THE INTEGRATED USE OF PHYSICOCHEMICAL AND *IN VITRO* DATA FOR PREDICTING CHEMICAL TOXICITY

ANDREW PAUL WORTH

A thesis submitted in partial fulfilment of the requirements of
Liverpool John Moores University for the degree of
Doctor of Philosophy

This research programme was carried out in collaboration with the
European Centre for the Validation of Alternative Methods
(ECVAM) of the European Commission's Joint Research Centre

September 2000

ABSTRACT

The general aim of this PhD project was to investigate the integrated use of physicochemical and *in vitro* data for predicting the toxicological hazard of chemicals in animals. This was achieved in two stages: firstly, by developing two types of model for acute dermal and ocular toxicity - structure-activity relationships (SARs) based on easily calculated physicochemical properties, and prediction models (PMs) based on experimentally derived physicochemical or *in vitro* data; and secondly, by evaluating the tiered testing approach to hazard classification, in which different classification models (CMs) are applied sequentially before animal testing is conducted. The thesis therefore reports the development and assessment of CMs for skin irritation, skin corrosion and eye irritation, as well as the outcome of simulations in which these models were incorporated into tiered testing strategies for these toxicological endpoints. The results show that the tiered testing approach to hazard classification provides a reliable means of reducing and refining the use of animals, without compromising the ability to classify chemicals. In addition to developing the above-mentioned CMs, regression models for corneal permeability were developed, and the relationship between corneal permeability and eye irritation was investigated.

The thesis also describes the development and assessment of a novel statistical method called embedded cluster modelling (ECM), which generates elliptic models of biological activity from embedded data sets. The combined use of this method with the existing method of cluster significance analysis (CSA) is illustrated through the development of SARs for eye irritation potential.

Finally, novel applications of the bootstrap resampling method were investigated. In particular, algorithms based on this method are shown to provide a means of assessing: a) the minimal variability associated with the Draize rabbit tests for skin and eye irritation; and b) the variability in Cooper statistics (commonly used to summarise the performance of two-group CMs) that arises from chemical variation.

ACKNOWLEDGEMENTS

Firstly, I would like to acknowledge Liverpool John Moores University (LJMU) for giving me the opportunity to undertake this research project under the ægis of the university, while performing most of the work at the European Centre for the Validation of Alternative Methods (ECVAM, Italy). I am particularly grateful to Dr Mark Cronin, my main supervisor and the Director of Studies for this project, for the considerable advice, feedback and encouragement he has given me over the course of the past three years. I would also like to thank Professor John Dearden, for the crucial advisory role he has played throughout the project, even following his retirement. I am most appreciative of the sensitive manner with which Mark and John have guided this project, and I thank them for having made me feel a welcome member of the QSAR and Modelling Group, at LJMU's School of Pharmacy and Chemistry.

Secondly, I would like to acknowledge the European Commission's Joint Research Centre for awarding me a research training grant (Marie-Curie programme). I am very grateful to Professor Michael Balls, the head of ECVAM, for finding the time in his busy schedule to be my ECVAM supervisor during the second and third years of the project. In addition to offering encouragement and support, Michael has involved me in some of ECVAM's mainstream activities, which has provided an exciting and valuable context to the research. A special thanks goes to Dr Julia Fentem (Unilever, UK), who was my ECVAM supervisor during the first year of the project. Before leaving ECVAM in 1998, Julia played an important role in the conception and early development of this work, and has continued to show an interest in its progress. I extend a collective thanks to all my ECVAM colleagues, for creating such an amicable and stimulating work environment.

I am also indebted to a few people who have been particularly influential in my academic development, especially Dr Edward Gill (Worcester College, Oxford University), whose tuition and support when I was an undergraduate has played an important part in the fulfilment of my interests.

Finally, a heartfelt thanks goes to my parents, Adèle and Brian, whose constant support and encouragement means more to me than I can possibly express.

Andrew Worth

September 2000

THE INTEGRATED USE OF PHYSICOCHEMICAL AND IN VITRO DATA FOR PREDICTING CHEMICAL TOXICITY

CONTENTS

LIST OF ABBREVIATIONS

ANOVA	Analysis of variance
BCOP	Bovine corneal opacity and permeability
BLR	Binary logistic regression
C/NC	Corrosive/non-corrosive
CA	Cluster analysis
CM	Classification model
CSA	Cluster significance analysis
CT	Classification tree
DM	Dipole moment
dV1	First-order difference valence molecular connectivity index
ECVAM	European Centre for the Validation of Alternative Methods
EMS	Explained mean square
ESS	Explained sum of squares
EC	European Community
EU	European Union
FL-20	Fluorescein leakage (20%)
HETCAM	Hen's egg test on the chorioallantoic membrane
HOMO	Highest occupied molecular orbital
ICE	Isolated chicken eye
IRE	Isolated rabbit eye
I/NI	Irritant/non-irritant
LDA	Linear discriminant analysis
LogP	Logarithm (to the base 10) of a partition coefficient (usually the octanol-water partition coefficient)
LOO	Leave-one-out
LUMO	Lowest unoccupied molecular orbital
MLR	Multiple linear regression
MP	Melting point
MSA	Molecular surface area
MSD	Mean squared distance
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide
MV	Molecular volume
MW	Molecular weight
NRU	Neutral red uptake
OECD	Organisation for Economic Cooperation and Development
PC	Principal component
PCA	Principal components analysis
pH	Negative logarithm of the aqueous proton (H^+) concentration
pKa	Negative logarithm of the acid dissociation constant (Ka)
PM	Prediction model
(Q)SAR	(Quantitative) structure-activity relationship
RMS	Residual mean square
RSS	Residual sum of squares
SPR	Structure-permeability relationship
TER	Transcutaneous electrical resistance
TSS	Total sum of squares

CHAPTER 1

INTRODUCTION AND OVERVIEW

1.1 BACKGROUND TO THE PROJECT

In the attempt to protect human health against the potentially adverse effects of chemicals, numerous national and international laws require toxicity tests to be performed on products of various kinds, including industrial chemicals, pesticides, foodstuffs, cosmetics and pharmaceuticals. A toxicity test is any procedure for determining toxicological hazard, which is the inherent potential of a chemical to cause toxicity. The risk of a given chemical to a group of society (e.g. consumers) is assessed by combining information on chemical hazard with information on the exposure of that group to the chemical in question.

Traditionally, the potentially adverse effects of chemicals on human health have been assessed by testing on living animals, even though this practice has been widely criticised on both scientific and ethical grounds. Nowadays, the availability of alternative methods for toxicity testing, which do not rely on the use of living animals, promises not only to improve the scientific basis of toxicological assessments, but also to protect animals from painful or distressing procedures. The 'alternatives' concept is attributed to Russell & Burch (1959), who defined three types of alternative procedure: reduction alternatives, which obtain a comparable level of information from the use of fewer animals, or more information from the same number of animals; refinement alternatives, which alleviate or minimise potential pain, suffering and distress; and replacement alternatives, which permit a given purpose to be achieved without using animals. The three types of alternative procedure are referred to collectively as the 'Three Rs'.

The Three Rs were enshrined in European law in 1986, when the European Council passed *Directive 86/609/EEC* (EC, 1986), requiring that alternative methods are used instead of living animals whenever scientifically feasible. This directive, which was incorporated into British law by means of the *Animals (Scientific Procedures) Act* (Home Office, 1986), applies not only to the use of animals for toxicity and potency testing, but also to the use of animals in fundamental biological and medical research.

To overcome the limitations of *individual* alternative methods, it is widely accepted that the reduction, refinement and eventual replacement of animal use in toxicology will depend on the development of suitable integrated testing strategies. These are strategically-designed combinations of methods whose application in a stepwise and/or

parallel fashion is intended to provide the most effective way of predicting toxicological hazard, while at the same time minimising the use of animals. While the need for integrated toxicity testing may be widely accepted, a considerable amount of research is still required to develop and evaluate specific approaches. The general purpose of this PhD project was to contribute to the conceptual development of this field by investigating the integrated use of physicochemical and *in vitro* data for predicting *in vivo* endpoints for skin and eye toxicity. This was achieved in two stages: firstly, by developing two types of model for acute dermal and ocular toxicity - structure-activity relationships (SARs) based on easily calculated physicochemical properties, and prediction models (PMs) based on experimentally derived physicochemical or *in vitro* data; and secondly, by evaluating the tiered testing approach to hazard classification. The project was carried out under the ægis of the School of Pharmacy and Chemistry at Liverpool John Moores University, and was funded by the European Centre for the Validation of European Methods (ECVAM), which is part of the European Commission's Joint Research Centre.

1.2 OVERVIEW OF THE THESIS

The thesis is divided into three parts: Part I contains a general introduction to the field of research; Part II summarises the research, with each chapter taking the format of a scientific article; and the appendices in Part III contain various computer routines written during the project (Appendix A), and a list of the author's publications (Appendix B). All figures and tables are located, in that order, at the end of their respective chapters, since many of them were considered to be too large to be incorporated comfortably within the body of the text.

The first chapter in Part I, chapter 2, describes the animal methods used for the regulatory assessment of acute local toxicity to the skin and eye, and reviews the various computer-based, physicochemical and *in vitro* methods that have been developed to predict the results of the animal tests. Chapter 2 also explains the concept of integrated testing strategy.

Since some of the models developed in this project were based on physicochemical properties calculated with specialised software packages, chapter 3 provides a survey of these properties, and describes the computer algorithms used for their estimation. In addition to the models based on physicochemical data, additional models were

developed from *in vitro* data, generated by methods described in chapter 2. All models were developed with statistical methods, which are described in chapter 4, along with various data resampling methods.

In Part II of this thesis, chapter 5 illustrates how a particular data resampling method, bootstrap resampling, can be applied in the field of alternative method development. In particular, bootstrap resampling was used to estimate the uncertainty inherent in Cooper statistics, which are commonly used to assess the performance of two-group classification models, and to analyse the variability of animal data, which places an upper limit on the predictive ability of any model.

Chapter 6 describes a novel statistical method, called embedded cluster modelling, which was developed in this project to derive elliptic models of biological activity, i.e. classification models that separate toxic from non-toxic chemicals by means of an elliptic boundary. The usefulness of this method is illustrated by its application to an eye irritation data set.

With a view to developing an integrated testing strategy for acute local toxicity, new models for predicting skin irritation and corrosion potential are described in chapter 7, and new models for eye irritation are described in chapter 9. In addition to these classification models for toxic potential, regression models were developed to predict the ability of chemicals to penetrate across the cornea. This work is described in chapter 8, along with an investigation into the relationship between corneal penetration and eye irritation.

Some of the models presented in chapters 7 and 9 were integrated in the form of stepwise testing strategies for skin irritation and corrosion, and for eye irritation, respectively. Chapter 10 describes how these strategies were evaluated by simulating the possible outcomes obtained when the strategies were applied to real data. In addition to simulating the strategic approach now adopted by the Organisation for Economic Cooperation and Development (OECD), described in chapter 2, an alternative approach was also investigated.

Finally, chapter 11 summarises the findings and conclusions of this project in the context of existing knowledge. In addition, some perspectives are offered regarding the conceptual basis of integrated testing, and an assessment is made of current research needs and future prospects.

CHAPTER 2

THE ASSESSMENT AND PREDICTION OF ACUTE DERMAL AND OCULAR TOXICITY

2.1 INTRODUCTION

In this project, the use of physicochemical and *in vitro* data for predicting toxicological hazard was investigated by focusing on the prediction of acute local effects to the skin and to the eye. Two types of acute skin toxicity were considered: skin irritation (irritant contact dermatitis [ICD]), a reversible inflammatory response, and skin corrosion, the irreversible destruction of the skin tissue. In addition, acute eye irritation/corrosion was considered.

This chapter describes the animal methods used to assess, and the alternative methods used to predict, acute dermal and ocular toxicity. In addition, a summary is given of animal and alternative methods for skin and eye penetration, since these processes are also relevant to the prediction of acute dermal and ocular toxicity.

2.2 THE BIOLOGICAL BASIS OF ACUTE LOCAL TOXICITY

2.2.1 Physiology of the skin

As illustrated in Figure 2.1, the skin is composed of three layers: the epidermis (outer layer), the dermis (middle layer), and the subcutaneous tissue (inner layer).

The epidermis

The epidermis is about 30-300 μm thick, depending on anatomical site, and is divided into four anatomical layers: the basal layer (stratum germinativum), the spiny (or prickle cell) layer (stratum spinosum), the granular layer (stratum granulosum), and the horny layer (stratum corneum). The basal layer is the innermost epidermal layer and contains the stem cells from which all keratinocytes are derived. Lying above the basal layer is the spiny layer, which acquires its name from the spiky appearance produced by the intercellular bridges (desmosomes) that connect adjacent cells. The next layer is called the granular layer because the cells contain keratohyalin granules. The outermost epidermal layer, the stratum corneum, is composed of corneocytes, dead epithelial cells that are filled with a protein, keratin, which waterproofs and toughens the skin. The corneocytes are flat and hexagonal in shape, and overlap to form the ‘bricks’ in a ‘bricks-and-mortar’ arrangement of cells. The ‘mortar’ is a complex mixture of lipids,

including ceramides, free fatty acids, and cholesterol and its sulphate derivative (Bronaugh *et al.*, 1982). The lipids form structured arrays of bilayers, with characteristic phase transition temperatures (Bronaugh & Maibach, 1985). In addition to keratinocytes, which account for approximately 95% of the epidermal cell population, the epidermis also contains melanocytes, Langerhans cells and Merkel cells. Melanocytes, which are interspersed among the basal cells, are dendritic cells that produce melanin, a pigment which imparts colour to the skin and protects it from the effects of ultraviolet radiation; Langerhans cells, which are also dendritic cells, are scattered throughout the spiny layer, where they play a role in the immune responses characteristic of skin irritation (and skin sensitisation); and Merkel cells are found in the basal layer, where they play a role in sensory perception.

The dermis

The dermis (or corium), which is 20-30 times thicker than the epidermis, is composed of connective tissue through which sensory nerves, blood vessels and lymph vessels thread. In humans, the dermis projects into the overlying epidermis in ridges called papillæ. Nerves which extend through the dermis and end in the papillæ are sensitive to pressure, pain, cold and warmth. The dermis also provides the base for sweat glands, oil glands, hair follicles, and nail beds.

The major component of dermal connective tissue is collagen, an extra-cellular fibrous protein which provides strength to the skin. Another fibrous protein, elastin, provides elasticity to the skin. The fibrous bundles are surrounded by an extra-cellular gel-like matrix called ground substance, consisting primarily of water, ions, and complex carbohydrates attached to proteins (proteoglycans).

The dermis contains at least three types of cells: fibrocytes, histiocytes and mastocytes. Fibrocytes synthesise the elastic fibres and ground substance. Histiocytes are macrophages, i.e. cells that engulf and destroy cellular debris and invading micro-organisms. Mastocytes, or mast cells, are located near blood vessels, and contain substances (e.g. histamine) that are released during inflammation.

The subcutaneous tissue

The subcutaneous tissue is a layer of loose connective tissue, consisting mostly of adipocytes arranged into collagen-bound lobules. This layer cushions the skin, provides insulation, and serves as a nutritional storage depot.

2.2.2 Physiology of the eye

The eye is a complex structure (Figure 2.2). This description will focus on the tissues that are most commonly referred to in toxicological studies.

The outermost covering of the eyeball is composed of two dense fibrous tissues that are continuous with each other: the transparent cornea, which covers the front part of the eye, and the opaque sclera, which covers the rest of the eye. The transparency of the cornea is due to a regular arrangement of collagen fibres (see below). Certain chemicals may disrupt this regular structure, making the cornea opaque and impairing vision. The surface of the cornea (and of the eyelids) is covered with a thin mucous membrane, the conjunctiva. The conjunctiva and the sclera are both vascularised, and may therefore become inflamed upon contact with irritant chemicals. In contrast, the cornea contains no blood vessels (except at its margins), but it contains many nerves, making it sensitive to pain or touch. The conjunctiva is lubricated by a tear film made up of three layers: a) an inner mucous (mucin) layer, which coats the surface of the cornea; b) a middle aqueous layer, which keeps the eyes moist; and c) an outer lipid layer, which reduces the evaporation of water from the surface of the eye.

Behind the cornea is the anterior chamber, filled with a clear, transparent fluid called the aqueous humour. At the back of the anterior chamber is the iris, a circular muscle which opens and closes to regulate the amount of light entering the eye through the pupil, an opening at the centre of the iris. The aqueous humour flows through the pupil into the posterior chamber, a small space between the iris and the lens. The aqueous humour provides nutrition for the lens and the cornea, both of which are devoid of blood vessels, and removes waste products. Behind the lens is a much larger cavity filled with a clear jelly-like substance, the vitreous humour, which maintains the spherical shape of the eyeball.

The cornea

The cornea, which is about 20 mm in diameter, contains five distinguishable layers: the epithelium (outer covering), Bowman's membrane, the stroma (substantia propria), Descemet's membrane, and the endothelium (inner lining). The corneal epithelium, which is a continuation of the epithelium of the conjunctiva, is made up of about six layers of cells. The superficial layer is shed continuously, and the layers are renewed by multiplication of the cells in the basal layer of the epithelium. The stroma accounts for about 80 percent of the thickness of the cornea, and is composed of collagen fibres which run parallel to form sheets (lamellæ). In the human eye, the lamellae are about 1.5 to 2.5 μm thick, and there are about 200 lamellæ in the stroma. Between the lamellae lie the corneal corpuscles, cells which synthesise new collagen. Bowman's membrane, which lies above the stroma and below the epithelium, is really a part of the stroma, in which the collagen fibrils are arranged in a less orderly fashion. Beneath the stroma is Descemet's membrane, which is about 5 to 10 μm thick. Descemet's membrane is composed of a different type of collagen, secreted by the single layer of flattened cells that make up the endothelium.

2.2.3 Mechanistic basis of acute skin toxicity

Skin penetration

The barrier properties of the skin are considered to reside in its outer layer, the stratum corneum, and chemicals are thought to diffuse through the intercellular lipid-containing spaces in this layer, rather than across the corneocytes (Howes *et al.*, 1996). Chemicals penetrating by this route will pass sequentially across various bilayers, and will therefore come into contact with regions which are both hydrophilic and lipophilic. Therefore, it is not surprising that the predominant physicochemical factors that have been found to describe skin permeability are partitioning behaviour and molecular size (see below, § 2.4.9).

Passive diffusion through intercellular lipid channels is not the only means of skin penetration. Other pathways include hair follicles (Hueber *et al.*, 1994) and sweat ducts (Cullander & Guy, 1991). In addition, some corrosive chemicals penetrate the skin by lysing or solubilising the cells and macromolecules in their path.

Skin irritation

Chemically induced skin irritation is mainly associated with inflammation (erythema and œdema) in the dermis, although epidermal effects may also be observed, such as pigmentary alterations, scaling (shedding of flakes), and fissuring (formation of cracks).

Skin irritation is said to result from an *indirect* chemical action, since irritant chemicals act by triggering the release of inflammatory mediators from corneocytes in the lower stratum corneum and from keratinocytes, Langerhans cells and melanocytes in the viable epidermis. It is then the inflammatory mediators that initiate and control the inflammatory response. The initial release of inflammatory mediators can be the result of direct cytotoxic action, but some chemicals elicit inflammatory mediator release in the absence of cell damage (Lawrence *et al.*, 1997).

While the biochemical details of the inflammatory response are not completely understood, the release of interleukin-1 alpha (IL-1α) and tumour necrosis factor alpha (TNF-α) are considered to be primary events. The release of these 'primary' cytokines leads to the sythesis and release of 'secondary' cytokines which play a role in maintaining the inflammatory response. The biochemistry of skin inflammation is reviewed elsewhere (Williams & Kupper, 1996; Corsini & Galli, 2000).

Skin corrosion

Chemically induced skin corrosion is associated with the destruction and irreversible alteration of the skin at the site of contact. It is manifested by ulceration (the loss of epidermis and part or all of the dermis), bleeding and subsequent scar formation (Emmett, 1986).

Most corrosive chemicals are thought to act by solubilisation of macromolecules in the stratum corneum, which may lead to cell lysis (Gordon *et al.*, 1994). A minority of corrosive chemicals are thought to act as irritants, with the corrosive reaction developing as a consequence of biochemically mediated keratinocyte necrosis (Lewis & Botham, 1994). Finally, surfactants, which are commonly used as ingredients in cosmetic formulations, are thought to act by solubilisation of the stratum corneum (Barratt *et al.*, 1998).

2.2.4 Mechanistic basis of acute eye toxicity

The mechanisms which lead to eye irritation and corrosion are less well documented than the corresponding mechanisms for skin irritation and corrosion. Nevertheless, there is evidence that the inflammatory response of the human conjunctiva is mediated by cytokines (Gamache *et al.*, 1997). The human cornea (which contains blood vessels only at its edges) is also reported to release cytokines during inflammation (Torres *et al.*, 1994). The destruction of cells that occurs in eye corrosion probably occurs by the same direct chemical mechanisms that account for skin corrosion. Also, just as some skin corrosives appear to act by an irritant-like mechanism, there is also evidence that corrosion in the cornea is associated with cytokine expression (Planck *et al.*, 1997), although this could be a consequence rather than a cause. Despite these similarities, the mechanisms by which chemicals penetrate into the eye are likely to be different from the mechanisms of skin penetration, due to the presence of a tear film, and to anatomical differences between the conjunctiva and cornea on the one hand, and the stratum corneum on the other.

2.3 ANIMAL METHODS

2.3.1 Percutaneous absorption

Although many animal species have been used for studying skin absorption, the rat has probably been used most frequently (Howes *et al.*, 1996). The animal is killed at the end of the experiment, and the extent of percutaneous absorption is estimated from the known amount of chemical applied to the skin, and from determination of the total amount excreted and of the amounts left on the skin and in the body.

In Europe, the assessment of percutaneous absorption for regulatory purposes is carried out in the context of acute dermal toxicity testing. At the international level, a draft OECD guideline is currently being considered (OECD, 1994). This guideline recommends using the rat, but the use of other species is not excluded.

2.3.2 The Draize rabbit skin test

Data on skin irritation and corrosion are required by various regulatory authorities. Current regulations include the European Union (EU) directive on dangerous substances (EC, 1993), and the United Nations regulation on the labelling of packaged chemicals for international transport purposes. Current guidelines include OECD Test Guideline 404 (OECD, 1992), and the EC guideline for non-clinical dermal tolerance testing of medicinal products (EC, 1998).

These regulations and guidelines are based on the Draize rabbit skin test (Draize *et al.*, 1944). In this test, the dorsal area of the trunk of the experimental animal is clipped free of hair or fur, and 0.5 ml of liquid (or 0.5 g of solid) test substance is applied, under a semi-occlusive patch, to a surface area of about 6 cm^2. After an exposure period of up to 4 hours, the test substance is removed from the skin. In the case of pharmaceuticals, the duration of administration is determined by the proposed conditions of administration in clinical use (EC, 1998). The exposed skin is evaluated at specified time intervals for up to 21 days, using the formation of œdema (grades 0-4) and erythema/eschar (grades 0-4) as endpoints for irritation, and the production of irreversible full-thickness necrosis as the endpoint for corrosion.

In the EU and OECD classification systems, the erythema/eschar and œdema tissue scores are used directly to classify chemicals as skin irritants or severe skin irritants (see below, § 2.3.4). Within the cosmetics industry, however, chemicals are often ranked for skin irritation potential on the basis of a composite tissue score, the primary irritation index (PII), defined as follows:

$$PII = [\Sigma \text{ erythema } (24, 48, 72h) + \Sigma \text{ œdema } (24, 48, 72h)] / 3 N \qquad \text{(Equation 2.1)}$$

where N is the number of animals in the experimental group, and the PII takes values from 0 to 8.

2.3.3 The Draize rabbit eye test

For regulatory purposes, the Draize rabbit eye test (Draize *et al.*, 1944) is the method of choice for the assessment of eye irritation hazard (EC, 1993; OECD, 1987). In this test, 0.1ml of liquid (or, in the case of solids, a maximum of 0.1g) is instilled into the rabbit eye, and effects on the conjunctiva, cornea and iris are observed at fixed time intervals for up to 21 days. At each observation point, the following endpoints are scored subjectively on a scale of 0 to a maximum of 4: conjunctival redness, oedema, and discharge; corneal opacity and area involved; and iritis.

In the EU (EC, 1983) and OECD (1998) classification systems, some of the tissue scores are used directly to classify chemicals as eye irritants or severe eye irritants (see below, § 2.3.4). Within the cosmetics industry, however, it is more usual to use a composite tissue score, called the modified maximum average score (MMAS), which has a minimum value of zero (no irritancy) and a maximum of 110 (extreme irritancy). The MMAS is calculated from tissue scores in the conjunctiva, cornea and iris, observed in three or more rabbits, at 24, 48 and 72 hours following instillation of the test substance. For each rabbit, a weighted Draize score (WDS) is calculated at each time-point, using the following weighting scheme devised by Draize *et al.* (1944):

$$WDS = 2(R+C+D) + 5(O \times A) + 5(I) \qquad \text{(Equation 2.2)}$$

where R is conjunctival redness, C is conjunctival chemosis (œdema), D is conjunctival discharge, O is corneal opacity, A is the area of cornea involved, and I is iritis. R and D are scored on a scale of 0-3; C, O and A from 0-4; and I from 0-2. The maximum WDS for a given rabbit, observed at any of the three time-points (24, 48 and 72 hours), is the total Draize score (TDS), and the average of these maximum scores over all rabbits in the experimental group is the maximum average score (MAS). The term ‘modified’ in MMAS refers to the fact the first observations were carried out 24 hours after instillation, rather than sooner.

2.3.4 Hazard classification of chemicals

Classification of skin irritants

The EU classification system (EC, 1983) has a single class of skin irritancy, denoted by the risk phrase R38. A chemical is classified as R38 if it causes 'significant' inflammation which persists for at least 24 hours following an exposure period of up to 4 hours. The inflammation is treated as significant if one or more criteria are fulfilled, one of them being that the mean value of the erythema scores, or of the œdema scores, calculated at 24, 48 and 72 hours, over all animals tested, should be ≥ 2. An additional criterion allows an R38 classification to be assigned, if an irritant effect (e.g. scaling or discolouration) is observed in at least two animals at the end of the observation period.

The OECD classification system (OECD, 1998) distinguishes between irritants (class 2) and mild irritants (class 3). Class 2 corresponds approximately with R38, except that the threshold is a little higher (the mean erythema/eschar or œdema score should be $\underline{\geq}$ 2.3), and the mean values can be calculated from the scores on any three consecutive days (to account for delayed reactions). As in the EU system, 'significant' inflammation at the end of the observation period can be taken into account. Less irritating chemicals are placed into class 3, if their mean erythema or œdema scores have a mean value > 1.5, but less than 2.3, on three consecutive days.

Classification of skin corrosives

European regulations (EC, 1983) require classification according to risk phrases, determined according to whether a chemical causes corrosion after it has been applied for 3 minutes (R35) or 4 hours (R34). In the OECD classification system (OECD, 1998), R35 and R34 correspond to corrosive subclasses 1A and 1C, respectively. The OECD system also allows for an intermediate classification 1B, for chemicals which cause corrosion after following an application of 1 hour.

Classification of eye irritants and corrosives

Chemicals are classified as eye irritants if they produce reversible effects in the rabbit eye, and as seriously damaging to the eyes if they produce irreversible effects. Irreversible effects include not only corrosive effects, such as the destruction of cornea or conjunctiva, but also persistent indication of impairment of sight. It should be noted

that chemicals that have been classified as skin corrosives are not tested for ocular effects, since they are also assumed to be corrosive to the eye.

The European Union classification system (EC, 1993) allows for two classes of acute eye toxicity: R36 for (moderate) irritants, and R41 for severe irritants and corrosives. The criteria for assigning these risk phrases are awkward, since they differ according to whether three or more animals have been used in the Draize test (Table 2.1).

The OECD classification system (OECD, 1998) also has two classes, but these are simpler to assign (Table 2.2). Class A is assigned to chemicals with irreversible effects (severe irritants and corrosives), and Class B to chemicals with reversible effects. An additional Class B1 is optional for chemicals whose irritant effects reverse within 7 days of application. The OECD scheme is also more humane, being based on the use of a maximum of three animals per test substance.

Finally, it should be noted that regulatory classification systems are subject to changes, which is an important consideration when developing or evaluating models for predicting animal-based classifications. For example, the 17th Amendment to the Dangerous Substances Directive (EC, 1992) introduced a change in the criteria for serious eye damage: the observation of irreversible eye damage at the end of a 21-day observation period became sufficient for an R41 classification to be assigned, irrespective of the extent of that damage, whereas previously, R41 was only assigned if the appropriate thresholds were exceeded (Table 2.1).

2.4 ALTERNATIVE METHODS

2.4.1 Terminology

Alternative methods and prediction models

An alternative method is any method which serves to replace, reduce or refine the use of an existing animal procedure. Such methods include: a) computer-based methods (mathematical models and expert systems; b) physicochemical methods, in which physical or chemical effects are assessed in systems lacking cells; and, most typically, c) *in vitro* methods, in which biological effects are observed in cell cultures, tissues or

organs. To obtain meaningful results with physicochemical and *in vitro* test systems, it is necessary to have a means of extrapolating the physicochemical or *in vitro* data to the *in vivo* level. To achieve this, Bruner *et al.* (1996) introduced the concept of prediction model (PM), defined as an algorithm which converts the results of one or more alternative methods into a prediction of *in vivo* toxicity. A PM could be a classification model for predicting toxic potential (e.g. EU risk phrases), or it could be a regression model for predicting toxic potency (e.g. Draize scores).

The usefulness of an alternative method for regulatory purposes is formally assessed by performing an interlaboratory validation study. An alternative method is judged valid for a specific purpose (e.g. the classification of chemicals on the basis of skin corrosivity) if it meets pre-defined criteria of reliability and relevance (Balls & Karcher, 1995). In this context, 'reliable' means that the data generated by the alternative method are reproducible (within and between laboratories), and 'relevant' means that the method is associated with a PM of sufficient predictive ability.

Structure-activity relationships and integrated models

One type of mathematical model discussed in this thesis is the structure-activity relationship (SAR), defined as any mathematical model for predicting biological activity from the structure and/or physicochemical properties of a chemical. In the literature, a distinction is generally drawn between quantitative SARs (QSARs) and (qualitative) SARs. Some workers use the term QSAR to indicate that a quantitative measure of chemical structure is used, in contrast to the SAR, which is simply an association between a specific molecular (sub)structure and biological activity (e.g. Livingstone, 1995). Confusingly, other workers use the modifier 'quantitative' to denote the fact that the biological activity is based on a quantitative, rather than a qualitative, scale (e.g. McKinney *et al*., 2000). In this thesis, the term SAR will be used without distinction to refer to any method which relates physicochemical properties to biological activity.

A subtle distinction can be made between SARs and the PMs associated with physicochemical tests. The distinction is that while any PM (associated with a physicochemical test) could also be called a SAR, not all SARs could also be called PMs. For example, SARs can also be based on theoretical descriptors (e.g. topological

indices) or on experimental properties that are themselves more easily predicted than measured (e.g. the octanol-water partition coefficient).

The term ‘structure-property relationship’ (SPR) is sometimes used in distinction to SAR, to denote models that predict physicochemical properties (e.g. aqueous solubility), rather than biological activities. However, some workers use the term ‘SAR’ for both kinds of model (e.g. Barratt, 1995a). Another term which is often used is ‘structure-permeability relationship’ (also abbreviated SPR; e.g. Cronin *et al.*, 1999). In this thesis, SPR will be used to denote a structure-permeability relationship, i.e. any mathematical model that enables a permeability coefficient (K_p) to be predicted from one or more physicochemical properties.

Finally, it should be noted that some workers have reported SARs for *in vivo* effects, based not only on chemical structure/properties, but also on *in vitro* endpoints (e.g. Barratt *et al.*, 1996). The present author suggests that such models are distinguished from PMs and SARs by the term ‘integrated model’ (IM).

2.4.2 *In vitro* methods for skin penetration

Percutaneous absorption can be studied *in vitro* by measuring the diffusion of chemicals across excised (human or animal) skin. A draft guideline for *in vitro* percutaneous absorption is being considered by the OECD (1994), since there is good evidence that *in vitro* data are predictive of both human and animal data (Scott *et al.*, 1992; ECETOC, 1993). Nevertheless, for some chemicals, the *in vitro/in vivo* correlations have been poor, with the *in vitro* method usually underestimating absorption (e.g. Bronaugh & Stewart, 1984). *In vitro* methods for percutaneous absorption are reviewed in Howes *et al.* (1996).

2.4.3 *In vitro* methods for skin corrosion

The use of *in vitro* methods to assess skin corrosion is allowed for in several regulations and international guidelines. For example, the Dangerous Substances Directive (EC, 1993) states that: “...classification can be based on the results of validated *in vitro* tests.” Similarly, OECD Test Guideline 404 (OECD, 1992) includes the following statement:

"...it may not be necessary to test *in vivo* materials for which corrosive properties are predicted on the basis of results from *in vitro* tests." This section will describe three *in vitro* methods for skin corrosion, two of which are referred to in Part II of this thesis.

Transcutaneous electrical resistance assay

The transcutaneous electrical resistance (TER) assay was developed by Oliver and coworkers (Barlow *et al.*, 1991; Oliver *et al.*, 1986, 1988). In this method, test materials are applied for up to 24 hours to the epidermal surfaces of skin discs taken from the pelts of shaved and humanely killed young rats. Corrosive materials are identified by their ability to produce a loss of normal stratum corneum integrity and barrier function, which is measured as a reduction in the inherent TER below a predetermined threshold level (5kΩ); irritant and non-irritant substances do not reduce the TER below the threshold level.

Skin^2

The Skin^{2TM} model ZK1350 (Advanced Tissue Sciences Inc., La Jolla, CA, USA) is a three-dimensional human skin model with dermal, epidermal and corneal layers. The test material is applied to the stratum corneum for 10 seconds, and the extent of cell damage, as determined by the MTT reduction assay, is used to assess the degree of corrosivity (Perkins & Osborne, 1993; De Wever & Rheins, 1994; Perkins *et al.*, 1996). MTT is a derivative of tetrazolium bromide, and its colour change upon reduction is used as an assay of mitochondrial function (and therefore cell viability).

EPISKIN

EPISKIN^{TM} (SADUC, Biomatériaux Imedex, Chaponost, France; the production rights now belong to L'Oréal) is a three-dimensional human skin model comprising a reconstructed epidermis with a functional stratum corneum. Its use for skin corrosivity testing involves topical application of test materials to the surface of the skin for 3, 60 and 240 minutes, and the subsequent assessment of their effects on cell viability, using the MTT assay.

Validation status of in vitro tests for skin corrosion

The abilities of the three *in vitro* skin corrosivity methods (and one physicochemical method, CORROSITEX^{TM}; see below) to distinguish between corrosive and non-corrosive chemicals were assessed during an ECVAM validation study (Barratt *et al.*, 1998; Fentem *et al.*, 1998). All four tests showed acceptable reproducibilities, but only the TER and EPISKIN assays met the criteria for relevance. The regulatory acceptance of EPISKIN and TER at the European level took place in February 2000, and the protocols for these methods were incorporated into Annex V (Testing Methods) of the Dangerous Substances Directive in June 2000 (EC, 2000).

2.4.4 *In vitro* methods for skin irritation

This section will provide a brief description of the three main types of *in vitro* method used for skin irritation testing: keratinocyte cultures, reconstituted human skin models, and organ cultures. The current status of these methods is reviewed in Robinson *et al.* (1999) and in van de Sandt *et al.* (1999).

Keratinocyte cultures

Keratinocyte cultures (based on isolated cells or cell lines) have been used with a variety of endpoints for studying the cytotoxic effects of chemicals on skin cells. In early studies, endpoints for basal cytotoxicity were used, such as MTT reduction and neutral red uptake. More recently, attention has focussed on the use of more specific, mechanistically based endpoints, such as the release of IL-1 and prostaglandin (PG)-E_2. In general, it appears that reasonable agreement between *in vitro* cytotoxicity data and *in vivo* irritation scores/classifications is only observed with specific groups of related compounds, such as surfactants (Osborne & Perkins) and N-alkyl sulphates (Wilhelm *et al.*, 1994). However, due to absence of a corneum stratum in keratinocyte cultures, they tend to be over-predictive of skin irritation measured *in vivo*.

Reconstituted human skin models

In contrast to simple monolayer keratinocyte cultures, reconstituted human skin models are multi-layered, differentiated keratinocyte cultures, which exhibit the main characteristic features of the native epidermis, including the cuboidal appearance of the basal cell layer, and the presence of the spiny, granular and horny layers. The

intercellular space of the stratum corneum has a lipid composition similar to that of its native counterpart (van de Sandt *et al.*, 1999). The epidermal surface of these systems is exposed to the air, so test substances can be applied directly. In this respect, human skin models have an advantage over monolayer keratinocyte cultures, which are immersed in culture medium, since some test substances are insoluble in water. A variety of endpoints have been used in reconstructed systems, including MTT reduction and cytokine release (Coquette *et al.*, 1999).

Currently, three human skin models are commercially available: a) EPISKIN; b) EpiDerm™ (MatTek, Ashland, MA, USA); and c) SKINETHIC™ (Skinethic Laboratories, Nice, France). EPISKIN is based on the skin model developed by Tinois *et al*. (1991); EpiDerm was developed by Cannon *et al*. (1994); and SKINETHIC is based on the model developed by Rosdy & Clauss (1990). Other human skin models have also been reported (e.g. Boyce *et al.*, 1990; Bell *et al.*, 1991).

Organ cultures

Organ cultures based on skin explants from animals and humans represent the most relevant model for skin irritation testing, since all skin layers and cell types are present. The dermal side of the organ culture is immersed in cell culture medium, while the surface of the stratum corneum is exposed to the air, thus permitting the topical application of test chemicals. Several toxicological endpoints have been employed, including cell viability, the release of inflammatory mediators, histomorphology, and epidermal cell proliferation. A limitation of these systems is that they can only be used for short exposures (up to 3 days) to test substances, because of the brief survival time of the tissue *in vitro* (van de Sandt *et al.*, 1999).

Validation status of in vitro methods for skin irritation

Progress in the development and validation of *in vitro* methods for skin irritation has been slower than in the case of methods for skin corrosion. This is probably due to the greater complexity of the irritant response *in vivo*: inflammatory effects (erythema and œdema) are the final manifestations of a chain of molecular and cellular events, and depend on the endothelial cells of surface blood vessels, which do not exist *in vitro*. Nevertheless, a number of promising models based on ‘earlier’ endpoints are currently being evaluated in an ECVAM-funded prevalidation study. The models being evaluated

are EPISKIN, two other human skins models (EpiDerm and PREDISKIN^{TM}), and the non-perfused pig ear model.

2.4.5 *In vitro* methods for eye penetration

The penetration of chemicals into the eye is generally studied in the context of ocular pharmacology (rather than ocular toxicology), the aim of such studies being to maximise the delivery of ocular drugs (e.g. anti-glaucoma drugs) to the appropriate target tissues in the eye. Most of these studies have employed isolated corneas or scleral tissue (a compilation of studies is given by Prausnitz & Noonan [1998]), although a limited number of studies have used isolated conjunctivas (e.g. Kompella *et al.*, 1999).

2.4.6 *In vitro* methods for eye irritation

A number of *in vitro* methods for eye irritation have been developed and evaluated in multi-centre validation studies. At present, however, no single method, or combination of methods, has been found to meet all the requirements of the regulatory authorities. This section will provide a brief introduction to those methods which are referred to in Part II of the thesis. Detailed descriptions of these tests are given in Balls *et al.* (1995), and their current validation status is described in Balls *et al.* (1999).

A number of *in vitro* tests for eye irritation are based on simple cell cultures, including: a) the detection of hæmolysis and protein denaturation in red blood cells (RBCs); b) the determination of neutral red uptake (NRU) in fibroblasts (e.g. murine 3T3 cells); c) the determination of fluorescein leakage (FL) across an epithelial monolayer (Madin-Darby canine kidney [MDCK] cells); and d) the measurement of metabolic activity (production of acid metabolites) in fibroblasts (e.g. murine L929 cells), using the silicon microphysiometer (SM).

Another group of tests are referred to as *ex vivo* tests, since they are based on isolated tissues. These tests include: a) the assessment of effects in the isolated rabbit eye (IRE) and the isolated chicken eye (ICE); b) the measurement of opacity and permeability in the bovine cornea (the bovine cornea opacity and permeability [BCOP] test); and c) the subjective scoring of hæmorrhage, lysis and coagulation in the vascularised

chorioallantoic membrane (CAM) of embryonated chicken eggs (the hen's egg test on the chorioallantoic membrane [HETCAM]). The endpoints used in all of these tests are summarised in Table 2.3.

Finally, for completeness, it should be noted that there is a third group of *in vitro* tests: the human tissue equivalents. For example, the EpiOcular^{TM} system (MatTek Corporation, MA, USA) consists of human-derived epidermal keratinocytes cultured to form a stratified epithelium similar to that found in the human cornea (Stern *et al.*, 1998).

