for this purpose. However, it has been recognised that many *in vitro* systems are not suitable for studying sub-chronic and chronic toxicity due to their short lifespan. The ECVAM workshop on long-term toxicity identified that the long-term maintenance of *in vitro* systems will require the use of new cell culture methods that allow for repeat or continuous administration of test compounds.[54]

The use of human-derived proteins, cells and tissues in these circumstances can provide valuable data on the response of humans. For each of the above uses of these tissue-specific *in vitro* models, prior knowledge that the chemical (or class of chemicals) causes toxicity in that particular organ is needed. Data from conventional animal tests or evidence of toxicity in humans from epidemiological or clinical trial data are currently used for this purpose.

While these *in vitro* models can provide invaluable information on a case-by-case basis as indicated above, the large number of endpoints and the biological complexity of the intact organism in toxicological testing mean that replacement of animal tests for the evaluation of long-term toxicity is difficult to envisage at present. In the context of complete replacement of animals in repeat-dose toxicity studies, we will be asking different questions from those currently addressed using *in vitro* models for target organ toxicity such as those listed above.

If absorption studies have indicated that systemic exposure to the test material will occur, the initial question to ask will then be ‘what target organs are affected by the chemical?’ Table 4 contains a list of tissues that can potentially be examined in repeat dose animal studies. At present, histological examination of these tissues together with clinical chemistry identifies which organs have been affected. In the absence of animal experimentation, could we rely on a battery of *in vitro* tests to do this? This is an important point to address before concentrating on how the individual target organs could be studied. It is not feasible to expect that, for each new chemical, work would be undertaken *in vitro* on all of the potential target organs. How then would we identify, for a novel chemical, which target organs it could potentially affect in humans? For example, if reliable models existed for predicting adverse effects in the kidney, liver and brain, but not for the pancreas and bladder, how would it be possible to perform an adequate risk assessment as to whether the likely human exposure to the chemical was safe?

In addition to identifying which target organs should be studied, *in vivo* toxicity is often the sequential result of a primary toxicity to the vascular endothelium allowing the toxin to reach a secondary target cell, *e.g*. pulmonary epithelial and hepatocellular toxicity following vascular endothelial damage. This also includes chronic effects such as pulmonary fibrosis due to vascular damage, *etc*. This demonstrates the importance of modelling organ structures that are critical to target organ toxicity and cell interactions as part of the pathogenesis of the toxicity in question.

Table 4 *Organs that can be examined histologically following a repeat dose toxicity study*

Adrenals	Jejunum	Skin with mammary glands
Aorta	Kidneys	Spinal column with cord
Brain	Larynx and oropharynx	Spleen
Caecum	Liver	Sternum with marrow
Cervical Lymph Node	Lungs	Stifle joint/femur/bone marrow
Colon	Mesenteric lymph node	Stomach
Duodenum	Oesophagus	Testes
Epididymides	Ovaries	Thymus
Eyes and optic nerves	Pancreas	Thyroid and parathyroids
Harderian gland	Peripheral nerve	Tongue
Middle/inner ear	Pituitary	Trachea
Nasal chambers	Prostate	Tracheal bifurcation
Nasopharynx	Rectum	Tracheobronchial lymph node
Zymbals gland	Salivary glands	Urinary bladder
Heart	Seminal vesicles	Uterus
Ileum	Skeletal muscle	Vagina

There is perhaps a role for QSARs in the future for identification of which target organs, a chemical may affect, based on structural similarity to chemicals where the toxicity profile is already well defined. It is also possible that physiologically-based biokinetic (PBBK) models might be useful to determine how a chemical is likely to be distributed within a body and therefore give some potential information on likely target organs where toxicity may occur. PBBK models are mathematical models that can describe the biokinetic parameters of a molecule within an organism based on knowledge about both the test chemical, in terms of *in vitro* data (*e.g*. clearance rate from the liver) and physicochemical parameters (*e.g*. octanol/water partition coefficient), as well as physiological information on the organism (*e.g*. body weight, liver blood flow, *etc*.).[55] Such models are in the early stages of development but may, in the future, provide the ability to predict tissue/organ levels of chemicals and allow the integration of *in vitro*-derived ADME data with specifically tailored *in vitro*-derived target organ toxicity data.