2.4.7 Physicochemical methods for skin irritation and corrosion

Use of pH, pK_a and acid/alkali reserve measurements

One approach to the determination of skin irritation and corrosion has been based on the determination of pH (a measure of the acidity/alkalinity of substances in aqueous solution, defined as the negative logarithm of the hydrated proton concentration), or on the combined use of pH and acid/alkali reserve, the latter being a purported measure of buffering capacity. The acid reserve of a substance is defined as the amount (in grams) of sodium hydroxide required to bring 100g of acidic test substance (in a 10% solution or suspension) up to a pH of 4. The alkali reserve is defined as the amount (in grams) of sodium hydroxide equivalent to the amount of sulphuric acid required to bring a 100g of alkaline test substance (in a 10% solution or suspension) down to a pH of 10. Further details on the generation and use of acid/alkali reserve measurements are given in Young *et al.* (1988).

To predict skin corrosion potential from pH alone, OECD Test Guideline 404 (1992) recommends the following PM:

If the pH of a substance ≤ 2, or if pH ≥ 11.5, it is likely to be corrosive.

No guidance is provided in the 1987 version of this guideline on the concentration at which the pH should be measured.

Regarding the combined use of pH and acid/alkali reserve measurements, Young & How (1994) recommended the following PMs:

If pH - acid reserve/12 ≤ -0.5, or if pH + alkali reserve/12 ≥ 14.5, the substance is likely to be corrosive.

If pH - acid reserve/6 ≤ 1, or if pH + alkali reserve/6 ≥ 13, the substance is likely to be irritant.

To apply this PM, it was recommended that pH measurements are carried out on a 10% (w/w) solution for solid test substances, and on the undiluted liquid for liquid substances. An evaluation carried out by the author (Worth *et al.*, 1998) indicates that the PM based on pH and acid/alkali reserve measurements is no more predictive of skin corrosion potential than the PM based on pH measurements alone, at least in the case of single chemicals. The acid/alkali reserve method may, however, be useful for predicting the corrosion potential of mixtures, especially those containing ionic salts (A.P. Walker, personal communication).

The use of the acid dissociation constant (pK_a) for predicting skin irritation has been reported by Nangia *et al.* (1996). These workers used a modified Draize scale (with tissue scores ranging from 0 to 5) to rank the human skin irritancy of 12 basic compounds having similar *in vitro* penetration rates. The following PM was reported:

$\log (VIS) = 0.17\ pK_a - 1.44$ (Equation 2.3)
n = 12, r = 0.84, s and F not given

where VIS is the visual irritation score (the modified Draize score). Similarly, in a study by Berner *et al.* (1988), the human skin irritancy of a series of four benzoic acidic compounds was reported to correlate with their pKa values.

CORROSITEX

CORROSITEX™ (InVitro International, Irvine, CA, USA) is a physicochemical method for skin corrosion based upon the determination of the time required for a test material to pass through a biobarrier membrane (a reconstituted collagen matrix,

constructed to have physicochemical properties similar to rat skin). If the test substance alters the biobarrier sufficiently to be able to pass through it, a colour change is produced in a liquid (the 'chemical detection system'), located directly below the biobarrier. The time required for this change to occur (the 'breakthrough time') is reported to be inversely proportional to the degree of corrosivity of the test material (Fentem *et al.*, 1998).

2.4.8 Physicochemical methods for eye irritation

Use of pH, pK_a and acid/alkali reserve measurements

The combined use of pH and acid/alkali reserve measurements for predicting eye irritation potential was reported to be useful by Régnier & Imbert (1992). In contrast, an evaluation by the author (Worth & Fentem, 1999) found no evidence for the additional usefulness of acid/alkali reserve measurements over pH measurements alone. However, the two studies applied different PMs to different data sets, which could account for the different conclusions.

The use of the acid dissociation constant (pK_a) for predicting eye irritation has been reported by Sugai *et al.* (1991). These workers studied a homogeneous set of 10 salicylates (2-hydroxybenzoic acids) and reported a significant correlation between pKa and the weighted cornea score (WCS):

$WCS = -31\ pK_a + 144.1$ (Equation 2.4)

$n = 10$, $r = -0.73$, s and F not given

The correlation between pK_a and the weighted conjunctival score was not found to be significant.

EYTEX

EYTEX™ is a physicochemical method for eye irritation based on the detection of protein precipitation in an artificial matrix, which is assumed to mimic opacity formation in the cornea. Protein precipitation is assayed as the turbidity of the EYTEX reagent, using a spectrophotometer.

Human cornea model

Recently, the development of an artificial human cornea, containing epithelial, stromal and endothelial layers, was reported (Griffith *et al.*, 1999). This system could prove to be a valuable model for predicting eye irritation.

2.4.9 Mathematical models

Membrane penetration as a partition-diffusion process

The transport of chemicals across biological membranes is generally modelled as a partition-diffusion process: i.e. the penetration rate of a chemical is considered to depend partly on the extent to which it partitions into the membrane phase, and partly on its ability to diffuse through that phase.

Experimentally, transmembrane flux is generally determined by fixing the concentration gradient across a membrane of known surface area. Initially, there will be a non-linear concentration gradient across the membrane, but following a lag-time (t), the concentration gradient will be linear, and a steady-state flux will be attained. From Fick's laws of diffusion, it can be shown that the steady-state flux across an isotropic (homogeneous) membrane is given by Equation 2.5 (see Flynn, 1990, for a mathematical derivation):

$$J = K_m . D . (A/L) . \Delta C \qquad \text{(Equation 2.5)}$$

where J is the transmembrane flux (mg/s), K_m is the (dimensionless) membrane-solvent partition coefficient, D is the diffusion coefficient (diffusivity) of the penetrant (cm^2/s), A is the area of membrane (cm^2), L is diffusion path length (cm), and ΔC is the concentration gradient across the membrane (mg/cm^3). The permeability coefficient, K_p, is defined by Equation 2.6:

$$K_p = K_m . (D/L) \qquad \text{(Equation 2.6)}$$

where K_p is the permeability coefficient of the penetrant (cm/s). Thus, if a biological membrane is treated (for simplicity) as an isotropic membrane, and the diffusion path

length is approximated by the membrane thickness h, Equation 2.5 can be rewritten as follows:

$J = K_p . D . (A/h) . \Delta C$ (Equation 2.7)

where A, h and ΔC are under experimental control, and D can be calculated from the lag-time t (the time taken for a steady-state flux to be achieved) by using Equation 2.8:

$D = h^2/6.t$ (Equation 2.8)

In some studies, it is assumed that the diffusion coefficient is constant for all permeants, in which case D can be calculated from the intercept [equal to log(D/h)] of a plot of $\log K_m$ against $\log K_p$. However, according to Potts and Guy (1992), this can lead to an overestimation of K_p for small substances, which have relatively high values of K_m, due to the inverse relationship that exists between the diffusion coefficient and permeant size (Lieb & Stein, 1971).

Finally, it should be noted that the permeability coefficient K_p is dependent on the solvent used in the diffusion experiment (due to the dependence of K_p on K_m). Therefore, SARs should ideally be developed from data obtained with a common solvent. This is usually chosen to be water, because this does not damage the membrane. However, water is not a suitable solvent for highly lipophilic substances, which is why cosmetic and pharmaceutical formulations often contain penetration enhancers (Guy *et al.*, 1990).

Structure-permeability relationships for skin penetration

Early attempts to derive SPRs were based simply on the logarithm of the octanol-water partition coefficient (logP), which produced reasonable models for certain homologous series, such as alcohols (Scheuplein, 1965), steroids (Scheuplein *et al.*, 1969), and phenols (Roberts *et al.*, 1977). More recently, Kirchner *et al.* (1997) reported SPRs based on logP with r^2 values > 0.7, but these only applied to subsets of chemicals falling within defined ranges of molar volume. Notably, the model based on a heterogeneous data set of 114 chemicals showed a poor fit to the data (r^2=0.32).

Since it appears that reliable SPRs based on logP alone cannot be derived for heterogeneous groups of chemicals, most recent studies have considered additional predictor variables. A common approach has been to develop SPRs based on logP and molecular weight (MW). Perhaps the most commonly cited SPR based on logP and MW is the one derived by Potts & Guy (1992), who analysed a data set of *in vitro* skin permeability coefficients published by Flynn (1990):

$\log K_p = 0.71 \log P - 0.0061 MW - 6.3$ (Equation 2.9)
$n = 93, s = 0.710, r = 0.82, r^2 = 0.67$

where $\log K_p$ is the logarithm of the permeability coefficient across excised human skin (in units of cm/sec).

Similar models have been reported by other workers (Kasting *et al.*, 1987; Bronaugh & Barton, 1993; Cronin *et al.*, 1999). An SPR based on logP and molecular volume (MV) has also been reported (Barratt, 1995a), which is essentially the same, given the high collinearity between MV and MW.

A different approach has been to derive SPRs based on fragmental constants, in which organic molecules are broken into fragments, with each fragment then being assigned a value for its contribution to the K_p (Pugh & Hadgraft, 1994). This approach provides a way of assessing how the introduction of specific functional groups will affect skin permeability.

Abraham *et al.* (1995) applied the solvatochromic approach to the prediction of K_p, i.e. K_p was treated as a function of the so-called 'solvatochromic parameters', which express the steric, electronic and hydrogen-bonding properties of a solute (Equation 2.10):

$\log K_p = -5.33 - 0.57 \pi_2 - 0.51 \Sigma\alpha_2^H - 3.37 \Sigma\beta_2^H + 1.77 V_x$ (Equation 2.10)
$n = 25, s = 0.26, r = 0.98, r^2 = 0.96$

where π_2 is the dipolarity/polarisability; $\Sigma\alpha_2^H$ and $\Sigma\beta_2^H$ are descriptors for hydrogen bond acidity and basicity; and Vx is the McGowan characteristic volume.

This approach was further developed by Pugh *et al.* (1996), who separated K_p into two terms, the partition coefficient between water and the stratum corneum (K_{sc}), and a diffusion term (D/h). They then showed that the partition coefficient could be predicted from logP (Equation 2.11), whereas the diffusion term could be predicted from the hydrogen bond donor (α) and acceptor (β) abilities (Equation 2.12):

$\log K_{sc} = -0.024 + 0.59 \log P$ (Equation 2.11)
$n = 45$, s not given, $r^2 = 0.84$

$\log (D/h) = -1.327 - 1.30 \alpha - 2.57 \beta$ (Equation 2.12)
$n = 31$, $s = 0.179$, $r^2 = 0.85$

Alternatively, the permeability coefficient can be predicted directly by using Equation 2.13 (Pugh *et al.*, 1996):

$\log K_p = -1.09 + 0.508 \log P - 1.26 \alpha - 2.84 \beta$ (Equation 2.13)
$n = 31$, $s = 0.155$, $r^2 = 0.96$

Structure-permeability relationships for eye penetration

SPRs for predicting the transport of substances across ocular tissues do not appear to be as widely reported as SPRs for skin penetration. A number of SPRs for corneal permeability have been derived for congeneric series, including β-blocking agents (Schoenwald & Huang, 1983) and steroids (Schoenwald & Ward, 1978; Grass *et al.*, 1988). In these studies, a parabolic relationship between the corneal permeability coefficient and logP (or logD, the logarithm of the distribution coefficient) was observed. However, some chemicals within these congeneric series deviated from the parabolic curves.

More recently, Yoshida & Topliss (1996) published an SPR (Equation 2.14) with a high correlation coefficient (r = 0.9), derived from a data set consisting mostly of steroids and β-blockers:

$\log K_p = -3.55 - 0.478\ \Delta\log P + 0.105 \log D$ (Equation 2.14)

$n = 26$, $s = 0.282$, $r = 0.92$, $r^2 = 0.84$

where logD is the logarithm of the distribution coefficient at pH 7.65, and ΔlogP is $\log P_{octanol} - \log P_{alkane}$. In cases where experimental data for ΔlogP were missing, Yoshida & Topliss used a solvation equation to predict ΔlogP from five solvatochromic parameters (R_2, a term representing molar refraction; π_2, a descriptor for dipolarity/polarisability; $\Sigma\alpha_2^H$ and $\Sigma\beta_2^H$, descriptors for hydrogen bond acidity and basicity; and V_x, the McGowan characteristic volume). To apply this SPR to most chemicals of interest, it would therefore be necessary to obtain values for the five solute descriptors, in addition to logD values. In this project, an attempt was made to derive SPRs for corneal permeability which are based on more readily accessible descriptor variables. This work is described in Chapter 8.

Structure-activity relationships for skin irritation and corrosion

Relatively few SAR studies for skin irritation have been reported in the literature. Barratt (1996a) reported a SAR for predicting the PII of organic chemicals, but this had little predictive value ($r^2 = 0.42$). In the same study, discriminant analysis was shown to discriminate between irritant and non-irritant chemicals, defined by EU classification criteria, with an accuracy of 67%. However, an explicit classification model was not reported.

Hayashi *et al.* (1999) reported two SARs for predicting the molar-weighted PII of phenols. One model, based on absolute hardness, was proposed for chemicals with negative lowest unoccupied molecular orbital (LUMO) energies, whereas the other model, based on the logarithm of the octanol-water partition coefficient (logP), was proposed for chemicals with positive LUMO energies. These models had correlation coefficients of 0.72 and 0.82, respectively (i.e. r^2 values of 0.52 and 0.67).

In a recent study by Smith *et al*. (2000), best subsets regression and linear discriminant analysis were applied to a data set of 42 esters, for which human skin irritation data were obtained and 19 physicochemical variables were calculated. Best subsets regression was used to select variables for subsequent inclusion in discriminant models. A total of 10 models were reported, each one being derived from the data for the 13 irritant esters and the data for 13 non-irritant esters, these having been randomly selected with replacement from the set of 26 non-irritant esters. The best (most frequently selected) variables were water solubility (lower for irritants than non-irritants), a dispersion parameter (higher for irritants), a hydrogen-bonding parameter (higher for irritants), the sum of partial positive charges (lower for irritants), and density (lower for irritants). The discriminant model based on all five parameters had a sensitivity of 85% and a specificity of 92%.

Various SARs for skin corrosion have been reported by Barratt and colleagues (Barratt, 1996a; Barratt, 1996b; Whittle *et al.*, 1996; Barratt *et al.*, 1998). On the whole, the SARs presented in these studies take the form of principal component (PC) plots, which are based on physicochemical properties and show a separation between corrosive (C) and non-corrosive (NC) chemicals. Explicit classification models were not presented. Rather than modelling a heterogeneous group of chemicals, separate analyses were performed for acids, bases, electrophiles and neutral organics (defined as uncharged molecules which lack the potential to react covalently and which do not ionise under biological conditions [M.D. Barratt, personal communication]). The most recent presentation of this approach is given in Barratt *et al.* (1998). In addition to PC analyses, discriminant analysis and neural network analysis were also applied to the neutral and electrophilic chemicals (Barratt, 1996a) and to the acids, bases and phenols (Barratt, 1996b). Finally, in another study (Barratt *et al.*, 1996), PC plots for acids were based not only on physicochemical propertes, but also on *in vitro* cytotoxicity measurements in mouse 3T3 cells. The additional inclusion of the cytotoxicity variable helped to separate moderate (R34) corrosives from severe (R35) corrosives.

Structure-activity relationships for eye irritation

Cronin *et al.* (1994) applied linear regression analysis to a heterogeneous data set of 53 organic liquids, to derive SARs for predicting the molar eye score (MES). This is the usual MMAS corrected for the number of molecules applied to the rabbit eye, and its

use as the response variable is justified on the grounds that biological responses produced by pure chemicals are expected to be proportional to the number of active molecules, rather than to the amount (in g or ml) of active substance. Although 23 physicochemical properties were examined, it was found that statistically significant regression models could not be obtained for the complete set of chemicals. However, reasonable models could be obtained for the alcohols (Equation 2.15) and the acetates (Equation 2.16):

$$MES = 2.06\ logP + 4.13 \qquad \text{(Equation 2.15)}$$
$$n = 9,\ s = 1.89,\ r^2 = 0.72$$

$$MES = -1.37\ logP + 2.97 \qquad \text{(Equation 2.16)}$$
$$n = 7,\ s = 0.52,\ r^2 = 0.74$$

where MES = molar eye score (molar MMAS).

Abraham *et al.* (1998a) applied the solvatochromic approach to a data set comprising 38 of the 53 organic liquids previously analysed by Cronin *et al.* (1994). Again, the MES was used as the response variable, but the authors argued that logP is not the most appropriate variable for describing the transfer of a liquid into an organic biophase. Instead, it was argued that this transfer process is equal to the product of the solubility of a vapour into hexadecane (L) and the saturated vapour pressure (P^o), measured in ppm. Thus, instead of deriving a SAR based on logP, a model was based on logL and $logP^o$. The solvatochromic parameters were also used as descriptor variables, giving the following SAR:

$$\log (MES/P^o) = -6.97 - 0.17R_2 + 0.88\pi_2 + 3.83\Sigma\alpha_2^H + 1.41\Sigma\beta_2^H + \log L$$
$$n = 38,\ s = 0.46,\ r^2 = 0.89 \qquad \text{(Equation 2.17)}$$

where R_2 is the excess molar refraction (a term representing the tendency of a solute to interact through n or π electrons); π_2 is the dipolarity/polarisability; and $\Sigma\alpha_2^H$ and $\Sigma\beta_2^H$ are descriptors for hydrogen bond acidity and basicity. The authors chose to represent the response variable as $\log(MES/P^o)$, instead of using log(MES) and adding $logP^o$ to the right-hand side of the equation. This may have been done to emphasise the fact that

the response is not the MES resulting from direct application of a liquid to the eye, but the MES resulting from exposure to the corresponding vapour. If $\log P^o$ is added to both sides of Equation 2.17, it becomes apparent that six variables are being used to make predictions for 38 chemicals (i.e. there is a 16% chance that Equation 2.17 is a spurious correlation). In a subsequent study, Abraham *et al.* (1998b) claimed that the eye irritation threshold in humans (EIT), expressed in parts per million (ppm), can be calculated from the MES in rabbits by using the following equation:

$\log (1/EIT) = \log (MES/P^o) - 0.66$ (Equation 2.18)
n = 8, r and s not given

Equation 2.18 was used to calculate the EIT for 37 compounds for which human data were not available, and the 37 predicted EIT values were combined with 17 measured values to derive the following SAR for the EIT:

$\log (1/ EIT) = -7.92 - 0.48R_2 + 1.42\pi_2 + 4.03\Sigma\alpha_2^H + 1.22\Sigma\beta_2^H + 0.85\log L$
n = 54, s =0.36, r^2 =0.93 (Equation 2.19)

A different approach to the prediction of the molar eye score was adopted by Kulkarni & Hopfinger (1999). In addition to calculating the usual intramolecular solute properties, these workers also used a molecular dynamics method to generate intermolecular membrane-solute interaction properties. SARs based on these properties were then derived by using a genetic algorithm, instead of traditional multiple linear regression. The following SAR had the highest r^2 value:

$MES = -0.81 - 0.07\ E - 0.48\ F(H_2O) + 0.35\ \kappa 3$ (Equation 2.20)
n = 16, s =0.41, r^2 =0.92

where E is the sum of the electrostatic and van der Waals interaction energies between the solute and the membrane (modelled as a dimyristoylphosphatidylcholine monolayer); $F(H_2O)$ is the free energy of solvation of the solute in water; and $\kappa 3$ is the third-order kappa index. The 16 chemicals used to derive the model were aliphatic and aromatic hydrocarbons, and aliphatic ketones, alcohols and acetates.

In addition to the derivation of regression models, a number of attempts have been made to derive classification models for eye irritation. For example, Cronin *et al.* (1994) applied linear discriminant analysis to the data set of 53 organic liquids, but found no linear combination of physicochemical properties capable of discriminating between irritant (I) and non-irritant (NI) chemicals (as defined by EU classification criteria). However, in a PC plot based on all 23 variables, the I chemicals appeared to form an embedded cluster within the NI chemicals.

The phenomenon of embedded clustering of irritant chemicals was investigated further by Cronin (1996), this time by using the technique of cluster significance analysis (CSA) to determine whether the embedded clustering was statistically significant. Out of a total of 23 physicochemical descriptors, it was reported that the five most significant, in terms of the embedded clustering of 52 organic liquids, were logP, $logP^2$, the heat of formation, dipole moment, and the zero-order valence connectivity index. Similarly, an embedded cluster of eye irritants was apparent in a PC plot of 57 neutral organic chemicals, based on logP, dipole moment, and the second and third principal inertial axes (Barratt, 1997).

The physicochemical determinants of eye irritation potential were also investigated by Rosenkranz *et al.* (1998). In comparison with non-irritant chemicals, these workers reported that irritant chemicals have significantly lower molecular weights, higher aqueous solubilities, lower logP values, and greater molecular orbital energy gaps (absolute hardness values). On the basis of the last-named observation, the authors concluded that chemical reactivity does not appear to be a requirement for eye irritation. Rosenkranz *et al.* also used their Multi-CASE expert system to identify biophores (substructures which occur with a significantly greater frequency in irritants than in non-irritants) and biophobes (substructures which occur significantly more frequently in non-irritants). The major structural determinants included primary, secondary and tertiary amine groups (i.e. basic groups), as well as carboxylate, organosulphate and sulphonate groups (i.e. acidic groups). Other groups were also reported, but their mechanistic significance is unclear, and they may even represent statistical artefacts. In this study, 202 chemicals were classified as I or NI according to the system of Smyth & Carpenter (1946), but an additional 94 chemicals were classified according to the EU system, so an imbalance was introduced into the analyses.

The analyses described above aimed to predict the eye irritation potency or potential of non-surfactant chemicals, so it is of interest that a recent investigation has focused on surfactants (Patlewicz *et al*., 2000). In this study, neural network analysis was used to show that the MAS of cationic surfactants shows a positive, non-linear correlation with surfactant concentration and critical micelle concentration, and a negative non-linear correlation with logP. Contrary to expectation, a positive correlation with molecular volume was also found. The trained neural network was able to predict about 70% of the variance in the training set, which is probably as much as could be expected, given the variability in the rabbit data. However, apart from highlighting a few useful predictor variables, this study is of limited usefulness, since it does not provide a prediction model for converting such variables into predictions of the MAS.

2.5 INTEGRATED TESTING STRATEGIES

Due to the limitations of individual alternative methods for predicting toxicological hazard, there is a growing emphasis on the use of integrated approaches (Barratt *et al.*, 1995). This has led to the concept of the 'integrated testing strategy', which has been defined as follows (Blaauboer *et al.*, 1999):

"An integrated testing strategy is any approach to the evaluation of toxicity which serves to reduce, refine or replace an existing animal procedure, and which is based on the use of two or more of the following: physicochemical, *in vitro*, human (e.g. epidemiological, clinical case reports), and animal data (where unavoidable), and computational methods, such as (quantitative) structure-activity relationships ([Q]SAR) and biokinetic models."

Since integrated testing strategies are based on the use of different types of information, they are expected to be particularly successful at predicting *in vivo* endpoints which are too complex in biochemical and physiological terms for any single method to reproduce.

A particular type of integrated testing strategy is the so-called tiered (stepwise or hierarchical) testing approach to hazard classification. This is based on the sequential use of data derived from alternative methods (structure-activity relationships, biokinetic models, physicochemical methods and *in vitro* tests), and existing information, before

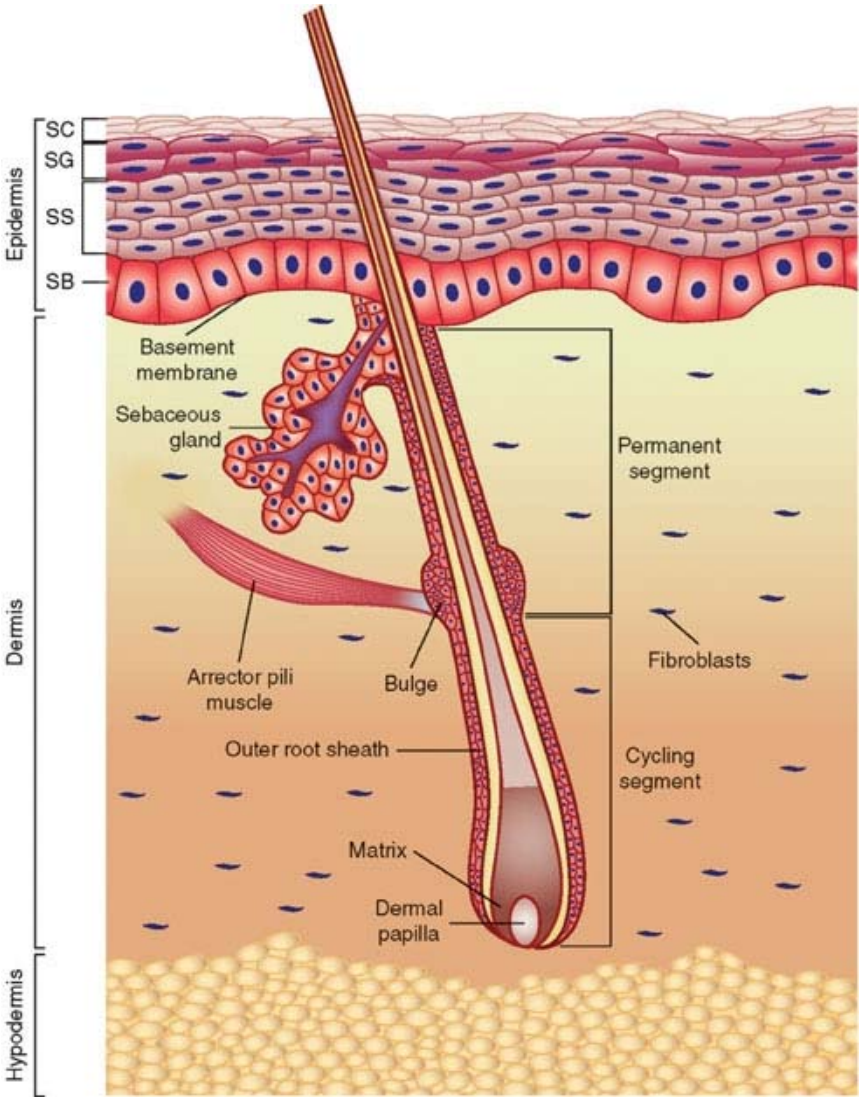
any animal testing is performed. In addition to providing a means of implementing the Three Rs, tiered testing schemes are intended to optimise the use of existing knowledge and resources, while maximising the accuracy of predicted classifications. An important feature in the design of such strategies is that chemicals which are predicted to be toxic in an early step are classified without further assessment. Conversely, chemicals which are predicted to be non-toxic proceed to the next step for further assessment. In this way, it is intended that toxic chemicals will be identified by non-animal methods, while the animal test(s) performed at the end of the stepwise procedure will merely serve to confirm predictions of non-toxicity made in previous steps.

Various tiered testing strategies been proposed in the literature for acute local toxicity, including skin irritation/corrosion (Basketter *et al.*, 1994; Botham *et al.*, 1998), skin sensitisation (Basketter *et al.*, 1995), and eye irritation (Spielmann *et al.*, 1996). One proposal combines the prediction of acute local and systemic toxicity into a single strategy (Barratt *et al.*, 1995).

At the regulatory level, stepwise approaches for classifying skin irritants and corrosives, and for classifying eye irritants and corrosives, are allowed for in OECD Test Guidelines 404 (OECD, 1992) and 405 (OECD, 1987), respectively, although no particular strategies are specified in these guidelines. Testing strategies proposed at an OECD workshop in January 1996 (OECD, 1996) were adopted, with minor amendments, by the OECD in November 1998 (OECD, 1998). The OECD testing strategies for skin irritation/corrosion, and for eye irritation/corrosion, are reproduced in Figures 2.3 and 2.4, respectively.

The author has evaluated the tiered testing approach to hazard classification, as it applies to the prediction of acute dermal and ocular toxicity (Worth *et al.*, 1998; Worth & Fentem, 1999). A development of this work is reported in Chapter 10.

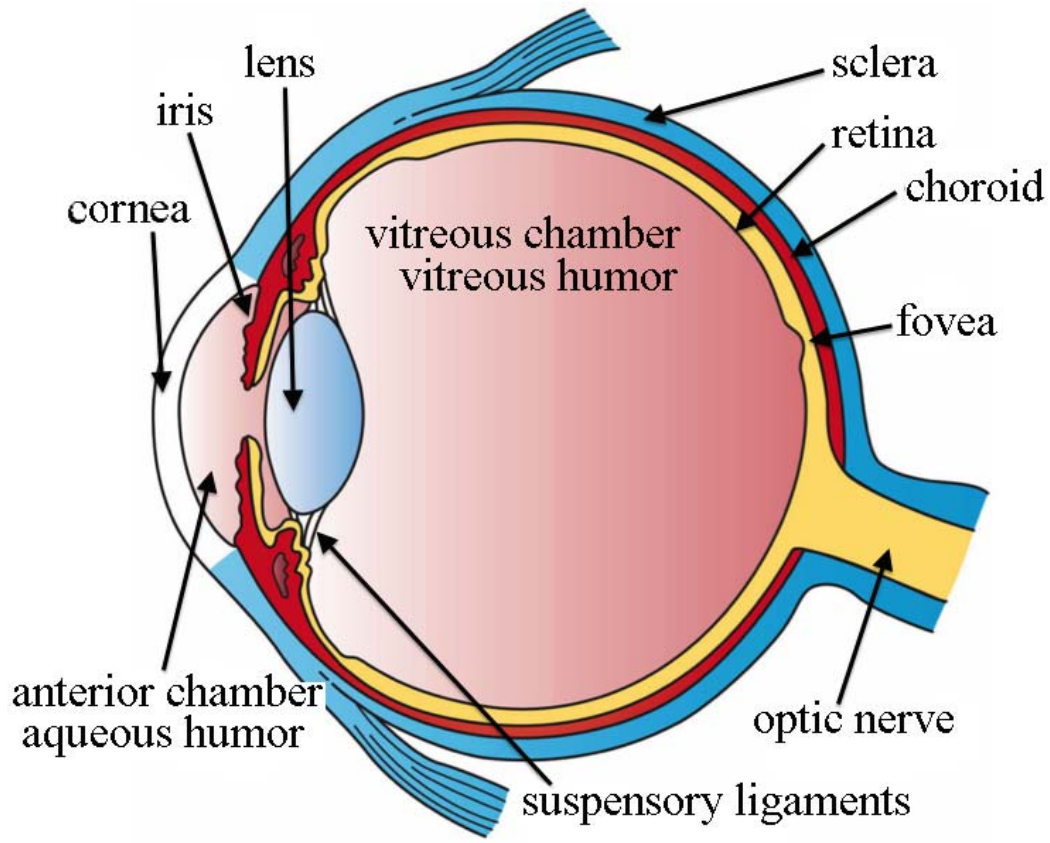
Figure 2.1 Anatomy of the skin[a]



Footnote to Figure 2.1

[a]Taken from Wikimedia Commons (https://commons.wikimedia.org/wiki/Category:Human_skin#/media/File:Anatomy_of_the_skin.jpg) under the Creative Commons Attribution-ShareAlike 3.0 Unported (CC BY-SA 3.0) licence.

Figure 2.2 Anatomy of the eye[a]



Footnote to Figure 2.2

[a]Taken from Wikimedia Commons (https://commons.wikimedia.org/wiki/File:Three_Main_Layers_of_the_Eye.png) under the Creative Commons Attribution-ShareAlike 3.0 Unported (CC BY-SA 3.0) licence.

Figure 2.3 The OECD tiered testing strategy for skin irritation and corrosion[a]

Step	Parameter		Finding	Conclusion
1a	Existing human or animal experience	⇒	Corrosive	Classify as corrosive
	⇓ Not corrosive or no data ⇓			
1b	Existing human or animal experience	⇒	Irritant	Classify as irritant
	⇓ Not irritant or no data ⇓			
1c	Existing human or animal experience	⇒	Not corrosive or irritant	No further testing
	⇓ No data ⇓			
2a	Structure-activity relationships or structure-property relationships	⇒	Corrosive	Classify as corrosive
	⇓ Not corrosive or no data ⇓			
2b	Structure-activity relationships or structure-property relationships	⇒	Irritant	Classify as irritant
	⇓ Not irritating or no data ⇓			
3	pH with buffering	⇒	pH $\leq$ 2 or $\geq$11.5	Classify as corrosive
	⇓ Not pH extreme or no data ⇓			
4	Existing dermal data in animals indicate no need for animal testing	⇒	Yes	Classify as corrosive/irritant
	⇓ No indication or no data ⇓			
5	Valid and accepted *in vitro* dermal corrosion test	⇒	Positive response	Classify as corrosive
	⇓ Negative response or no data ⇓			

Figure 2.3 cont'd. The OECD tiered testing strategy for skin irritation and corrosion[a]

Step	Parameter		Finding	Conclusion
6	Valid and accepted *in vitro* dermal irritation test	⇒	Positive response	Classify as irritant
	⇓			
	Negative response or no data			
	⇓			
7	*In vivo* dermal corrosion test (1 animal)	⇒	Corrosive response	Classify as corrosive
	⇓			
	Negative response			
	⇓			
8	*In vivo* dermal irritation test (3 animals total)	⇒	Irritant response	Classify as irritant
	⇓			
	Negative response	⇒	No further testing	
	⇓			
9	When it is ethical to perform human patch testing	⇒	Irritant response	Classify as irritant
	⇓			
	Not as above	⇒	Non-irritant response	No further testing

Footnote to Figure 2.3

[a]Taken from OECD (1998) with minor modifications.

Figure 2.4 The OECD tiered testing strategy for eye irritation and corrosion[a]

Step	Parameter		Finding	Conclusion
1a	Data relating to historical human or animal experience	⇒ ⇒	Severe eye damage Eye irritant	Classify in Class A Classify in Class B
	⇓ No or don't know ⇓			
1b	Data relating to historical human or animal experience	⇒	Skin corrosive	Classify in Class A
	⇓ No or don't know ⇓			
1c	Data relating to historical human or animal experience	⇒	Skin irritant	Classify in Class B
	⇓ No or don't know ⇓			
2a	Structure-activity relationships or structure-property relationships	⇒	Severe eye damage	Classify in Class A
	⇓ No or don't know ⇓			
2b	Structure-activity relationships or structure-property relationships	⇒	Eye irritant	Classify in Class B
	⇓ No or don't know ⇓			
2c	Structure-activity relationships or structure-property relationships	⇒	Skin corrosive	Classify in Class A
	⇓ No or don't know ⇓			
3	pH/acid or alkaline reserve	⇒	pH $\leq$ 2 or pH $\geq$11.5	Classify in Class B
	⇓ 2 < pH < 11.5 (no buffering potential) ⇓			
4	Other information indicating the material is a dermal corrosive	⇒	Yes	Classify in Class A
	⇓ No ⇓			
5	Is a valid *in vitro* test available to assess severe damage to eyes	⇒	No	Go to step 6
	⇓ Negative response or no data ⇓			
5a	*In vitro* test for severe eye irritation	⇒	Severe eye damage	Classify in Class A
	⇓ Not a severe eye irritant ⇓			

Figure 2.4 cont'd. The OECD tiered testing strategy for eye irritation and corrosion[a]

Step	Parameter		Finding	Conclusion
6	Is a valid *in vitro* test available for eye irritation	⇒	No, but *in vitro* test for severe eye irritancy was negative	Go to step 8
	⇓			
		⇒	No, in the absence of any *in vitro* test	Go to step 7
	⇓			
	Yes			
	⇓			
6a	*In vitro* eye irritation test	⇒	Eye irritant	Classify in Class B
	⇓			
	No indication of eye irritant properties			
	⇓			
7	Experimentally assess skin corrosion potential (see testing strategy for skin irritation/corrosion)			
	⇓			
8	1 rabbit eye test	⇒	Serious eye damage	Classify in Class A
	⇓			
	No indication of eye irritant properties			
	⇓			
9	1 or 2 further rabbits	⇒	Eye irritant	Classify in Class B
	⇓			
	Not an eye irritant			

Footnote to Figure 2.4

[a]Taken from OECD (1998) with minor modifications. Chemicals in Class A have irreversible effects, whereas the effects of Class B chemicals are reversible.

Table 2.1 European Union classification system for eye irritation/corrosion

	R36 (Irritating to eyes)		R41 (Risk of serious damage to eyes)[c]	
	3 animals[a]	4-6 animals[b]	3 animals[a]	4-6 animals[b]
Corneal opacity	≥ 2.0 but <3.0	≥ 2.0 but <3.0	≥ 3.0	≥ 3.0
Iris lesion	≥ 1.0 but <2.0	≥ 1.0 but ≤1.5	≥ 2.0	> 1.5
Conjunctiva redness	≥ 2.5	≥ 2.5		
Conjunctiva chemosis	≥ 2.0	≥ 2.0		

[a]The classification is assigned if the mean tissue effect (averaged over the 24h, 48h and 72h time-points) is greater than or equal to the threshold value in at least two of the three animals.

[b]The classification is assigned if the mean tissue effect (averaged over the three time-points and over the six animals) is greater than or equal to the threshold value.

[c]A classification of R41 is also assigned if, in at least one animal, one of the eye effects has not reversed at the end of the observation period.

Table 2.2 OECD classification system for eye irritation/corrosion

	Category B[a]	Category A[b]
Corneal opacity	≥ 1.0	≥ 3.0
Iris lesion	≥ 1.0	> 1.5
Conjunctiva redness	≥ 2.0	
Conjunctiva chemosis	≥ 2.0	

Classifications A and B are assigned if the mean tissue effect (averaged over the 24h, 48h and 72h time-points) is greater than or equal to the threshold value in at least two of three tested animals.

[a]All effects have to be reversible within 21 days of treatment. The subcategory of B1 can be used for chemicals considered to be mildly irritating to the eyes, i.e. chemicals whose eye effects are reversible within 7 days of treatment.

[b]Category B is also applicable if, in at least one animal, an eye effect has not reversed, or is not expected to reverse, within 21 days of treatment.

Table 2.3 ***In vitro*** **endpoints for eye irritation**

Test	Endpoint(s)
Silicon microphysiometer test	**MRD_{50}** - the concentration of test substance required to produce a 50% reduction in the metabolic acidification rate of L929 fibroblasts
Red blood cell haemolysis test	**D_{low}** - the lowest concentration causing denaturation of hæmoglobin **D_{max}** - the maximum percentage denaturation seen at any concentration tested **H_{50}** - the concentration causing 50% hæmolysis
Neutral red uptake test	**NRU_{50}** - the concentration which causes 50% inhibition of neutral red uptake into 3T3 cells
Fluorescein leakage test	**FL_{20}** - the concentration which causes 20-% fluorescein leakage across a layer of MDCK cells
HETCAM method	subjective measures of **hæmorrhage**, **lysis** and **coagulation** in the chorioallantoic membrane of embryonated chicken eggs
Bovine corneal opacity/ permeability (BCOP) test	**BCOP score** - subjective measure of opacity and permeability in bovine corneas
Isolated, enucleated chicken eye test	Measures of **corneal swelling** (% of untreated control), **corneal opacity** (score) and **fluorescein retention** (score) in the cornea of enucleated chicken eyes
Isolated, enucleated rabbit eye test	Measures of **corneal swelling** (% of untreated control) and **corneal opacity** (score) in the cornea of enucleated rabbit eyes, taken at two time points

CHAPTER 3

METHODS I: CALCULATION OF PHYSICOCHEMICAL PROPERTIES

3.1 INTRODUCTION

A wide variety of physicochemical properties have been used as descriptor variables in structure-activity relationships (SARs), to express the shape, size, partitioning behaviour, and reactivity of chemicals (Dearden, 1990). In many cases, these properties can be determined experimentally or estimated by calculation. A special group of properties are the so-called topological descriptors, which can be derived from chemical graph theory alone. The emphasis of this chapter is on the algorithms and computer methods that can be used to calculate physicochemical descriptor variables, and on the interpretation of these properties.

3.2 SUBSTITUENT PROPERTIES

Substituent constants can be used as predictor variables when the goal of the investigation is to derive a SAR for a congeneric series of substances, i.e. a set of molecules whose structures differ only in terms of a few substituents on a parent molecule. The use of substituent constants is based on the assumption that the type of biological activity is determined by the common substructure, and that the substituents merely serve to modulate the size of the biological effect. A compilation of substituent constants is given in Hansch *et al.* (1995), and illustrations of their use are given in Hansch & Leo (1995).

3.2.1 Hansch constant

The hydrophobic constant (π) was introduced by Hansch *et al.* (1962) in an attempt to quantify the effect of substituents on the partitioning behaviour of phenoxyacetic acids:

$$\pi = \log P_{(R)} - \log P_{(H)} \qquad \text{(Equation 3.1)}$$

where $\log P_{(R)}$ and $\log P_{(H)}$ are the logarithms of the octanol-water partition coefficient for the substituted phenoxyacetic acid (R-$PhOCH_2COOH$) and the unsubstituted acid ($PhOCH_2COOH$), respectively.

3.2.2 Hammett constant

The Hammett electronic constant (σ) quantifies the electronic effect of a substituent on the dissociation of benzoic acid (Reaction 3.1), and is defined by Equation 3.2:

$PhCOOH \leftrightarrow PhCOO^- + H^+$ (Reaction 3.1)

$\sigma = \log Ka_{(R)} - \log Ka_{(H)} = \log (Ka_{(R)} / Ka_{(H)})$ (Equation 3.2)

where $Ka_{(R)}$ = equilibrium constant of the substituted acid, and $Ka_{(H)}$ = equilibrium constant of the unsubstituted acid. A generalised form of this equation, known as the Hammett equation (Hammett, 1937), is used to generalise the use of σ to other chemical reactions involving aromatic species:

$\log (K_{(R)} / K_{(H)}) = \rho\sigma$ (Equation 3.3)

where ρ is a constant for a given reaction (equal to unity for the dissociation of benzoic acids), σ is the substituent constant, and K is the equilibrium constant for the reaction of interest.

A given substituent is likely to have different σ values, depending on whether it is in the ortho, meta or para position to the ionisation centre. Therefore, σ_o, σ_m and σ_p values have been produced to describe ortho, meta and para effects, respectively. The σ_o values have proven more difficult to obtain reliably, due to variability in short-range steric and hydrogen-bonding effects, and therefore tend to be situation specific.

3.2.3 Taft constant

The Taft steric constant (E_s) quantifies the effect of a substituent (R) on the hydrolysis (Reaction 3.2) of a methyl ester (RCOOMe), and is defined by Equation 3.4:

$RCOOMe + H_2O \leftrightarrow RCOOH + MeOH$ (Reaction 3.2)

$E_s = \log k_{(R)} - \log k_{(Me)} = \log (k_{(R)} / k_{(Me)})$ (Equation 3.4)

where $k_{(R)}$ = rate constant for the hydrolysis of RCOOMe, and log $k_{(Me)}$ = rate constant for the hydrolysis of the MeCOOMe.

3.2.3 Sterimol parameters

The sterimol parameters, developed by Verloop *et al.* (1976), characterise the shape and volume of a substituent: L is the length of the substituent; B1-B4 are measures of width in four directions perpendicular to L, defined in such a way that they are mutually perpendicular and that B1≤ B2≤ B3≤ B4; B5 is the maximum width perpendicular to L.

3.3 WHOLE-MOLECULE PROPERTIES

If a SAR is to be derived for a heterogeneous group of substances, i.e. a set of substances with no common, parent structure, it will be necessary to use whole-molecule properties as the predictor variables. In addition, whole–molecule properties may be necessary to model the activity of a congeneric series, if the classical substituent constants prove to be irrelevant or insufficient.

3.3.1 Octanol-water partition coefficient

The octanol-water partition coefficient (P) expresses the relative affinity of two immiscible phases (octanol and water) for a given solute, and is widely used as a measure of hydrophobicity. By definition, P is the ratio of the solute concentration in the non-polar phase (octanol) to the solute concentration in the polar phase (water), at equilibrium:

$$P = [solute]_o / [solute]_w \qquad \text{(Equation 3.5)}$$

where $[solute]_o$ is the equilibrium concentration of the solute in octanol; and $[solute]_w$ is the equilibrium concentration of the solute in water.

The partition coefficient is generally expressed as its common logarithm (logP). A substance with a logP of zero has an equal affinity for both phases, whereas a negative

logP value indicates that the solute has a greater affinity for water, and a positive value indicates a greater affinity for octanol.

$$\log P = \log \{[\text{solute}]_o / [\text{solute}]_w\} = \log [\text{solute}]_o - \log [\text{solute}]_w \qquad \text{(Equation 3.6)}$$

When used as a predictor variable, logP is commonly interpreted as a measure of the ease with which the solute can traverse lipid membranes. If a substance has a low logP value, it does not readily enter membranes, whereas if it has a high logP value, it has a tendency to stay inside them. Thus, only substances with intermediate logP values can readily traverse membranes (by passive diffusion).

The first method for calculating logP, developed by Hansch and Fujita (1964), was based on the addition of the hydrophobic substituent constant (π) to the measured logP of a parent compound. This approach was based on two assumptions: a) that π values are additive; and b) that the hydrogen atom has a π value of zero. Subsequently, it was shown that π values are not strictly additive (Fujita *et al.*, 1964), and that the π value of hydrogen is not zero, which means that the error in the calculated logP increases with the number of π values summed. Furthermore, the method was limited to the derivation of logP from a parent structure whose logP was already known. To address these shortcomings, subsequent methods have adopted a fragmental approach, in which the molecule of interest is broken down into a number of substructures, and the hydrophobic constants for these fragments are summed, taking non-additivity into account by using correction factors. The first of these fragmental methods was published by Nys & Rekker (1973). Nowadays, a number of commercially-available software packages have implemented (in different ways) the fragmental approach. Examples include the TSAR package (Oxford Molecular, UK), which uses the atomic fragment method of Viswanadhan (Viswanadhan *et al.*, 1989); and the KOWWIN software (Syracuse Research Corporation, NY, USA), which uses a method based on atomic and larger group fragments (Meylan & Howard, 1995). A different approach, based on quantum mechanical calculations, is the molecular orbital approach of Bodor (Bodor *et al.*, 1989). A comprehensive survey of different methods for predicting logP is given in Buchwald & Bodor (1998).

3.3.2 Dissociation constant

Strictly speaking, calculations of logP are only accurate for non-ionisable substances, or for ionisable substances in their uncharged forms. If the substance of interest contains *n* ionisable groups, it will exist as a mixture of 2^n different ionic forms (microspecies), each of which will have its own partition coefficient. In such cases, a more appropriate predictor variable is the octanol-water distribution coefficient (logD), which is defined as the ratio of the total concentration in octanol to the total concentration in water:

$$\log D = \log \left(\frac{\sum c_{io}}{\sum c_{iw}} \right) \qquad \text{(Equation 3.7)}$$

where c_{io} is the concentration of the *i*-th microspecies in octanol, and c_{iw} is the concentration of the *i*-th microspecies in water.

For substances containing a single ionisable group, logD can be calculated from the pK_a of the substance and from the pH of the aqueous solution. A simple algorithm for predicting the logD of acids is given by Equation 3.8 (Scherrer & Howard, 1977):

$$\log D = \log P - \log (1+10^{\,pH-pKa}) \qquad \text{(Equation 3.8)}$$

For substances containing multiple ionisation groups, the logD calculation needs to be based on the partition coefficients and ionisation constants of all microspecies. Several commercially-available software packages perform such calculations by using empirically-derived equations; examples include ACD/LogD Suite and ACD/logD Batch (Advanced Chemistry Development Inc., Canada), and PrologD (CompuDrug Chemistry Ltd, Hungary). The PrologD algorithm is described in Csizmadia *et al.* (1997).

3.3.3 Acid-base dissociation constant

The acid-base dissociation (ionisation) constant (K_a) is a measure of the strength of a weak acid - the larger the K_a value, the stronger the acid. It is defined (Equation 3.9) as the equilibrium constant for the dissociation of the acid (HA) in water (Reaction 3.3):

$HA\ (aq) \leftrightarrow H^+\ (aq) + A^-\ (aq)$ (Reaction 3.3)

$$K_a = \frac{[H^+(aq)][A^-(aq)]}{[HA(aq)]}$$ (Equation 3.9)

Since K_a is generally a large negative number, it is more usual to use pK_a, the negative logarithm of K_a. Thus, the smaller the pK_a value of a substance, the more it is acidic:

$pK_a = - \log K_a$ (Equation 3.10)

Algorithms for predicting pKa adopt a fragmental approach. For example, the ACD/pKa algorithm, implemented in the ACD/pKa software (Advanced Chemistry Development Inc., Canada), decomposes the molecule of interest into ionisable substructures, and the pKa value of each substructure is calculated by using empirically-derived Hammett-type equations (equations in which pKa is derived from electronic substituent constants), taking the molecular environment of each ionisable group into account. Another fragmental approach, based on the method of Perrin (Perrin *et al.*, 1981), has been implemented in the pKalc software (CompuDrug Chemistry Ltd, Hungary).

3.3.4 Aqueous solubility

The aqueous solubility (S) of a substance, defined as the maximum concentration of the substance that will dissolve in pure water at a specified temperature, provides a measure of hydrophilicity. The logarithm of the aqueous solubility at 25°C for neutral species can be calculated from logP and the melting point (MP). For example, Yalkowsky & Valvani (1980) reported the following equation:

$\log S = 0.87 - 1.05 \log P - 0.012\ MP$ (Equation 3.11)
$n = 155,\ s = 0.308,\ r^2 = 0.98$

where S is the aqueous solubility (mol/l), and MP is the melting point in degrees Celsius.

The WSKOWIN software (Syracuse Research Corporation, NY, USA) uses two equations:

$$\log S = 0.796 - 0.854 \log P - 0.00728 MW + C \quad \text{(Equation 3.12)}$$

$$\log S = 0.693 - 0.960 \log P - 0.0092(MP-25) - 0.00314 MW + C \quad \text{(Equation 3.13)}$$

where S is the aqueous solubility (mol/l), MW is the molecular weight (g/mol), MP is the melting point, and C is a correction term. Different corrections are applied for different chemical classes. Equation 3.12 is used in the absence of a measured melting point, whereas Equation 3.13 is used when a measured melting point is available. The WSKOWIN methodology is described in more detail in Meylan *et al.* (1996). Because aqueous solubilty is (inversely) correlated with logP to such a large extent, statistical models only need to be based on one of these descriptors, to avoid collinearity problems (redundancy and parameter instability).

3.3.5 Hydrogen-bonding

The simplest way of modelling hydrogen bonding, first reported by Fujita *et al.* (1977), is to use an indicator variable, set to one if the substance (or substituent) is capable of forming a hydrogen bond, and set to zero otherwise. Another approach, suggested by Charton & Charton (1982), is to count the total number of hydrogen bonds a substance (or substitutent) is capable of making, with a possible subdivision into hydrogen bond donors and acceptors. Hydrogen bond counts for whole molecules can be calculated with the MOLCONN-Z software (Hall Associates Consulting, MA, USA), in which case the number of hydrogen bond donors is the number of hydrogen atoms attached to an electronegative atom (O or N), and the number of hydrogen bond acceptors is the number of atoms carrying a lone pair.

Various quantitative descriptors of hydrogen-bonding potential have also been proposed. Kamlet *et al.* (1983) introduced terms for hydrogen-bond donor acidity (α) and hydrogen-bond acceptor basicity (β), derived from the effects of solvents on the ultraviolet and visible spectra of solutes. A different approach, developed by Abraham and coworkers (Abraham, 1993), is based on scales of hydrogen-bond acidity (α_2^H and

$\Sigma\alpha_2^H$) and basicity (β_2^H and $\Sigma\beta_2^H$), obtained by measuring hydrogen-bonding equilibrium constants in tetrachloromethane.

To avoid the need for experimentally determined descriptors, Cramer *et al.* (1993) proposed the use of four hydrogen-bonding descriptors, based on quantum-mechanical calculations: a) the most negative partial charge on any atom in the molecule (Q^-); b) the most positive partial charge on any hydrogen atom (Q^+); c) the energy of the highest occupied molecular orbital (E_{HOMO}); and d) the energy of the lowest unoccupied molecular orbital (E_{LUMO}). The first two descriptors represent the ionic contribution to the hydrogen bond, whereas the second two descriptors represent the covalent contribution. The hydrogen-bonding donor capacity of a substance is described Q^+ and E_{LUMO}, whereas the hydrogen-bonding acceptor capacity is described by Q^- and E_{HOMO}.

3.4 TOPOLOGICAL INDICES

Topological indices are numerical descriptors of molecular topology. They represent an attempt to quantify some aspect of the two-dimensional structure of a molecule, so that this can be related to its physicochemical and biological properties. A wide variety of topological indices have been proposed in the literature (reviewed by Sabljic, 1990), of which the most widely used have probably been the molecular connectivity indices.

The concepts of the ‘molecular graph’ and the ‘distance matrix’ are of particular importance in understanding the derivation of topological indices. A molecular graph is a topological object that models the connectivity of atoms (vertices) in a molecule. If the distance (Dij) between any two atoms (i and j) is taken to be the number of bonds separating them, the distance matrix of the molecular graph is a symmetric N x N matrix containing all of the Dij distances, where N is the number of atoms. In general, only heavy (non-hydrogen) atoms are considered. For example, the hydrogen-suppressed graph of 2-methyl butane (Figure 3.1) can be used to generate its corresponding distance matrix (Table 3.1).

3.4.1 Molecular connectivity indices

Molecular connectivity (χ) indices are topological descriptors that encode structural features such as size, branching, unsaturation, heteroatom content and cyclicity (Kier & Hall 1986; Hall & Kier, 1991). They are derived by decomposing the molecular structure into fragments, and are represented by the following generic symbol:

$$ {}^{m}\chi_{t}^{v} $$

where *m* is the order of the index (number of bonds in the structural fragment), and *t* is the type of fragment into which the structure has been decomposed. Four types of fragment are distinguished by the right-side subscript *t*: continuous path-type fragments (p), clusters (c), path clusters (pc), and chains (ch). If there is no right-side subscript, the index is assumed to be a path-type index. The presence of the right-side superscript *v* indicates that the index is of the valence type, and its absence indicates that the index is of the simple type. In valence indices, the presence of heteroatoms is taken into account, whereas in simple indices, heteroatoms are treated as carbon atoms.

To calculate the molecular connectivity indices, the molecule is first represented by its hydrogen-suppressed graph. Each atom is assigned two descriptors: a sigma electron descriptor (δ) and a valence electron descriptor (δ^v). The sigma electron descriptor (simple delta value) is equal to the number of neighboring non-hydrogen atoms, and is given by:

$$\delta = \sigma - h \qquad \text{(Equation 3.14)}$$

where σ is the number of sigma electrons and h is the number of hydrogen atoms. The simple delta values of the heavy atoms in a molecule are used to calculate the simple molecular connectivity indices. The valence electron descriptor is given by:

$$\delta^v = (Z^v - h) / (Z - Z^v - 1) \qquad \text{(Equation 3.15)}$$

where Z^v is the number of the valence electrons, Z is the number of all electrons (i.e. the atomic number), and h is the number of hydrogen atoms. The valence delta values are used to calculate the valence molecular connectivity indices.

The next step is to decompose the molecular skeleton into fragments of a particular size (order) and type (path, cluster, path/cluster, or chain), depending on the connectivity index being calculated. The zero-order connectivity indices, ${}^0\chi$ and ${}^0\chi^v$, are based on single-atom fragments, and are defined as:

$${}^0\chi = \sum (\delta_i)^{-0.5} \qquad \text{(Equation 3.16)}$$

$${}^0\chi^v = \sum (\delta_i^v)^{-0.5} \qquad \text{(Equation 3.17)}$$

where the summation is over all heavy (non-hydrogen) atoms (*i*). The zero-order valence connectivity index (${}^0\chi^v$) is reported to correlate with molecular weight (Protic & Sabljic, 1989).

The first-order connectivity indices are calculated by decomposing the molecular skeleton into two-atom fragments, and are defined as:

$${}^1\chi = \sum (\delta_i * \delta_j)^{-0.5} \qquad \text{(Equation 3.18)}$$

$${}^1\chi^v = \sum (\delta_i^v * \delta_j^v)^{-0.5} \qquad \text{(Equation 3.19)}$$

where *i* and *j* refer to two adjacent atoms. It is implicit that these indices are of the path type, since there is no other way of connecting two atoms but in a linear fashion. The simple first-order index, ${}^1\chi$, is reported to correlate with molecular surface area (Sabljic, 1987).

The second-order (${}^2\chi$) molecular connectivity indices are calculated by:

$${}^2\chi = \sum (\delta_i * \delta_j * \delta_k)^{-0.5} \qquad \text{(Equation 3.20)}$$

$${}^2\chi^v = \sum (\delta_i^v * \delta_j^v * \delta_k^v)^{-0.5} \qquad \text{(Equation 3.21)}$$

where i, j and k correspond to three consecutive non-hydrogen atoms. Again, these are necessarily path-type indices.

For molecular connectivity indices having an order of three or more, it is necessary to specify whether the index is of the path-type (t=p), cluster-type (t=c), or chain type (t=ch). In cluster indices, all bonds are connected to a common central atom. For example, the third-order cluster molecular connectivity indices (m=3, t=c) are calculated by decomposing the molecular structure into *tert*-butane-like fragments, which have three bonds joined to a common atom:

$$^{3}\chi_{C} = \sum (\delta_i * \delta_j * \delta_k * \delta_l)^{-0.5} \qquad \text{(Equation 3.22)}$$

where i, j, k and l are non-hydrogen atoms in the subgraph, and the summation is over all t-butane type subgraphs in the molecule. The cluster indices describe the degree of branching in a molecule: their values increase with the degree of branching.

The path/cluster indices also describe the degree of branching in a molecule, their values increasing with the degree of branching. The simplest of these indices is the fourth-order path/cluster molecular connectivity index ($^{4}\chi_{pc}$). It is calculated by decomposing the molecular structure into isopentane-like fragments, which consist of four adjacent bonds between non-hydrogen atoms, three of which are joined to the same non-hydrogen atom:

$$^{4}\chi_{PC} = \sum (\delta_i * \delta_j * \delta_k * \delta_l * \delta_m)^{-0.5} \qquad \text{(Equation 3.23)}$$

where i, j, k, l and m are non-hydrogen atoms in the subgraph, and the summation is over all isopentane type subgraphs in the molecule.

The chain molecular connectivity indices (χ_{ch}) provide a measure of cyclicity and the degree of substitution present on any rings. The simplest chain indices are third-order indices, calculated by decomposing the chemical graph into cyclopropane-like subgraphs. The sixth-order chain indices are derived from benzene-like fragments, whereas the seventh-order indices correspond to mono-substituted benzene rings.

3.4.2 Difference molecular connectivity indices

It is desirable that the descriptors used in a SAR model are orthogonal. In the case of molecular connectivity indices, collinearity is unlikely to pose a significant problem if the data set consists of diverse molecular structures. However, if the set of molecules covers a large range in molecular size and/or there is not a high degree of structure diversity, a significant degree of collinearity could arise. The main cause of collinearity is the significant contribution of molecular size to the constitution of the chi indices. Therefore, to address the collinearity problem, Kier & Hall introduced the difference connectivity indices, i.e. simple and valence path indices from which the contribution of size is subtracted, so that the salient aspects of structure (branching, cyclisation, heteroatom content, heteroatom position) are emphasised (Kier & Hall, 1991). The difference connectivity indices are defined by Equation 3.24 for simple path indices, and by Equation 3.25 for valence path indices:

$$d^{m}\chi_n = {}^{m}\chi_p - {}^{m}\chi_n \qquad \text{(Equation 3.24)}$$

$$d^{m}\chi_n{}^{v} = {}^{m}\chi_p{}^{v} - {}^{m}\chi_n{}^{v} \qquad \text{(Equation 3.25)}$$

where ${}^{m}\chi_n$ is the simple path index, and ${}^{m}\chi_n{}^{v}$ is the valence path index, for the unbranched, acyclic graph that has the same size as the given molecule. The indices ${}^{m}\chi_n$ and ${}^{m}\chi_n{}^{v}$ are defined for each order of path index.

3.4.3 Delta connectivity and sum connectivity indices

The simple connectivity indices encode only sigma electrons, whereas the valence connectivity indices encode pi and lone pair electrons in addition to sigma electrons. Therefore, subtraction of the simple connectivity index from the corresponding valence index of the same order results in an index that encodes only pi and lone pair elections. This index is called the delta connectivity ($\Delta\chi$) index:

$$\Delta\,{}^{m}\chi_t = {}^{m}\chi_t{}^{v} - {}^{m}\chi_t \qquad \text{(Equation 3.26)}$$

where t denotes the four types of connectivity indices (p, c, pc and ch). The sum connectivity ($\Sigma\chi$) index is given by Equation 3.27:

$$\sum {}^{m}\chi_t = {}^{m}\chi_t{}^{v} + {}^{m}\chi_t \qquad \text{(Equation 3.27)}$$

In general, the sum connectivity index is orthogonal to the delta connectivity index of the same order, so the combined use of the delta and sum connectivity indices represents another way of reducing collinearity in the data set.

3.4.4 Molecular shape (kappa) indices

Attributes of molecular shape are encoded into shape indices (kappa values), derived from the counts of one-bond (^{1}P), two-bond (^{2}P) and three-bond (^{3}P) fragments in the molecular structure (Kier, 1985 & 1990). The values of the kappa indices are calculated from the path count ^{m}P and from the number of atoms (A) in the molecular structure. The first-order and second-order kappa indices are defined by Equations 3.28 and 3.29, respectively. The definition of the third-order kappa index depends on whether the number of atoms is even (Equation 3.30) or odd (Equation 3.31):

$$^{1}\kappa = A\,(A\text{-}l)^2 / (^{1}P)^2 \qquad \text{(Equation 3.28)}$$

$$^{2}\kappa = (A\text{-}1)(A\text{-}2)^2 / (^{2}P)^2 \qquad \text{(Equation 3.29)}$$

$$^{3}\kappa = (A\text{-}2)(A\text{-}3)^2/(^{3}P)^2 \qquad \text{(Equation 3.30)}$$

$$^{3}\kappa = (A\text{-}1)(A\text{-}3)^2/(^{3}P)^2 \qquad \text{(Equation 3.31)}$$

The first-order kappa index is claimed to encode cyclicity, whereas the second-order and third-order indices are said to encode spatial density (star-graph likedness), and centrality of branching, respectively (Kier, 1990).

The presence of atoms other than sp^3-hybridised carbon can be taken into consideration in the above equations by replacing ^{m}P with $^{m}P+\alpha$, and A with $A+\alpha$, where:

$$\alpha = \{r(x) / r(C[sp3])\} - 1 \qquad \text{(Equation 3.32)}$$

r(x) is the covalent radius of atom x, and r(C[sp3]) is the covalent radius of carbon in the sp^3-hybridised state. The resulting kappa values are designated as ${}^{m}\kappa\alpha$.

In the case of small molecules, certain ${}^{m}P$ quantities may be equal to zero, which is problematic in that the kappa indices calculated by the above equations would be infinite. To overcome this problem, the kappa values of such molecules are arbitrarily fixed. For example, the ${}^{1}P$ value is zero for any molecule represented by a single point in its chemical graph, such as methane. In this case, a ${}^{1}\kappa$ value of 1 is adopted, since, for straight chain molecules, ${}^{1}\kappa$ is generally equal to A. The calculation of ${}^{2}\kappa$ values leads to zero values for molecules comprising just one or two heavy atoms, such as methane and ethane. In these cases, ${}^{2}\kappa$ values of 0 and 1 are proposed for methane and ethane, respectively. Similarly, the ${}^{3}\kappa$ values calculated from the above equations are zero for methane, ethane, and propane, while the value for butane is 4.000, which is the same as for pentane. More useful values of ${}^{3}\kappa$ values for these molecules are defined as: zero for methane, 1.450 for ethane, 2.000 for propane, and 3.378 for butane.

3.5 MOLECULAR MODELLING

A number of physicochemical properties that are useful predictor variables cannot be estimated from one-dimensional or two-dimensional representations of molecular structure, but need to be estimated from the three-dimensional structures of molecules. Examples of such properties include solvent-accessible surface areas and volumes, and electronic properties, such as dipole moment, frontier orbital energies, and partial atomic charges. For simplicity, the calculation of such properties is generally performed on an energy-minimised molecular structure (i.e. a structure in which the net effect of all intra-molecular forces is a minimum), even though the active forms of molecules may actually be higher-energy conformations. To further simplify these calculations (called single-point calculations), the minimum-energy structure is often determined for the gas-phase (i.e. solvent effects are ignored).

There are two main approaches to energy minimisation and single point calculation - molecular mechanics and quantum mechanics. Molecular mechanics is mathematically the simpler (and computationally the quicker) procedure, but it is also less accurate. It is appropriate for the calculation of properties that are not dependent on the distribution of

electronic charge in the molecule (e.g. surface area and volume). For the calculation of electronic properties, it is necessary to resort to quantum mechanical methods, among which semi-empirical methods are the simplest, and *ab initio* methods are the more complex. To reduce the time needed for a quantum mechanical minimisation, it is common to perform a molecular mechanical minimisation first, to produce an approximate conformation which can be optimised by the quantum mechanical method.

3.5.1 Molecular mechanics

The methods of molecular mechanics treat molecules as hard spherical particles (atoms) connected by springs (bonds) of different elasticities. Nuclei and electrons are not considered, so electronic properties cannot be calculated. The energy of the molecule is calculated as the sum of energies corresponding to bond stretching, bending (bond angle deformation), torsion (rotation of adjacent groups about a common bond), and non-bonding interactions (van der Waals interactions and coulombic interactions). In addition, the energy due to intra-molecular hydrogen bonding and out-of-plane bending may also be considered. Each of these energies is described by a potential energy function. The sum of all energy terms, parameterised with respect to a well-defined data set, is called a force field.

Force fields can be divided into two main groups. In the first type of force field, all atoms are considered in the energy calculation, including hydrogen atoms. This is called an ‘all atom’ approach, and is appropriate for the modelling of small molecules. In the second type of force field, most of the hydrogen atoms are not explicitly considered, but are accounted for implicitly by carbon atoms with an expanded van der Waals radius. The only hydrogen atoms considered explicitly are those having the potential to participate in hydrogen bonding. This is known as the ‘united atom’ approach, and is used to decrease the computational time needed to model large biological molecules (especially proteins and nucleic acids). A commonly used all-atom force field is MM2 (Allinger, 1977), and a commonly used united-atom force field is AMBER (Weiner *et al.* 1984).

To minimise the energy (optimise the geometry) of a molecule, a force field is chosen and applied to the input conformation. The total energy is calculated, the conformation

is then altered in an attempt to reduce energy, and the total energy is computed again. This process is repeated until a minimum energy is found. There are several mathematical algorithms for carrying out energy minimisation. For example, the steepest descents and conjugate gradients algorithms alter the molecular conformation by searching for the minimum value of the energy gradient (the first derivative of the total energy with respect to the atomic coordinates of the molecule). In contrast, the Newton-Raphson method also considers the second derivative of the energy function to find the minimum energy. The steepest descents and conjugate gradients methods are quicker, and are appropriate when the input conformation is far away from its minimum energy conformation. Newton-Raphson methods are slower, and are only appropriate for structures that are close to their energy minimum.

Having found a minimum energy conformation, it is not always certain whether this represents the global minimum or a local minimum. A number of methods address 'the problem of the global minimum' by generating an alternative starting structure, carrying out energy minimisation, and comparing the resulting 'minimum' energy with the original one. This process continues until conformation space has been searched with sufficient thoroughness (according to some predefined criterion). Methods that generate multiple starting geometries can be divided into deterministic searches, which cover conformation space systematically, and stochastic searches, which explore conformation space in a random manner. A number of these methods are compared in Saunders et al. (1990).

3.5.2 Quantum mechanics

Theoretical background

The energy of an isolated molecule can be obtained by solving the Schrödinger equation (Schrödinger, 1926), which (in its time-independent and non-relativistic form) can be written as:

$$\hat{H}\psi = E\psi \qquad \text{(Equation 3.33)}$$

where $\hat{H}$ is the Hamiltonian operator, ψ is the wavefunction, and E is the energy of the system (relative to the state in which the component atoms are infinitely separated and

at rest). The Hamiltonian operator expresses a mathematical procedure which is applied to the wavefunction to calculate the total energy of the molecule. The wavefunction itself describes the distribution of particles (nuclei and electrons) in space, and is therefore a function of spatial coordinates. According to the Born interpretation, the square of ψ (for a given particle) at a given point in space can be interpreted as the probability of finding the particle at that point.

Since the masses of the nuclei are much larger than those of the electrons, and their velocities much smaller, the problem of solving the Schrödinger equation can be simplified by separating it into two parts: an electronic part, describing the motions of the electrons in a field of fixed nuclei, and a nuclear part, describing the motions of the nuclei. The total molecular wavefunction is considered to be the product of an electronic and a nuclear wavefunction. The advantage of this approach, known as the Born-Oppenheimer approximation, is that only the electronic wavefunction needs to be solved. A further approximation, the independent electron approximation, simplifies the total electronic wavefunction for a many-electron system (molecule) as the sum of one-electron molecular wavefunctions. Thus, the one-electron Hamiltonian operator is a mathematical procedure for generating the kinetic and potential energies of an electron in a field of fixed nuclei.

In general, there is more than one acceptable solution to the Schrödinger equation, so it is usual to derive a set of different electronic wavefunctions, each of which is associated with an energy. Each wavefunction can be regarded as the mathematical description of a one-electron molecular orbital, and the energy associated with that wavefunction can be regarded as the energy required to remove an electron from that orbital.

The molecular wavefunction for an electron can be expressed as the product of a spatial function and a spin function. Neglecting electron-electron interactions, the spatial function is usually expressed as a linear combination of known one-electron wavefunctions, called a basis set. Typically, the one-electron wavefunctions in the basis set take the mathematical form of atomic orbitals, and the expansion is called a linear combination of atomic orbitals (LCAO):

$$\psi_i = \sum c_{ij} \phi_j \qquad \text{(Equation 3.34)}$$

where ψ_i = the (spatial) molecular wavefunction of electron *i*; ϕ_j are the one-electron wavefunctions (atomic orbitals) in the basis set; and c_{ij} are the coefficients that determine the contribution of each atomic wavefunction to the molecular wavefunction. The goal of LCAO methods is to find the coefficients c_{ij} that best approximate the molecular wavefunction ψ_i. According to the variational principle, this is equivalent to finding the coefficients that minimise the energy of the molecular wavefunction for a given choice of Hamiltonian and basis set. Without going into mathematical details, the energy E_i of electron *i* is calculated from a number of integrals involving the basis functions ϕ_j. Two types of integral are distinguished: coulomb integrals and overlap integrals. A coulomb integral can be interpreted as the ionisation energy for a one-electron molecular orbital, and an overlap integral as a measure of the spatial overlap between two atomic orbitals. Semi-empirical methods omit or approximate some of these integrals (to speed up the calculation), whereas *ab initio* methods evaluate all of the integrals (involving the chosen basis set) .

The independent electron approximation means that interactions between electrons are neglected, an approximation which becomes increasingly unreliable as the number of electrons in the system increases. Electron interactions include not only coulombic repulsions, but also Pauli interactions. i.e. the tendency for two electrons of opposite spin to attract each other, while electrons of the same spin repel each other.

To partially account for coulombic interactions, Hartree & Fock proposed that the total electronic wavefunction could still be regarded as the product of one-electron basis functions, except that the Hamiltonian should be replaced by an effective Hamiltonian (called a Fock operator), which takes into account the average field effect due to all other electrons in the molecule. Thus, the Fock operator describes a two-electron system, in contrast to the one-electron system described by the conventional Hamiltonian operator. The Hartree-Fock treatment introduces a third type of integral into the energy calculation, called the exchange (or resonance) integral, which has no simple physical interpretation. Again, the LCAO method is used to determine the coefficients c_{ij} that best determine ψ_i (i.e. the minimum value of E_i) However, a paradox arises because the Fock operator is dependent on the coefficients c_{ij}. The paradox is solved by using an iterative procedure in which the coefficients are first estimated and

used to solve the equations for ψ_i. A new Fock operator is determined from the derived coefficients c_{ij}, and the process is repeated until the coefficients c_{ij} generated during an iteration are identical to those used as input for that iteration. This is called the Hartree-Fock Self-Consistent Field (HF-SCF) procedure, of which there are two types: Restricted Hartree-Fock (RHF) and Unrestricted Hartree-Fock (UHF). In a RHF procedure, the two electrons in a particular molecular orbital are constrained to have the same molecular orbital coefficients. This is the appropriate method for closed shell molecules (in which all electrons are paired). In a UHF procedure, the two electrons in a particular molecular orbital are allowed to have different coefficients, which is appropriate for open shell molecules (which have unpaired electrons), such as free radicals.

The HF-SCF procedure does not fully account for all electron-electron interactions. In part, this is because each electron in the molecule is considered to experience the average electrical field of the remaining electrons, rather than the instantaneous forces due to electron movement. In addition, the Pauli effect is not considered. These effects are referred to collectively as electron correlation (a term expressing the idea that electrons move in a mutually dependent manner). The difference between the true energy and the Hartree-Fock energy is called the correlation energy.

A variety of post-Hartree Fock methods have been devised to calculate the correlation energy. For example, configuration interaction (CI) methods allow for the possibility that electrons are occasionally excited to energy levels higher than the ground state. These methods construct the molecular wavefunction from a multiple set of electron assignments in a manner analogous to the LCAO approach. Other post-Hartree Fock methods include Møller-Plesset Perturbation Theory, in which the difference between the correlated Hamiltonian and the Hartree-Fock Hamiltonian is expressed as a Taylor expansion of energy terms (Alberts & Handy, 1985), and Density Functional Theory, in which molecular orbital energy is calculated as a function of electron density integrated over all space (Labanowski & Andzelm, 1991).

Semi-empirical methods

Semi-empirical methods are based on a number of approximations to simplify the problem of solving the Schrödinger equation for a many-electron system: a) only

valence orbitals are included in their basis functions (i.e. core orbitals are ignored); b) some of the integrals are neglected or approximated; and c) electron correlation is accounted for implicitly through parameterisation, rather than explicitly. The methods differ in terms of: a) the basis sets they use; b) the parameterisation of their basis functions; and c) the specific integrals they neglect or approximate.

The first molecular orbital method to be proposed, Hückel Theory (HT), is an LCAO method in which σ electrons, σ-π interactions, and electron-electron repulsions are ignored (Hückel, 1931). Integrals that are not ignored are given fixed empirical values. HT is limited to predicting the energies of conjugated systems and gives no information about molecular structure. Extended Hückel Theory (EHT), developed by Hoffman (1963), can be used to produce qualitative descriptions of molecular orbitals. However, for most quantitative purposes, EHT has been superseded by more-recent methods.

More-recent semi-empirical methods are based on the 'neglect of differential overlap' (NDO) approximation, i.e. they neglect two-electron integrals involving different atomic orbitals. In decreasing order of approximation, NDO methods include: a) Complete Neglect of Differential Overlap (CNDO; Pople & Beveridge, 1970); b) Intermediate Neglect of Differential Overlap (INDO; Pople *et al.*, 1967); and c) Modified Intermediate Neglect of Differential Overlap (MINDO; Bingham *et al.*, 1975).

Current semi-empirical methods are based on the 'neglect of diatomic differential overlap' (NDDO) approximation, in which two-electron integrals are considered if they involve atomic orbitals on the same atom. Such methods include: a) Modified Neglect of Differential Overlap (MNDO; Dewar and Thiel, 1977); b) Austin Model 1 (AM1; Dewar *et al.*, 1985); and c) Parametric Model 3 (PM3), a reparameterised version of AM1 (Stewart, 1989a & 1989b).

Ab initio methods

Like the semi-empirical methods discussed above, the most common *ab initio* methods are based on the LCAO approximation. Unlike semi-empirical methods, *ab initio* methods use basis functions that take account of core electrons in addition to valence electrons, and they evaluate all integrals in the calculation of the molecular

wavefunction. For this reason, *ab initio* methods take longer than semi-empirical methods and, at present, are limited to modelling smaller molecules.

There are two types of basis function - Slater-type orbitals (STOs), which provide an accurate representation of hydrogen-like orbitals, and Gaussian-type orbitals (GTOs), which are less accurate. In STOs, the electron density decays as an exponential function of distance, whereas in GTOs, the electron density decays as an exponential function of the squared distance. Ideally, the basis functions used by *ab initio* methods would be STOs, similar to those used by semi-empirical methods. However, since *ab initio* calculations are more computer-intensive than semi-empirical ones, and GTOs are easier to manipulate mathematically than STOs, the usual approach is to approximate the latter by sums of the former. The approximation of STOs by GTOs has given rise to the so-called STO-*N*G basis sets, in which each basis function is constructed from *N* GTOs. A unique feature of the STO-*N*G method is that the exponential coefficients for all atomic orbitals in a given shell are constrained to be identical, to simplify the calculation of integrals involving the basis functions. A minimal basis set contains only one basis function per atomic orbital; for example, there is just one basis function for H and He (corresponding to the 1s orbital), and 5 basis functions for Li to Ne (corresponding to the 1s, 2s and three 2p orbitals). STO-3G has become a standard for minimal basis calculations.

More-elaborate basis sets are the so-called split-valence basis sets, which have different basis functions for core orbitals and for valence orbitals (Binkley *et al.*, 1980). For example, the 6-31G basis set uses six GTOs to approximate core orbitals, three GTOs for the 2s and 2p (inner) valence orbitals, and one GTO for the 3s and 3p (outer) valence orbitals. The 6-31G basis set is said to be a valence-double-zeta (VDZ) basis set, because two basis functions are used for each valence atomic orbital. Triple zeta (TZ) and quadruple zeta (QZ) basis sets have also been developed, representing valence atomic orbitals with three and four basis functions, respectively.

For even greater accuracy, it is necessary to incorporate basis functions capable of representing orbital polarisation. These are denoted with asterisks; for example, the 6-31G* is the basis set for 6-31G to which d-orbital functions have been added to account for the polarisation of heavy atoms, and the 6-31G** basis set is the 6-31G* basis set

supplemented with p-orbital functions to account for the polarisation of hydrogen atoms (as well as heavy atoms). Finally, augmented basis sets contain so-called diffuse functions to accommodate loosely held electrons. These basis sets are designated by the addition of a plus sign to basis set designator (e.g. 6-31G+)

Thus, the success of an *ab initio* calculation depends upon the choice of an appropriate basis set for the molecule of interest. Ideally, the largest available basis set, with the most extended set of polarisation and diffuse functions, would be used. In practice, however, a compromise is required, since the time required for a calculation increases with the fourth power of the number of basis functions.

3.5.3 Conformation-dependent properties

Solvent-accessible surface area and molecular volume

The solvent-accessible surface is defined by the loci of the centre of a solvent molecule (typically water), when this molecule is 'rolled over' the molecular surface defined by the van der Waals radii of the molecule's constituent atoms. The area of the solvent-accessible surface is generally expressed in $Å^2$, and the volume enclosed by this surface in $Å^3$. Both of these properties may be regarded as measures of molecular size.

Dipole moment

The dipole moment of an electric dipole (pair of charges of opposite sign [± q] separated by a distance [r]) is a vector having the magnitude q.r and the direction of the line from the negative to the positive charge. Dipole moment is measured in coulomb metres (Cm), or in debyes (D), where $1D=3.338x10^{-30}$ Cm.

The dipole moment of a molecule ($\boldsymbol{\mu}$) is the vector sum of the dipole moments of its constituent electric dipoles:

$$\boldsymbol{\mu} = \sum q_i \mathbf{r_i} \qquad \text{(Equation 3.35)}$$

where q_i is the charge on atom *i*, and r_i is the distance of atom *i* from the origin (centre of charge or centre of mass). The bold font is used to denote a vector quantity.

It is generally the magnitude of this vector (μ), referred to hereafter as DM, that is used as a descriptor in SARs, although the magnitude of one or more of the vector's components (along the x, y and z cartesian axes) can also be used, if all molecules in the data set have been aligned with their first inertial axis parallel to the x axis, and their second and third inertial axes parallel to y and z axes, respectively.

Apart from being a measure of polarity, the physicochemical interpretation of DM is not well-established. One possibility is that DM expresses the extent to which a molecule, having entered a lipid membrane, diminishes the electric field across it. The rationale for this is that electric dipoles are known to align themselves anti-parallel to externally applied electric fields. Any reduction of the electrical gradient across a cell membrane could affect the transport of charged species into and out of the cell. An alternative hypothesis is that DM expresses the tendency of molecules to accumulate at the interface of the lipid bilayer. Such an accumulation has been proposed to increase the lateral pressure within the bilayer, thereby altering the ratio of open to closed ion channels (Cantor, 1997). In support of this idea, molecular dynamics simulations indicate that molecules of increasing polarity are more likely to accumulation at the bilayer interface (Pohorille & Wilson, 1996).

Frontier orbital energies

Molecular orbital methods yield a set of energy levels in which all the available electrons are accommodated. The highest filled energy level is called the Highest Occupied Molecular Orbital (HOMO). The next higher energy level, which is unoccupied because no more electrons are available, is the Lowest Unoccupied Molecular Orbital (LUMO). On the basis of Koopmans theorem, the HOMO and LUMO of a molecule can be approximated as its ionisation energy and electron affinity, respectively (Koopmans, 1933). In addition, HOMO can be used as a descriptor of basicity or nucleophilicity, and LUMO as a descriptor of acidity or electrophilicity.

Frontier orbital energies can be used to derive other descriptors. For example, the absolute hardness (η) of a molecule is defined as half the difference between the energies of the LUMO and the HOMO:

$$\eta = \tfrac{1}{2} (E_{LUMO} - E_{HOMO}) \qquad \text{(Equation 3.36)}$$

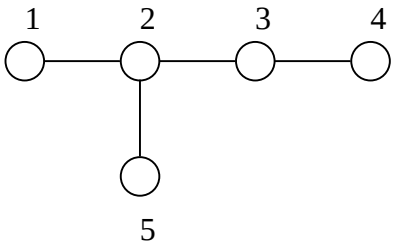
Absolute hardness has been inversely correlated with chemical reactivity (Pearson, 1993). Similarly, the absolute electronegativity (χ) is half the sum of the HOMO and LUMO energies:

$$\chi = ½ (E_{HOMO} + E_{LUMO}) \quad \text{(Equation 3.37)}$$

Electron density and related properties

Quantum-mechanical methods can be used to display the electron density contours of a molecule as a means of visualising possible sites of electrophilic attack. For quantitative information, partial atomic charges can be calculated. These properties can also be used to calculate the dipole moment of a molecule, or to determine the hydrogen-bonding descriptors (Q^+ and Q^-).