If it is possible to identify the target organs to study *in vitro* for a new ingredient, a decision would then need to be taken regarding which markers of toxicity are relevant to the chemical. Traditionally this has been done with reference to known compounds in the same (or similar) chemical class. For example, in the liver, when studying peroxisome proliferators, toxicity endpoints such as palmitoyl coA oxidation have been measured *in vitro*, when studying hepatotoxins that cause steatosis, toxicity endpoints such as triglyceride secretion have been studied *in vitro*, and when studying hepatotoxins that cause porphyria, haem synthesis and ferrochelatase activity have been measured *in vitro*. To cover all possible effects in a specific organ, a large battery of endpoints would be needed. This could indicate that there is a need to move away from traditional biochemical-based endpoints that are tailored to detecting specific types of toxicity. Transcriptomic/proteomic changes that cover all

aspects of the target-cell biology to identify basic pathways/networks that a chemical affects could provide valuable new tools in this area.

Another important point to address will be to determine the relevance of concentrations of the chemicals tested *in vitro*. For example, for an untested chemical that is a suspected nephrotoxin, what concentrations should be tested in an *in vitro* test? The eventual replacement of repeat-dose toxicity tests must involve some integration of *in vitro* data on target organ toxicity with *in vitro*/*in silico* data on ADME properties, *e.g*. a new chemical may be structurally related to others known to cause nephrotoxicity. In a rat study following oral dosing, however, this compound may not be nephrotoxic because it is not absorbed through the gut into the systemic circulation. If this compound were tested *in vitro* (*e.g*. in LLCPK1 cells) it may be toxic at high concentrations. This sort of example does not mean that the *in vitro* data on the compound are wrong, but that the kinetics of the chemical in the whole organism mean that the nephrotoxic potential of the compound does not translate to the *in vivo* situation. It is important therefore that the concentrations of the chemical tested *in vitro* relate to the predicted tissue levels for that organ. In this respect integration of toxicokinetic data (*e.g*. from PBBK modelling) into any replacement test strategy needs to be ensured.

The most important point to address when considering the use of alternative tests for repeat dose toxicity is how data from *in vitro* target organ toxicity tests could be used in risk assessment. As mentioned above, the hazard assessment from repeat dose animal studies will usually be expressed in terms of an NOAEL. For the subsequent risk assessment regarding human safety, the required margin of safety over the NOAEL observed in animals is dependent upon the nature of the adverse effect seen during the animal study. An appropriate figure is agreed on in discussion with experts on the effects seen. Commonly, a 100x margin of safety is applied between the worst case predicted dose in humans (mg/kg) and the NOAEL in animals (mg/kg). In some cases, a 100-fold margin of safety may not be sufficient due to the nature of the toxic effects seen during the study (*e.g*. if neurotoxic or reproductive toxicity effects were noted).

For ranking of chemicals and understanding mechanisms of toxicity where *in vitro* models of target organ toxicity have successfully been used in the past, IC50 values (the predicted concentration of a test compound that reduces the measured parameter by 50%) have been used extensively. If data from *in vitro* tests can ever be used (even in part) to replace repeat dose toxicity studies in animals, new risk assessment paradigms must be developed to use the data from these studies in assessing risk to humans.

The questions outlined above are only a representation of some of the enormous challenges faced for replacing animal tests for systemic toxicity and toxicokinetcs.

7 New Approaches to Risk Assessment without Animals

In the future it is plausible that new paradigms could be developed to enable risk assessment to support decisions on consumer safety of cosmetic and personal care products without the need to generate new data in animals.[56] Such new paradigms may not rely on direct one-for-one ‘replacement’ of the current animal studies used to provide data to make risk assessment decisions. The availability of technologies that did not exist 10 years ago makes this new approach plausible. Such techniques

include both improved *in vitro* models (*e.g*. recent advances in tissue engineering primarily within the field of regenerative medicine) and technologies that allow far greater insights into changes within biological systems at a molecular level (*e.g*. transcriptomics, proteomics and metabonomics).