Figure 3.1. Hydrogen-suppressed graph of 2-methyl butane[a]



Footnote to Figure 3.1

[a]Structural Formula is $CH_3CH(CH_3)CH_2CH_3$

Table 3.1. Distance matrix of the hydrogen-suppressed graph of 2-methyl butane

	1	2	3	4	5
1	0	1	2	3	2
2	1	0	1	2	1
3	2	1	0	1	2
4	3	2	1	0	3
5	2	1	2	3	0

CHAPTER 4

METHODS II: STATISTICAL ANALYSIS

4.1 INTRODUCTION

The first step in the development of models for predicting chemical toxicity is to obtain or compile a suitable toxicological data set. A toxicological data set can be thought of as a 2 x 2 data matrix, in which the objects (rows) are chemicals, and the variables (columns) are physicochemical and/or *in vitro* properties, along with one or more *in vivo* endpoints of toxicity. Due to the presence of multiple variables, such data sets are said to be multivariate, and their analysis requires the use of multivariate statistical methods. The first step in a multivariate analysis usually consists of an investigation into the structure of the data matrix, to determine whether any of the variables correlate strongly with other variables, and whether any observations are similar to other observations with respect to their values along one or more variables. This step, which is sometimes called 'exploratory data analysis', enables redundancies to be removed from the data set before any models are derived. Several methods for performing exploratory data analysis are described in Section 4.2. Having removed redundancies from the toxicological data set, it is possible to derive models for predicting chemical toxicity from physicochemical and/or *in vitro* properties. In this thesis, such models are either classification models (CMs) or regression models, depending on whether the toxicological effect of interest is categorical or continuous, respectively. Statistical methods for developing classification models are described in Section 4.3, and methods for developing regression models are described in Section 4.4. Section 4.5 provides a summary of several resampling methods that are sometimes used in conjunction with classification and regression methods. Numerous textbooks explain in more detail the methods of univariate statistics (e.g. Caulcutt, 1991) and multivariate statistics (e.g. Tabachnick & Fidell, 1996). Other texts review the applications of statistical analysis in chemistry and pharmacology (e.g. Livingstone, 1995), and in toxicology (e.g. ECETOC, 1998b).

4.2 EXPLORATORY DATA ANALYSIS

4.2.1 Principal components analysis and factor analysis

Principal components analysis (PCA) and factor analysis (FA) are used to convert a data set having three or more variables into a data set having fewer variables, while

minimising the loss of explained variance. The new variables are called principal components (PCs) in the case of PCA, and factors in the case of FA, although in the following discussion, only the term factor is used, for simplicity. In PCA and FA, the factors are derived in such a way that: a) the factors are linear combinations of the original variables, the first factor being the one that explains the largest proportion of variance in the original data set, with subsequent factors explaining decreasing parts of the remaining variance; and b) all factors are orthogonal to (uncorrelated with) one another. If the objects are plotted as a cluster in multi-dimensional space, the first factor can be thought of as the longest straight line that crosses the cluster, and the second factor can be thought of as the second longest straight line traversing the cluster that is orthogonal to the first factor. The numbers in the transformed data set are called principal component scores (in the case of PCA) or factor scores (in the case of FA). The loadings on a given factor, which reflect the relative contributions made by the original variables to the factor, are simply the Pearson correlation coefficients between each variable and that factor.

In terms of matrix algebra, PCA and FA can be regarded as processes in which the original variable matrix (*N* objects [rows] x *P* variables [columns]) is transformed into a factor scores matrix (*N* objects [rows] against *Q* factors [columns]) and a factor loadings matrix (*P* variables [rows] against *Q* factors [columns]). The factor loadings matrix is the matrix of correlations between variables and factors, and can be used to interpret the meaning of each factor. The factor scores matrix (**F**) is generated from the original variables matrix (**Z**), according to Equation 4.1:

$$\mathbf{F} = \mathbf{Z} \cdot \mathbf{B} \qquad \text{(Equation 4.1)}$$

where **B** is a matrix of regression weights called the factor score coefficients matrix (variables x factors). In this equation, the factor scores in **F** and the variables in **Z** are autoscaled, i.e. standardised by subtraction of the mean and division by the standard deviation.

The condition of orthogonality places a restriction on the constitution of the factors, but it has the advantage that there is no multi-collinearity in the transformed data set, i.e. the different factors explain different parts of the variance and could therefore be used as

truly independent variables in subsequent analyses. For example, the use of principal components as predictor variables in multiple regression (principal components regression [PCR]) is sometimes favoured, because the addition (or removal) of a PC variable from the regression model does not alter the coefficients of the other variables. The disadvantage of PCR, however, is that the PC variables are less interpretable than the original variables, and the PC loadings based on a single sample might not reflect those in the corresponding population.

Factors may be difficult to interpret when they are associated with a large number of original variables. To overcome this difficulty, the factors can be rotated, in the hope that they become more closely aligned with a smaller number of original variables, and therefore more amenable to interpretation. Rotation does not lead to a reduction in the proportion of variance explained by the factors. A commonly used type of rotation is varimax rotation, which maximises the variance of the loadings within factors (high loadings become higher and low loadings become lower). This is accompanied by a redistribution of variance between factors, so that successive factors account for similar proportions of the total variance. Another type of rotation, quartimax rotation, does for variables what varimax rotation does for factors, i.e. it maximises the variance of the loadings within variables. Thus, varimax rotation operates on the columns of the loading matrix (variables x factors), whereas quartimax rotation operates on the rows. Finally, a third type of orthogonal rotation, called equamax, simultaneously operates on both the columns and rows of the loading matrix.

Assessing the importance of principal components or factors

The number of factors extracted from a given data set is either the number of objects (rows) or the number of variables (columns) in the data matrix, whichever is the smaller. Before factors are extracted from the data set, the original variables can be autoscaled (standardised by subtraction of the mean then division by the standard deviation). Each autoscaled original variable has a mean of zero and a variance of one, and therefore contributes one unit of variance to the overall variance of the autoscaled data set. The matrix of the autoscaled variables is called the correlation matrix (in distinction to the variance/covariance matrix of unscaled original variables). By performing PCA or FA on the correlation matrix, the importance of a factor can be judged from its eigenvalue (proportion of variance explained). According to one rule-of-

thumb, the Kaiser criterion, a factor is important (worth retaining) if its eigenvalue is greater than one; otherwise, the factor does not account for more variance than any one of the original variables.

Differences between principal components analysis and factor analysis

The main difference between PCA and FA is that in PCA, all of the variance in the original variables is used in the analysis, whereas in FA, only the variance that each variable has in common with the other variables is used, with unique variance being excluded from the analysis. The proportion of variance that a particular variable shares with all other variables is called the communality (a value between 0 and 1), so the unique variance of a variable is its total variance minus its communality. The communalities in FA are determined by an iterative procedure, typically by using the squared multiple correlation coefficient of a variable with all other variables as the starting values in the iteration.

4.2.2 Cluster analysis

Cluster analysis (CA) refers to a variety of methods for visualising and quantifying the similarity between different objects (cluster analysis of objects) or variables (cluster analysis of variables). The methods work by placing the objects (or variables) in multi-dimensional space and by grouping objects (or variables) according to the distance between them. A number of distance measures can be used, such as the Euclidean distance (the geometric distance in multi-dimensional space) for objects, and the Pearson correlation coefficient for variables. The results of cluster analysis are generally visualised in the form of a dendrogram.

One way in which clustering algorithms are distinguished is by the number of links they allow. The simplest types of clustering algorithm proceed by the formation of single links, i.e. only two objects (or variables) are joined together in any one step.

Clustering algorithms also differ in the way in which objects (or variables) are grouped. In non-hierarchical clustering, the links between objects (or variables) can be broken, whereas in hierarchical clustering, the links are never broken. An example of non-hierarchical clustering is K-means clustering. In this method, the user determines the

number (K) of clusters to be formed, and the algorithm assigns objects to the K clusters in such a way that the variability between clusters is maximised, and the variability within clusters is minimised. The K objects used to start the clustering process can either be randomly selected (by the program), or specifically chosen (by the user). Hierarchical clustering algorithms can either be agglomerative, in which case the individual objects (or variables) initially form their own clusters, before being grouped into larger ones, or they can be divisive, in which case the first cluster, containing all of the objects (or variables), splits into smaller clusters.

The choice of clustering method depends on the purpose of the investigation. If it is known *a priori* that there should be a specific number of final clusters, then K-means clustering would be the preferred method, whereas hierarchical clustering is more appropriate when the final number of clusters is unknown. Finally, a less commonly used type of clustering method is two-way joining, in which objects and variables are clustered simultaneously.

4.2.3 Correspondence analysis

Correspondence analysis is a technique for determining whether there is a relationship between two or more categorical variables in a frequency (contingency) table. If there are two categorical variables, the rows and columns in the two-way frequency table could be regarded as the predictor and response variables, respectively. Two-way tables are analysed by simple correspondence analysis, whereas multi-way frequency tables, composed of three or more variables, are analysed by multiple correspondence analysis. A comprehensive description of correspondence analysis and its applications is provided in Greenacre (1984, 1993).

By analogy with physical objects, the frequency table is considered to have a certain mass, distributed across the cells in proportion to their relative frequencies. The total row and column frequencies are therefore referred to as the row mass and column mass, respectively. The term ‘inertia’ is used to denote the moment of inertia, which is a measure of the distribution of mass in a physical object, defined as the integral (over distance) of the mass times the squared distance to the centroid. Bodies with a small moment of inertia have their mass distributed close the centroid, whereas bodies of the

same mass with a large moment of inertia have their mass distributed further away from the centroid. In correspondence analysis, the inertia of a table is defined by the following equation:

$$\text{Inertia} = \frac{\chi^2}{N} \qquad \text{(Equation 4.2)}$$

where χ^2 is the Pearson chi-square value; and N is the total sum of frequencies in the table (grand total). The chi-square statistic is defined as:

$$\chi^2 = \sum \frac{(O - E)^2}{E} \qquad \text{(Equation 4.3)}$$

where *O* is the observed frequency in a cell; and *E* is the expected frequency in a cell. If there is no relationship between the rows and columns, the expected frequency in each cell is equal to the respective column total times the row total, divided by the grand total. The chi-square value is a measure of the extent to which the actual frequencies depart from their expected values.

Correspondence analysis works by decomposing the total inertia in a frequency table into a number of dimensions. The dimensions are extracted in such a way that: a) the distances between the row or column coordinates are maximised; and b) successive dimensions are orthogonal to one another and account for decreasing proportions of the total chi-square. The relative importance of each dimension is denoted by its eigenvalue. The maximum number of dimensions extracted from a two-way table is equal to the number of columns minus 1, or the number of rows minus 1, depending on which is the smaller.

The first step in correspondence analysis is to standardise the contingency table. The frequency in each cell is divided by the grand total, so that the sum of frequencies (now called relative frequencies) across all cells is equal to one. The next step depends on whether the goal of the analysis is to: a) compare rows in terms of columns; b) compare columns in terms of rows; or c) examine the relationship between columns and rows. In the first case, the row profile matrix is used, which means that the relative frequencies

in each row are scaled so that their sum across all columns is equal to one. Each entry n_{ij} in the row profile matrix can be interpreted as the conditional probability that a case belongs to column *j*, given that it also belongs to row *i*. In the second case, the column profile matrix is analysed, in which case the relative frequencies in each column are scaled so that their sum across all rows equals one. In both cases, a two-dimensional plot of the first two factors can be used to display the similarity between rows or columns, which is represented by the proximity of the points corresponding to the row or column profiles. Because there is a geometric correspondence between the row and column profile plots, the two plots can be merged into a biplot, containing points for both row and column profiles. In the biplot representation, row and column points lying in the same direction from the origin are positively correlated, while those lying in opposite directions are negatively associated (Devillers & Karcher, 1990).

The quality of a point (row or column profile) is defined as the ratio of the squared distance of the point from the origin in the chosen number of dimensions, over the squared distance from the origin in the space defined by the maximum number of dimensions. The quality of a point is a measure of how well the inertia of that point is accounted for by the chosen number of dimensions. The proportion of the total inertia accounted for by the respective point (which is independent of the chosen number of dimensions) is given by the relative inertia.

The cosine² value ($\cos^2 \theta$) for a given point along a given dimension is a measure of the correlation between the point and the dimension. If $\cos^2 \theta$ is high, this means that the angle between the profile vector and the given dimension is low, i.e. the profile vector lies in the direction of the dimension and therefore correlates well with it. An alternative way of expressing this is to say that the dimension contributes highly to the inertia of the point. This is referred to as a relative contribution, because $\cos^2 \theta$ is independent of the mass of the point. If the mass of the point is taken into account by multiplying this with its squared distance (along the dimension) from the origin, this gives the absolute contribution of the point to the inertia of the axis. The meaning of a given dimension can be interpreted from the profile vectors that have high absolute contributions.

4.3 DEVELOPMENT OF CLASSIFICATION MODELS

In this project, classification models were derived for predicting categories of skin and eye irritation potential, such as EU classifications. This section describes three methods for deriving classification models: linear discriminant analysis, logistic regression and classification tree analysis.

4.3.1 Linear discriminant analysis

Linear discriminant analysis (LDA) is used to determine which variables discriminate between two or more pre-defined groups, and to derive classification models for predicting the group membership of new observations.

The simplest type of LDA, two-group LDA, can be thought of as a method that generates a linear boundary between two groups of objects. The nature of the boundary depends on how many predictor variables are used: one variable generates a point boundary (a cut-off value); two variables give rise to a straight line boundary; and three variables result in a planar boundary. In the case of four or more variables, the boundary will be a hyperplane.

To discriminate between two groups, a linear discriminant function (LDF) that passes through the centroids (geometric centres) of the two groups is generated:

$$LDF = a + b_1.x_1 + b_2.x_2 + ... + b_p.x_p \qquad \text{(Equation 4.4)}$$

where *a* is a constant, and b_1 to *bp* are the regression coefficients for *p* variables. The LDF can also be written in standardised form:

$$D = beta_1.z_1 + beta_2.z_2 + ... + beta_p.z_p \qquad \text{(Equation 4.5)}$$

where *D* is the standardised LDF, $beta_1$ to $beta_p$ are the standardised regression coefficients, and z_1 to z_p are the standardised variables. The variables that contribute most to the prediction of group membership are ones with the largest, standardised regression coefficients. The mean of *D* over all objects is zero, because the mean of

each variable, when standardised, is zero. Therefore, an object can be classified into one group if its D score is greater than zero, and into the other group if its D score is less than zero.

The linear boundary between the two groups can be thought of as a function that is orthogonal (perpendicular) to the LDF. Since there are an infinite number of orthogonal surfaces, the most obvious choice is the boundary that is equidistant from the two centroids; this is called 'LDA with equal prior probabilities'. Alternatively, the boundary can be shifted along the direction of the LDF towards one of the two centroids so that objects of unknown group membership are more likely to be assigned to one group rather than the other. This is called 'LDA with unequal prior probabilities', and is appropriate when it is known, *a priori*, that there are different proportions of objects in the different groups (in the population).

When LDA is used to distinguish between three or more groups, it is likely that more than a single LDF will be required. For example, in three-group LDA, one of the LDFs could be used to discriminate between group one and groups two and three combined, whereas the other LDF could be used to discriminate between groups two and three. In this case, the two linear boundaries would be non-parallel. In the less usual situation where a single LDF is sufficient to distinguish between three groups, the two boundaries would be parallel to one another (but would dissect the LDF at different points).

In principle, standardised LDFs (D scores) can be used to assign objects to three or more groups, although in practice, classification functions are generally used instead. There will be as many classification functions as there are groups, and each object will be assigned to the group for which it has the highest classification score. Classification scores are computed according the following equation:

$$S_i = c_i + w_{i1}.x_1 + w_{i2}.x_2 + ... + w_{ip}.x_p \qquad \text{(Equation 4.6)}$$

where the subscript *i* denotes the respective group; the subscripts *1, 2, ..., p* denote the *p* variables; c_i is a constant for the *i*'th group; w_{ij} is the weight for the j^{th} variable in the computation of the classification score for the i^{th} group; x_j is the observed value for the respective object for the j^{th} variable; and S_i is the resultant classification score.

To calculate the probability that a given object belongs to a given group, the Mahalanobis distance is used. This is the distance of the object from the group centroid in the multi-dimensional space defined by the predictor variables. The Mahalanobis distance is an appropriate measure of distance when the variables making up the multidimensional space are correlated; if the predictor variables are uncorrelated, it is the same as the Euclidean distance. The calculation of this 'posterior probability' is based not only on the Mahalanobis distance, but also on the prior probabilities.

Assumptions of linear discriminant analysis

The LDA method makes the following assumptions:

1) the predictor variables are interval or ratio variables (i.e. they are either measured on an interval scale, which enables a ranking of objects and a comparison of the sizes of differences between them, or they are measured on a ratio scale, which has the same properties as an interval scale, and the additional property that the scale contains an absolute zero).
2) the assumption of multivariate normality: the scores on the predictor variables have been independently and randomly sampled from a population having a multivariate normal distribution.
3) the assumption of equal variance/covariance matrices: the different groups have the same variance/covariance matrix. In graphical terms, this means that the objects in each group should form a multidimensional 'cloud' of the same shape, size and orientation.

If the assumption of multivariate normality is violated, this does not so much invalidate the classification model as reduce its predictive power, i.e. departures from normality do not necessarily lead to type I errors (in which the null hypothesis of no discrimination between groups is falsely rejected) but could lead to a reduced ability to classify objects when the model is applied to independent data. In contrast, heterogeneity of the variance/covariance matrices is likely to be more important, since objects are more likely to be classified into the group with greatest dispersion.

Canonical discriminant analysis

Canonical discriminant analysis is a variant of LDA in which the LDFs are derived in such a way that they are orthogonal (independent). The first function provides the largest contribution to the discrimination between groups, the second function provides second largest and non-overlapping discrimination, and so on. The maximum number of functions will be equal to the number of groups minus one, or the number of variables in the analysis, whichever is smaller.

Stepwise discriminant analysis

When numerous variables are available for analysis, stepwise discriminant analysis can be used to identify the best subset of variables for classifying objects, and to build a discriminant model based on these variables. In forward stepwise analysis, all variables are evaluated in the first step to determine which one provides the most significant and unique discrimination between groups. Once this variable has been included in the model, all remaining variables are evaluated to determine which one provides the next best discrimination. The procedure continues until the addition of a new variable does not significantly improve the discrimination between groups. Backward stepwise analysis starts by including all variables in the model and proceeds by eliminating, at each step, the variable that contributes least to the prediction of group membership. The stepwise procedure is directed by statistical criteria, such as the user-specified F-to-enter and F-to-remove values. The F value for a variable is a measure of the extent to which it makes a unique contribution to the prediction of group membership. During forward stepping, a variable will only be added to the model if its F value is greater than the F-to-enter value, and during backward stepping, a variable will only be removed from the model if its F value is less than the F-to-remove value. A variable that has entered a model during forward stepping will be removed if the subsequent entry of another variable results in the F value of the former falling below the F-to-remove value. Similarly, a variable that has been removed during backward stepping will be re-entered, if the removal of another variable results in the F value of the former being greater than the F-to-enter value.

It is important to bear in mind that stepwise discriminant analysis capitalises on chance effects, since the greater the number of variables that can be included or excluded, the greater the possibility that one of these variables may appear significant or insignificant

simply as a result of the particular object values for that variable. Therefore, the significance level associated with the F statistic at each step does not reflect the type I error rate.

4.3.2 Logistic regression

Logistic regression can be used to predict, on the basis of one or more predictor variables, the probability that a given object will belong to two or more groups. If there are two groups (e.g. toxic and non-toxic), binary logistic regression is used, whereas if there are three or more groups, a choice has to be made between nominal and ordinal logistic regression. Nominal logistic regression is used when there is no natural ordering to the groups (e.g. irritant, neurotoxic and embryotoxic), whereas ordinal logistic regression is used when there is a rank ordering (e.g. non-toxic, weakly toxic and strongly toxic).

Binary logistic regression

In the simplest type of logistic regression, binary logistic regression, the probability (p_1) of an object belonging to group 1, and the probability (p_2) of it belonging to group 2, is given by:

$$\ln (p_1/p_2) = b_0 + b_1.x_1 + b_2.x_2 + ... + b_n.x_n \qquad \text{(Equation 4.7)}$$

where (p_1/p_2) is called the odds ratio and ln (p_1/p_2) the logit transform of p_1; x_n is the nth predictor variable; and b_n is the coefficient of the nth predictor variable. In this equation, the logit transform is being used to relate the probabilities of group membership to a linear function of the predictor variables. The parameters of the logistic model (b_0 to b_n) are derived by the method of maximum likelihood, i.e. the parameters are chosen in such a way that, given that the 'true model' is assumed to be a logistic function, the chosen (population) parameters are the ones that maximise the probability (i.e. likelihood) that the actual response values would have been generated by taking a random sample from the population. Provided that all observations are independent of one another (i.e. there is no serial correlation between them), the likelihood of observing the sample of response values is the geometric sum of the probabilities of observing the individual response values. Instead of expressing the likelihood of the model in terms of

a geometric sum (product of probabilities), it is customary to take the natural logarithm of the likelihood (the log-likelihood), which then becomes a function of the arithmetic sum of individual probabilities. Thus, algorithms for logistic regression work by maximising the log likelihood of a model, which is equivalent to minimising its negative log-likelihood.

The probabilities of group membership can also be expressed directly in terms of the predictor variables:

$$p_1 = \exp(b_0 + b_1.x_1 + ... + b_n.x_n) / \{1 + \exp(b_0 + b_1.x_1 + ... + b_n.x_n)\} \quad \text{(Equation 4.8)}$$

$$p_2 = 1 /\{1 + \exp(b_0 + b_1.x_1 + ... + b_n.x_n)\} \quad \text{(Equation 4.9)}$$

The regression coefficients b_1 to b_n can be used in combination with the constant b_0 to define a model for classifying objects into one of the two groups. The basis of such a model is that an object with an equal probability of belonging to the two groups has $p_1=p_2$, which means that $\ln (p_1/p_2) = 0$, and (from Equation 4.7) that:

$$b_0 + b_1.x_1 + ... + b_n.x_n = 0 \quad \text{(Equation 4.10)}$$

Therefore, the surface defined by Equation 4.7 in the multi-dimensional space defined by x_1 to x_n is the linear boundary between objects in group 1 and objects in group 2. An object will be assigned to group 1 or to group 2 according to the following rules:

Classify into group 1 if: $b_0 + b_1.x_1 + ... + b_n.x_n > 0$ (Equation 4.11)

Classify into group 2 if: $b_0 + b_1.x_1 + ... + b_n.x_n < 0$ (Equation 4.12)

Assumptions of binary logistic regression

Logistic regression makes the assumption that a link function (in this case, the logit transform) can be used to relate the probabilities of group membership to a linear function of the predictor variables. It is also assumed that the observations are independent, but (unlike in the case of linear regression) it is not assumed that they are normally distributed.

Ordinal and nominal logistic regression

Ordinal logistic regression is similar to binary logistic regression except that the link function is:

$$\ln (p_{ij} / (1-p_{ij})) \qquad \text{(Equation 4.13)}$$

where p_j is the cumulative probability that object *i* is in group *j* or lower. A single logistic function is derived with *k* constants (where *k* is the total number of groups minus one). For example, if there are three groups, there will be two constants. The first constant can be used to define the cut-off value (along the linear combination of predictor variables) between group 1 and groups 2 and 3 combined, whereas the second constant can be used to define the cut-off value between groups 1 and 2 combined and group 3.

In the case of nominal logistic regression, the number of logistic functions is one less than the number of groups. One of the groups is taken to be a reference group (group zero), so that the first logistic function can be used to predict the probability that an object will belong to group 1 rather than group zero, and the second logistic function can be used to predict the probability that an object will belong to group two rather than to group zero.

4.3.3 Classification tree analysis

A classification tree (CT) is used to predict the group membership of objects on the basis of one or more predictor variables. The tree consists of a set of decision rules, applied in a sequential manner, until each object has been assigned to a specific group. The first decision rule, applied at the 'root node' of the tree to the values of all objects along one or more predictor variables, has two possible outcomes: objects are either sent to a terminal node (leaf), which assigns a class, or to an intermediate node, which applies another decision rule. Ultimately, all objects are sent to a terminal node and assigned a class. In the simplest type of classification tree, the splits are binary (each parent node is attached to two daughter nodes) and the decision rules are univariate (based on a single variable). Classification trees can be based on continuous or discrete predictor variables, or on a mixture of both (when univariate splits are used), and the trees are generally constructed by recursive partitioning (i.e. a given predictor variable

can be used in more than one decision rule). One of the advantages of CTs is that they are non-parametric classifiers, i.e. no assumptions are made about the distributions of the variables. Thus, CT analysis can be used as a method of last resort, when the assumptions of LDA and BLR have not been satisfied.

Classification tree algorithms

Two commonly used algorithms for constructing classification trees are CART (Classification and Regression Trees), developed by Breiman *et al*. (1984), and QUEST (Quick, Unbiased, Efficient Statistical Trees), developed by Loh & Shih (1997). CART finds optimal univariate splits by carrying out an exhaustive search of all possible splits, whereas QUEST finds optimal univariate or multivariate splits by applying a modified form of discriminant analysis. It has been claimed (Loh & Shih, 1997) that the CART algorithm is biased toward selecting predictor variables having more levels, whereas QUEST lacks this bias, and is therefore more appropriate when some predictor variables have few levels and other predictor variables have many levels.

Criteria for predictive accuracy

The optimal classification tree is one that minimises costs. If the prior probabilities of objects belonging to the different classes are set proportional to the class sizes (i.e. proportion of objects belonging to each class in the sample is considered to reflect the corresponding proportion in the population), and if misclassification costs are set equal for every class, 'minimising costs' is equivalent to minimising the overall proportion of misclassified objects. If, however, the prior probabilities are set equal or set (differently) according to previous knowledge, or if different misclassification costs are used, then minimising costs does not correspond exactly with minimising the misclassification rate. Unequal misclassification costs are used when one kind of misclassification is considered 'worse' than another; for example, incorrectly predicting a toxic chemical to be non-toxic might be considered worse than incorrectly predicting a non-toxic chemical to be toxic. In such a case, misclassifications of the former type could be penalised more than misclassifications of the latter, when assessing the accuracy of classification.

Selecting optimal splits

Decision tree classifiers differ in the way they select splits for partitioning the training set into subsets.

If QUEST-style discriminant-based univariate splitting is used, the algorithm first selects the best terminal node to split and determines which predictor to use, and then determines a cut-off value along that predictor. For each terminal node, p-levels are computed for significance tests of the relationship between class membership and the levels of each predictor variable. For categorical predictors, the chi-square test of independence is used, whereas for ordered predictors, analysis of variance (ANOVA) is used. The predictor variable producing the smallest p-level is chosen to split the corresponding node. Further details are given in Loh & Shih (1997). To determine the cut-off value of the split, a two-means clustering algorithm (Hartigan & Wong, 1978) is used for ordered predictors. This determines the means of two 'superclasses' for the node, and chooses the split closest to a superclass mean. For categorical predictors, dummy-coded variables representing the levels of the categorical predictor are transformed into a set of non-redundant ordered predictors. The procedures for ordered predictors are then applied and the obtained split is 'mapped back' onto the original levels of the categorical variable.

If QUEST-style discriminant-based multivariate (linear combination) splits are used, the predictors must be on at least interval scales. The method works by treating the continuous predictors from which linear combinations are formed in a manner similar to the way categorical predictors are treated in discriminant-based univariate splitting. The continuous predictors are transformed into a new set of non-redundant (continuous) predictors, 'superclasses' are found, and the split closest to a superclass mean is obtained. Finally, the results are 'mapped back' onto the original continuous predictors to produce a univariate split on a linear combination of predictor variables.

If CART-style univariate splitting is used (on categorical or ordered predictor variables), all possible splits for each predictor variable at each node are examined to find the split producing the largest improvement in goodness-of-fit. The more homogeneous (less heterogeneous) in terms of class distribution the daughter nodes

compared to the parent node, the better the improvement in goodness-of-fit, i.e. the greater the reduction in node heterogeneity.

A commonly used measure of homogeneity is deviance. A perfectly homogenous node, in which all objects are of the same class, has a deviance of zero, whereas nodes of increasing heterogeneity have increasingly positive deviances. The deviance of a node *i* is given by:

$$D_i = -2 \sum_k n_{ik} \log p_{ik} \qquad \text{(Equation 4.14)}$$

where n_{ik} is the number of observations in terminal node *i* of class *k*; and p_{ik} is the probability of an observation at node *i* belonging to class *k* (the relative frequency of class *k* at node *i*). Another measure of homogeneity, the Gini index, is defined by the following equation:

$$Gini_i = 1 - \sum_k (p_{ik})^2 \qquad \text{(Equation 4.15)}$$

where $Gini_i$ = the Gini index of node *i*; and p_{ik} is the probability of an observation at node *i* belonging to class *k*. This equals zero for a node containing objects all of the same class, and becomes increasingly positive for nodes of increasing impurity.

Determining when to stop splitting

If a classification tree is grown until all terminal nodes are pure, the resulting tree is likely to overfit the data, and will therefore have a lower accuracy of classification when applied to new objects. The user can therefore apply a stopping rule so that splitting stops at nodes that are either pure, or (in the case of nodes containing more than one class) have no more than a specified number of objects. Alternatively, splitting can be stopped at nodes that are either pure or contain no more objects than a specified fraction of the objects. Other strategies include specifying the size of the tree to be grown (i.e. the number of terminal nodes), or defining the minimum heterogeneity that a node must have in order to be split.

Selecting the 'best' tree

Even when the user applies a stopping rule, the resulting tree may not be the 'best' tree, in the sense of maximising accuracy of classification while at the same time minimising complexity. Thus, procedures have been devised for pruning trees, i.e. for successively 'snipping off' the least important splits until the best tree is produced. In minimal cost-complexity pruning, a nested sequence of optimally pruned subtrees is generated as the tree of maximum size (determined by the stopping criterion) is pruned to the root node. The sequence is optimally pruned since, for every size of tree in the sequence, there is no other tree of the same size with lower costs. The learning sample (resubstitution) cost decreases as the size of the tree increases. However, the cost generally decreases slowly as the first terminal nodes are removed, until a point is reached when the cost rises rapidly upon removal of additional nodes. This turning point can be used to define the best-sized tree. Alternatively, if cross-validation is performed at each step of the pruning process (minimal cost-complexity cross-validation pruning), the cross-validated costs can be used to identify the best tree in the sequence. Generally, the cross-validated (CV) cost falls slowly to a minimum value as terminal nodes are removed, and then rises rapidly as the last few nodes are removed. Thus, the best tree can be defined as the tree closest to the minimum, i.e. the tree with the minimum CV cost. Alternatively, Breiman *et al.* (1984) suggested that the best-sized tree can be identified as the smallest tree whose CV cost does not exceed the cost of the minimum CV cost tree plus one standard error of this tree's CV cost.

4.3.4 Assessing the goodness-of-fit of classification models

The goodness-of-fit of a two-group CMs can be assessed in terms of its Cooper statistics (Cooper, 1979), which can be calculated from a 2 x 2 contingency table (Table 4.1), using the definitions given in Table 4.2. A Minitab macro for calculating the Cooper statistics is given in Appendix A1.

The statistics sensitivity, specificity and concordance are generally of most interest when assessing the performance of a CM, since they provide measures of its ability to detect known toxic chemicals (sensitivity), non-toxic chemicals (specificity) and all chemicals (accuracy or concordance). The false positive and false negative rates can be calculated from the specificity and sensitivity, as shown in Table 4.2.

The other two statistics, the positive and negative predictivities, are of more interest when focussing on the effects of individual chemicals. These statistics can be thought of as conditional probabilities: if a chemical is predicted to be toxic, the positive predictivity gives the probability that it really is toxic; similarly, if a chemical is predicted to be non-toxic, the negative predictivity gives the probability that it really is non-toxic.

Strictly, the Cooper statistics should only be used for two-group CMs. If there are three or more levels of the categorical response, other statistics should be used. An example is the kappa (κ) statistic, a chance-corrected accuracy that takes a value of zero when there is no agreement, and a value of one when there is perfect agreement, and is defined as follows:

$$\kappa = (O - E) / (1-E) \qquad \text{(Equation 4.16)}$$

where O is the observed accuracy; and E is the accuracy expected by chance.

4.4 DEVELOPMENT OF REGRESSION MODELS

Linear regression models probably account for most of the SARs that have been published. This section describes three types of linear regression analysis: simple and multiple linear regression, which have been the traditional methods of choice, and canonical correlation and partial least squares (PLS) regression, which have gained popularity more recently.

4.4.1 Simple and multiple linear regression

In simple (ordinary) linear regression, the response variable is treated as a linear function of a single predictor variable:

$$\hat{y} = mx + c \qquad \text{(Equation 4.17)}$$

where $\hat{y}$ is the predicted value of the response variable; x is the value of predictor variable; m is the slope of straight line; and c is the intercept of straight line on y axis. This equation, which is used to predict raw scores of the response variable, can be rewritten in standardised form, in which the predictor and response variables have means of zero and standard deviations of one:

$$\hat{z}_y = r.z_x \qquad \text{(Equation 4.18)}$$

where $\hat{z}_y$ is the predicted value of the standardised response z_y; r is the Pearson correlation coefficient between x and y; and z_x is the standardised form the predictor variable x.

In multiple linear regression (MLR), the response variable (y) is treated as a linear function of two or more predictor variables (x_1-x_n). The linear models derived by MLR are the equations of a plane or a hyperplane, depending on whether there are two or more predictor variables, respectively:

$$y = m_1x_1 + m_2x_2 + ...+ m_nx_n + c \qquad \text{(Equation 4.19)}$$

When this equation is written in its standardised form, the regression coefficients (m_1-m_n) are called beta weights or partial regression coefficients. The advantage of expressing the equation in standardised form is that the absolute values of the beta coefficients give the rank order of importance of the predictor variables. The relative importance of any two predictor variables can be obtained by taking the ratio of the squares of their respective beta weights. In contrast, the regression coefficients in the raw score form of the equation do not indicate the relative contributions of different predictor variables to the response variable.

Assumptions of linear regression analysis

Apart from the obvious assumption of a linear relationship existing between the response variable and the predictor variable(s), linear regression analysis is based on the following assumptions:

1) the predictor variables are on an interval or ratio scale
2) there is no error in the predictor variables
3) the residuals (differences between predicted and observed response values) are normally distributed about each predicted value of the response variable (the assumption of normality)
4) the variance of the residuals is the same at different predicted values of the response variable (the assumption of homoscedasticity or absence of heteroscedasticity)
5) the residuals at different values of the response variable should not depend on the order in which observations were made (the assumption of independence or absence of autocorrelation).

The assumptions of normality and homoscedasticity can be checked by various kinds of residual analysis. For example, a histogram of residuals should indicate whether the residuals are approximately normally distributed, and a plot of residuals against predicted values should indicate whether heteroscedasticity occurs. The assumption of independence may or may not need to be checked depending on how the data are collated. At one extreme, if a series of observations are performed on a single object or individual (e.g. a human volunteer), there is a possibility that later observations are affected by earlier ones. At the other extreme, if measurements relating to a series of chemicals are taken from different literature sources, it would be reasonable to assume that the data were independent. The degree of autocorrelation can be quantified by using the Durbin-Watson statistic, which takes on a value of 2 (in a range of 0 to 4) in the ideal case of no autocorrelation.

In practice, it is not always possible to determine whether an assumption has been satisfied (for example, if there are fewer than 20 observations, normality tests may fail to detect non-normality even when it exists). In other cases, violations may be detected, but not considered to be of any practical importance (for example, if there are a large number of observations [100 or more], normality tests may be oversensitive and reject the assumption of normality even though a frequency histogram looks normal). Moreover, when there is a large number of observations, statistical tests such as the t-test and F-test (which are used to assess the significance of linear regression models) are considered to be robust to departures from normality (Lindman, 1974). In practice, then, one or more assumptions may appear to be violated, so the question to be addressed is

not so much whether the assumptions have been obeyed, but to what extent they have been violated, and whether, as a consequence, the model is invalid (i.e. the mathematical structure of the model is 'incorrect'), or simply of reduced predictive quality (i.e. the model structure is 'correct', but the parameters have not been optimised).

Assessing the goodness-of-fit of linear regression models

The variance in the response variable is expressed as the total sum of squares (TSS), which can be divided into the variance attributed to the model (the explained sum of squares [ESS]), and the variance attributed to the prediction error (the residual sum of squares [RSS]). The TSS is defined as the sum of the squared differences between the observed values of the dependent variable (y_i) and their mean ($\overline{y}$); the ESS is the sum of the squared differences between the predicted values of the dependent variable ($\hat{y}_i$) and the mean of the observed values ($\overline{y}$); and the RSS is the sum of the squared differences between the observed values (y_i) and the predicted values ($\hat{y}_i$). These statistics are summarised in Table 4.3, along with the explained mean square (EMS) and the residual mean square (RMS). Linear regression models are usually derived by the method of least squares, which optimises the goodness-of-fit by maximising the ESS and minimising the RSS.

The strength of linear correlation between the descriptor variable(s) and the response variable is expressed by the Pearson correlation coefficient (r) for simple regressions, and by the coefficient of determination (R^2) for multiple regressions:

$$r = \sqrt{\frac{ESS}{TSS}} \qquad \text{(Equation 4.20)}$$

$$R^2 = ESS/TSS = 1- (RSS/TSS) \qquad \text{(Equation 4.21)}$$

Because the R^2 value is never negative, all chance fluctuations of the sample R around the population R are converted into positive fluctuations of the sample R^2 above the population R^2. In other words, R^2 tends to be overestimated, and the degree of overestimation increases with decreasing sample size and increasing number of

variables. To correct for this chance-mediated inflation, the "adjusted R^2 value" can be used:

$$R^2(adj) = 1- [(1-R^2) (n-1/n-p-1)] \quad \text{(Equation 4.22)}$$

where $R^2(adj)$ is the adjusted R^2; n is the number of observations; and p is the number of predictor variables in the equation.

The prediction error is expressed by the standard error of the estimate (s):

$$s = \sqrt{RMS} = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n - p - 1}} \quad \text{(Equation 4.23)}$$

where n is the number of observations; and p is the number of predictor variables in the equation. The standard error of the estimate is not a standard error in the conventional sense of the term (standard deviation of a sample divided by the square root of the number of observations), but is a measure of the standard deviation about the regression line (or surface). The standard error of the estimate can be used to define a prediction band about the regression line (or surface); for example, a 95% prediction interval is the region in which 95% of the observations in the population are expected to lie. In contrast, the 95% confidence interval is the region around the regression line (surface) where the 'true' regression line is expected to lie with a 95% probability. The confidence interval depends not only on the prediction error (i.e. the standard error of the estimate), but also on the error in estimating the intercept of the true regression line along the response variable, and the error in estimating the slope of the true regression line. The latter source of error becomes more pronounced as the predictor values deviate more from their average value, which results in the confidence band being bowed about the sample regression line.

Assessing the statistical significance of linear regression models

The statistical significance of a regression model can be assessed by means of the Fisher (F) statistic, which is defined as the ratio of the explained mean square (EMS) to the residual mean square (RMS):

$$F = \frac{EMS}{RMS} = \frac{R^2(n-p-1)}{p(1-R^2)} \qquad \text{(Equation 4.24)}$$

where EMS = (ESS / p) and RMS = (RSS / n-p-1). The greater the F value of an equation, the greater its statistical significance. An equation is considered to be statistically significant if the observed F value is greater than the tabulated value for the chosen level of significance (typically, the 95% level). The tabulated value depends not only on the significance level, but also on the two degrees of freedom for F, υ_1 and υ_2, where υ_1 = p, and υ_2 = n-p-1. If an equation is reported to be significant at the 95% level, this means that there is a 5% chance (or less) that the equation is not significant, i.e. does not explain any more variance in the response variable than is explained by the random scatter of the dependent variable about its mean value.

The statistical significance of an individual variable in a regression can be assessed by means of the t statistic, which is defined as the ratio of the variable's regression coefficient to the standard deviation of the coefficient:

$$t = b/s_b = \frac{R\sqrt{n-2}}{\sqrt{1-R^2}} \qquad \text{(Equation 4.25)}$$

where b is the regression coefficient of the variable; and s_b is the standard deviation of the regression coefficient, given by:

$$s_b = \frac{s}{\sqrt{\sum (x_i - \bar{x})^2}} \qquad \text{(Equation 4.26)}$$

The greater the t value of a predictor variable, the greater its statistical significance. The t statistic is determined in a similar way to the F statistic, except that t tables have only one degree of freedom, corresponding to the RMS (i.e. υ = n-p-1). If a variable is said to be statistically significant at the 95% level, this means that there is a 5% (or smaller) chance that the regression coefficient of the variable is not significantly different from zero.

Strictly speaking, the F and t statistics are only meaningful if the assumptions of normality, independence and homoscedasticity are obeyed, although departures become less important as the number of degrees of freedom increases (Lindman, 1974).

Even if an equation and its predictor variables are judged to be significant by the t and F statistics, there is still a possibility that the equation is a 'chance relationship'. For example, if there are fives times as many objects as variables, there is a 20% (1 in 5) chance that the relationship is based on chance alone. Clearly, to avoid the derivation of chance relationships, the ratio of objects to variables should be kept as high as possible. Topliss & Costello (1972) recommended a minimum ratio of five objects for every variable.

Stepwise multiple regression

If there is a choice of predictor variables for inclusion in a multiple regression equation, several techniques can be used to select the ones that maximise the goodness-of-fit. An example is stepwise regression, which can proceed by forward stepping or backward stepping. In forward stepping, the initial model contains only one predictor variable, all other variables being excluded. In the first step, one of the excluded variables is added, this being the variable that improves the fit of the model (i.e. increases the r^2 value) more than the other excluded variables. This process is repeated in subsequent steps until a fixed number of variables have been included, or until a pre-defined fit has been achieved. In backward stepping, the initial model contains all possible predictor variables, and each step results in the loss of one variable, this being the variable that minimises the reduction in the fit of the model. Again, the procedure is continued until a fixed number of variables are present in the final equation or until a desired fit has been achieved. If a particular variable is considered to be an important predictor (on the basis of prior knowledge), it can be 'forced' into the final equation, irrespective of its contribution to the explained variance in the data set under investigation.

4.4.2 Canonical correlation

Canonical correlation is a method for relating a set of continuous predictor variables to a set of continuous response variables. The method works by correlating a linear combination (weighted sum) of the predictor variables with a linear combination of the

response variables. Each linear combination is called a canonical variate, and the two canonical variates that are correlated form a pair of canonical variates. The weights in the two canonical variates forming a pair are chosen in such a way that the canonical correlation is maximised. The squared correlation coefficient between the two canonical variates is called the root or eigenvalue (λ), and can be interpreted as the proportion of overlapping variance between the two canonical variates. There will be as many pairs of canonical variates extracted as there are predictor variables or response values, depending on which is the smaller. Successively extracted pairs of canonical variates are orthogonal, and have decreasing eigenvalues.

Statistical significance of canonical variates

To assess the statistical significance of canonical correlations, the significance of all canonical pairs is assessed first. The significance of all roots except the first one is then assessed, and so on, until the significance of the final root is assessed. Thus, if the combination of all roots is found to be significant, but the set of roots remaining after removal of the first one is insignificant, it would be concluded that only the first root is significant. In general, the statistical test used to determine whether a set of roots is significant is the chi square test:

$$\chi^2 = - [n-1-(p+q+1/2)] \ln \Lambda \qquad \text{(Equation 4.27)}$$

where n is the number of objects; p is the number of predictor variables; q is the number of response variables; and Λ is the product ($\prod$) of differences between eigenvalues (λ) and one, taken across all canonical correlations:

$$\Lambda = \prod (1-\lambda) \qquad \text{(Equation 4.28)}$$

Interpretation of canonical variates

The meaning of a given canonical variate can be interpreted from its canonical weights (the coefficients in the linear combination of variables that defines the canonical variate). The larger the absolute value of a canonical weight, the greater is the respective variable's unique positive or negative contribution to the canonical variate. Alternatively, a canonical variate can be interpreted from its canonical factor loadings (the correlation coefficients between the canonical variate and the variables from which

it is derived). The proportion of variance extracted from a set of variables by a canonical variate is the sum of squared canonical factor loadings divided by the number of variables in the set. If the proportion of variance extracted by a canonical variate is multiplied by the squared correlation coefficient for the pair (the eigenvalue), this gives the redundancy in the canonical variate. This is the proportion of variance extracted by the canonical variate from the variables in the other set of variables.

To obtain reliable estimates of the canonical factor loadings for the most significant canonical pair only, Stevens (1986) recommended that the ratio of objects to variables should be at least 20. To obtain reliable estimates for two canonical roots, Barcikowski & Stevens (1975) recommended a ratio of objects to variables of at least 40.

Assumptions of canonical correlation

The significance test applied to the canonical correlations is based on the assumption that the variables in the population (from which the sample was drawn) have a multivariate normal distribution. Multivariate normality is the assumption that all variables and all linear combinations of variables are normally distributed. It is also assumed that the canonical variate scores show linearity and homoscedasticity.

4.4.3 Partial least squares regression

PLS regression is a linear regression method for predicting one or more response variables from two or more predictor variables. The method works by forming two sets of components (latent variables or factors), one set being derived from the predictor variables, and the other set from the response variables. The components are derived in such a way that: a) the components in each set are orthogonal to one another; and b) the response components are maximally correlated with the predictor components. The purported advantages of PLS regression over traditional multiple linear regression are that it can be used to derive regression models when there are fewer observations than predictor variables, and when predictor variables are highly correlated. A description of the PLS algorithm is provided in Geladi & Kowalski (1986).

In terms of matrix algebra, PLS regression derives a linear model:

$$\boldsymbol{Y=XB+E} \qquad \text{(Equation 4.29)}$$

where ***Y*** is the response variable matrix (*n* objects by *q* variables); ***X*** is the predictor variable matrix (*n* objects by *p* variables); ***B*** is the regression coefficient matrix (*p* by *q* variables); and ***E*** is a matrix of residuals having the same dimensions as ***Y***. The variables in ***X*** and ***Y*** are usually standardised.

The *t* factors extracted from ***X*** can be related to the predictor variables by Equation 4.30, and the *u* factors extracted from ***Y*** can be related to the response variables by Equation 4.31:

$$\boldsymbol{T=XW} \qquad \text{(Equation 4.30)}$$

$$\boldsymbol{U=YC} \qquad \text{(Equation 4.31)}$$

where ***T*** is the (*n* by *t*) factor score matrix derived from the predictor variables; ***W*** is a (*p* by *t*) matrix of *T* factor loadings; ***U*** is the (*n* by *u*) factor score matrix derived from the response variables; and ***C*** is a (*q* by *u*) matrix of *U* factor loadings. The weights in ***W*** are computed so that they maximise the correlation between the predictor variables and the *U* scores, and the weights in ***C*** are computed to maximise the correlation between the response variables and the *T* scores. This is achieved by an iterative procedure (e.g. NIPALS [non-linear iterative partial least squares]), in which factors (*t* and *u*) are calculated one pair at a time.

4.5 RESAMPLING METHODS

In resampling methods, the data points of a toxicological data set are sampled (at random, in many cases), with a view to assessing, for example, the statistical significance or predictivity of a model. Resampling methods are iterative procedures. During each iteration, objects are selected from the data set, and a statistic of interest is calculated on the basis of data for the chosen objects. The statistic calculated during each iteration is generally stored, so that at the end of the iterative procedure, a sampling distribution of the statistic can be generated, and inferences can be made about the population for which the data set represents a sample. If the resampling process is

carried out *with replacement*, this means that *during each iteration*, the selected objects are not removed from the data set, and can therefore be selected more than once. Conversely, if resampling is performed *without replacement*, this means that once an object has been selected, it is removed from the data set, and cannot be selected again during the *same* iteration.

4.5.1 Cross-validation

The goodness-of-fit of a statistical model provides a measure of how well the model fits the data in its training set. The predictivity of the model (i.e. its ability to make predictions for an independent [external] set of data) will almost certainly be worse than the goodness-of-fit. Thus, statistical models should ideally be assessed by using independent data (external validation). However, a shortage of reliable data often means that all available data are used for model development, leaving none for external validation. Therefore, various cross-validation techniques have been developed to enable a more realistic (but probably still optimistic) performance of the predictive capacity of statistical models to be assessed by resampling a single training set. In the simplest type of cross-validation, called the leave-one-out (LOO) procedure, each object is omitted once in turn, and predicted from a model derived from the remainder of the data set. The predicted values of the dependent variable are then compared with the measured values so that a variety of cross-validated statistics can be derived, such as the predicted residual sum of squares (PRESS) and the cross-validated R^2 (Q^2):

$$\text{PRESS} = \sum (\hat{y} - y_i)^2 \qquad \text{(Equation 4.32)}$$

$$Q^2 = 1 - (\text{PRESS} / \text{TSS}) \qquad \text{(Equation 4.33)}$$

Instead of leaving out every object in turn, it is possible to omit a larger proportion of the objects, either in a fixed or random manner.

Cross-validation provides more realistic measures not only of model performance, but also of model robustness (stability); if a model is robust, its regression coefficients are insensitive to the omission of data points. For the purposes of this project, Minitab

macros were written to calculate the cross-validated R^2 for simple and multiple linear regressions (Appendix A2).

4.5.2 Cluster significance analysis

If chemicals are classified into two groups (active/inactive or toxic/non-toxic) and are plotted on a box plot (one variable), or on a scatter plot (two or more variables), it sometimes appears that the chemicals in one group form an 'embedded cluster' surrounded by the chemicals in the other group, the 'diffuse cluster'. Typically, the active (or toxic) chemicals form the embedded cluster, and the inactive (or non-toxic) chemicals form the diffuse cluster, indicating that for each of the chosen variables, there is an optimum range of values for the biological response to occur.

The statistical significance of an embedded cluster can be assessed by cluster significance analysis (CSA). The first algorithms for CSA, proposed by McFarland & Gans (1986a, 1986b), were based on an index of cluster tightness called the 'mean squared distance' (MSD), defined as the sum of the squared (Euclidean) distances between all possible pairs of data points, divided by the number of such pairs. Since the MSD of a cluster is proportional to its variance, Rose & Wood (1998) proposed an algorithm for CSA that uses variance as the measure of cluster tightness.

Algorithm for Cluster Significance Analysis

The first step in CSA is to calculate the tightness of the observed cluster, using its variance as the index of cluster tightness. The variance of a single variable is given by:

$$\text{variance} = \frac{\sum (x - \bar{x})^2}{n - 1} \qquad \text{(Equation 4.34)}$$

where x is the variable under consideration, $\bar{x}$ is the mean value of x, and n is the number of data points in the cluster. In the case of two or three variables, the total variance is calculated as the sum of the variances of the individual variables, which is valid as long as the variables are measured on the same scale. This is achieved by standardisation (subtraction of the mean, and division of the difference by the standard deviation).

The second step is the generation of the sampling distribution of variances (or weighted variances) for the data set. To avoid the need to identify all possible combinations of chemicals, random sampling is carried out irrespective of whether the data set is small or large. The number of active chemicals in the data set is counted, and this number is randomly sampled, without replacement, from the total set of chemicals. The variance of each random combination (subset of chemicals) is then calculated. If the variance of the random subset is less than or equal to the variance of the embedded cluster, a counter (which starts the sampling procedure with a value of zero) is increased by one. Thus, when the specified number of random subsets has been sampled, the value of the counter corresponds to the number of times that a random subset had a variance less than or equal to the variance of the embedded cluster. It is important to note that while the chemicals in each subset are sampled without replacement (to ensure that a given chemical can only be selected once), the subsets themselves are selected with replacement, so a given subset of chemicals could be selected more than once.

In the third step, the probability (p-value) that the observed embedded cluster could have arisen by chance is given by:

$$\text{p-value} = n / N \qquad \text{(Equation 4.35)}$$

where n is the number of randomly generated subsets that are at least as tight as the observed embedded cluster; and N is the total number of randomly selected clusters. If the p-value ≤ 0.05, it is concluded (at the 95% significance level) that the observed cluster has not arisen by chance.

The fourth step is the calculation of the confidence interval (CI) for the p-value, which is simply the CI for a proportion, given by:

$$\text{CI} = p - k\sqrt{\frac{p(1-p)}{n}} - 0.5/n \text{ to } p + k\sqrt{\frac{p(1-p)}{n}} + 0.5/n \qquad \text{(Equation 4.36)}$$

where k is the critical value for a two-tailed significance test based on the normal distribution. A 95% CI is calculated by using k=1.96, and a 99% CI by using k=2.58.

For the purposes of this project, the CSA algorithm was coded in the form of three Minitab macros, appropriate for carrying out CSA in one, two and three dimensions, respectively (Appendix A3).

Variable selection in cluster significance analysis

In extensions of their original algorithms, McFarland & Gans (1990; 1994), introduced stepwise procedures for automatically selecting the most significant combinations of q variables (where $q \geq 2$) from a set of n variables (where $n > q$). For example, to evaluate the significance of seven physicochemical descriptors for predicting the sweetness of sulphamate compounds (McFarland & Gans, 1990), the p-value for the seven-variable combination was determined first, and then variables were deleted one at a time. If the removal of a single variable from a given q-variable combination reduced the p-value, the remaining combination of (q-1) variables would be entered into the subsequent step, whereas if the p-value increased or stayed the same, the effect of removing a different variable would be assessed instead.

Stepwise procedures such as this enable the relative significance of different q-variable combinations to be determined (where q is a fixed number), but they do not indicate whether the addition or removal of a variable results in a significant change in the significance of clustering. To overcome this shortcoming, Rose & Wood (1998) proposed a forward stepwise algorithm in which the significance of adding a new variable is assessed in terms of its 'conditional p-value'. This algorithm, called 'CSA with conditional probabilities', is an extension of the authors' simplified algorithm for CSA. It works by randomly generating a large number of subsets (e.g. 100,000) of the total set of compounds, and by storing the variance of each subset (along each variable) in a matrix. This matrix, which summarises the sampling distribution of the variances, has as many rows as there are randomly selected subsets of compounds, and as many columns as there are variables. The first variable to be selected is simply the variable which has the smallest p-value. To select the second variable, the variances in the column vector corresponding to the first variable are arranged in order of increasing size, with the remaining column vectors (corresponding to the remaining variables) being sorted in parallel. This results in row sN of the first column vector (where s is the significance level and N is the number of random subsets) having a variance

approximately equal to the observed variance. Then, a fixed number of rows on either side of this row are taken as a subset from the matrix corresponding to the sorted sampling distribution. The second variable to be selected is simply the variable in this sub-matrix that has the lowest p-value. This process is continued until the addition of a variable results in a conditional p-value that exceeds the chosen significance level (typically, 0.05). At this point, the addition of new variables is no longer considered to contribute significantly to the tightness of the embedded cluster. In effect, stepwise CSA with conditional probabilities works by 'zooming in' on subsets of the sampling distribution matrix where previously selected variables are not significant.

CSA with conditional probabilities, performed according to the Rose & Wood algorithm (1998), requires a large number of random subsets to be generated. The current author suggests that a simpler and quicker algorithm could be based directly on the p-values for the different variable combinations, without the need to generate a large number of random subsets. To determine whether the addition a new variable results in a significant difference of clustering tightness, a 'conditional' probability could be calculated by determining whether the difference between the p-value obtained in n dimensions is significantly different from the p-value obtained in n+1 dimensions. The null hypothesis is that the two p-values are the same, and the test statistic is given by the following equation:

$$\text{Test statistic} = \frac{p_1 - p_2}{\sqrt{p(1-p)[(1/n_1) + (1/n_2)]}} \qquad \text{(Equation 4.37)}$$

where p_1 is the larger p-value (presumably for n variables) and p_2 is the smaller p-value (presumably for n+1 variables); $p = (p_1 + p_2)/2$ (i.e. the mean p-value); n_1 is the number of random samples generated for the calculation of p_1; and n_2 is the number of random samples generated for the calculation of p_2. If the test statistic ≥ 1.96, then it can be concluded at the 95% confidence level that there is a significant difference between the two p-values, and if the statistic ≥ 2.58, the same conclusion can be reached at the 99% significance level. These values are the critical values taken from the normal distribution. Clearly, the method could operate in either a forward or backward manner, with systematic addition (in a forward procedure) or deletion (in a backward procedure) of one variable at a time.

Variants of cluster significance analysis

For the analysis of embedded data sets in which the biological activity is measured on a continuous scale rather than a dichotomous scale, Rose & Wood (1998) introduced the concept of 'generalized' CSA (GCSA). Algorithms for GCSA work by using weighted variances for each variable, i.e. by taking all chemicals into consideration, rather than a subset, and by allowing each chemical to contribute to the variance in proportion to its activity. A Minitab macro for performing GCSA is given in Appendix A4.

Another type of CSA, developed by Rose *et al*. (1991), is 'single class discrimination using principal component analysis' (SCD-PCA). This method maximises the variance of the diffuse cluster with respect to the variance of the embedded cluster, and therefore improves the ability to discriminate between compounds in the two clusters.

Finally, another variant of CSA, called 'rank distance clustering' (RDC) simplifies the calculation of cluster tightness by ranking the distances of all compounds from the centroid, and by comparing the sum of ranks for the active cluster with the sum of ranks that would be expected by chance (Rose & Wood, 1998)

4.5.3 Bootstrap resampling

Bootstrap resampling is a statistical method for making inferences about an unknown population on the basis of a single sample (Efron & Tibshirani, 1993). The basic premise of the method is that the sample is representative of the population from which it was drawn, i.e. it is assumed that each observation in the sample had an equal probability of being selected from the population. Since there is only one sample, bootstrapping simulates what would happen if the population were resampled by randomly resampling the sample. The sample is sampled with replacement, i.e. if there are *n* observations in the sample, *n* observations are randomly selected from the sample in such a way that any given observation can be picked any number of times between zero and *n*. Every time a random selection of observations is made, a statistic of interest is calculated on the basis of these observations. If the resampling procedure is performed a large number of times, the resulting bootstrap distribution of the statistic is

assumed to approximate its true sampling distribution. Therefore, the bootstrap distribution can be used to estimate confidence limits for population parameters.

Standard bootstrap method

The simplest method for obtaining bootstrap confidence limits is called the standard bootstrap method. According to this method, if the sample statistic is denoted by T, and the corresponding population parameter is denoted by θ, then the 95% confidence interval (CI) for θ is given by Equation 4.38:

$$95\%\ \text{CI for } \theta = \text{bootstrap mean} \pm 1.96\ s_B \qquad \text{(Equation 4.38)}$$

where 'bootstrap mean' is the mean value of the bootstrap distribution of T; and s_B is the standard deviation of the bootstrap distribution of T. For this equation to be valid, it is necessary that: a) the bootstrap distribution of T is approximately normal; and b) T is an unbiased estimate of θ, i.e. the bootstrap mean equals the estimate based on the original sample. No assumptions are made about the distribution of observations in the population.

The bias in the estimator T can be calculated as the difference between the bootstrap mean and the original sample mean:

$$\text{Bias (T)} = \text{bootstrap mean} - \text{sample mean} \qquad \text{(Equation 4.39)}$$

If the sample statistic (T) turns out to be a biased estimate of the population parameter (θ), the bias can be accounted for by centering the 95% CI for θ on the bias-corrected estimate of the population mean, i.e. Equation 4.40 would be used instead of Equation 4.38:

$$\text{Bias-corrected 95\% CI for } \theta = (\text{sample mean} - \text{bias}) \pm 1.96\ s_B \qquad \text{(Equation 4.40)}$$

Percentile-based bootstrap methods

If the bootstrap distribution of estimates (T) shows a significant departure from the normal distribution (for example, if there is strong skewness or kurtosis), the standard bootstrap method will not give accurate confidence limits. In such a case, a better approach is to calculate percentiles from the bootstrap distribution. The 95% CI for the population parameter θ is then given by two values that encompass 95% of the bootstrap distribution. According to the simple percentile method, these would be the values that exceed 2.5% and 97.5%, respectively, of the bootstrap distribution. The simple percentile method makes the assumption that a transformation of T exists such that the transformed variable is normally distributed. Although the mathematical form of the transformation does not need to be known, the transformation is assumed to be monotonic, i.e. the ordering of the transformed bootstrap estimates from the smallest to the largest value is assumed to correspond to the ordering of the untransformed bootstrap estimates.

If T is a biased estimator of θ, i.e. if the mean of the bootstrap distribution is not equal to the estimate based on the original sample, then it is more appropriate to use the bias-corrected percentile method. In this method, the number of times (n) that the bootstrap estimate (T_B) exceeds the sample estimate (T) in N bootstrap samples is counted, so that the probability (p) of T_B being greater than T is given by n/N. The next step is the calculation of Z_p, the value of the standard normal distribution that is exceeded with a probability of p. The value of Z_p is used to calculate two proportions, P_L and P_H:

$$P_L = \phi (2Z_p - 1.96) \qquad \text{(Equation 4.41)}$$

$$P_H = \phi (2Z_p + 1.96) \qquad \text{(Equation 4.42)}$$

where ϕ (z) is the proportion of the standard normal distribution that is less than z.

Finally, the values of P_L and P_H are used to calculate the 2.5^{th} and 97.5^{th} percentiles, i.e. the lower and upper values in the bootstrap distribution corresponding to cumulative probabilities of 0.025 and 0.975, respectively:

Lower 95% CL = INVCDF{$\phi(2Z_p - 1.96)$} (Equation 4.43)

Upper 95% CL = INVCDF{$\phi(2Z_p + 1.96)$} (Equation 4.44)

where CL is the confidence limit; and INVCDF(p) is the value in the bootstrap distribution corresponding to a cumulative probability of p.

Table 4.1 A 2 x 2 contingency table

		Predicted class		
		Toxic	Non-toxic	Marginal totals
Observed (*in vivo*) class	Toxic	a	b	a+b
	Non-toxic	c	d	c+d
	Marginal totals	a+c	b+d	a+b+c+d

Table 4.2 Definitions of the Cooper statistics

Statistic	**Definition: "the proportion (or percentage) of the …**	
sensitivity	toxic chemicals (chemicals that give positive results *in vivo*) which the CM predicts to be toxic."	= a/(a+b)
specificity	non-toxic chemicals (chemicals that give negative results *in vivo*) which the CM predicts to be non-toxic."	= d/(c+d)
concordance or accuracy	chemicals which the CM classifies correctly."	= (a+d)/(a+b+c+d)
positive predictivity	chemicals predicted to be toxic by the CM that give positive results *in vivo*."	= a/(a+c)
negative predictivity	chemicals predicted to be non-toxic by the CM that give negative results *in vivo*."	= d/(b+d)
false positive (over-classification) rate	non-toxic chemicals that are falsely predicted to be toxic by the CM."	= c/(c+d) = 1 - specificity
false negative (under-classification) rate	toxic chemicals that are falsely predicted to be non-toxic by the CM."	= b/(a+b) = 1 - sensitivity

Footnote to Table 4.2

CM = classification model

Table 4.3 Variance components used in linear regression analysis

Statistic		**Definition**	**Degrees of freedom[a]**
Total sum of squares	(TSS)	$\sum(y_i - \bar{y})^2$	n-1
Explained sum of squares	(ESS)	$\sum(\hat{y}_i - \bar{y})^2$	p
Residual sum of squares	(RSS)	$\sum(y_i - \hat{y}_i)^2$	n-p-1
Explained mean square	(EMS)	ESS/p	
Residual mean square	(RMS)	RSS/n-p-1	

Footnote to Table 4.3

[a] n = number of observations; p = number of variables

CHAPTER 5

NOVEL APPLICATIONS OF THE BOOTSTRAP RESAMPLING METHOD

5.1 INTRODUCTION

As described in chapter 4, bootstrap resampling is a statistical method for making inferences about an unknown population on the basis of a single sample. This chapter describes the use of novel bootstrap resampling algorithms to analyse the variability in: a) Cooper statistics, which vary according to the chemicals included in the classification model (CM) test set; and b) Draize test endpoints, which are subject to biological variability (between rabbits) and to temporal variability (between observation points). The bootstrap method can also be used to assess the variability in model parameters, as illustrated in Chapters 6 and 8 for CMs and regression models, respectively.

5.2 BOOTSTRAP ANALYSIS OF COOPER STATISTICS

5.2.1 Introduction

As described in chapter 4, the predictive abilities of two-group CMs are often expressed in terms of their Cooper statistics (Cooper *et al*., 1979). These statistics are often reported without any indication of their uncertainty, making it impossible to judge whether the predicted classifications are significantly better than would be expected by chance, or significantly better than the predictions made by a different CM. To address this problem, the aim of the work described in this section was to develop a means of expressing the uncertainty associated with Cooper statistics. As described below, this was achieved by formulating bootstrap resampling routines that derive 95% confidence intervals (CIs) for the various Cooper statistics. The usefulness of these bootstrapping routines is illustrated by constructing CIs for the Cooper statistics of four alternative tests for skin corrosion (the rat skin transcutaneous electrical resistance [TER] assay, EPISKIN^{TM}, Skin^{2TM} and CORROSITEX^{TM}), and for four two-step sequences in which each *in vitro* test is used in combination with a physicochemical test for skin corrosion based on pH measurements.

5.2.2 Method

Coding of the bootstrap resampling algorithms

As described in chapter 4, bootstrap CIs can be derived by various methods, including the standard method, the simple percentile method, and the bias-corrected percentile method. To compare the use of these methods, bootstrap algorithms were coded in the form of two macros for use with the Minitab software package (Minitab Inc., State College, PA).

The first macro, the 'bootstrapping macro' (Appendix A5), generates bootstrap distributions and calculates the standard 95% CIs of three Cooper statistics: concordance, sensitivity and specificity. To reduce the time taken for the macro to run, bootstrap distributions are not calculated for the false positive and false negative rates, since these distributions can be derived from the corresponding bootstrap distributions for specificity and sensitivity, respectively (see chapter 4). To carry out the simple percentile method, it is sufficient to calculate the 2.5^{th} and 97.5^{th} percentiles of the bootstrap distributions generated by the bootstrapping macro.

The bootstrapping macro works by counting the total number of chemicals in the data set, and by randomly sampling, with replacement, this number of chemicals from the data set. This procedure is carried out N times (where N is a number defined by the user) to generate N bootstrap samples. Each bootstrap sample consists of: a) the known classification for each chemical selected; and b) predicted classifications made by different laboratories, possibly on different occasions (experimental runs). The two sets of classifications (known and predicted) are used to calculate the Cooper statistics. The distribution of a given statistic over the N bootstrap samples gives the bootstrap distribution of that statistic, and the 95% confidence interval of each Cooper statistic is calculated from the bootstrap distribution by using Equation 4.38. Since the bootstrapping macro generates a different bootstrap distribution for each laboratory/run combination, the user can calculate the Cooper statistics obtained by individual laboratories (by averaging over all experimental runs with that laboratory) or by all laboratories taken together (by averaging over all of the laboratory/run combinations). If only one set of predicted classifications is available, this is treated as the first laboratory/run combination.

A second Minitab macro, the 'percentile macro' (Appendix A6), performs the bias-corrected percentile method. This macro applies the algorithm described in Chapter 4 to calculate bias-corrected percentiles, which can then be applied to the bootstrap distributions already generated by the bootstrapping macro to derive the bias-corrected confidence limits.

Calculation of Cooper statistics for four in vitro skin corrosion tests

To calculate 95% CIs for the Cooper statistics of the TER, EPISKIN, Skin^2 and CORROSITEX assays, the Minitab macros were applied to a heterogeneous data set of 60 chemicals (Table 5.1). These chemicals, which were the test chemicals in the ECVAM Skin Corrosivity Validation Study (Barratt *et al.*, 1998; Fentem *et al.*, 1998), cover a range of physical forms and chemical classes. The validation study was designed so that the TER, Skin^2 and CORROSITEX tests would classify the 60 chemicals on two occasions (runs) in each of three laboratories, i.e. a total of 360 (60x2x3) predicted classifications, whereas the EPISKIN assay would classify the chemicals on three occasions in each of three laboratories, i.e. a total of 540 (60x3x3) predicted classifications. In practice, the actual number of *in vitro* classifications for Skin^2 and CORROSITEX differed from 360. In the case of Skin^2, 50 of the 60 chemicals were tested on just one occasion by one of the laboratories, which resulted in 310 Skin^{2}classifications being available for comparison with the *in vivo* classifications. In the case of the CORROSITEX assay, 26 of the 60 chemicals were found to be incompatible with the test system by one or more laboratories, and were therefore not tested by these laboratories. As a result, 229 CORROSITEX classifications were available for comparison with the *in vivo* classifications. The bootstrapping macro was written to automatically exclude missing and non-qualifying observations during the calculation of Cooper statistics.

To compare the various bootstrap methods, the 95% CIs for the Cooper statistics of TER, EPISKIN, Skin^2 and CORROSITEX were also calculated, not only by the standard bootstrap method, but also by the simple and bias-corrected percentile methods. In each case, random resampling was carried out 1000 times. When the standard bootstrap method was applied, the bias in each bootstrap mean was calculated by using Equation 4.39 (see chapter 4).

Calculation of Cooper statistics for four two-step sequences of tests

The standard bootstrap method was used to calculate CIs for the Cooper statistics of two-step sequences based on the use of a pH test followed by one of the following methods: a) the TER assay; b) the EPISKIN assay; c) the Skin^2 assay; or d) the CORROSITEX assay. Predicted classifications of skin corrosion potential based on pH measurements were taken from Worth *et al.* (1998) and are reproduced in Table 5.1. To determine whether the four stepwise sequences are significantly more predictive than the four stand-alone *in vitro* tests, t-tests for independent samples were used to compare the 'sequential' bootstrap distributions with their corresponding 'stand-alone' distributions, using a p-level of 0.01.

5.2.3 Results

The data set of 60 chemicals is presented in Table 5.1, which also gives the *in vivo* classifications of skin corrosion potential (C/NC) and the average (most frequent) *in vitro* classifications made by each test method.

Bootstrap analysis of the Cooper statistics for four in vitro tests

A separate bootstrap distribution for each Cooper statistic of each *in vitro* test was generated (i.e. a total of 20 distributions). An illustrative example is the bootstrap distribution of the concordance of EPISKIN, represented by a frequency histogram in Figure 5.1, and by a cumulative probability distribution in Figure 5.2.

The results obtained by applying the standard bootstrap method to the 20 distributions are presented in Table 5.2, which also gives the Cooper statistics calculated from the original data set of *in vitro* and *in vivo* classifications. The 'original values' correspond very closely with the mean values of the bootstrap distributions, which means that there is little or no bias (bootstrap mean – original value) in the bootstrap mean values. For example, there is a bias of -1% in the concordance, sensitivity and specificity of the TER assay, which means that the corresponding population parameters are likely to be 1% higher than the values based on the sample of 60 chemicals. In the case of EPISKIN and CORROSITEX, the Cooper statistics show no bias, whereas in the case of Skin^2,

the concordance and sensitivity are positively biased, and the false negative rate is negatively biased.

The 20 bootstrap distributions were used to derive 95% CIs by the simple and bias-corrected percentile methods (Table 5.3). The median values obtained by the simple percentile method are identical to the mean values obtained by the standard bootstrap method (Table 5.2) in all but six cases. In these cases (specificity and false positive rate of TER; sensitivity and false negative rate of Skin^2; sensitivity and false negative rate of CORROSITEX), the median value differs from the mean value by just one percent. The high level of agreement between the median and mean values means that the bootstrap distributions are not skewed to a significant extent.

The bias-corrected percentile CIs can be compared with the standard CIs by comparing the bias-corrected lower and upper percentiles with 0.025 and 0.975, respectively. It can be seen that the bias-corrected percentiles for EPISKIN and CORROSITEX are closer to the 0.025 and 0.975 than are the bias-corrected percentiles for TER and Skin^2. This indicates that the Cooper statistics for EPISKIN and CORROSITEX are less biased than are the corresponding statistics for TER and Skin^2, which is in accordance with the estimates of bias made by subtracting the original sample values from the bootstrap mean values (Table 5.2). In the case of TER, the lower and upper bias-corrected percentiles for concordance, sensitivity and specificity are higher than 0.025 and 0.975, respectively, which means that the 95% CIs for the Cooper statistics are shifted to higher values. This is in accordance with the negative bias estimated by the standard bootstrap method. In the case of Skin^2, the lower and upper confidence limits for concordance and sensitivity are shifted to lower values, which is consistent with the positive bias estimated by the standard bootstrap method. In the case of Skin^2, bias-corrected percentiles for specificity could not be calculated. The reason for this is that all of the values in the bootstrap distribution are 100%, so the proportion of bootstrap estimates exceeding the specificity derived from the original sample (100%) is zero. Since one of the steps in the algorithm calculates the inverse cumulative density function of this proportion for a standard normal distribution, which is not defined when the proportion is zero, the algorithm is not applicable in this case (and the macro terminates with an error message).

Bootstrap analysis of the Cooper statistics for four sequences of alternative methods

The bootstrap estimates of the Cooper statistics, obtained by applying the standard bootstrap method to the four stepwise sequences are presented in Table 5.4, which also gives the Cooper statistics based on the original sample of 60 chemicals. Again, there is little or no bias in the Cooper statistics, indicating that the values based on the original sample are good approximations of their 'true' values.

In Table 5.5, a comparison is made, for each of the four *in vitro* tests, between the Cooper statistics (the bootstrap mean values) relating to its stand-alone use and the corresponding statistics relating to its use in sequence with the pH test. Although, in practice, the pH test would be carried out before the *in vitro* test, these statistics would be the same irrespective of the order in which the two tests are carried out, since the decision rule is effectively 'if at least one of the two tests predicts a chemical to be corrosive, then assume it is corrosive; if both tests predict the chemical to be non-corrosive, then assume it is non-corrosive'. The data in Table 5.5 show that the performances of the TER and EPISKIN tests in combination with the pH test are better than their individual performances, as judged by statistically significant increases in the concordance, sensitivity and specificity. The increases in the latter two statistics are paralleled by decreases in the false negative and positive rates, respectively. In the case of the Skin^2 test, the concordance and sensitivity are increased when used in combination with the pH test, but the specificity is decreased. However, the increase is sensitivity is much greater (25%) than the decrease in specificity (6%). In the case of the CORROSITEX assay, the apparent benefits of combining its use with the pH test are not statistically significant, except for a slight increase in specificity from 76% to 78%.

5.2.4 Discussion

The uncertainty in the Cooper statistics of a CM can be attributed to several sources, relating to the choice of: a) test chemicals; b) laboratories; c) parameters in the CM; and d) mathematical form of the CM. In this study, the bootstrap distributions were generated to reflect the variability arising from the choice of test chemicals only. This is justified on the basis that the majority of the uncertainty in the predictions made by the TER, EPISKIN, Skin^2 and CORROSITEX assays results from chemical variation, rather than from inter-laboratory and intra-laboratory variation, as demonstrated by the

statistical analyses carried out during the ECVAM validation study (Fentem *et al.*, 1998). This approach assumes that the mathematical form and the parameters (cut-off values) of the CM are 'correct'. Bootstrap resampling can also be used to assess the uncertainty in CM parameters by reparameterising the model of interest on the basis of each random sample, thereby generating a bootstrap distribution of the parameter. This approach has been adopted, for example, in the development of SARs for eye irritation potential (Worth & Cronin, 2000a). The most intangible source of uncertainty, however, lies in the mathematical form of the model.

The key assumption made by bootstrap resampling methods is that the sample is representative of the population. In this study, the data set of 60 chemicals is assumed to represent a population of industrial chemicals and consumer product ingredients comprising acids, bases, neutral organics, phenols, inorganic salts, electrophiles and surfactants. It is difficult, if not impossible, to determine the validity of this assumption, but it is more likely to be reasonable for chemicals that are more numerous in the data set (e.g. the organic acids) than for less numerous chemicals (e.g. the surfactants).

Several bootstrap algorithms were compared in this study. The simplest algorithm is the standard bootstrap method, which is applicable if the bootstrap distribution is approximately normal. If the bootstrap distribution shows a significant departure from normality, one of the percentile-based methods should be used instead. In general, the calculation of percentile confidence limits requires more bootstrap samples than does the calculation of standard confidence limits, because of the need to estimate accurate percentage points for the bootstrap distribution, instead of just estimating the mean and standard deviation of the bootstrap distribution (Efron & Tibshirani, 1993). In this study, the three bootstrap algorithms were applied to bootstrap distributions containing 1000 samples. It was noticed, however, that 100 random samples gave essentially the same results with the standard bootstrap method. The CIs obtained by the three bootstrap methods were comparable, because the individual bootstrap distributions were approximately normal. In such cases, the standard bootstrap method is clearly the simplest method of choice.

In validation studies for alternative methods, the bootstrap distributions of Cooper statistics could be used to compare the predictive abilities of different *in vitro* tests,

provided that the different tests have been applied to a common set of chemicals, and have been shown to be reproducible. In the case of two *in vitro* tests (i.e. two CMs), the statistics could be compared by using the t-test, whereas in the case of three or more PMs, one-way analysis of variance (ANOVA) could be used. In principle, the t-test and ANOVA are only applicable to variables that are normally distributed. However, since the number of random samples making up each bootstrap distribution can be set at 100 or more, the Cooper statistics derived from these distributions can be expected to follow a normal distribution, in accordance with the central limit theorem.

Another, perhaps more meaningful, application of bootstrap resampling in a validation study would be to compare the predictive abilities of individual test methods with a set of pre-defined performance criteria. In the ECVAM validation study on *in vitro* tests for skin corrosion (Barratt *et al.*, 1998; Fentem *et al.*, 1998), the management team of the study decided that for a given test to be valid, it should correctly identify at least 70% of the corrosive chemicals in the test set, i.e. the sensitivity of the test should be ≥ 70%. In 'bootstrap language', this criterion could be expressed as "at least 95% of the bootstrap estimates of sensitivity should be greater than or equal to 70%". To take an example, the sensitivity of EPISKIN is 83%, but the 95% CI ranges from 69% to 96% (Table 5.2). It turns out, however, that 96% of the values in this range have a sensitivity greater than or equal to 70%, so the EPISKIN test could be said to satisfy the 'bootstrap criterion for sensitivity'. Conversely, only 58% of the bootstrap estimates for sensitivity of CORROSITEX have a value greater than or equal to 70%, so this test would not satisfy the 'bootstrap criterion for sensitivity'.

Similarly, the Cooper statistics in Table 5.4 show that the four two-step sequences of tests all have a sensitivity greater then or equal to 70%. However, it is also apparent that the sensitivities of the sequences involving Skin^2 and CORROSITEX have lower 95% confidence limits that are less than 70%. It turns out that just 55% of the values in bootstrap distribution for the sensitivity of the Skin^2 sequence are greater than or equal to 70%. Similarly, in the case of the CORROSITEX sequence, just 60% of the values are greater than or equal to the threshold value. Thus, neither Skin^2 nor CORROSITEX used in combination with the pH test could be said to satisfy the 'bootstrap criterion for sensitivity'.

5.2.5 Conclusions

In conclusion, the bootstrap resampling method can be used to assess at least some of the uncertainty associated with Cooper statistics. It can therefore be used to compare the predictive capacities of different CMs with each other, or with benchmark values that define an acceptable predictive performance. The standard bootstrap method could be a useful tool during the validation of alternative methods and strategies that predict toxicological hazard. In particular, it could help judgements to be made when the observed Cooper statistics for an alternative method or strategy are only marginally better than the acceptance criteria established for the study. The results of this study regarding four *in vitro* tests for skin corrosion concur with those made by the management team of an ECVAM validation study. In addition, the results indicate that the performances of three tests, the TER, EPISKIN and Skin^2 assays, are enhanced when they are used in combination with a simple pH test.

5.3 BOOTSTRAP ANALYSIS OF DRAIZE TEST ENDPOINTS

5.3.1 Introduction

As described in chapter 2, the acute dermal and ocular effects of chemicals are generally assessed by performing the Draize skin and eye tests, respectively. Since the animal data obtained in these tests are used to develop and validate alternative methods for skin and eye irritation, it is important to assess the inherent variability of the data, since this places an upper limit on the predictive performance that can be expected of any model. Therefore, the first objective of the study reported in this section was to estimate the variability arising from the use of different animals and time-points, and to use the estimates of variability to determine the capacities of the Draize skin and eye irritation tests to 'predict their own results'.

The results of the Draize skin and eye irritation tests are often expressed in terms of the PII and the MMAS, respectively. Since these are both composite measures of toxicity, derived by averaging specific tissue effects, alternative methods could be designed to reproduce the more fundamental endpoints, rather than the more complex ones. However, since there may be a certain amount of redundancy between the specific

tissue effects, it may be unnecessary to develop models for predicting all of them. Therefore, the second objective of this study was to examine the relationships between: a) the two skin irritation effects (erythema and œdema) from which the primary irritation index (PII) is derived; and b) the six eye irritation effects (conjunctival erythema, œdema and discharge; corneal opacity and area; iritis) from which the weighted Draize score (WDS) and modified maximum average score (MMAS) are derived.

5.3.2 Method

Variability of Draize skin scores

A data set of 143 chemicals (Table 5.6) was compiled from the ECETOC skin irritation data bank (ECETOC, 1995). For each chemical, the Draize skin scores for erythema and œdema, obtained at three time-points (24h, 48h & 72h), were entered into an Excel spreadsheet. Subsequently, the data were copied into Minitab spreadsheets so that there was one Minitab spreadsheet per chemical. The data in each Minitab spreadsheet took the following format, with a maximum of 36 columns (2 skin effects x 3 time-points x 6 rabbits) containing the animal data:

Rabbit 1						Rabbit 2						Rabbit 3 …
Erythema			Œdema			Erythema			Œdema			
24h	48h	24h	48h	72h	24h	48h	72h	72h	24h	48h	72h	

A Minitab macro was written (Appendix A7) to bootstrap the tissue scores in this format, taking into account the number of rabbits for which there were data. For example, if tissue scores are available for three rabbits, each iteration of the macro selects, at random and with replacement, 9 erythema and 9 œdema scores. Bootstrap estimates of erythema and œdema are then generated by averaging the appropriate set of 9 scores. The bootstrap estimate of the PII is calculated by adding the bootstrap estimates of erythema and œdema, in accordance with Equation 2.1 (see chapter 2). This procedure is repeated N times, where N is a number specified by the user. In this study, the bootstrap routine was performed 1000 times.

The bootstrap estimates of the PII were compared with the 'conventional' PII values calculated from the sample data, to determine whether any of the sample PII values were biased, i.e. needed a correction to better approximate their 'true' population mean values.

The variability in the erythema, œdema and PII values were measured in terms of the bootstrap standard deviation (SD) and the bootstrap coefficient of variation (CV), the latter being equal to the bootstrap SD divided by the bootstrap mean. To assess the ability of the Draize skin test to 'predict itself', a simulation was performed in which 1000 PII values, taking values from 0 to 8 Draize units, were randomly generated from a uniform distribution. From these 'observed' PII values, 'predicted' PII values were generated according to Equation 5.1.