Clinical medicine is using the opportunities offered by these new 'omics' technologies to advance the understanding of disease. The application of these technologies in clinical medicine will generate massive amounts of data that will need processing and interpretation to allow clinicians to better diagnose disease and understand the patients' responses to therapeutic intervention. A new approach to support decisions on consumer safety may be based on the concept that data and information derived from applying existing and new technologies to non-animal models can be interpreted in terms of disease and harm in man. A prerequisite is that similar data and information generated in a clinical setting are available to permit this 'translation'. If these clinical data and information can be made generally accessible in a legal and ethical way they should also permit the translation of experimental, non-animal data, so that they can be used in risk assessment.

The processing, interpretation and translation of these data need to be supported by powerful informatics capabilities and statistical tools. The use of integrated 'systems biology' approaches will further support the interpretation by providing an increased understanding of the underlying biological complexity and mechanisms of toxicity.

Acknowledgements

I would like to acknowledge Dr Julia Fentem and many of my colleagues at Unilever for their valuable advice and discussions as part of Unilever's ongoing effort to develop novel ways of delivering consumer safety.

Abbreviations

ADME	Absorption, distribution, metabolism and excretion
BCOP	Bovine corneal opacity and permeability
DEREK	Deductive Estimation of Risk from Existing Knowledge
ECVAM	European Centre for the Validation of Alternative Methods
LLNA	Local lymph node assay
NOAEL	No observed adverse effect level
PBBK	Physiologically-based biokinetic
QSAR	Quantitative structure activity relationship
SCCNFP	Scientific Committee on Cosmetic Products and Non-Food Products (Currently known as SCCP)

References

1. European Commission, *Off. J. Eur. Commun.*, 1976, **L262**(27.9), 169.
2. D. Howes, R. Guy, J. Hadgraft, J. Heylings, U. Hoeck, F. Kemper, H. Maibach, J.-P. Marty, H. Merk, J. Parra, D. Rekkas, I. Rondelli, H. Schaefer, U. Tauber and N. Verbiese, *ATLA*, 1996, **24**, 81.

3. OECD guidelines for the testing of chemicals No. 428, Paris, France: Organisation for Economic Cooperation and Development, 2003.
4. H.A. Benson, *Am. J. Clin. Dermatol*., 2000, **1**, 217.
5. S. Mitragotri, *J. Control. Release*, 2003, **86**, 69.
6. G.P. Moss, J.C. Dearden, H. Patel and M.T.D. Cronin, *Toxicol. In Vitro*, 2002, **16**, 299.
7. European Commission, *Off. J. Eur. Commun.*, 1967, **196**(16.8), 1.
8. W.M.S. Russell and R.L. Burch, *The Principles of Humane Experimental Techniques*, Meuthen and Co., London, 1959.
9. European Commission, *Off. J. Eur. Union*, 2003, **L66**(11.3), 26.
10. Anon, *ATLA*, 2005, **33**(Suppl. 1), 7.
11. A.P. Worth and M. Balls, *ATLA*, 2002, **30**(Suppl. 2), 15.
12. OECD guidelines for the testing of chemicals No. 430, Paris, France: Organisation for Economic Cooperation and Development, 2002.
13. OECD guidelines for the testing of chemicals No. 431, Paris, France: Organisation for Economic Cooperation and Development, 2002.
14. J.H. Fentem, G.E.B. Archer, M. Balls, P.A. Botham, R.D. Curren, L.K. Earl, D.J. Esdaile, H.G. Holzhütter and M. Liebsch, *Toxicol. In Vitro*, 1998, **12**, 483.
15. J.H. Fentem and P.A. Botham, *ATLA*, 2002, **30**(Suppl 2), 61.
16. OECD guidelines for the testing of chemicals No. 432, Paris, France: Organisation for Economic Cooperation and Development, 2002.
17. P.A. Jones, A.V. King, L.K. Earl and R.S. Lawrence, *Toxicol In Vitro*, 2003, **17**, 471.
18. H.J. Duesing, M. Akhiani, A. Schepky, H. Wenck and W. Diembeck, *ALTEX*, 2005, **22,** 153.
19. J.H. Draize, G. Woodward and H.O. Calvery, *J. Pharmacol. Exp. Ther*., 1944, **82,** 377.
20. OECD guidelines for the testing of chemicals No. 404, Paris, France: Organisation for Economic Cooperation and Development, 2002.
21. P.A. Botham, L.K. Earl, J.H. Fentem, R. Roguet and J.J.M. van de Sandt, *ATLA*, 1998, **26**, 195.
22. M.K. Robinson, C. Cohen, A.B. de Fraissinette, M. Ponec, E. Whittle and J.H. Fentem, *Food Chem. Toxicol.*, 2002, **40,** 573.
23. World Medical Association (1964) Declaration of Helsinki, Recommendation guiding physicians in biomedical research involving human subjects. Adopted by the 18th World Medical Assembly, Helsinki, June 1964, amended by the 29th World Medical Assembly, Tokyo, October 1975, the 35th World Medical Assembly, Venice, October 1983 and the 41st World Medical Assembly, Hong Kong, September 1989. *Proceedings of the XXVIth Conference*, Geneva, 1993.
24. M. York, H.A. Griffiths, E. Whittle and D.A. Basketter, *Contact Dermatitis*, 1996, **34**, 204.
25. V. Zuang, M. Balls, P.A. Botham, A. Coquette, E. Corsini, R.D. Curren, G.R. Elliott, J.H. Fentem, J.R. Heylings, M. Liebsch, J. Medina, R. Roguet, J.J.M. van de Sandt, C. Wiemann and A.P. Worth, *ATLA*, 2002, **30**, 109.
26. O. de Silva and D.A. Basketter, *ATLA*, 2004, **32**(Suppl. 1), 679.
27. A.P. Worth and M. Balls, *ATLA*, 2002, **30**(Suppl. 1), 1.

28. OECD guidelines for the testing of chemicals No. 405, Paris, France: Organisation for Economic Cooperation and Development, 1987.
29. L.A. Lambert, W.A. Chambers, S. Green, K.C. Gupta, R.N. Hill, P. Hurley, C.C. Lee, J.K. Lee, P.T. Lui, D.K. Lowther, C.D. Roberts, J.A. Springer and N.L. Wilcox, *Food Chem. Toxicol.*, 1993, **31**, 99.
30. M. Balls, N. Berg, L.H. Bruner, R.D. Curren, O. de Silva, L. Earl, D.J. Esdaile, J.H. Fentem, M. Liebsch, Y. Ohno, M.K. Prinsen, H. Spielman and A.P. Worth, *ATLA*, 1999, **27**, 53.
31. P. McNamee, D. Bagley, C. Faller, P. Jones, B. Le Varlet, I. Manou, W. Pape, C. van den Berghe and F. van Goetham, *ALTEX*, 2005, **22**, 237.
32. OECD guidelines for the testing of chemicals No. 429, Paris, France: Organisation for Economic Cooperation and Development, 2002.
33. D.A. Basketter, P. Evans, R.J. Fielder, R.J. Dearman and I. Kimber, *Food Chem. Toxicol.*, 2002, **40**, 593.
34. D.A. Basketter, L. Blaikie, R. Dearman, I. Kimber, C.A. Ryan, G.F. Gerberick, P. Harvey, P. Evans, I.R. White and R.J. Rycroft, *Contact Dermatitis*, 2000, **42**, 344.
35. D.W. Roberts and G. Patlewicz, *Environ. Res*., 2002, **13,** 145.
36. J.C. Dearden, M.D. Barratt, R. Benigni, D.W. Bristol, R.D. Combes, M.T.D. Cronin, P.N. Judson, M.P. Rayne, A.M. Richard, M. Tichy, A.P. Worth and J.J. Yourick, *ATLA*, 1997, **25**, 223.
37. I. Kimber, J.S. Pichowski, C.J. Betts, M. Cumberbatch, D.A. Basketter and R.J. Dearman, *Toxicol. In Vitro*, 2001, **311**, 77.
38. G.F. Gerberick, J.D. Vassallo, R.E. Bailey, J.G. Chaney, S.W. Morrall and J.P. Lepoittevin, *Toxicol. Sci.*, 2005, **81**, 332.
39. J.M. Ovigne, F. Rousset, P. Aeby, F. Python and B. LeVarlet, *ALTEX*, 2005, **22**, 145.
40. C. Ryan, H. Sakaguchi, T. Ashikaga, F. Rousset, K. Schroeder and D. Basketter, *ALTEX*, 2005, **22**, 146.
41. D.J. Kirkland, *Basic Mutagenicity Tests: UKEMS Recommended Procedures*, Cambridge University Press, Cambridge, UK, 1990.
42. D. Kirkland, M. Aardema, L. Henderson and L. Muller L, *Mutat. Res.*, 2005, **584**, 1.
43. P. Carthew, H. Griffiths, S. Keech and P. Hartop, *Inhal. Toxicol.*, 2002, **14**, 401.
44. G. Oberdorster, *Regul. Toxicol. Pharmacol.*, 1995, **21**, 123.
45. C.R. Lambre, M. Aufderheide, R.E. Bolton, B. Fubini, H.P. Haagsman, P.M. Hext, M. Jorissen, Y. Landry, J.P. Morin, B. Nemery, P. Nettesheim, J. Pauluhn, R.J. Richards, A.E.M. Vickers and R. Wu, *ATLA*, 1996, **24**, 671.
46. B.E. van Vranken, H.M. Romanska, J.M. Polak, H.J. Rippon, J.M. Shannon and A.E. Bishop, *Tissue Eng.*, 2005, **11**, 1177.
47. B. Forbes and C. Ehrhardt, *Eur. J. Pharm. Biopharm.*, 2005, **60**, 193.
48. S. Coecke, V. Rogiers, M. Bayliss, J. Castell, J. Doehmer, G. Fabre, J. Fry, A. Kern and C. Westmoreland, *ATLA*, 1999, **27**, 579.
49. P.W. Ehrhardt, Drug metabolism databases and high throughput testing during drug design and development, *IUPAC*, Blackwell Science, Malden, MA, 1999.
50. B.J. Blaauboer, A.R. Boobis, J.V. Castell, S. Coecke, Groothuis, A. Guillouzo, T.J. Hall, G.M. Hawksworth, G. Lorenzon, H.G. Miltenburger, V. Rogiers, P. Skett, P. Villa and F.J. Wiebel, *ATLA*, 1994, **22**, 231.