Predicted PII = observed PII + 2 E (Equation 5.1)

where E is the (positive or negative) error, randomly generated from a normal distribution with a SD equal to the maximum value of the bootstrap SD. The Pearson correlation coefficient (r) and the coefficient of determination (r^2) were then calculated for the linear relationship between the predicted and the observed PII values.

Relationships between Draize skin endpoints

To examine the relationships between six skin irritation endpoints (2 types of skin effect for each of 3 time-points), a data set of 176 objects (143 chemicals, plus 33 multiple observations) was used (Table 5.6). For each object, mean tissue scores were obtained by averaging, across all rabbits, the tissue scores for each effect/time-point combination. The six skin irritation endpoints were used to construct a matrix plot, and were subjected to Spearman correlation analysis.

Variability of Draize eye scores

A data set of 92 chemicals (Table 5.7) was compiled from the ECETOC eye irritation data bank (ECETOC, 1998a). For each chemical, the Draize eye scores for conjunctival erythema, œdema and discharge, for corneal opacity and area, and for iritis, obtained at three time-points (24h, 48h & 72h), were entered into an Excel spreadsheet. Subsequently, the data were copied into Minitab spreadsheets so that there was one

Minitab spreadsheet per chemical. The data in each Minitab spreadsheet took the following format, with a maximum of 108 columns (6 eye irritation effects x 3 time-points x 6 rabbits) containing the animal data:

Conjunctival erythema		Conjunctival œdema		Conjunctival discharge …
Rabbit 1 24h 48h 72h	Rabbit 2 …	Rabbit 1 24h 48h 72h	Rabbit 2 …	

A Minitab macro was written (Appendix A8) to bootstrap the tissue scores in this format, taking into account the number of rabbits for which there were data. For example, if there are three rabbits, each iteration of the macro samples, at random and with replacement, 9 scores of each eye effect. Bootstrap estimates of the six eye effects are then generated by averaging the appropriate set of 9 scores. The bootstrap estimate of the WDS is calculated as a weighted sum of the six endpoints, in accordance with Equation 2.2 (see chapter 2). This procedure is repeated N times, where N is a number specified by the user. A maximal WDS, intended to provide a theoretical upper limit for the MMAS, is estimated by applying Equation 2.2 to the maximal values of the bootstrap distributions for the six eye effects.

The variability in the six eye irritation effects and in the WDS was measured in terms of the bootstrap SD and CV. To assess the ability of the Draize eye test to 'predict itself', a simulation was performed in which 1000 WDS values, taking values from 0 to 110, were randomly generated from a uniform distribution. From these 'observed' WDS values, 'predicted' WDS values were generated according to Equation 5.2.

Predicted WDS = observed WDS + 2 E (Equation 5.2)

where E is the (positive or negative) error, randomly generated from a normal distribution with a SD equal to the maximum value of the bootstrap SD. The correlation between the predicted and the observed WDS values was assessed in terms of the Pearson correlation coefficient (r) and the coefficient of determination (r^2).

Relationships between Draize eye endpoints

To examine the relationships between 18 eye irritation endpoints (6 types of effect for each of 3 time-points), a data set of 94 objects (92 chemicals plus two replicates) was used. For each object, mean tissue scores were obtained by averaging, across all rabbits, the tissue scores for each effect/time-point combination. The 18 eye irritation endpoints were used to examine the relationships between: a) the six eye effects (averaged over the three time-points); and b) the weighted conjunctival, corneal and iridial endpoints at each time-point (see chapter 2 for details of the Draize weighting scheme).

5.3.3 Results

Variability of Draize skin scores

The results obtained by bootstrapping the Draize skin scores are summarised in Table 5.6. The bootstrap SD varies from 0 to 1 Draize unit, corresponding to 95% CIs of up to ± 2 Draize units. The bootstrap CV ranges from 0 to 97%. Very few of the bootstrap distributions are skewed, as indicated by the fact that nearly all bootstrap mean values are identical to the corresponding median values.

For 23 of the chemicals taken from the ECETOC skin irritation data bank (ECETOC, 1995), there was more than one data sheet, which meant that more than one set of Draize skin scores could be derived. For example, there are four data sheets for hexyl salicylate, associated with PII values of 3.4, 3.7, 4.2 and 3.3 (in order of presentation). All four data sheets were supplied to ECETOC by the same industrial company. The application of one-way ANOVA to the four bootstrap distributions for the PII of hexyl salicylate revealed that these values are significantly different ($p<0.05$). Subsequent analysis of the four erythema distributions, and of the four œdema distributions, revealed that the variability in the PII of hexyl salicylate is related to the variability in œdema, but not erythema. The œdema scores in the fourth data sheet were the least variable. The variability in the PII of hexyl salicylate, obtained by bootstrapping the tissue scores in the fourth data sheet, is illustrated in Figure 5.3

Comparison of the bootstrap estimates of the PII with the PII values calculated from the sample data revealed that the PII values for six of the 143 chemicals are biased (Table 5.8), i.e. there is a difference between the bootstrap PII and the sample PII. This suggests that the sample PII values of these six chemicals should be corrected to more accurately represent their 'true' (population) values (see Table 5.8). The correlation between the observed and predicted PII values is illustrated in Figure 5.4; the r value was found to be 0.7, which corresponds to an r^2 value of 0.57.

Relationships between Draize skin endpoints

A matrix plot (Figure 5.5) of the six skin endpoints indicates that rectilinear relationships exist between the erythema endpoints, and between the œdema endpoints. As for the relationship between erythema and œdema at each time-point, it appears that any given value of erythema corresponds to a range of œdema values falling below a certain threshold. Conversely, for any given œdema value, there appears to be a range of œdema values lying above a certain threshold. In other words, if one knows the erythema value at a given time-point, one could predict the maximal œdema value at that time-point, whereas knowledge of the œdema value would permit prediction of a minimal erythema value.

Spearman correlation analysis of the six skin irritation endpoints for 176 observations (Table 5.9) revealed that each pairwise combination of variables is significantly correlated ($p<0.001$). In particular, the three erythema variables are highly correlated with each other ($r = 0.86$-0.96), as are the œdema variables ($r=0.87$-0.94). The correlations between the erythema and œdema variables tend to be lower ($r=0.74$-0.86).

Variability of Draize eye scores

The results obtained by bootstrapping the Draize eye scores are summarised in Table 5.7. The bootstrap SD of the WDS varies from 0 to 8 Draize units, corresponding to 95% CIs of up to ± 16 units. The bootstrap CV of the WDS ranges from 3 to 96%. Some of the bootstrap distributions are skewed, as indicated by the fact that their mean values differ from their median values.

The maximal WDS, obtained by bootstrap resampling 1000 times, was generally found to be greater than or equal to the MMAS, calculated from original sample data. However, contrary to expectation, for 9 chemicals (shaded in Table 5.7), the maximal WDS was slightly smaller than the MMAS. The tissue scores for these chemicals were therefore bootstrapped again, this time by randomly sampling 5000 times. In each case, this resulted in the maximal WDS being equal to the MMAS. The fact that the number of bootstrap samples needed to be increased for the MMAS values of these chemicals to fall within the range of their resampled WDS is an indication of how unrepresentative their MMAS values are. In other words, the MMAS for these chemicals is calculated from a set of tissue scores that has a small probability of being selected during the bootstrapping procedure.

In the cases of two chemicals (xylene and methyl amyl ketone) taken from the eye irritation data bank (ECETOC, 1998a), there was more than one data sheet, which meant that more than one set of Draize eye scores could be derived. For example, the two data sheets for heptanone (methyl amyl ketone), correspond to mean WDS values of 4.7 and 7.5, respectively. The application of t-tests (for independent samples) to the two bootstrap distributions for the WDS revealed that these values are significantly different ($p<0.05$). Analysis of the unpaired distributions for the six eye effects showed that the difference between the two WDS values is related to differences in the assessment of conjunctival effects (erythema, œdema and discharge). However, there are no significant differences, between the two data sheets, in the assessments of corneal effects (opacity and area) and iritis. The tissue scores in the first data sheet are more variable than those in the second. The bootstrap distribution of the WDS for heptanone, obtained using the data in the second data sheet, is illustrated in Figure 5.6. The correlation between the observed and predicted WDS values is illustrated in Figure 5.7; the r value was found to be 0.90, which corresponds to an r^2 value of 0.81.

Relationships between Draize eye endpoints

A matrix plot (Figure 5.8) of the six eye irritation effects, in which each effect is averaged over three time-points and all rabbits in the experimental group, indicates that while there is relatively little scatter between the three conjunctival effects (erythema, œdema and discharge), the scatter is more pronounced between the two corneal effects (opacity and area), and between iritis and all other effects. Spearman correlation

analysis (Table 5.10) revealed that each pairwise combination of effects is significantly correlated ($r \geq 0.75$; $p<0.01$).

A matrix plot (Figure 5.9) of the weighted tissue scores for the conjunctiva, cornea and iris, in which each weighted score is averaged over all rabbits separately for each time-point, indicates that rectilinear relationships exist between the weighted conjunctival scores, and between the weighted corneal scores, obtained at the three time-points. However, the scatter between the weighted iris scores is so great that no relationship can be ascertained. Spearman correlation analysis (Table 5.11) revealed that the three conjunctival endpoints are strongly and significantly correlated ($r=0.85$-0.90; $p<0.01$), as are the three corneal endpoints ($r=0.86$-0.93; $p<0.01$). The correlations between the three iridial endpoints are slighly weaker ($r=0.67$-0.85), albeit statistically significant ($p<0.01$).

The WDS for each time-point was calculated by summing the weighted conjunctival, corneal and iridial scores for the same time-point. The matrix plot in Figure 5.10 and the Spearman correlations in Table 5.12 show that these are strongly and significantly related (0.86-0.99; $p<0.01$), with each other and with the maximal value of the WDS, obtained at any of the three time-points. Furthermore, the strongest correlation is between the WDS obtained at 24h and the maximal WDS, suggesting that most of the chemicals investigated exert their maximal effects in the rabbit eye within the 24 hours following exposure.

5.3.4 Discussion

The PII and the MMAS are both composite measures of acute local toxicity in that they are derived from two or more ‘fundamental’ types of tissue effect. It is interesting to note, however, that the two measures are based on a different rationale. The PII is intended to reflect an average skin effect, where the average is taken over three time-points and over all rabbits in an experimental group. In contrast, the MMAS is intended to reflect the maximal value of the WDS that can be obtained at any time-point. The reason for the different rationale may be that toxicologists are interested in the worst possible effects of chemicals, when assessing eye irritation, but are satisfied with the

more representative, mean effect when assessing skin irritation Alternatively, the different rationale may be based on nothing more than historical precedent.

A more consistent approach to the assessment of acute local toxicity would be to use the PII (which by definition is an average over three time-points and all rabbits) and the mean WDS (i.e. the WDS averaged over three time-points and all rabbits). In fact, the variability of the Draize eye test, which has been widely commented on (e.g. Earl *et al.*, 1997) may in part be due to the fact the MMAS is unrepresentative of the spectrum of possible WDS values. The use of bootstrap resampling to explore the variability in the WDS shows that the MMAS is a value located towards the far end of the right tail of the bootstrap distribution. In the case of a few chemicals, the MMAS was so extreme that resampling had to be carried out 5000 times to be captured within the bootstrap distribution. Another criticism of the MMAS (and the WDS) is the Draize weighting scheme itself - since corneal effects account for 80 out of a total of 110 Draize units, the variability in the corneal scores is amplified. The use of an alternative measure of eye irritation, based on an unweighted sum of tissue effects, would be analogous to the use of the PII for skin irritation.

In this study, the SD associated with the PII was found to range from 0 to 1 Draize units, whereas the SD associated with the WDS was found to be as large as 8 Draize units. These values reflect the variability in the Draize skin and eye tests arising from the use of different animals and time-points within a single experiment, and not from differences within or between laboratories, which are expected to result in larger variabilities. Indeed, in a study of the intra-laboratory results of four chemicals, Blein *et al.* (1991) reported SDs for the MMAS of 0-14, and in a study of the inter-laboratory variability of nine chemicals, Weil & Scala (1971) reported SD values for the MMAS of 0-30. Thus, the results of this study can be considered to reflect the minimal variability inherent in the Draize skin and eye tests.

Using the bootstrap SD values to simulate the abilities of the skin and eye tests to predict their own results, the skin test predicted the PII with an r^2 of 0.57, and the eye test predicted the mean WDS with an r^2 of 0.81. These values define upper limits to the performances of models for the PII and the WDS (and presumably the MMAS), respectively. Thus, if a model for the PII has an r^2 greater than 0.57, it is probably over-

fitting its training set. Similarly, the best possible model for the WDS is expected to have an r^2 of 0.81.

In the case of a few chemicals, multiple data sheets of tissue scores were available in the ECETOC reports, so analysis of the bootstrap distributions has enabled conclusions to be drawn about the source of intra-laboratory variability, in terms of observations on specific tissues. For example, it has been shown that the significant intra-laboratory variability between two independent measurements of the PII of hexyl salicylate is related to the variability in assessments of œdema, but not erythema.

Finally, it should be noted that the bootstrap distributions of Draize tissue scores can also be used to determine the certainty with which chemicals have been classified on the basis of animal data, which is especially useful in the case of borderline chemicals. For example, in a comparison of the EU and OECD systems for classifying eye irritants, it was noted that ethanol is classified as a non-irritant by the EU system, but as an irritant by the OECD system (Prinsen, 1999). The reason is that there are differences between the EU and OECD systems in the cut-off values for conjunctival erythema and corneal opacity (see chapter 2). In the EU system, a chemical is classified as an eye irritant if its conjunctival erythema ≥ 2.5, or if its corneal opacity ≥ 2, whereas in the OECD system, the lower thresholds of 2.0 and 1.0 are used for erythema and opacity, respectively. The bootstrap resampling of the tissue scores for ethanol generated bootstrap mean values of 2.1 and 1.1 for erythema and opacity, respectively. While these values just exceed the OECD cut-offs, they fall short of the EU cut-offs. Examination of the bootstrap distributions for ethanol revealed that 76% of the bootstrap estimates of erythema and opacity were greater than or equal to the OECD thresholds, whereas only 0.5% of the erythema estimates, and 0% of the opacity estimates, were greater than or equal to the EU thresholds. The identification of borderline chemicals (in terms of the boundaries imposed by a particular classification system) could be useful in the selection of validation study test chemicals, and in the definition of ‘reference standards’, i.e. chemicals of known toxic potency that are tested alongside chemicals of unknown potency with a view to rank ordering the test chemicals.

5.3.5 Conclusions

On the basis of the results obtained in this study, it is concluded that:

1) the bootstrap resampling of Draize tissue scores provides a useful means of exploring the variability resulting from the use of different animals (biological variability) and time-points (temporal variability);
2) since tissue effects at different time-points are strongly correlated, it appears that most of the variability results from the use of different animals;
3) the bootstrap distributions of Draize tissue scores can be used to assess the certainty with which chemicals have been classified;
4) the Draize skin test could be refined by assessing œdema and erythema at the 24-hour time-point only;
5) the variability in Draize skin scores is such that no model for predicting the PII can be expected to have an r^2 greater than 0.57;
6) the use of a mean WDS in the Draize eye test would provide a more representative composite measure of eye irritation than the MMAS;
7) the Draize eye test could be refined by assessing the conjunctival effects (erythema, œdema and discharge) and the corneal effects (opacity and area) at the 24-hour time-point only; it is unclear when the iris lesion should be assessed (if at all), since the correlations for this effect are associated with more scatter;
8) the variability in Draize eye scores is such that no model for predicting the WDS or MMAS can be expected to have an r^2 greater than 0.81.

Figure 5.1 Bootstrap distribution of the concordance of the EPISKIN test for skin corrosion

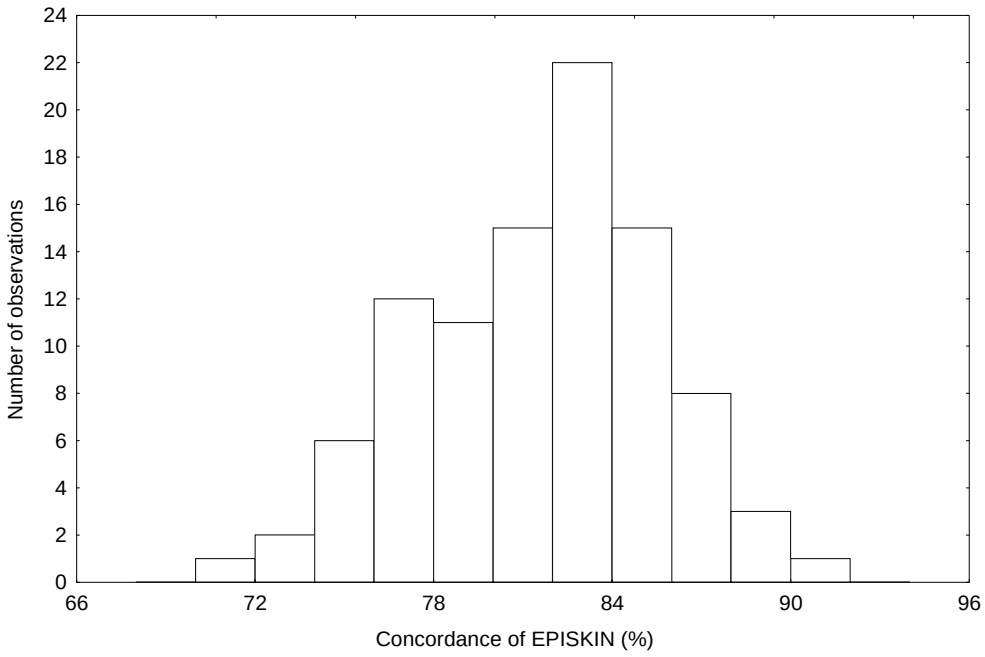


Figure 5.2 Cumulative probability distribution of the concordance of the EPISKIN test for skin corrosion

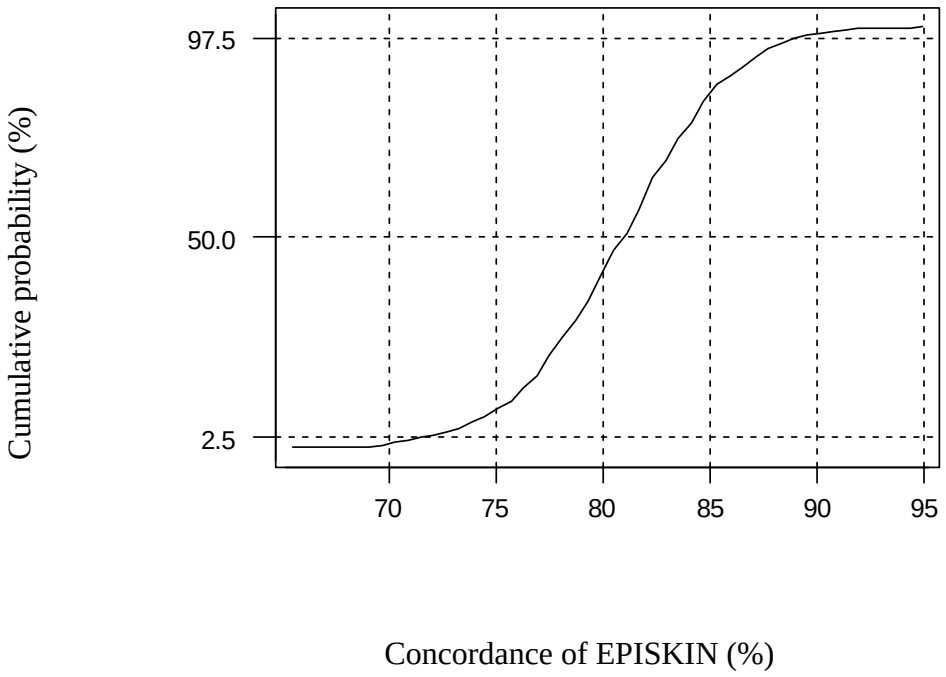


Figure 5.3 Bootstrap distribution of the primary irritation index of hexyl salicylate

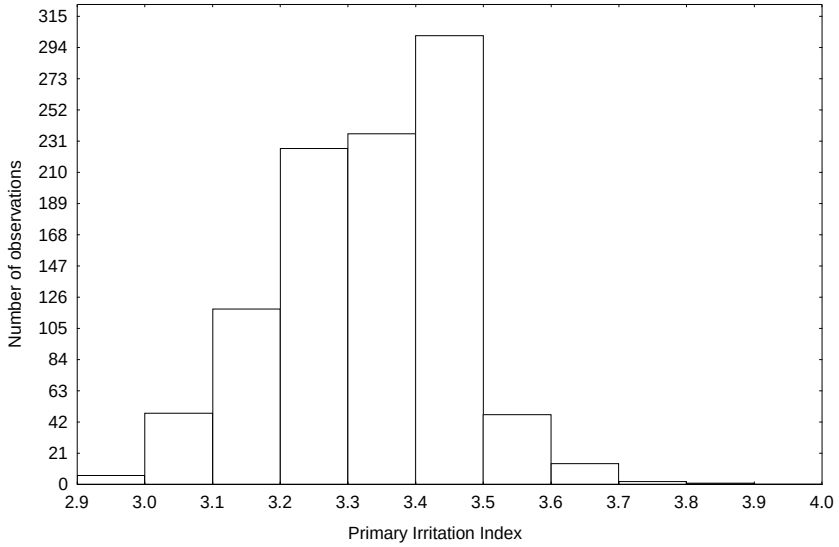


Figure 5.4 Simulation of the ability of the Draize skin test to 'predict itself'

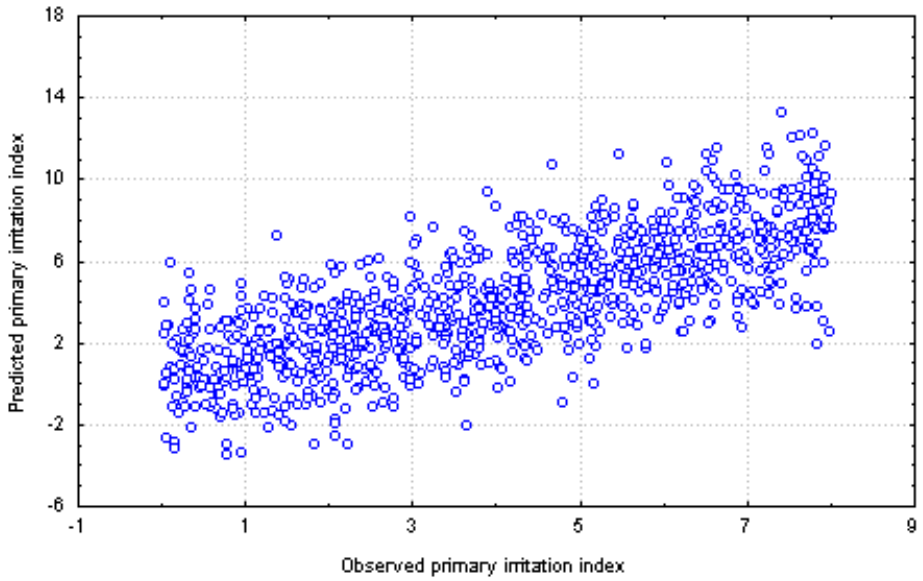
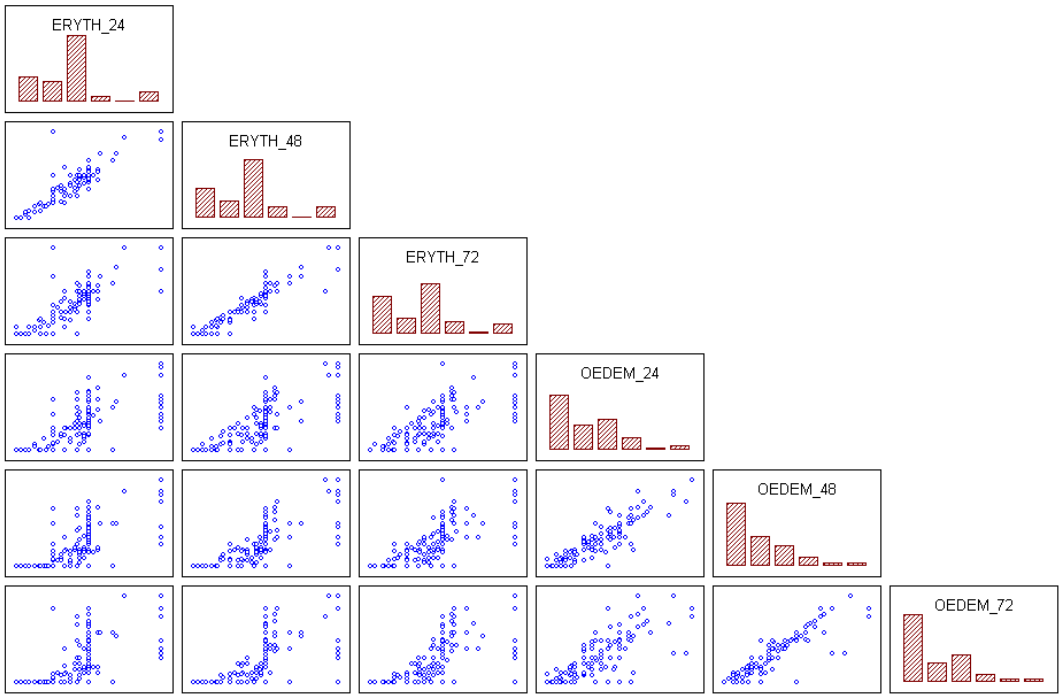


Figure 5.5 Matrix plot of 6 skin irritation endpoints obtained in the Draize skin test



Footnote to Figure 5.5

ERYTH_24 = erythema at 24h; ERYTH_48 = erythema at 48h; ERYTH_72 = erythema at 78h; OEDEM_24 = œdema at 24h; OEDEM_48 = œdema at 48h; OEDEM_72 = œdema at 72h.

Each tissue score is a mean value, taken over all rabbits in the experimental group.

Figure 5.6 Bootstrap distribution of the weighted Draize score of heptanone

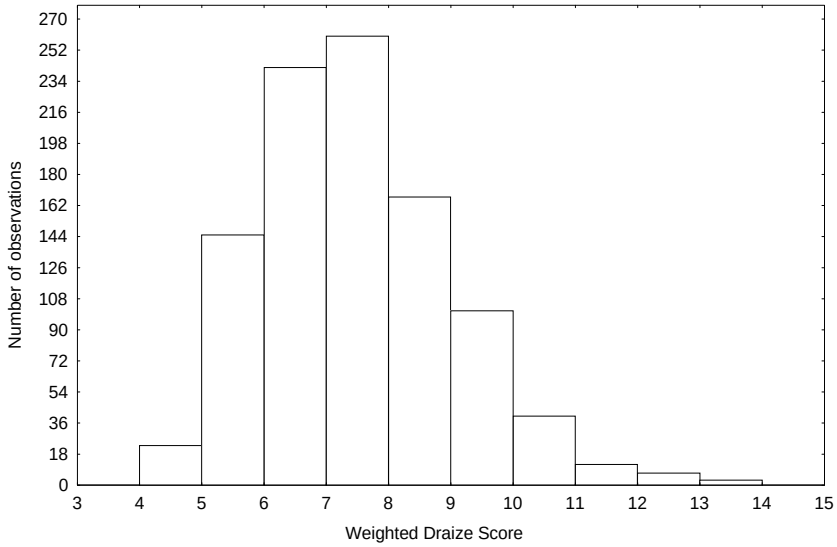


Figure 5.7 Simulation of the ability of the Draize eye test to 'predict itself'

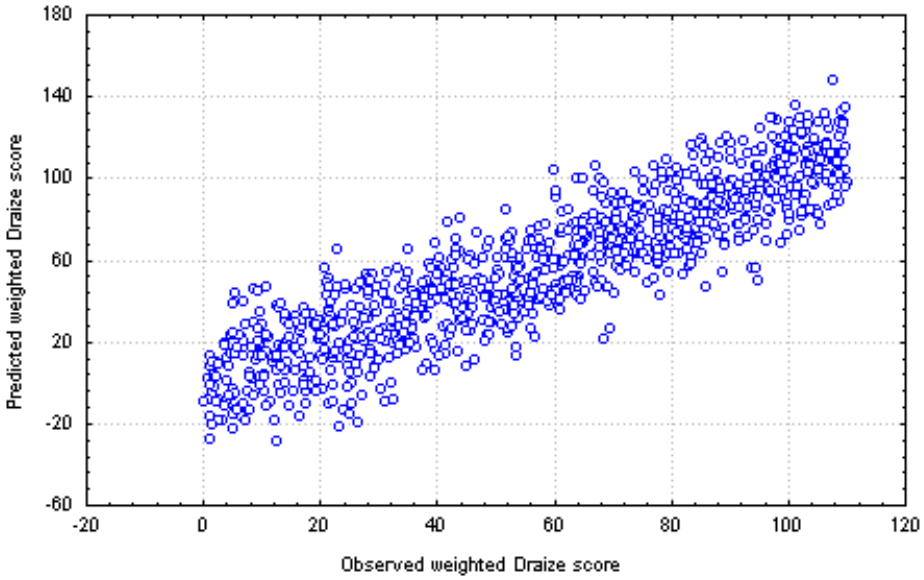
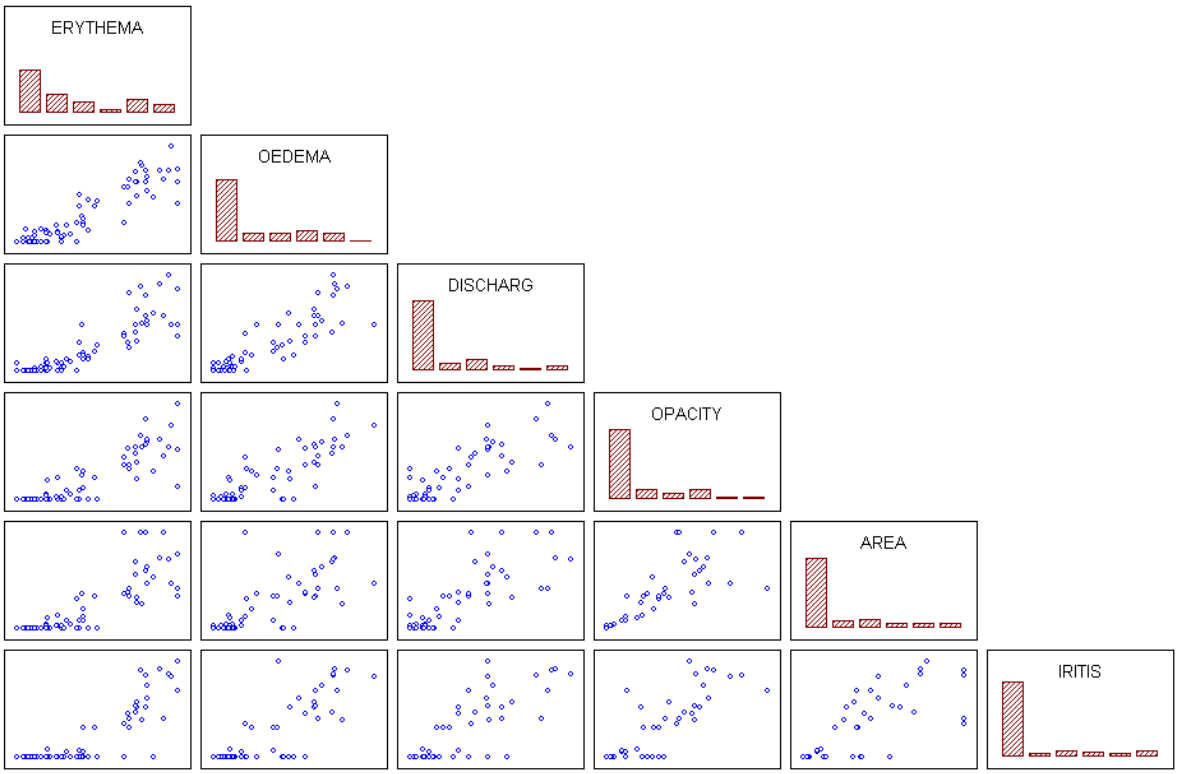


Figure 5.8 Matrix plot of 6 eye irritation effects determined in the Draize eye test

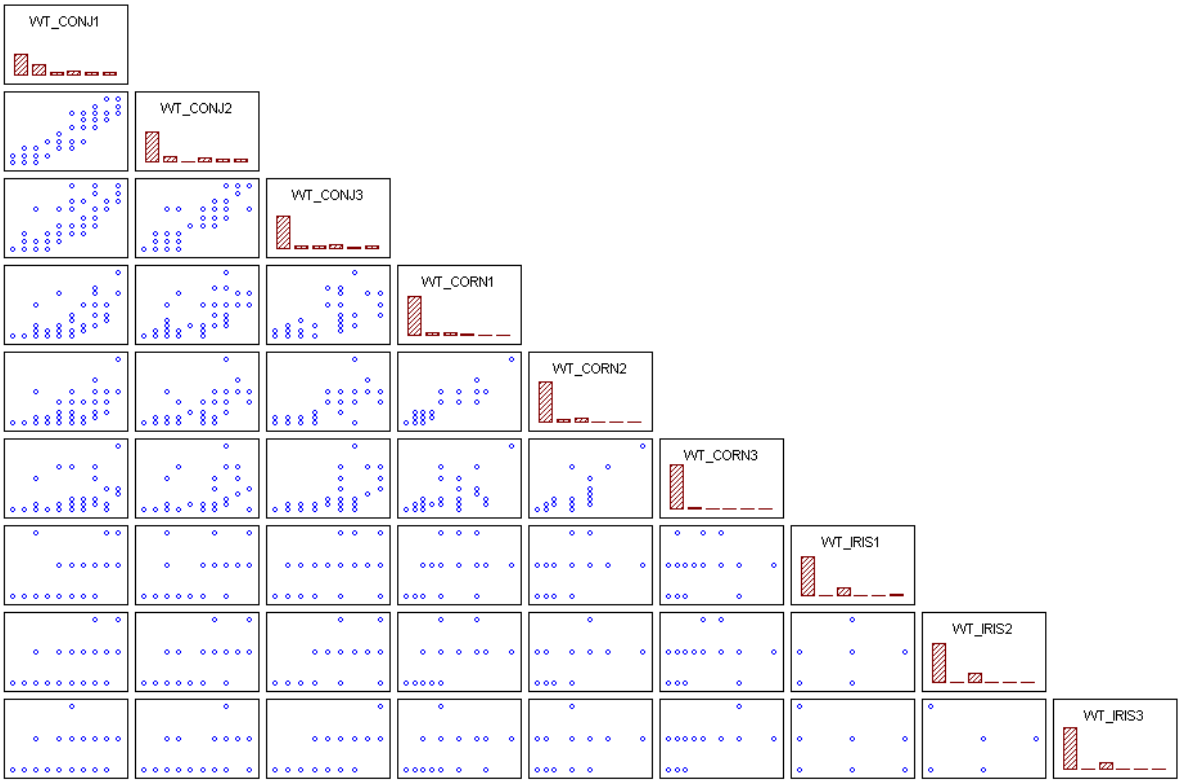


Footnote to Figure 5.8

ERYTHEMA = conjunctival erythema (redness); OEDEMA = conjunctival œdema (chemosis); DISCHARG = conjunctival discharge; OPACITY = corneal opacity; AREA = area corneal lesion; IRITIS = iris lesion.

Each tissue score is a mean value, taken over three time-points and all rabbits in the experimental group.

Figure 5.9 Matrix plot of 9 weighted tissue scores obtained in the Draize eye test

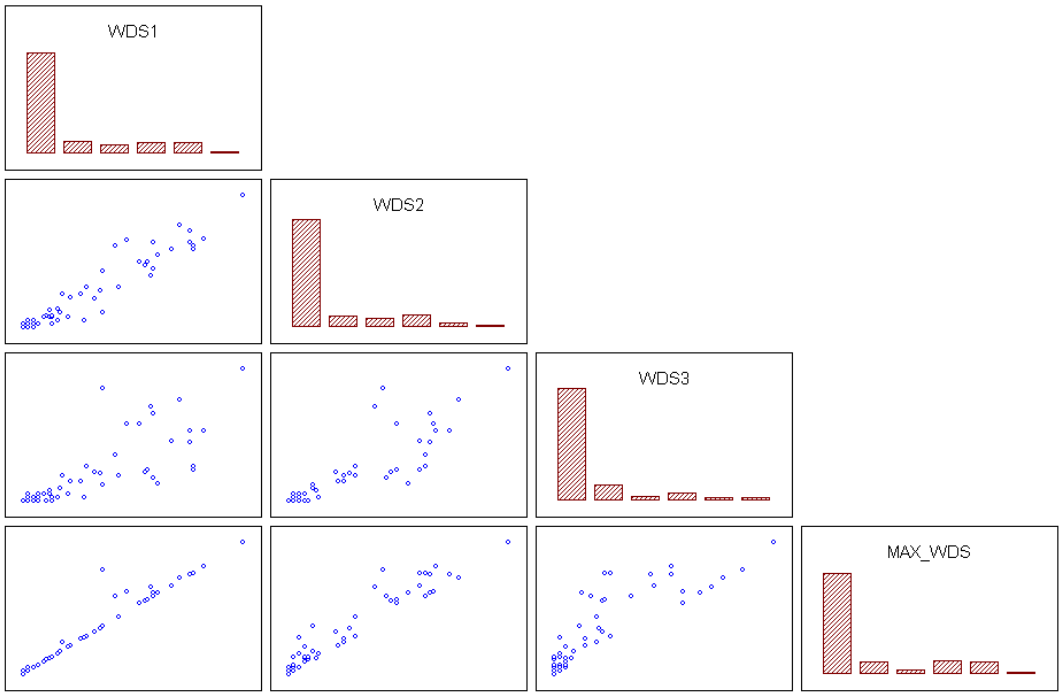


Footnote to Figure 5.9

WT_CONJ1 = weighted conjunctival score (2 x erythema x œdema x discharge) at 24h
WT_CONJ2 = weighted conjunctival score (2 x erythema x œdema x discharge) at 48h
WT_CONJ3 = weighted conjunctival score (2 x erythema x œdema x discharge) at 72h
WT_CORN1 = weighted cornea score (5 x opacity x area) at 24h
WT_CORN2 = weighted cornea score (5 x opacity x area) at 48h
WT_CORN3 = weighted cornea score (5 x opacity x area) at 72h
WT_IRIS1 = weighted iris score (5 x iritis) at 24h
WT_IRIS2 = weighted iris score (5 x iritis) at 48h
WT_IRIS3 = weighted iris score (5 x iritis) at 72h

Each tissue score is a mean value, taken over all rabbits in the experimental group.

Figure 5.10 Matrix plot of 4 weighted tissue scores obtained in the Draize eye test



Footnote to Figure 5.10

WDS1 = weighted Draize score at 24h
WDS2 = weighted Draize score at 48h
WDS3 = weighted Draize score at 72h
MAX_WDS = maximal value of the WDS obtained at any of the three time-points.

The weighted Draize score is the sum of weighted conjunctival, corneal and iritis scores (see chapter 2)

Each tissue score is a mean value, taken over all rabbits in the experimental group.