51. G.M. Hawksworth, P.H. Bach, J.F. Nagelkerke, W. Dekant, J.E. Diezi, E. Harpur, E.A. Lock, C. MacDonald, J.P. Morin, W. Pfaller, F.A.J.J.L. Rutten, M.P. Ryan, H.J. Toutain and A. Trevisan, *ATLA*, 1995, **23**, 713.
52. L. Gribaldo, J. Bueren, A. Deldar, P. Hokland, C. Meredith, D. Moneta, P. Mosesso, R. Parchment, D. Parent-Massin, A. Pessina, J. San Roman and G. Schoeters, *ATLA*, 1996, **24**, 211.
53. C. Atterwill, A. Bruinink, J. Drejer, E. Duarte, M. McFarlane, E. Abdulla, C. Meredith, P. Nicotera, C. Regan, E. Rodriguez-Farre, M.G. Simpson, R. Smith, B. Veronesi, H. Vijverberg, E. Walum and C.D. Williams, *ATLA*, 1994, **22**, 350.
54. W. Pfaller, M. Balls, R. Clothier, P. Cierickx, B. Ekwall, B.A. Hanley, T. Hartung, P. Prieto, M.P. Ryan, G. Schmuck, D. Sladowski, J.A. Vericat, A. Wendel, A. Wolf and J. Zimmer, *ATLA*, 2001, **29**, 393.
55. G.D. Loizou, *Toxicol. Lett*., 2001, **124**, 59.
56. J. Fentem, M. Chamberlain and B. Sangster, *ATLA*, 2004, **32**, 617.

Subject Index

Adrenal gland function, 60
Alternative
  methods for chemical testing, 101
  tests, 31, 100
    with current regulatory acceptance, 101
Alternatives to animal toxicology studies, 81
Ames test, 108
Androgen receptor, 54
Animal
  screens, 68
  testing
    of cosmetic ingredients, 100
    problems, 23
  welfare, 69, 92
    considerations, 3
As low as is reasonably practicable, 6
Assays for EDs, 68

Benchmark Dose (BMD), 16
Best available technology, 6
Beuhler test, 106
Bioaccumulation potential, 78
Biokinetic modelling, 43
Broadly acceptable risk, 17