Table 5.1 ***In vitro*** **and** ***in vivo*** **classifications of skin corrosion potential for 60 chemicals**

CHEMICAL		PREDICTED CLASSIFICATION[a]					IN VIVO
No	Name	pH	TER	EPISKIN	Skin[2]	CORROSITEX	CLASS[b]
1	Hexanoic acid	C	C	C	C	NC	C
2	1,2-Diaminopropane	C	C	C	C	C	C
3	Carvacrol	NC	C	C	NC	NQ	C
4	Boron trifluoride dihydrate	C	C	C	C	C	C
5	Methacrolein	NC	NC	C	C	NC	C
6	Phenethyl bromide	NC	NC	NC	NC	NQ	NC
7	3,3'-Dithiodipropionic acid	NC	NC	NC	NC	NC	NC
8	Isopropanol	NC	NC	NC	NC	NQ	NC
9	2-Methoxyphenol (guaiacol)	NC	NC	C	NC	NQ	NC
10	2,4-Xylidine (2,4-dimethylaniline)	NC	C	C	NC	NC	NC
11	2-Phenylethanol (phenylethyl alcohol)	NC	NC	NC	NC	NQ	NC
12	Dodecanoic (lauric) acid	NC	NC	NC	NC	NQ	NC
13	3-Methoxypropylamine	C	C	C	C	C	C
14	Allyl bromide	NC	C	C	NC	NC	C
15	Dimethyldipropylenetriamine	NC	C	C	C	C	C
16	Methyl trimethylacetate	NC	NC	NC	NC	NQ	NC
17	Dimethylisopropylamine	C	C	C	C	C	C
18	Potassium hydroxide (10% aq.)	C	C	C	C	C	C
19	Tetrachloroethylene	NC	NC	NC	NC	NC	NC
20	Ferric [iron (III)] chloride	C	C	C	NC	C	C
21	Potassium hydroxide (5% aq.)	C	C	C	NC	C	NC
22	Butyl propionate	NC	NC	NC	NC	NQ	NC
23	2-*tert*-Butylphenol	NC	C	C	C	NQ	C
24	Sodium carbonate (50% aq.)	NC	C	NC	NC	C	NC
25	Sulphuric acid (10% wt.)	C	C	C	NC	C	C
26	Isostearic acid	NC	NC	NC	NC	NQ	NC
27	Methyl palmitate	NC	NC	NC	NC	NQ	NC
28	Phosphorus tribromide	C	C	C	C	C	C
29	65/35 Octanoic/decanoic acids	NC	C	C	NC	NC	C
30	4,4'-Methylene-bis-(2,6-ditertbutylphenol)	NC	NC	NC	NC	NQ	NC

Table 5.1 continued

CHEMICAL		PREDICTED CLASSIFICATION[a]					*IN VIVO*
No	Name	pH	TER	EPISKIN	Skin[2]	CORROSITEX	CLASS[b]
31	2-Bromobutane	NC	C	NC	NC	NQ	NC
32	Phosphorus pentachloride	C	C	C	C	C	C
33	4-(Methylthio)-benzaldehyde	NC	NC	NC	NC	NQ	NC
34	70/30 Oleine/octanoic acid	NC	NC	NC	NC	NC	NC
35	Hydrogenated tallow amine	NC	NC	NC	NC	NC	NC
36	2-Methylbutyric acid	NC	C	C	NC	C	C
37	Sodium undecylenate (34% aq.)	NC	C	C	NC	NC	NC
38	Tallow amine	NC	C	NC	NC	NC	C
39	2-Ethoxyethyl methacrylate	NC	NC	NC	NC	NQ	NC
40	Octanoic acid (caprylic acid)	NC	C	C	NC	NC	C
41	20/80 Coconut/palm soap	NC	NC	NC	NC	NC	NC
42	2-Mercaptoethanol, Na salt (46% aq.)	C	C	NC	NC	C	C
43	Hydrochloric acid (14.4% wt.)	C	C	C	NC	C	C
44	Benzyl acetone	NC	NC	NC	NC	NQ	NC
45	Heptylamine	C	C	NC	C	C	C
46	Cinnamaldehyde	NC	NC	NC	NC	NC	NC
47	60/40 Octanoic/decanoic acids	NC	C	C	NC	NC	C
48	Glycol bromoacetate (85%)	C	NC	C	C	C	C
49	Eugenol	NC	NC	NC	NC	NQ	NC
50	55/45 Octanoic/decanoic acids	NC	C	C	NC	NC	C
51	Methyl laurate	NC	NC	NC	NC	NQ	NC
52	Sodium bicarbonate	NC	C	NC	NC	NC	NC
53	Sulphamic acid	C	C	C	NC	C	NC
54	Sodium bisulphite	NC	C	NC	NC	NC	NC
55	1-(2-Aminoethyl)piperazine	C	C	NC	NC	C	C
56	1,9-Decadiene	NC	NC	NC	NC	NQ	NC
57	Phosphoric acid	C	C	C	NC	C	C
58	10-Undecenoic acid	NC	NC	C	NC	NC	NC
59	4-Amino-1,2,4-triazole	NC	NC	NC	NC	NQ	NC
60	Sodium lauryl sulphate (20% aq.)	NC	C	NC	NC	NQ	NC

Footnote to Table 5.1

[a]For the pH test, predicted classifications were taken from Worth *et al*. (1998); for the four *in vitro* tests, the most frequently occurring *in vitro* classification, made across laboratories and runs, was taken from Fentem *et al*. (1998); [b]Based on the application of European Union classification criteria to Draize test data.

C = corrosive; NC = non-corrosive; NQ = non-qualifying chemical; TER = transcutaneous electrical resistance assay.

Table 5.2 Predictive abilities (%) of the TER, EPISKIN, Skin^2 and CORROSITEX tests calculated by the standard bootstrap method

In vitro test	Cooper statistic	Bootstrap estimates[a]			Estimate from original data set[b]	Bias	Bias-corrected estimate
		Mean value	Lower limit	Upper limit			
TER	concordance	78	69	87	79	-1	80
	sensitivity	87	76	97	88	-1	89
	specificity	71	57	85	72	-1	73
	false positives	29	15	43	28	1	27
	false negatives	13	3	24	12	1	11
EPISKIN	concordance	81	73	90	81	0	81
	sensitivity	83	69	96	83	0	83
	specificity	80	69	91	80	0	80
	false positives	20	9	31	20	0	20
	false negatives	17	4	31	17	0	17
Skin^2	concordance	75	64	86	74	1	73
	sensitivity	45	27	63	43	2	41
	specificity	100	100	100	100	0	100
	false positives	0	0	0	0	0	0
	false negatives	55	37	73	57	-2	59
CORROSITEX	concordance	73	60	87	73	0	73
	sensitivity	71	55	88	71	0	71
	specificity	76	55	98	76	0	76
	false positives	24	2	45	24	0	24
	false negatives	29	12	45	29	0	45

Footnote to Table 5.2

[a]Each set of bootstrap estimates (mean, lower value, upper value) is calculated from a bootstrap distribution based on 1000 random samples taken from the original data set of 60 chemicals.

[b]The set of 60 test chemicals used in the ECVAM Skin Corrosivity Validation Study (Fentem *et al*., 1998).

Table 5.3 Predictive abilities (%) of the TER, EPISKIN, Skin^2 and CORROSITEX tests calculated by the simple and bias-corrected percentile bootstrap methods

In vitro test	Cooper statistic	Bootstrap estimates[a]						
		Simple percentile method			Bias-corrected percentile method			
		Median value	Lower limit[b]	Upper limit[c]	Lower percentile	Lower limit	Upper percentile	Upper limit
TER	concordance	78	68	86	0.044	69	0.987	88
	sensitivity	87	75	95	0.061	78	0.991	96
	specificity	72	56	83	0.033	57	0.981	85
	false positives	28	17	44	0.019	15	0.967	43
	false negatives	13	5	25	0.009	4	0.941	22
EPISKIN	concordance	81	72	89	0.020	71	0.969	89
	sensitivity	83	68	95	0.028	69	0.978	95
	specificity	80	68	91	0.027	68	0.977	91
	false positives	20	9	32	0.024	9	0.974	32
	false negatives	17	5	32	0.022	5	0.972	31
Skin^2	concordance	75	64	85	0.009	61	0.941	83
	sensitivity	46	27	63	0.007	22	0.930	59
	specificity	100	100	100	d	d	d	d
	false positives	0	0	0	d	d	d	d
	false negatives	54	37	73	0.073	42	0.993	80
CORROSITEX	concordance	73	60	86	0.020	59	0.969	85
	sensitivity	72	53	86	0.016	52	0.962	85
	specificity	76	55	97	0.020	53	0.970	97
	false positives	24	3	45	0.030	3	0.980	47
	false negatives	28	14	47	0.038	15	0.984	48

Footnote to Table 5.3

[a]Each set of bootstrap estimates (median, lower value, upper value) is calculated from a bootstrap distribution based on 1000 random samples taken from the original data set of 60 chemicals.
[b]Based on the 2.5th percentile (0.025) of the bootstrap distribution.
[c]Based on the 97.5th percentile (0.975) of the bootstrap distribution.
[d]The bias-corrected percentile method does not work because the proportion of bootstrap estimates exceeding the specificity or false positive rate calculated from the original sample is 0. The inverse cumulative probability function is not defined for a probability of 0.

Table 5.4 Predictive abilities (%) of four two-step sequences, calculated by the standard bootstrap method

Sequence of *in vitro* tests	Cooper statistic	Bootstrap estimates[a]			Estimate from original data set[b]	Bias	Bias-corrected estimate
		Mean value	Lower limit	Upper limit			
Sequence 1[c]	concordance	83	74	93	83	0	83
	sensitivity	96	89	100	96	0	96
	specificity	73	58	88	73	0	73
	false positives	27	12	42	27	0	27
	false negatives	4	0	11	4	0	4
Sequence 2[d]	concordance	88	80	96	88	0	88
	sensitivity	96	89	100	96	0	96
	specificity	82	69	94	82	0	82
	false positives	18	6	31	18	0	18
	false negatives	4	0	11	4	0	4
Sequence 3[e]	concordance	83	74	93	83	0	83
	sensitivity	70	53	88	70	0	70
	specificity	94	86	100	94	0	94
	false positives	6	0	14	6	0	6
	false negatives	30	12	47	30	0	30
Sequence 4[f]	concordance	74	60	88	74	0	74
	sensitivity	72	54	90	72	0	72
	specificity	78	54	100	78	0	78
	false positives	22	0	46	21	1	20
	false negatives	28	10	46	28	0	28

Footnote to Table 5.4

[a]Each set of bootstrap estimates (mean, lower value, upper value) is calculated from a bootstrap distribution based on 1000 random samples taken from the original data set of 60 chemicals.
[b]The set of 60 test chemicals used in the ECVAM Skin Corrosivity Validation Study.
[c]Based on the sequential application of the pH test and the TER assay.
[d]Based on the sequential application of the pH test and the EPISKIN assay.
[e]Based on the sequential application of the pH test and the Skin^2 assay.
[f] Based on the sequential application of the pH test and the CORROSITEX assay.

Table 5.5 Comparison between the stand-alone and sequential use of each alternative test

In vitro test	Cooper statistic	Bootstrap mean estimates		Significant difference[b] ?
		Stand-alone test	Sequence[a]	
TER	concordance	78	83	yes
	sensitivity	87	96	yes
	specificity	71	73	yes
	false positives	29	27	yes
	false negatives	13	4	yes
EPISKIN	concordance	81	88	yes
	sensitivity	83	96	yes
	specificity	80	82	yes
	false positives	20	18	yes
	false negatives	17	4	yes
Skin[2]	concordance	75	83	yes
	sensitivity	45	70	yes
	specificity	100	94	yes
	false positives	0	6	yes
	false negatives	55	30	yes
CORROSITEX	concordance	73	74	no
	sensitivity	71	72	no
	specificity	76	78	yes
	false positives	24	22	yes
	false negatives	29	28	no

Footnote to Table 5.5

[a]Statistics refer to the use of each *in vitro* test in sequence with the pH test
[b]As judged by a t-test for independent samples ($p<0.01$), applied to the appropriate bootstrap distributions.

Table 5.6 Skin irritation data for 143 chemicals

Chemical No	Name	PII[a]	PII[b] mean	median	SD	CV	Erythema[b] mean	SD	Œdema[b] mean	SD
1	2-Methylbutyric acid	>4	5.0	5.0	0.2	4.5	2.3	0.1	2.7	0.2
2	10-Undecenoic Acid	2.4	2.4	2.4	0.2	6.8	1.8	0.1	0.6	0.1
3	Ethyltriglycol methacrylate	0.2	0.2	0.2	0.1	61.1	0.2	0.1	0.0	0.0
4	Ethylthioethyl methacrylate	0.6	0.6	0.6	0.2	28.7	0.6	0.2	0.0	0.0
5	2-Ethoxyethyl methacrylate	1.6	1.6	1.6	0.2	10.9	1.6	0.2	0.0	0.0
6	2-Methoxyethyl acrylate	NG	7.1	7.2	0.7	9.7	3.6	0.3	3.5	0.4
7	Benzyl alcohol [1]	1.6	1.6	1.6	0.3	17.5	1.3	0.2	0.2	0.1
7	Benzyl alcohol [2]	1.8	1.8	1.8	0.2	12.1	1.4	0.1	0.4	0.1
8a	dl-Citronellol [1]	4.2	4.2	4.2	0.1	3.4	2.0	0.0	2.2	0.1
8b	dl-Citronellol [2]	4.0	4.0	4.0	0.2	5.9	2.0	0.0	2.0	0.2
8c	dl-Citronellol [3]	3.7	3.3	3.3	0.2	6.4	2.2	0.1	1.2	0.2
9	Decanol	3.3	3.3	3.3	0.2	6.0	2.2	0.1	1.2	0.1
10a	Dihydromercenol [1]	3.7	1.6	1.6	0.2	14.4	1.3	0.2	0.2	0.1
10b	Dihydromercenol [2]	2.0	2.0	2.0	0.3	14.8	1.5	0.1	0.5	0.2
11	2,6-Dimethyl-4-heptanol	0.0	0.0	0.0	0.0	ND	0.0	0.0	0.0	0.0
12a	Dipropylene glycol [1]	0.3	0.3	0.3	0.2	47.3	0.3	0.2	0.0	0.0
12b	Dipropylene glycol [2]	0.0	0.0	0.0	0.0	ND	0.0	0.0	0.0	0.0
13a	Geraniol [1]	3.7	3.7	3.7	0.2	4.3	2.0	0.0	1.7	0.2
13b	Geraniol [2]	3.3	3.3	3.3	0.3	9.1	2.1	0.1	1.2	0.2
13c	Geraniol [3]	2.9	2.9	2.9	0.1	4.8	1.9	0.1	1.0	0.1
14	Geranyl dihydrolinalol	2.3	2.3	2.3	0.2	7.4	1.4	0.1	0.8	0.1
15	Geranyl linalol	4.3	4.3	4.3	0.2	3.6	2.2	0.1	2.1	0.1
16	α-Ionol	1.3	1.3	1.3	0.2	13.5	0.9	0.1	0.4	0.1
17	β-Ionol	1.9	1.9	1.9	0.2	9.0	1.2	0.1	0.7	0.1
18a	Linalol [1]	3.3	3.3	3.3	0.2	6.6	1.9	0.1	1.9	0.1
18b	Linalol [2]	3.4	3.4	3.4	0.1	4.2	2.0	0.0	1.4	0.1
18c	Linalol [3]	2.1	2.1	2.1	0.2	11.5	1.7	0.1	0.4	0.1
19	p-Mentha-1,8-dien-7-ol	3.3	3.3	3.3	0.2	5.8	2.0	0.0	1.3	0.2
20	2-Methyl-4-phenyl-2-butanol	1.6	1.6	1.6	0.2	12.4	1.3	0.1	0.3	0.1
21	Phenylethyl alcohol [1]	2.2	2.2	2.2	0.3	12.1	1.6	0.2	0.7	0.2
21b	Phenylethyl alcohol [2]	0.9	0.9	0.9	0.2	20.3	0.7	0.1	0.2	0.1
22	Isopropanol	0.8	0.8	0.8	0.3	33.8	0.8	0.3	0.0	0.0
23	Isostearyl alcohol	2.6	2.6	2.6	0.4	16.4	1.6	0.2	1.0	0.3
24a	α-Terpineol [1]	4.4	4.4	4.4	0.3	6.2	2.0	0.2	2.4	0.2
24b	α-Terpineol [2]	4.8	4.7	4.8	0.2	4.3	2.2	0.1	2.6	0.2
24c	α-Terpineol [3]	4.0	4.0	4.0	0.3	8.1	1.9	0.1	2.1	0.3
25	p-*tert*-Butyl dihydrocinnamaldehyde	2.4	2.4	2.4	0.4	15.1	1.6	0.2	0.8	0.2
26	Isobutyraldehyde	0.1	0.1	0.1	0.1	49.4	0.1	0.1	0.0	0.0
27	Cinnamaldehyde	3.7	3.7	3.7	0.2	4.2	2.0	0.0	1.7	0.1
28	Citrathal	3.6	3.6	3.6	0.2	6.3	2.0	0.0	1.7	0.2
29a	Cyclamen aldehyde [1]	5.1	5.1	5.1	0.2	3.6	2.2	0.1	2.9	0.1
29b	Cyclamen aldehyde [2]	4.2	4.2	4.2	0.2	3.9	2.0	0.0	2.2	0.2
29c	Cyclamen aldehyde [3]	4.8	4.6	4.6	0.2	3.7	2.0	0.0	2.6	0.2
29d	Cyclamen aldehyde [4]	3.4	3.4	3.4	0.1	4.0	2.0	0.0	1.4	0.1
30	2,4-Decadienal	4.8	4.8	4.8	0.2	5.0	2.3	0.1	2.5	0.2
31	2,4-Dimethyl-3-cyclohexene-1-carboxaldehyde	3.2	3.2	3.2	0.1	3.3	2.0	0.0	1.2	0.1

Table 5.6 continued

Chemical No	Name	PII[a]	PII[b] mean	median	SD	CV	Erythema[b] mean	SD	Œdema[b] mean	SD
32	3,7-Dimethyl-2,6-nondien-1-al (Ethyl citral)	3.8	3.8	3.8	0.1	2.9	2.0	0.0	1.8	0.1
33	2,4-Dimethyltetrahydro benzaldehyde	2.8	2.8	2.8	0.2	8.2	1.9	0.1	0.9	0.2
34	2-Ethylhexanal	3.9	3.9	3.9	0.1	3.6	2.0	0.0	1.9	0.1
35	Heptanal	5.0	5.0	5.0	0.2	4.9	3.0	0.2	2.0	0.1
36	2,4-Hexadienal	7.1	7.1	7.1	0.2	2.6	3.6	0.1	3.5	0.1
37a	α-Hexyl cinnamic aldehyde [1]	4.0	4.0	4.0	0.2	5.4	2.0	0.0	2.0	0.2
37b	α-Hexyl cinnamic aldehyde [2]	4.0	4.0	4.0	0.1	2.8	2.0	0.0	2.0	0.0
37c	α-Hexyl cinnamic aldehyde [3]	2.6	2.6	2.6	0.2	8.5	1.8	0.1	0.8	0.1
38	Hydroxycitronellal [1]	1.1	1.1	1.1	0.2	16.2	0.9	0.1	0.2	0.1
38	Hydroxycitronellal [2]	0.9	0.9	0.9	0.2	26.8	0.8	0.2	0.1	0.1
39a	Lilestralis (lilial) [1]	4.6	4.6	4.6	0.3	6.0	2.0	0.2	2.6	0.2
39b	Lilestralis (lilial) [2]	3.6	3.6	3.6	0.2	6.1	1.9	0.1	1.7	0.2
40	3-Methylbutyraldehyde	2.8	2.8	2.9	0.3	11.6	1.6	0.2	1.3	0.2
41	2,5-Methylene-6-propyl-3-cyclo-hexen-carbaldehyde	2.4	2.4	2.4	0.2	9.6	1.8	0.1	0.6	0.2
42	Nonanal	3.5	3.4	3.5	0.2	4.7	2.0	0.0	1.4	0.1
43	2-Phenylpropionaldehyde	2.9	2.9	2.9	0.2	5.9	2.0	0.0	0.9	0.2
44	p-Isopropylphenylacetaldehyde	2.3	2.3	2.3	0.4	19.0	1.3	0.2	0.9	0.3
45	Salicylaldehyde	2.5	2.5	2.5	0.3	11.9	1.5	0.1	1.0	0.2
46	Tetrahydrogeranial	2.6	2.6	2.6	0.2	6.2	1.9	0.1	0.7	0.1
47	4-Tricyclo-decylindene-8-butanal	3.3	3.3	3.3	0.3	9.9	2.1	0.1	1.2	0.3
48	Methacrolein	4.1	4.1	4.1	0.2	4.5	2.0	0.0	2.1	0.2
49	Erucamide	0.0	0.0	0.0	0.0	ND	0.0	0.0	0.0	0.0
50	N,N-Diethylaminopropylamine	NG	c	c	c	c	c	c	c	c
51	N,N-Dimethylbenzylamine	NG	c	c	c	c	c	c	c	c
52	Dimethylbutylamine	5.1	5.1	5.1	0.1	1.4	4.0	0.0	1.1	0.1
53	Dimethylisopropylamine	5.6	5.6	5.6	0.1	2.5	4.0	0.0	1.6	0.1
54	Dimethyldipropylenetriamine	NG	c	c	c	c	c	c	c	c
55	Heptylamine	6.7	6.7	6.7	0.3	4.4	4.0	0.0	2.7	0.3
56	3-Methoxypropylamine	6.7	6.7	6.7	0.3	4.1	4.0	0.0	2.7	0.3
57	Oleyl propylene diamine dioleate	3.7	3.7	3.7	0.7	20.0	2.1	0.3	1.5	0.5
58	2,4-Xylidine	1.4	1.4	1.4	0.2	11.6	1.0	0.0	0.4	0.2
59	Allyl bromide	7.2	7.2	7.2	0.3	3.9	4.0	0.0	3.2	0.3
60	2-Bromobutane	2.4	2.5	2.4	0.4	14.6	1.8	0.1	0.7	0.3
61	1-Bromo-4-chlorobutane	0.0	0.0	0.0	0.0	ND	0.0	0.0	0.0	0.0
62	1-Bromo-2-chloroethane	2.3	2.3	2.3	0.4	15.7	1.8	0.1	0.5	0.3
63	1-Bromo-4-fluorobenzene	0.3	0.3	0.3	0.2	47.1	0.3	0.2	0.0	0.0
64	1-Bromohexane	4.0	4.0	4.0	0.4	9.7	2.4	0.2	2.4	0.2
65	1-Bromopentane	4.4	4.5	4.4	0.8	18.0	2.7	0.3	1.8	0.6
66	2-Bromopropane	1.4	1.4	1.4	0.2	15.8	1.4	0.2	0.0	0.0
67	1,6-Dibromohexane	0.9	0.9	0.9	0.2	27.5	0.9	0.2	0.0	0.0
68	1,3-Dibromopropane	1.9	1.9	1.9	0.1	5.5	1.9	0.1	0.0	0.0
69	Phenethyl bromide	0.0	0.0	0.0	0.0	ND	0.0	0.0	0.0	0.0
70	Dichloromethane	5.7	5.7	5.7	0.2	2.7	4.0	0.0	1.7	0.2
71	Tetrachloroethylene	5.7	5.7	5.7	0.2	3.9	4.0	0.0	1.7	0.2
72	1,1,1-Trichloroethane	5.2	5.2	5.2	0.4	8.3	3.7	0.3	1.5	0.2
73	Trichloroethylene	5.4	5.4	5.4	0.2	3.1	4.0	0.0	1.4	0.2

Table 5.6 continued

Chemical No	Name	PII[a]	PII[b] mean	median	SD	CV	Erythema[b] mean	SD	Œdema[b] mean	SD
74	Allyl heptanoate	2.1	2.1	2.1	0.2	7.7	1.7	0.1	0.4	0.1
75	Allyl phenoxyacetate	0.4	0.4	0.4	0.1	28.1	0.3	0.1	0.1	0.1
76a	Benzyl acetate [1]	1.6	1.6	1.6	0.4	26.6	1.0	0.3	0.6	0.2
76b	Benzyl acetate [2]	0.8	0.8	0.8	0.2	23.9	0.8	0.2	0.0	0.0
77a	Benzyl benzoate [1]	0.0	0.0	0.0	0.0	ND	0.0	0.0	0.0	0.0
77b	Benzyl benzoate [2]	1.6	1.1	1.1	0.3	29.1	0.9	0.2	0.2	0.1
78a	Benzyl salicylate [1]	0.3	0.3	0.3	0.2	46.7	0.3	0.2	0.0	0.0
78b	Benzyl salicylate [2]	0.8	0.7	0.8	0.2	23.6	0.6	0.1	0.2	0.1
79a	Isobornyl acetate [1]	3.9	3.9	3.9	0.1	0.0	1.9	0.1	2.0	0.0
79b	Isobornyl acetate [2]	3.5	3.5	3.5	0.2	5.3	1.9	0.1	1.6	0.1
80	Butyl propionate	1.1	1.1	1.1	0.1	13.3	1.1	0.1	0.0	0.0
81a	Diethyl phthalate [1]	0.0	0.0	0.0	0.0	ND	0.0	0.0	0.0	0.0
81b	Diethyl phthalate [2]	0.2	0.2	0.2	0.1	60.7	0.2	0.1	0.0	0.0
82	2-Ethylhexylcocoate	1.7	1.7	1.7	0.2	12.8	0.9	0.1	0.8	0.1
83	2-Ethylhexylpalmitate	0.6	0.6	0.6	0.2	30.7	0.6	0.2	0.0	0.0
84	Ethyl tiglate	1.2	1.2	1.1	0.2	16.7	1.0	0.1	0.1	0.1
85	Ethyl trimethyl acetate	0.5	0.4	0.4	0.1	25.2	0.4	0.1	0.0	0.0
86	Glycolbromoacetate	7.7	7.7	7.7	0.3	ND	4.0	0.0	3.7	0.3
87	Heptyl butyrate	1.8	1.7	1.8	0.2	12.4	1.5	0.2	0.2	0.1
88a	Hexyl salicylate [1]	3.4	3.4	3.4	0.2	5.0	2.0	0.0	1.4	0.2
88b	Hexyl salicylate [2]	3.7	3.7	3.7	0.2	4.9	2.0	0.0	2.0	0.0
88c	Hexyl salicylate [3]	4.2	4.2	4.2	0.2	4.6	2.0	0.0	2.2	0.2
88d	Hexyl salicylate [4]	3.3	3.3	3.3	0.1	4.0	2.0	0.0	1.3	0.1
89a	Linalyl acetate [1]	3.7	3.7	3.7	0.2	4.2	1.9	0.1	1.8	0.1
89b	Linalyl acetate [2]	2.9	2.9	2.9	0.2	7.9	1.9	0.1	1.0	0.2
90	Methyl caproate	2.8	2.8	2.8	0.2	7.5	1.7	0.2	1.1	0.1
91	Methyl laurate	3.9	3.9	3.9	0.1	2.5	2.0	0.0	1.9	0.1
92	Methyl linoleate	3.1	3.1	3.1	0.6	18.4	1.9	0.3	1.2	0.3
93	Methyl 2-methylbutyrate	0.7	0.7	0.7	0.1	18.4	0.6	0.1	0.0	0.0
94	Methyl palmitate	4.6	4.6	4.6	0.5	10.2	2.6	0.2	2.0	0.3
95	Methyl stearate	2.1	2.1	2.1	0.5	25.2	1.5	0.2	0.7	0.3
96	Methyl trimethyl acetate	0.0	0.0	0.0	0.0	ND	0.0	0.0	0.0	0.0
97	Isopropyl myristate	1.2	1.2	1.2	0.1	11.6	1.0	0.0	0.2	0.1
98	Isopropyl palmitate	1.4	1.4	1.4	0.3	18.9	1.2	0.1	0.2	0.1
99	Isopropyl isostearate	0.1	0.1	0.1	0.1	90.3	0.1	0.1	0.0	0.0
100a	α-Terpinyl acetate [1]	3.6	3.6	3.6	0.2	18.7	1.9	0.1	1.7	0.2
100b	α-Terpinyl acetate [2]	4.3	4.3	4.3	0.1	3.1	2.0	0.0	2.3	0.1
100c	α-Terpinyl acetate [3]	2.8	2.7	2.8	0.3	12.0	1.7	0.1	1.0	0.2
101	Eucalyptol	2.3	2.2	2.3	0.1	4.3	1.9	0.1	0.4	0.1
102	Octanoic acid	4.4	4.5	4.4	0.4	8.9	3.0	0.5	1.4	0.3
103	Isostearic acid	4.3	4.3	4.3	0.4	10.4	1.9	0.2	2.4	0.3
104	3-Chloro-4-fluoronitrobenzene	1.7	1.7	1.7	0.2	10.7	1.3	0.1	0.3	0.1
105	Fluorobenzene	0.1	0.1	0.1	0.1	97.4	0.1	0.1	0.0	0.0
106	2-Fluorotoluene	0.1	0.1	0.1	0.1	95.3	0.1	0.1	0.0	0.0
107	*cis*-Cyclooctene	1.9	2.1	2.1	0.4	19.4	1.7	0.3	0.4	0.2
108	1,9-Decadiene	3.0	3.0	3.0	0.0	0.0	2.0	0.0	1.0	0.0
109	1,5-Hexadiene	0.0	0.0	0.0	0.0	ND	0.0	0.0	0.0	0.0
110	1,13-Tetradecadiene	1.7	1.6	1.6	0.3	18.4	1.3	0.2	0.2	0.1
111	Benzyl acetone	1.2	1.0	1.0	0.4	36.5	0.8	0.3	0.2	0.1

Table 5.6 continued

Chemical No	Name	PII[a]	PII[b] mean	median	SD	CV	Erythema[b] mean	SD	Œdema[b] mean	SD
112	Diacetyl	0.6	0.6	0.6	0.1	19.4	0.6	0.1	0.0	0.0
113	*cis*-Jasmone	2.6	2.6	2.6	0.3	11.2	1.7	0.1	1.7	0.1
114	Isolongifolene ketone	3.0	3.0	3.0	0.5	15.5	1.9	0.1	1.1	0.4
115	Methyl lavender ketone (1-hydroxy-3-decanone)	3.8	3.8	3.8	0.2	6.1	2.2	0.1	1.6	0.2
116	2,3-Dichloropropionitrile	2.0	2.0	2.0	0.0	0.0	2.0	0.0	0.0	0.0
117	3-Diethylaminopropionitrile	0.0	0.0	0.0	0.0	ND	0.0	0.0	0.0	0.0
118	Carvacrol	>4	5.7	5.7	0.3	5.5	4.0	0.0	1.7	0.3
119	2-*tert*-Butyl phenol	5.7	5.7	5.7	0.1	2.0	4.0	0.0	1.7	0.1
120	Eugenol	2.9	2.9	2.9	0.1	4.2	1.9	0.1	1.0	0.1
121	2-Methoxyphenol (guaiacol)	2.4	2.4	2.4	0.2	8.6	1.8	0.1	0.6	0.1
122	4,4'-Methylene bis (2,6-ditert-butyl phenol)	0.0	0.0	0.0	0.0	ND	0.0	0.0	0.0	0.0
123	Dimethyl disulphide	3.0	3.0	3.0	0.3	8.4	1.8	0.1	1.2	0.2
124	Dipropyl disulphide	2.6	2.6	2.6	0.2	8.8	2.6	0.2	0.0	0.0
125	3-Mercapto-1-propanol	1.1	1.1	1.1	0.2	19.4	0.8	0.1	0.3	0.1
126	4-(Methylthio)benzaldehyde	0.9	0.9	0.9	0.2	20.8	0.9	0.2	0.0	0.0
127	Glycerol tri-isostearate	0.7	0.7	0.7	0.2	23.6	0.7	0.2	0.0	0.0
128	6-Butyl-2,4-dimethyldihydropyrane	2.0	2.0	2.0	0.2	8.9	1.7	0.1	0.3	0.1
129	Decylidene methyl anthranilate	2.1	2.1	2.1	0.2	8.7	1.5	0.1	0.5	0.1
130	2,6-Dimethyl-2,4,6-octatriene	3.0	3.0	3.0	0.2	6.0	2.0	0.0	1.0	0.2
131	1-Formyl-1-methyl-4(4-methyl-3-penten-1-yl)-3-cyclohexane	3.3	3.3	3.3	0.2	5.5	2.0	0.0	1.3	0.2
132a	d-Limonene [1]	3.6	3.6	3.6	0.2	4.7	2.0	0.0	1.6	0.2
132b	d-Limonene [2]	3.3	3.2	3.3	0.2	5.4	1.9	0.1	1.3	0.2
133	Linalol oxide	2.6	2.6	2.6	0.1	3.2	2.0	0.0	0.6	0.1
134a	Cinnamyl alcohol [1]	0.0	0.0	0.0	0.0	ND	0.0	0.0	0.0	0.0
134b	Cinnamyl alcohol [2]	0.5	0.5	0.5	0.2	34.6	0.5	0.1	0.1	0.1
135	2,4-Dinitromethylaniline	0.0	0.0	0.0	0.0	ND	0.0	0.0	0.0	0.0
136a	Dimethylbenzylcarbinyl acetate [1]	1.2	1.2	1.2	0.1	11.7	1.0	0.0	0.2	0.1
136b	Dimethylbenzylcarbinyl acetate [2]	1.4	1.4	1.4	0.3	21.5	0.9	0.2	0.5	0.2
137	Dodecanoic (lauric) acid	0.4	0.4	0.4	0.2	38.2	0.4	0.2	0.0	0.0
138	2-Chloronitrobenzene	0.0	0.0	0.0	0.0	ND	0.0	0.0	0.0	0.0
139	3-Methylphenol	0.0	0.0	0.0	0.0	ND	0.0	0.0	0.0	0.0
140	3,3'-Dithiopropionic acid	0.0	0.0	0.0	0.0	ND	0.0	0.0	0.0	0.0
141	Glycerol triundecanoate	0.0	0.0	0.0	0.0	ND	0.0	0.0	0.0	0.0
142	4-Amino-1,2,4-triazole	0.0	0.0	0.0	0.0	ND	0.0	0.0	0.0	0.0
143	Tonalid	0.0	0.0	0.0	0.0	ND	0.0	0.0	0.0	0.0

Footnote to Table 5.6

[a]The conventional PII, calculated from the original data (ECETOC, 1995).
[b]Bootstrap estimates, obtained by resampling the original data.

CV = coefficient of variation; SD = standard deviation; ND = not defined (denominator of zero); NG = not given in the ECETOC report (1995) since tissue scores were available for fewer than three animals; PII = primary irritation index.

Table 5.7 Eye irritation data for 92 chemicals

Chemical No	Name	MMAS[a]	WDS[b] mean	median	max	SD	CV	Mean Erythema	Mean Œdema	Mean Discharge	Mean Opacity	Mean Area	Mean Iritis
1	Methyl trimethyl acetate	2.7	1.9	1.9	3.1	0.4	22.2	0.6	0.3	0.1	0.0	0.0	0.0
2	Ethyl trimethyl acetate	3.8	2.3	2.3	4.5	0.6	26.5	0.6	0.4	0.1	0.2	0.2	0.0
3	Butyl acetate	7.5	3.2	3.1	8.1	0.7	21.0	1.0	0.3	0.3	0.2	0.2	0.0
4	Cellosolve acetate	15.0	6.1	5.7	21.8	2.2	35.8	1.2	0.7	0.5	0.4	0.5	0.0
5	Ethyl acetate	15.0	5.6	5.4	15.6	1.5	26.7	1.3	0.6	0.3	0.4	0.3	0.1
6	Ethyl-2-methyl acetoacetate	18.0	13.3	13.4	18.4	2.1	15.4	1.2	0.9	1.3	0.9	1.4	0.0
7	Methyl cyanoacetate	27.7	20.6	20.2	33.1	2.6	12.7	3.0	2.1	1.0	0.4	1.3	1.0
8	Methyl acetate	39.5	25.1	24.6	44.2	5.7	22.7	2.2	2.1	1.8	1.3	1.3	0.8
9	Isopropyl isostearate	1.3	0.4	0.4	1.3	0.3	64.1	0.2	0.0	0.0	0.0	0.0	0.0
10	2,2-Dimethylbutanoic acid	44.7	40.0	39.9	49.5	3.2	8.0	2.0	0.7	1.1	1.5	4.0	0.5
11	2-Ethoxyethyl methacrylate	0.0	0.0	0.0	0.0	0.0	ND	0.0	0.0	0.0	0.0	0.0	0.0
12	Ethyl thioethyl methacrylate	0.0	0.0	0.0	0.0	0.0	ND	0.0	0.0	0.0	0.0	0.0	0.0
13	Ethyl triglycol methacrylate	0.0	0.0	0.0	0.0	0.0	ND	0.0	0.0	0.0	0.0	0.0	0.0
14	Nonyl acrylate	0.0	0.0	0.0	0.0	0.0	ND	0.0	0.0	0.0	0.0	0.0	0.0
15	Nonyl methacrylate	0.0	0.0	0.0	0.0	0.0	ND	0.0	0.0	0.0	0.0	0.0	0.0
16	Iso-octyl acrylate	0.7	0.2	0.2	1.1	0.2	96.3	0.1	0.0	0.0	0.0	0.0	0.0
17	Heptyl methacrylate	1.3	0.4	0.4	1.6	0.3	61.6	0.2	0.0	0.0	0.0	0.0	0.0
18	Trifluoroethyl methacrylate	1.3	0.4	0.4	1.8	0.3	63.3	0.2	0.0	0.0	0.0	0.0	0.0
19	Allyl methacrylate	5.8	3.0	3.0	6.2	0.7	24.4	0.8	0.3	0.2	0.0	0.1	0.1
20	Glycidyl methacrylate	28.0	14.4	13.9	31.5	5.4	37.8	1.5	1.2	0.6	0.8	1.4	0.4
21	2-Methoxyethyl acrylate	45.0	25.5	25.0	49.0	6.5	25.5	2.3	2.7	1.4	1.9	1.0	0.6
22	2,6-Dichlorobenzoyl chloride	23.8	22.3	22.4	26.3	1.4	6.0	2.6	1.6	1.7	1.0	2.2	0.0
23	Isostearyl alcohol	0.0	0.0	0.0	0.0	0.0	ND	0.0	0.0	0.0	0.0	0.0	0.0
24	3-Methoxy-1,2-propanediol	0.0	0.0	0.0	0.0	0.0	ND	0.0	0.0	0.0	0.0	0.0	0.0
25	2,4-Pentanediol	1.3	0.9	0.9	1.8	0.3	36.9	0.4	0.0	0.0	0.0	0.0	0.0
26	Propylene glycol	1.3	0.6	0.6	1.4	0.2	36.5	0.3	0.0	0.0	0.0	0.0	0.0
27	Glycerol	1.7	0.6	0.6	1.3	0.2	38.4	0.3	0.0	0.0	0.0	0.0	0.0
28	Isomyristyl alcohol	4.0	2.0	2.0	5.1	0.9	42.9	0.5	0.5	0.1	0.0	0.0	0.0

Table 5.7 Eye irritation data for 92 chemicals

Chemical No	Name	MMAS[a]	WDS[b] mean	median	max	SD	CV	Mean Erythema	Mean Œdema	Mean Discharge	Mean Opacity	Mean Area	Mean Iritis
29	2,2-Dimethyl-3-pentanol	8.3	4.5	4.2	11.9	1.8	39.2	0.8	0.3	0.0	0.8	0.6	0.0
30	2-Methyl-1-pentanol	13.0	10.7	10.4	20.1	2.7	25.3	1.1	0.7	0.0	1.1	1.2	0.0
31	Cyclopentanol	21.7	18.7	18.6	24.8	1.9	10.0	2.0	1.9	1.0	1.2	1.4	0.0
32	Ethanol	24.0	18.0	17.7	40.7	5.3	29.4	2.1	1.3	0.7	1.1	1.3	0.4
33	Propan-2-ol (isopropanol)	28.3	17.6	17.0	35.3	4.6	26.1	2.3	1.6	0.8	0.7	1.1	0.8
34	Octanol	30.5	38.0	38.0	50.7	5.1	13.3	2.2	2.5	1.1	1.7	2.8	0.7
35	Furfuryl hexanol[c]	41.0	43.5	43.8	65.0	1.7	3.9	1.3	1.3	0.3	2.3	2.7	1.3
36	2-Ethyl-1-hexanol	44.0	27.2	26.6	54.1	6.1	22.4	2.1	1.9	0.8	1.8	1.5	0.7
37	Isobutanol	51.3	44.6	44.9	61.1	6.0	13.4	2.4	2.3	1.4	2.1	2.6	1.1
38	Butanol	60.3	40.3	40.4	63.5	6.8	16.8	2.4	2.1	1.6	2.0	2.4	0.7
39	Hexanol	64.8	53.7	53.8	68.6	5.2	9.8	2.7	2.5	2.5	2.2	2.9	1.3
40	Butyl cellosolve	68.7	68.1	68.2	88.0	4.5	6.6	2.3	2.8	2.4	2.3	4.0	1.3
41	Cyclohexanol	79.8	78.6	78.8	83.8	2.0	2.5	2.4	2.5	2.2	2.9	4.0	1.2
42	4-Carboxybenzaldehyde	50.3	46.1	46.1	60.4	4.1	8.9	3.0	1.3	1.3	1.8	3.1	1.4
43	4,4'-Methylene bis (2,6-ditertbutylphenol)	0.0	0.0	0.0	0.0	0.0	ND	0.0	0.0	0.0	0.0	0.0	0.0
44	4-Bromophenetole	1.3	0.4	0.4	1.3	0.3	61.0	0.2	0.0	0.0	0.0	0.0	0.0
45a	2-Xylene [1]	1.5	1.2	1.2	2.7	0.5	38.7	0.2	0.4	0.0	0.0	0.0	0.0
45b	2-Xylene [2]	9.0	6.5	6.5	9.0	0.6	9.1	1.2	1.7	0.4	0.0	0.0	0.0
46	3-Chloro-4-fluoronitrobenzene	1.7	0.8	0.8	1.8	0.3	36.4	0.3	0.1	0.0	0.0	0.0	0.0
47	1,3-Diisopropylbenzene	2.0	1.1	1.1	1.9	0.2	21.5	0.6	0.0	0.0	0.0	0.0	0.0
48	1-Methylpropylbenzene	2.0	0.7	0.7	1.3	0.2	32.0	0.3	0.0	0.0	0.0	0.0	0.0
49	3-Ethyltoluene	2.3	1.8	1.8	2.6	0.2	12.4	0.8	0.0	0.1	0.0	0.0	0.0
50	2,4-Difluoronitrobenzene	4.7	4.2	4.1	7.3	0.8	18.5	1.3	0.4	0.4	0.0	0.0	0.0
51	Styrene	6.8	3.2	3.2	5.4	0.7	21.1	0.7	0.6	0.3	0.1	0.1	0.0
52	Toluene	9.0	7.3	7.3	9.2	0.7	8.9	1.5	1.4	0.7	0.0	0.0	0.0
53	4-Fluoroaniline	69.8	47.2	47.2	63.0	4.6	11.0	2.7	2.2	1.7	1.6	4.0	0.6
54	Pyridine	48.0	46.1	46.3	55.1	3.3	7.1	2.9	3.3	1.3	2.7	1.9	1.2