Cancer, 53
Carcinogenesis, 108
Carcinogenicity, 13, 59
CASE expert system, 59
Cell line assays, 60
Classification (and labelling), 21
Computer
  automated structure evaluation, 108
  -generated models, 57
Conduct of studies and evaluations, 37
Control of substances hazardous to health (COSHH), 22
CORROSITEX, 102
Cosmetics, 53
  Directive, 3, 22, 29, 95, 99
Critical effect, 18
Cultural
  belief patterns, 9
  prototypes, 9

Dangerous
  Preparations Directive, 76–77, 86
  Substances Directive, 86
Data waivers, 82, 92
Decision tree integrated testing strategy, 66
Deductive Estimation of Risk for Existing Knowledge (DEREK), 107–108
Derived no effect level (DNEL), 12–13
Dermal absorption, 97
Deterministic or Stochastic effects, 13
DNEL, 13
Draize
  eye irritation test (DET), 29, 105
  rabbit skin irritation test, 103

ECVAM
validated tests, 82
validation process, 32
ED *see* Endocrine disruptor
EDSTAC tier testing strategy, 65
Endocrine
disrupting potential, 56
disruptor (ED), 39, 53, 55
Disruptor Screening and Testing Advisory Committee (EDSTAC), 64
disruptors affecting human health, 66
function, 55
system, 53, 55
Endogenous
sex steroid hormones, 56
steroid hormones, 55
Engineering/Actuarial approach to risk, 7
Environmental
effects of EDs, 61
risk assessment, 78
EpiDerm™, 102–104
EPISKIN™, 102, 104
Equity based criterion, 5
ER
binding
assay, 58
capacities, 58
pharmacore, 58
Estrogen receptor (ER), 54
EU
Cosmetics Directive, 36
notification studies, 91
European
Centre for the Validation of Alternative Methods (ECVAM), 82, 100
Cosmetics, Toiletries and Perfumery Association (Colipa), 105
Evaluating the reliability of an existing test, 80
Evident toxicity, 16
Existing Substances Regulation, 86
Expert
reports, 84
systems, 67
Exposure
assessment, 5, 76, 97
for possible data waiving of biocides, 84
Eye irritation, 105

Fund for the Replacement of Animals in Medical Experiments (FRAME), 24, 42

Gene expression, 60
Genetic toxicology, 108
Genotoxicity tests, 108
Gold standard, 36
fool's, 37
Guinea pig maximisation test, 106

H295R assay, 60
Hazard, 4
characterisation, 5, 21, 96
data, 98–99
identification, 5, 76, 95
and characterisation, 98
information, 15
prediction, 29
Hazardous properties assessment, 75
Hershberger
assays, 68
test, 57
Histological sections, 102
Hormonal disruption, 55
Hormone response elements, 54
Human health risk assessment, 78

In silico
approaches, 57
modelling, 56
models for metabolism prediction, 67
systems, 41
In vitro
assays, 63

methods, 31, 62
receptor binding assays, 59
skin penetration, 97
testing for thyroid effects, 61
tests of EDs, 59
Integrated testing
of chemicals, 44
strategies, 42
Intelligent toxicity testing, 92
Invalidation, 48
process, 46–47
Isolated eye tests, 105

Klimisch scoring scheme, 80

LC_{50}, 16
Legislation, 2
for the Testing of EDs, 56
Lifecycle stages of a substance, 77
Local lymph node assay (LLNA), 101, 106
Low volume eye test, 101, 105
Lowest Observed Adverse Effect Level (LOAEL), 16
Lung absorption, 111

Margin of
exposure (MoE),12, 20
safety, 12–13, 15, 20
Marketing and Use Directive, 86
Mathematical model of risk, 8
Medicines, 20
Metabolism, 63
Metabonomics, 115
Microarrays, 62
Mitogenesis (cell proliferation), 60
Models of human skin, 102
Molecular modelling of ligand-receptor complexes, 57
MULTICASE system, 57

No Observed Adverse Effect
Concentration, 16
Level (NOAEL), 12, 96
Non-animal
assays for detecting EDs, 57
methods, 67, 69
test methods, 64
testing strategy, 67
tests for endocrine disruptors (EDs), 56, 64
Notification
of New Substances Regulations, 2
testing, 88

Occupational exposure limits, 14
OECD, 39
test guidelines, 43
tiered testing scheme for endocrine disruptors, 65
Oestrogenic activity, 59
Oestrogenicity, 57–59
of xenobiotics, 61
Oestrogen-responding cells, 60
Oestrogens, 56–57
Outcomes of validation studies, 38

Pathways for endocrine disruption, 54
Perception of risks, 10
Personal care products, 3
Phototoxicity, 102
Physiologically-based biokinetic (PBBK) models, 113
Prague Declaration, 68
Precautionary Principle, 86
Predicted
environmental concentration (PEC), 78
no effect concentration (PNEC), 78
Predictive
risk assessment, 16
testing, 2
Prevalidation, 32
Proteomics, 115

(Quantitative) Structure-Activity Relationships, *see* (Q)SARs
(Q)SAR
approaches, 43
modelling, 57, 69

models, 41, 58
to prioritise chemicals, 58
systems, 35
/CoMFA model, 58
(Q)SARs, 29, 56, 81, 90, 98–99, 107–108, 113

Registration Evaluation, Authorisation and Restriction of Chemicals, *see* REACH
REACH, 3, 13–15, 17, 21–22, 28–29, 41–42, 46, 56, 85–87, 92
REACH Annex
V data, 88
VI data, 89
VII data, 89
VIII data, 90
registration safety data, 90
Read-across, 80–81, 99
approach, 45
Reducing risk, protecting people, 7
Reduction, refinement and replacement (the 3 Rs), 23 42, 81
Reference dose (RfD), 12, 17
Registration under REACH, 88
Regulators, 90
Regulatory
paradigm, 4
planning, 90
requirements, 2
testing of EDs, 69
Relevance of test, 79
Reliability of tests, 79
Repeat dose toxicity study, 113
Replacement alternative test methods, 31, 40
Reproductive
and developmental toxicity, 53
effects of chemicals, 59
Respirable
discharge, 98
dose, 98
Respiratory toxicity, 109
Reverse risk assessment, 46
Risk, 4
analysis, 9
assessment, 4, 34, 45–46, 75–77, 85, 95–96
for cosmetics, 103, 106
paradigm, 45
Technical Guidance Document, 79
without animals, 114
characterization, 5, 76
evaluation, 4–5
framework, 5
management, 4, 34
of death, 9
perception, 8, 10

Safety evaluation, 74
of Chemicals, 1
SAR/QSAR, 80
Scientific Committee on Cosmetic Products and Non-Food Products (SCCNFP), 3, 100
Screening of EDs, 56, 60
Seventh Amendment to the EU Cosmetics Directive, 99
Skin
absorption, 111
corrosion, 101
irritation, 103–104
penetration, 97, 103
sensitisation, 106–107
Stakeholder involvement, 14
Standardized risk evaluation, 14–15
Standards for comparison, 36
Steroid
hormone receptors (SHRs), 54
hormones, 53
Steroidgenesis, 60
Stimulation of mitogenesis, 60
Structural
alerts, 57
features of EDs, 56
Sunscreen chemicals, 97
Surrogate data, 80, 92
System toxicity, 111

Target organ toxicity, 112
Technology based criterion, 6

Test
batteries, 42
development, 35
and validation process, 49
Guidelines (TGs), 30
Testing strategies for EDs, 64
Three R's, 99
Thresholds of concern for human exposure for biocides, 83
Thyroid
gland, 61
hormone receptor, 61
Tolerability of risk, 7
Toxicity testing, 34
Toxicodymamics, 18
Toxicogenomic approach, 62
Toxicogenomics, 62
Toxicokinetics, 18, 111
Toxicological
endpoints, 96
risk assessment and evaluation, 24
Toxicology, 15, 34
and Risk Assessment, 11
studies for data waiving for biocides, 83
Transcriptomics, 115
Transcutaneous electrical resistance (TER), 101

Uncertainty Factors (UFs), 12, 19
Approach, 17
UNECE Globally Harmonized System (GHS), 3
US Interagency Co-ordinating Committee on the Validation of Alternative Methods (ICCVAM), 82
Uterotrophic assays, 68
Utility based criterion, 6

Validation
of
Alternative Tests, 28
in vitro tests for EDs, 68
process, 29, 30, 34, 35

Weight of evidence, 84
approach, 33
Welfare of animals, 2

Xenobiotics, 55