Table 5.7 Eye irritation data for 92 chemicals

Chemical No	Name	MMAS[a]	WDS[b] mean	median	max	SD	CV	Mean Erythema	Mean Œdema	Mean Discharge	Mean Opacity	Mean Area	Mean Iritis
55	2-Bromobutane	0.0	0.0	0.0	0.0	0.0	ND	0.0	0.0	0.0	0.0	0.0	0.0
56	1-Bromo-4-chlorobutane	0.0	0.0	0.0	0.0	0.0	ND	0.0	0.0	0.0	0.0	0.0	0.0
57	Octyl bromide	0.0	0.0	0.0	0.0	0.0	ND	0.0	0.0	0.0	0.0	0.0	0.0
58	Hexyl bromide	1.3	0.4	0.4	1.6	0.3	63.5	0.0	0.0	0.2	0.0	0.0	0.0
59	Amyl bromide	2.0	1.1	1.1	2.0	0.3	29.3	0.6	0.0	0.0	0.0	0.0	0.0
60	1,6-Dibromohexane	2.0	0.7	0.7	1.6	0.3	48.3	0.2	0.1	0.0	0.0	0.0	0.0
61	Isopropyl bromide	2.7	1.1	1.1	2.9	0.5	42.0	0.3	0.2	0.0	0.0	0.0	0.0
62	1,4-Dibromobutane	4.0	1.4	1.3	4.9	0.9	67.0	0.3	0.3	0.0	0.0	0.0	0.0
63	1,5-Dibromopentane	4.0	1.8	1.8	4.0	0.6	35.2	0.6	0.3	0.0	0.0	0.0	0.0
64	1,3-Dibromopropane	4.0	1.3	1.3	4.0	0.6	48.5	0.3	0.3	0.0	0.0	0.0	0.0
65	Ethyleneglycol diethylether	0.0	0.0	0.0	0.0	0.0	ND	0.0	0.0	0.0	0.0	0.0	0.0
66	Isostearic acid	3.3	1.5	1.6	3.6	0.6	39.5	0.6	0.0	0.2	0.0	0.0	0.0
67	3,3-Dimethylpentane	0.0	0.0	0.0	0.0	0.0	ND	0.0	0.0	0.0	0.0	0.0	0.0
68	3-Methylhexane	0.7	0.2	0.2	0.8	0.1	64.9	0.1	0.0	0.0	0.0	0.0	0.0
69	2-Methylpentane	2.0	0.8	0.8	2.0	0.3	35.8	0.3	0.0	0.1	0.0	0.0	0.0
70	1,9-Decadiene	2.0	0.9	0.9	1.6	0.2	26.2	0.4	0.0	0.0	0.0	0.0	0.0
71	Dodecane	2.0	0.8	0.8	2.1	0.3	34.4	0.4	0.0	0.0	0.0	0.0	0.0
72	1,5-Dimethylcyclooctadiene	2.8	1.3	1.3	3.3	0.3	24.0	0.6	0.0	0.1	0.1	0.1	0.0
73	*cis*-Cyclooctene	3.3	2.2	2.2	3.0	0.3	12.2	1.1	0.0	0.0	0.0	0.0	0.0
74	Methylcyclopentane	3.7	2.3	2.3	4.2	0.5	22.8	0.8	0.1	0.2	0.0	0.0	0.0
75	1,5-Hexadiene	4.7	2.4	2.4	4.7	0.5	21.2	0.9	0.2	0.1	0.0	0.0	0.0
76	Di-iso-butyl ketone (2,6-dimethyl-4-heptanone)	0.7	0.2	0.2	1.1	0.2	93.5	0.1	0.0	0.0	0.0	0.0	0.0
77	Methyl isobutyl ketone (4-methylpentan-2-one)	4.8	2.1	2.0	4.7	0.5	22.4	0.8	0.2	0.0	0.1	0.1	0.0
78a	Methyl amyl ketone [1]	10.5	4.7	4.5	12.1	1.9	40.4	0.9	0.6	0.3	0.3	0.3	0.1
78b	Methyl amyl ketone [2]	16.3	7.5	7.4	15.1	1.5	20.2	1.2	1.2	0.8	0.3	0.3	0.1
79	Methyl ethyl ketone (butanone)	50.0	27.0	26.7	53.2	6.0	22.1	2.1	2.2	2.3	1.3	1.7	0.6

Table 5.7 Eye irritation data for 92 chemicals

Chemical No	Name	MMAS[a]	WDS[b] mean	median	max	SD	CV	Mean Erythema	Mean Œdema	Mean Discharge	Mean Opacity	Mean Area	Mean Iritis
80	Acetone	65.8	50.6	50.7	64.5	4.8	9.3	2.8	2.5	2.7	1.9	2.9	1.3
81	3-Chloropropionitrile	13.7	4.5	4.3	17.0	1.8	41.1	0.6	0.3	0.2	0.7	0.4	0.1
82	Diethylaminopropionitrile	62.3	50.1	49.9	65.0	6.1	12.2	3.0	2.6	2.4	3.5	1.7	1.0
83	2-Ethylhexylthioglycolate	0.0	0.0	0.0	0.0	0.0	ND	0.0	0.0	0.0	0.0	0.0	0.0
84	4-Methylthiobenzaldehyde	0.0	0.4	0.4	2.2	0.4	95.6	0.1	0.1	0.0	0.0	0.0	0.0
85	Iso-octylthioglycolate	0.7	0.2	0.2	1.1	0.2	91.8	0.1	0.0	0.0	0.0	0.0	0.0
86	Dipropyl disulphide	1.3	0.9	0.9	2.0	0.3	36.7	0.4	0.0	0.0	0.0	0.0	0.0
87	Thiodiglycol	5.3	2.4	2.4	6.7	1.1	47.1	0.5	0.3	0.3	0.0	0.0	0.0
88	1,2,3-Trimercaptopropane	8.7	6.2	6.2	10.7	1.6	25.1	1.3	1.5	0.3	0.0	0.0	0.0
89	Ethylthioglycolate	24.7	11.6	11.0	33.4	5.7	48.8	1.2	0.8	0.4	1.0	0.8	0.5
90	Methylthioglycolate	53.0	36.4	36.2	57.7	7.7	21.1	2.4	1.8	1.3	2.2	1.9	0.9
91	Glycerol tri-isostearate	2.0	0.7	0.7	2.0	0.3	47.3	0.3	0.0	0.0	0.0	0.0	0.0
92	γ-Butyrolactone	43.0	36.2	36.2	52.8	4.5	12.3	2.2	2.1	1.3	1.9	2.2	0.8

Footnote to Table 5.7

[a]Modified maximum average score (MMAS) values taken from ECETOC eye irritation data bank (ECETOC, 1998a).
[b]Bootstrap estimates of the weighted Draize score.
[c]Data were only available for one rabbit in ECETOC (1998a).

CV = coefficient of variation; SD = standard deviation; ND = not defined (denominator of zero).
Shading indicates chemicals for which the maximal WDS < MMAS

Table 5.8 Chemicals for which the PII is biased

Chemical	Bootstrap PII[a]	Conventional PII[b]	Bias	Bias-corrected estimate of PII
Citronellol [3]	3.7	3.3	0.4	2.9
Cyclamen aldehyde	4.8	4.6	0.2	4.4
Benzyl benzoate	1.6	1.1	0.5	0.6
Cis-cyclooctene	1.9	2.1	-0.2	2.3
1,13-Tetradecadiene	1.7	1.6	0.1	1.5
Benzyl acetone	1.2	1.0	0.2	0.8

Footnote to Table 5.8

PII = primary irritation index

Citronellol [3] refers to the *third* data sheet for citronellol in the ECETOC report.
[a]The mean value of the bootstrap distribution for PII.
[b]The PII calculated from the original sample of data.

Table 5.9 Spearman correlations between six skin irritation endpoints shown in Figure 5.5

	Erythema 24h	Erythema 48h	Erythema 72h	Œdema 24h	Œdema 48h
Erythema 48h	0.90				
Erythema 72h	0.86	0.96			
Oedema 24h	0.79	0.83	0.84		
Oedema 48h	0.75	0.84	0.86	0.93	
Oedema 72h	0.74	0.80	0.84	0.87	0.94

Table 5.10 Spearman correlations between six eye irritation effects shown in Figure 5.8

	Conjunctival erythema	Conjunctival œdema	Conjunctival discharge	Corneal opacity	Corneal area
Conjunctival œdema	0.88				
Conjunctival discharge	0.88	0.88			
Corneal opacity	0.82	0.84	0.82		
Corneal area	0.83	0.83	0.83	0.98	
Iritis	0.75	0.77	0.79	0.86	0.84

Table 5.11 Spearman correlations between the weighted tissue scores shown in Figure 5.9

	Conjunctiva 24h	Conjunctiva 48h	Conjunctiva 72h	Cornea 24h	Cornea 48h	Cornea 72h	Iris 24h	Iris 48h
Conjunctiva 48h	0.90							
Conjunctiva 72h	0.85	0.90						
Cornea 24h	0.81	0.84	0.85					
Cornea 48h	0.74	0.79	0.80	0.92				
Cornea 72h	0.69	0.74	0.81	0.86	0.93			
Iris 24h	0.68	0.69	0.73	0.80	0.76	0.73		
Iris 48h	0.66	0.69	0.74	0.80	0.80	0.77	0.85	
Iris 72h	0.53	0.58	0.66	0.67	0.68	0.73	0.67	0.76

Table 5.12 Spearman correlations between the weighted Draize scores shown in Figure 5.10

	WDS 24h	WDS 48h	WDS 72h
WDS 48h	0.91		
WDS 72h	0.86	0.90	
max WDS	0.99	0.92	0.87

Footnote to Table 5.12

WDS = weighted Draize score.

CHAPTER 6

EMBEDDED CLUSTER MODELLING: A NOVEL METHOD FOR GENERATING ELLIPTIC CLASSIFICATION MODELS

6.1 INTRODUCTION

If chemicals are classified into two groups (active/inactive or toxic/non-toxic) and are plotted on a box plot (one variable), or on a scatter plot (two or more variables), sometimes it appears that the chemicals in one group form an 'embedded cluster' surrounded by the chemicals in the other group, the 'diffuse cluster'. Typically, the active (or toxic) chemicals form the embedded cluster, and the inactive (or non-toxic) chemicals form the diffuse cluster, indicating that for each of the chosen variables, there is an optimal range of values for the biological response to occur. An example of embedded clustering in toxicology is provided by chemicals that are irritating to the eye: embedded clusters of irritants have been observed among diffuse clusters of non-irritants in several SAR studies (Cronin *et al.*, 1994; Cronin, 1996; Barratt, 1997; Worth & Fentem, 1999).

Although a set of active (or toxic) chemicals may appear to form an embedded cluster when inspected visually on a scatter plot, the occurrence of embedded clustering can only be concluded with statistical significance if the method of cluster significance analysis (CSA) indicates that the apparent embedded cluster is unlikely to have arisen by chance. The method of CSA is described in detail in Chapter 4.

If the application of CSA indicates that the embedded cluster of active chemicals is statistically significant, it is then desirable to derive a classification model (CM) for classifying chemicals as active (toxic) or inactive (non-toxic). To derive such a CM, the technique of embedded cluster modelling (ECM) was devised by the author (Worth & Cronin, 1999). If only one variable is found to be significant by CSA, ECM can be used to identify the upper and lower cut-off values of the variable between which embedded chemicals are predicted to lie:

$$\text{CI for embedded chemicals} = \bar{x} \pm (1.96\,\sigma) \qquad \text{(Equation 6.1)}$$

where CI is the 95% confidence interval for embedded chemicals; $\bar{x}$ is the mean value of the physicochemical descriptor x for the active (toxic) chemicals; and σ is the corresponding standard deviation. In Equation 6.1, the factor of 1.96 is intended to

capture 95% of the active (toxic) chemicals, assuming that these are normally distributed in the *x* direction.

In the case of two or more significant variables, ECM treats the boundary of the embedded cluster as an ellipse in two or more dimensions, so that the active (toxic) chemicals can be identified as those lying inside the elliptic boundary, whereas the inactive (non-toxic) chemicals can be identified as those located outside the ellipse. The equation for a two-dimensional (2D) ellipse is given by Equation 6.2:

$$(x-h)^2 / \alpha^2 + (y-k)^2 / \beta^2 = 1 \qquad \text{(Equation 6.2)}$$

where *x* and *y* define all points on the 2D ellipse; *h* and *k* are the coordinates of the centre of the ellipse along the *x* and *y* cartesian axes, respectively; and α and β are the dimensions (radial axes) of the ellipse parallel with the *x* and *y* cartesian axes, respectively. The centre of the ellipse (*h*, *k*) is defined as the centroid of the embedded cluster, and the radial axes of the ellipse (α and β) are proportional to the standard deviations in the *x* and *y* directions (1.96 σ_x, and 1.96 σ_y). In Equation 6.2, the factor of 1.96 is intended to capture 95% of the active (toxic) chemicals within the ellipse, assuming that these are normally distributed in the *x* and *y* directions.

The aim of the work presented in this chapter was to illustrate how the methods of CSA and ECM can be used in combination to generate elliptic CMs from embedded data sets. This work was carried out as a follow-up to the original description of ECM (Worth & Cronin, 1999). It was presented at the 3rd World Congress on Alternatives and Animals in the Life Sciences (Worth & Cronin, 2000a), and subsequently at a meeting of the UK-QSAR and ChemoInformatics Group (October 20, 1999; SmithKline Beecham Pharmaceuticals, Essex, UK).

6.2 METHOD

6.2.1 Collection and treatment of *in vivo* data

A data set of 73 organic chemicals (Table 6.1) was constructed by using chemicals reported in the ECETOC reference chemicals data bank (ECETOC, 1998a). Chemicals were selected if they were organic liquids belonging to one of the following chemical classes: aliphatic hydrocarbons; aromatic compounds; aliphatic alcohols, ethers, esters, and ketones (including γ-butyrolactone); alkyl bromides; and sulphur-containing compounds. Solids were excluded, because the exposure of the rabbit eye to these substances is not comparable to that of organic liquids (Balls *et al.*, 1999). In addition, acids, acid precursors and alkalis were excluded, because these are generally applied as diluted aqueous solutions (rather than as neat liquids), and are likely to act by a pH-dependent mechanism. A CM for predicting eye irritation potential on the basis of pH is proposed in chapter 9.

Modified maximum average score (MMAS) values were taken directly from the ECETOC data bank. The method for calculating the MMAS is explained in Chapter 2. In the case of methyl amyl ketone (2-heptanone), there are two entries in the ECETOC data bank, so the average of the two MMAS values (10.5 and 16.3) was taken (i.e. 13.4).

Chemicals with a MMAS greater than, or equal to 4.0, were classified as irritant (I), and chemicals with MMAS values less than 4.0 were classified as non-irritant (NI). The cut-off value between I and NI chemicals was chosen as the median MMAS of the 73 chemicals, to produce a balanced training set (containing 37 I and 36 NI chemicals). Clearly, different threshold values could be employed, depending on the purpose of the investigation. For example, Kay & Calandra (1962) defined irritants as chemicals having a MMAS $\geq$ 0.5, whereas Spielmann *et al.* (1996) defined severe irritants as chemicals having a MMAS > 59.

6.2.2 Calculation of physicochemical descriptors

Various physicochemical properties were calculated from the molecular structures of the 73 chemicals, as represented by their one-dimensional SMILES notations (Weininger, 1988). The logarithm of the octanol-water partition coefficient (logP) was calculated by using KOWWIN for Windows (Syracuse Research Corporation, Syracuse, NY, USA), which uses a fragment-based method to predict logP (Meylan & Howard, 1995). In addition, the following molecular topology indices were calculated with the MS-DOS program, Molconn-Z (Hall Associates Consulting, Quincy, MA, USA): simple and valence path molecular connectivity indices of orders 0 to 4; simple and valence cluster molecular connectivity indices of orders 3 and 4; fourth-order simple and valence path-cluster molecular connectivity indices; differences in simple and valence path molecular connectivity indices; kappa indices of orders 1 to 3; kappa alpha indices of orders 0 to 3; and the flexibility index. These topological indices are considered to encode information about the size, shape, branching and heteroatom contents of molecules (Kier & Hall, 1986). Molecular weight (MW) was also calculated with the Molconn-Z program. In addition, dipole moment, molecular surface area and molecular volume were calculated by using HyperChem 5.1 for Windows (Hypercube Inc., Gainesville, FL, USA). The chemical structures were geometry-optimised by the successive application of HyperChem's MM+ molecular mechanics force field, which is based on the MM2 force field (Allinger, 1977), followed by the AM1 semi-empirical method (Dewar *et al.*, 1985).

6.2.3 Cluster significance analysis and embedded cluster modelling

A visual inspection of all pairwise combinations of descriptors, using matrix plots indicated that the embedded clustering of irritant chemicals is most apparent in the two-dimensional scatter plot of logP against dV1 (Figure 6.1). The second variable, dV1, stands for ‘the first-order difference valence connectivity index’, and is simplified notation for the more usual $d^1\chi^v$, which is defined as follows (Kier & Hall, 1986):

$$d^1\chi^v = {}^1\chi_P{}^v - {}^1\chi_N{}^v \qquad \text{(Equation 6.3)}$$

where $^{1}\chi_{P}^{v}$ is the first-order valence path index for the molecule of interest; and $^{1}\chi_{N}^{v}$ is the first-order valence path index for the unbranched, acyclic graph that has the same size as the given molecule. This topological index, which is claimed to be size-independent (Kier & Hall, 1986), appears to provide a measure of the degree of branching and/or cyclicity in a molecule - its numerical value is close to zero for linear, acyclic structures, and becomes increasingly negative as the number of branches or rings in the molecule increases (Figure 6.2).

To carry out CSA in the Minitab package (Minitab Inc., State College, PA, USA), three macros were written to perform CSA in one, two and three dimensions, respectively (see Appendix A3). The macro for 2D CSA was used to assess the statistical significance of the embedded cluster along logP and dV1, and a 2D elliptic model was derived by ECM, again by using a purpose-written macro (see Appendix A9). The variabilities in the model parameters (the centroid and radial axes of the ellipse) were assessed from their bootstrap distributions, derived from 1000 random samples of the logP and dV1 values for the irritant chemicals.

6.2.4 Assessment of model performance

The ability of the elliptic model to correctly distinguish between irritant and non-irritant chemicals was determined by comparing the predicted classifications (I/NI) with the observed (pre-defined) classifications for the 73 chemicals. Calculations of the sensitivity, specificity, concordance, the positive and negative predictivities, and the false positive and negative rates, were performed with Minitab 11 for Windows, using a purpose-written macro (see Appendix A1).

6.3 RESULTS

The application of CSA to the complete data set indicated that embedded clustering along logP and dV1 was highly significant ($p<0.01$), even without the removal of outliers. Following the removal of five irritant outliers (isomyristyl alcohol, 1,5-dibromopentane, 1,4-dibromobutane, 1,3-dibromopropane, and 2,4-dinitrofluoro benzene), the application of ECM generated the following PM:

classify an undiluted, organic liquid as an eye irritant if:

$$(\text{logP-}1.07)^2 / 2.06^2 + (\text{dV1} + 0.98)^2/0.99^2 \leq 1 \qquad \text{(Equation 6.4)}$$

The 2D ellipse described by this model is illustrated in Figure 6.1.

When the elliptic model was used to classify the 73 organic liquids, 18 chemicals were misclassified. Ten chemicals were false negatives (isomyristyl alcohol, 1,5-dibromobutane, 1,4-dibromobutane, 1,3-dibromopropane, ethanol, methyl cyanoacetate, cellosolve acetate [2-ethoxyethanol acetate], ethyl-2-methyl acetoacetate, 4-fluoroaniline, and 2,4-difluoronitrobenzene), and eight chemicals were false positives (2-xylene, di-isobutyl ketone, 4-methylthiobenzaldehyde, ethyl trimethyl acetate, methyl trimethyl acetate, ethylene glycol diethyl ether, 2,4-pentanediol, and propylene glycol). The Cooper statistics are summarised in Table 6.2.

The bootstrap distributions of the model parameters are approximately normal, as judged by subjective impression (for example, Figures 6.3 and 6.4), and by the Anderson-Darling normality test. Therefore, the 95% CIs (Table 6.3) could be calculated by using Equation 6.5:

$$95\%\ \text{CI} = \text{bootstrap mean} \pm (1.96\ \text{SD}) \qquad \text{(Equation 6.5)}$$

where SD is the standard deviation of the appropriate bootstrap distribution. The fact that the parameter estimates based on the full data set (i.e. the parameter values incorporated in the elliptic model) are so close (to within 0.04 units or less) to the mean values of the bootstrap distributions indicates that the parameter estimates are unbiased, i.e. the sample of chemicals is representative of the population from which it is drawn.

6.4 DISCUSSION

The aim of this work was to illustrate how the combined use of CSA and ECM can generate PMs from embedded data sets, by using an endpoint employed in eye irritation testing. For the purposes of this study, the distinction between eye irritants and non-irritants was based simply on the MMAS. Clearly, other distinctions could be used, such

as those specified in EU regulations (EC, 1993) or in the OECD guidelines on eye irritation testing (OECD, 1998).

The elliptic model of eye irritation potential is based on two variables (logP and dV1), and indicates that neat organic liquids (excluding acids and alkalis) are only irritant to the eye if their logP and dV1 values fall in a critical range. The first variable, logP, provides a measure of the tendency of chemicals undergo passive diffusion across biological membranes: if the logP of a substance is too high, it will remain preferentially within the membranes of the eye, whereas if the logP is too low, it will remain preferentially in the aqueous media surrounding membranes, e.g. in the tear film covering the eye. LogP was also found to be an important descriptor variable in the embedded clustering of eye irritants in previous SAR studies (Cronin *et al.*, 1994; Cronin, 1996; Barratt, 1997; Worth & Fentem, 1999).

The second variable, dV1, is a size-independent measure of molecular topology. Since it appears that dV1 is related to the degree of branching and/or cyclicity in a molecule (Figure 6.2), the importance of dV1 in the elliptic model indicates that organic liquids must have an intermediate degree of branching/cyclicity to be irritant to the eye. This interpretation is reminiscent of the observation by Barratt (1997) that the second and third principal inertial axes are important descriptors in the embedded clustering of neutral organic chemicals, which suggests that only chemicals with a certain cross-sectional area are irritant. While the two models are mechanistically similar, they are not exactly the same. The topological index dV1 is claimed to be size-independent (Kier & Hall,1986), whereas the second and third principal inertial axes are, by definition, size-dependent.

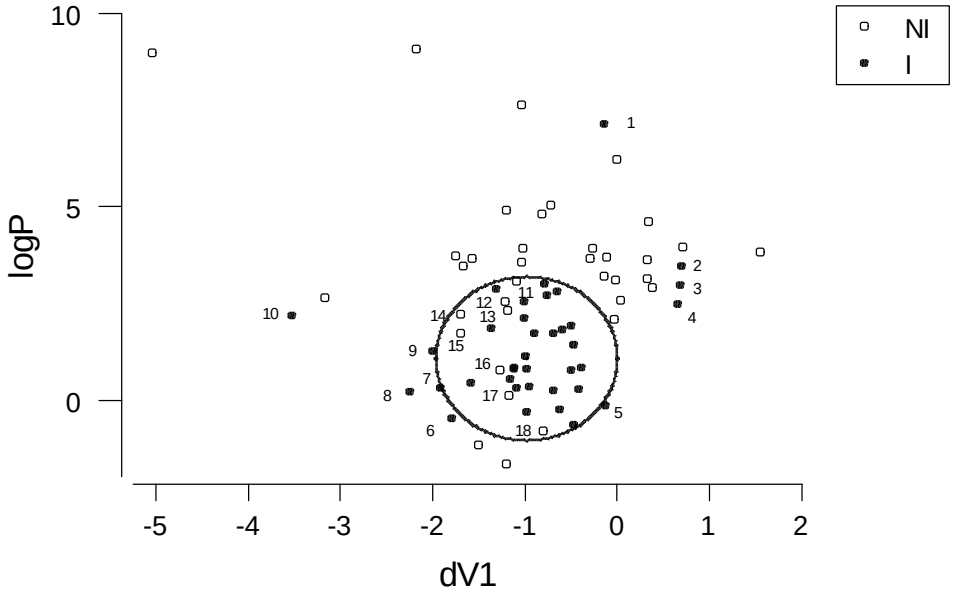
Given the physicochemical interpretations of logP and dV1, the elliptic model can be said to describe partition rather than reactivity. In other words, the model probably identifies the chemicals that are likely to cross the conjunctiva into the cornea, but does not necessarily identify which of those chemicals elicit a biological response in the cornea (effects in the cornea account for 73% of the MMAS). Consistent with this interpretation is the fact that the model predicts eight non-irritant chemicals to be irritant. Three of these chemicals (2-xylene, ethyl trimethyl acetate and propylene glycol) are close to the elliptic boundary, so their predicted classifications are best

regarded as equivocal. The model also predicts ten irritant chemicals to be non-irritant. Five of these chemicals (isomyristyl alcohol, 1,5-dibromopentane, 1,4-dibromobutane, 1,3-dibromopropane, and 2,4-dinitrofluorobenzene) were omitted from the training set, so it is not surprising that they are misclassified by the elliptic model. The other five false negatives (ethanol, methyl cyanoacetate, cellosolve acetate, ethyl-2-methyl acetoacetate, and 4-fluoroaniline) are close to the elliptic boundary. Incidentally, ethanol and ethyl-2-methyl acetoacetate are classified as non-irritant by the EU classification system (EC, 1993), but as irritant by the OECD system (OECD, 1998). If the elliptic model of eye irritation potential were employed in a tiered testing strategy, such as the one adopted recently by the OECD (1998), the occurrence of false negatives would not be problematic, since these chemicals would be correctly identified as irritant in a subsequent step. In contrast, the generation of false positives would be problematic, since these chemicals would be over-labelled without further testing. This suggests that the elliptic model would either have to be modified to incorporate one or more reactivity parameters, or used in combination with an SAR based on such parameters.

6.5 CONCLUSIONS

This study provides further evidence for the occurrence of embedded clustering in eye irritation data sets, and illustrates how the combined use of CSA and ECM can generate CMs from such data sets. In addition, the bootstrap resampling method is shown to provide a useful means of estimating the uncertainty in model parameters. CMs of eye irritation potential, derived by the combined use of CSA and ECM could, if adequately validated, be used in a tiered testing approach to eye irritation, such as the one adopted by the OECD (1998).

Figure 6.1 Scatter plot of the logarithm of the octanol-water partition coefficient (logP) against the first-order valence-corrected difference path index (dV1) for 73 organic liquids

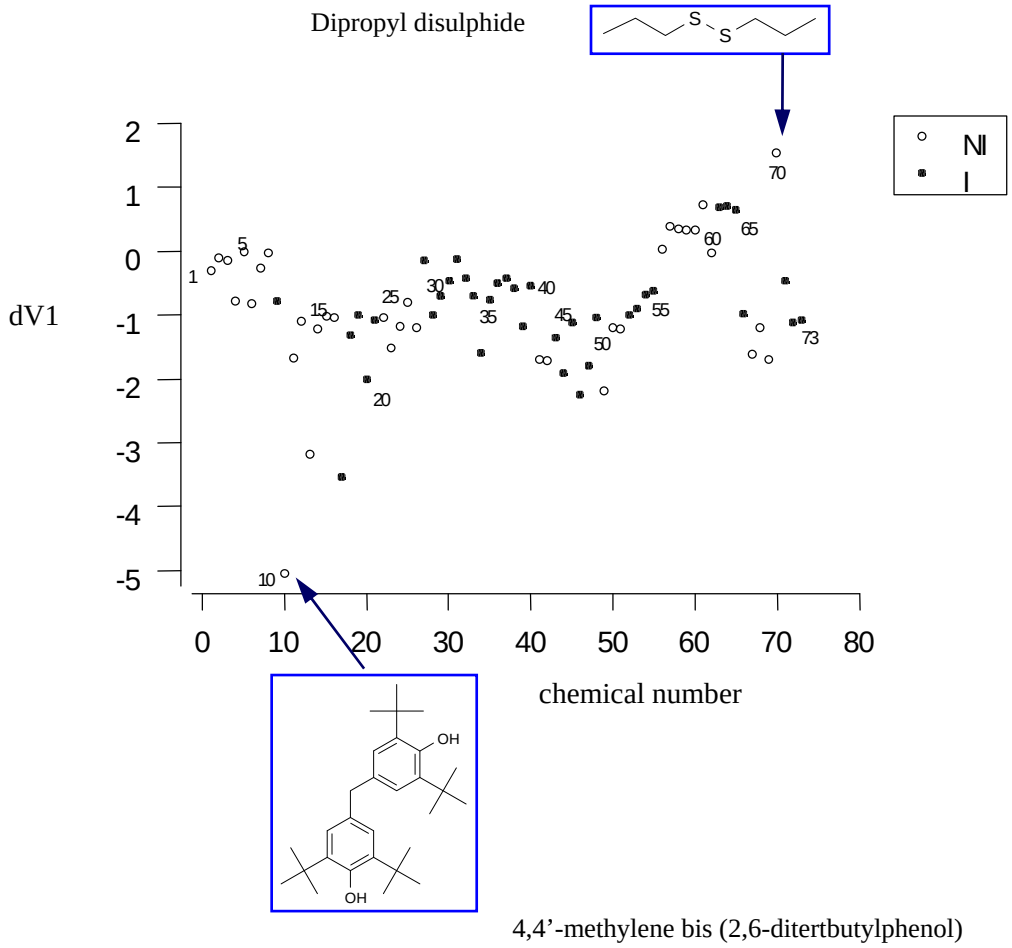


Footnote to Figure 6.1

The following chemicals are identified: 1) isomyristyl alcohol; 2) 1,5-dibromobutane; 3) 1,4-dibromobutane; 4) 1,3-dibromopropane; 5) ethanol; 6) methyl cyanoacetate; 7) cellosolve acetate; 8) ethyl-2-methyl acetoacetate; 9) 4-fluoroaniline; 10) 2,4-difluoronitrobenzene; 11) *o*-xylene; 12) di-isobutyl ketone; 13) 4-methylthiobenzaldehyde; 14) ethyl trimethyl acetate; 15) methyl trimethyl acetate; 16) ethylene glycol diethyl ether; 17) 2,4-pentanediol; and 18) propylene glycol.

I = irritant; NI = non-irritant.

Figure 6.2 Plot of the first-order difference valence connectivity index for 73 organic liquids



Footnote to Figure 6.2

The chemical numbers correspond to those in Table 6.1. Solid circles denote irritant chemicals, open circles non-irritants

Figure 6.3 Bootstrap distribution of the elliptic centroid along logP for 32 eye irritants

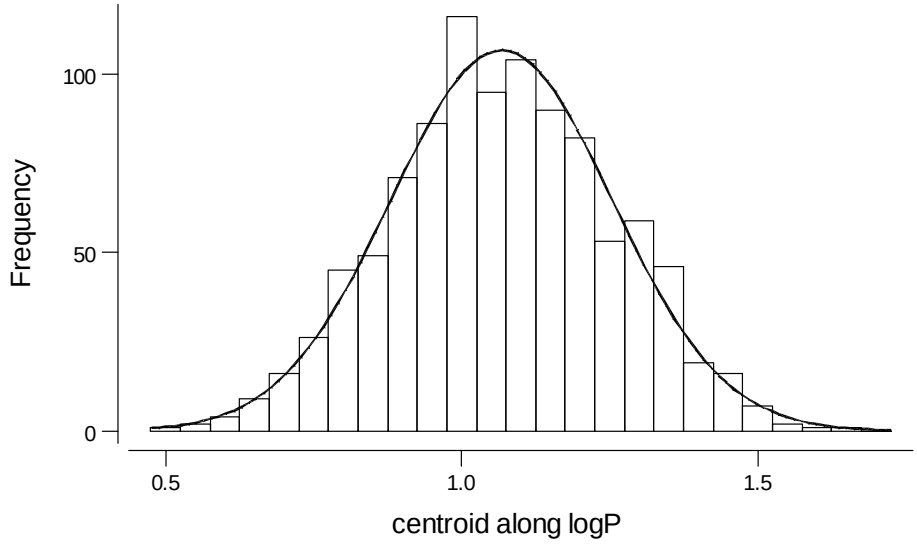


Figure 6.4 Bootstrap resampling distribution of the elliptic radius along logP for 32 eye irritants

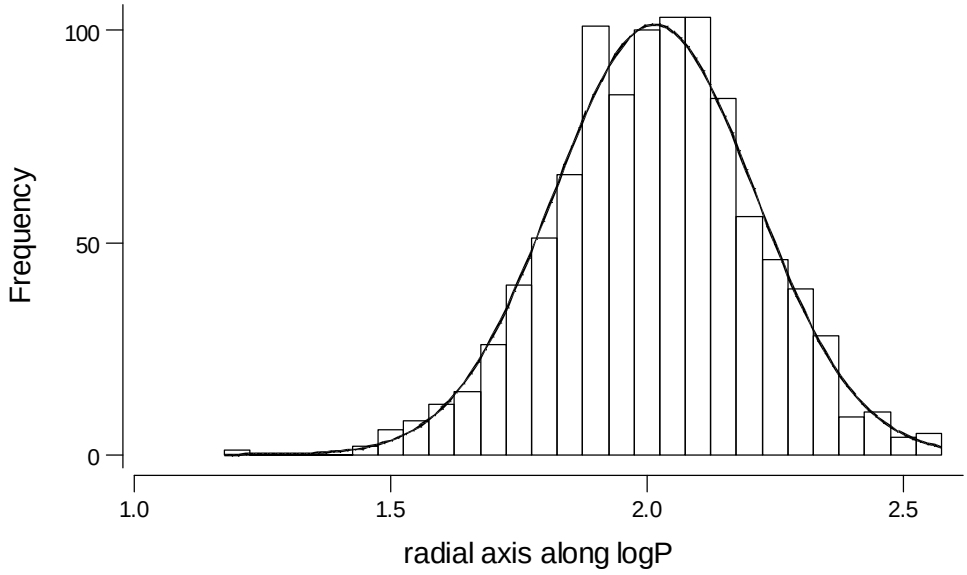


Table 6.1 Names, MMAS values and physicochemical properties for 73 organic chemicals

No.	Chemical	MMAS	Class	logP	dV1
	Hydrocarbons				
1	3,3-Dimethylpentane	0.0	NI	3.67	-0.293
2	3-Methylhexane	0.7	NI	3.71	-0.106
3	2-Methylpentane	2.0	NI	3.21	-0.144
4	1,9-Decadiene	2.0	NI	4.98	-0.781
5	Dodecane	2.0	NI	6.23	0.000
6	1,5-Dimethylcyclooctadiene	2.8	NI	4.82	-0.813
7	*cis*-Cyclooctene	3.3	NI	3.94	-0.264
8	Methylcyclopentane	3.7	NI	3.10	-0.020
9	1,5-Hexadiene	4.7	I	3.02	-0.781
	Aromatics				
10	4,4'-Methylene bis (2,6-ditertbutylphenol)	0.0	NI	8.99	-5.043
11	4-Bromophenetole	1.3	NI	3.46	-1.665
12	2-Xylene (1,2-dimethylbenzene)	1.5	NI	3.09	-1.087
13	3-Chloro-4-fluoronitrobenzene	1.7	NI	2.66	-3.169
14	1,3-Di-isopropylbenzene	2.0	NI	4.90	-1.206
15	1-Methylpropylbenzene (2-phenylbutane)	2.0	NI	3.94	-1.022
16	3-Ethyltoluene	2.3	NI	3.58	-1.032
17	2,4-Difluoronitrobenzene	4.7	I	2.21	-3.535
18	Styrene	6.8	I	2.89	-1.307
19	Toluene	9.0	I	2.54	-1.004
20	4-Fluoroaniline	69.8	I	1.28	-2.004
21	Pyridine	48.0	I	0.80	-1.075
	Alcohols				
22	Isostearyl alcohol (iso-octadecanol)	0.0	NI	7.64	-1.029
23	3-Methoxy-1,2-propanediol	0.0	NI	-1.15	-1.506
24	2,4-Pentanediol	1.3	NI	0.13	-1.169
25	Propylene glycol (1,2-propanediol)	1.3	NI	-0.78	-0.797
26	Glycerol	1.7	NI	-1.65	-1.196
27	Isomyristyl alcohol (iso-tetradecanol)	4.0	I	7.14	-0.144
28	2,2-Dimethyl-3-pentanol	8.3	I	2.13	-1.005
29	2-Methyl-1-pentanol	13.0	I	1.75	-0.695
30	Cyclopentanol	21.7	I	1.43	-0.465
31	Ethanol	24.0	I	-0.14	-0.130
32	Isopropanol (propan-2-ol)	28.3	I	0.28	-0.415
33	Octanol	30.5	I	2.81	-0.688
34	Furfuryl alcohol	41.0	I	0.45	-1.591
35	2-Ethyl-1-hexanol	44.0	I	2.73	-0.756
36	Isobutanol (butan-2-ol)	51.3	I	0.77	-0.500
37	Butanol	60.3	I	0.84	-0.428
38	Hexanol	64.8	I	1.82	-0.588
39	Butyl cellosolve (2-butoxyethanol)	68.7	I	0.57	-1.165
40	Cyclohexanol	79.8	I	1.92	-0.534

Table 6.1 continued

No.	Chemical	MMAS	Class	logP	dV1
	Esters				
41	Methyl trimethyl acetate	2.7	NI	1.74	-1.699
42	Ethyl trimethyl acetate	3.8	NI	2.23	-1.703
43	Butyl acetate	7.5	I	1.85	-1.362
44	Cellosolve acetate (2-ethoxyethyl acetate)	15.0	I	0.31	-1.913
45	Ethyl acetate	15.0	I	0.86	-1.119
46	Ethyl-2-methyl acetoacetate	18.0	I	0.21	-2.247
47	Methyl cyanoacetate	27.7	I	-0.47	-1.795
48	Methyl acetate	39.5	I	0.37	-1.041
49	Isopropyl isostearate	1.3	NI	9.07	-2.183
	Ether				
50	Ethylene glycol diethyl ether	0.0	NI	0.77	-1.197
	Ketones				
51	Di-iso-butyl ketone (2,6-dimethyl-4-heptanone)	0.7	NI	2.56	-1.211
52	Methyl isobutyl ketone (4-methylpentan-2-one)	4.8	I	1.16	-0.991
53	Methyl amyl ketone (2-heptanone)	13.4[a]	I	1.73	-0.902
54	Methyl ethyl ketone (butanone)	50.0	I	0.26	-0.686
55	Acetone	65.8	I	-0.24	-0.624
56	γ-Butyrolactone	43.0	I	-0.310	-0.972
	Alkyl bromides				
57	2-Bromobutane	0.0	NI	2.58	0.039
58	1-Bromo-4-chlorobutane	0.0	NI	2.90	0.392
59	Bromooctane	0.0	NI	4.61	0.348
60	Bromohexane	1.3	NI	3.63	0.335
61	Bromopentane	2.0	NI	3.14	0.324
62	1,6-Dibromohexane	2.0	NI	3.97	0.722
63	Isopropyl bromide	2.7	NI	2.08	-0.027
64	1,4-Dibromobutane	4.0	I	2.99	0.689
65	1,5-Dibromopentane	4.0	I	3.48	0.709
66	1,3-Dibromopropane	4.0	I	2.50	0.656
	Sulphur-containing compounds				
67	2-Ethylhexylthioglycolate	0.0	NI	3.68	-1.611
68	4-Methylthiobenzaldehyde	0.0	NI	2.31	-1.190
69	Iso-octylthioglycolate	0.7	NI	3.68	-1.687
70	Dipropyl disulphide	1.3	NI	3.84	1.553
71	Thiodiglycol	5.3	I	-0.62	-0.465
72	Ethylthioglycolate	24.7	I	0.81	-1.123
73	Methylthioglycolate	53.0	I	0.32	-1.083

Footnote to Table 6.1

[a]13.4 is the mean of 10.5 and 16.3, the replicate MMAS values reported in ECETOC (1998a).

dV1 = the first-order difference valence connectivity index; logP = the logarithm of the octanol-water partition coefficient; MMAS = modified maximum average score (taken from ECETOC, 1998a).

Table 6.2 Ability of the elliptic model of eye irritation potential to classify 73 organic chemicals

Statistic	Percent
Sensitivity	73
Specificity	78
Concordance	75
Positive predictivity	77
Negative predictivity	74
False positive rate	22
False negative rate	27

Table 6.3 Variability in the elliptic model parameters

Parameter	Estimate based on full data set	Bootstrap mean	Bootstrap standard deviation	Bootstrap lower 95% CL	Bootstrap upper 95% CL
centroid along logP	1.067	1.067	0.187	0.700	1.434
centroid along dV1	-0.983	-0.985	0.086	-1.154	-0.816
radial axis along logP	2.055	2.015	0.197	1.629	2.401
radial axis along dV1	0.986	0.958	0.129	0.705	1.211

CHAPTER 7

DEVELOPMENT OF NEW MODELS FOR ACUTE SKIN TOXICITY

7.1 OBJECTIVES

The aim of the work presented in this chapter was to develop the CMs for predicting the skin irritation or corrosion potential of organic chemicals. In particular, the following types of model were sought: a) SARs based on physicochemical properties that are easily calculated; b) PMs based on measured pH values; and c) PMs based on *in vitro* data. The emphasis of this chapter is to describe the development and assessment of individual models, whereas the integrated use of classification models is discussed in Chapter 10.

7.2 METHOD

7.2.1 Collection and treatment of *in vivo* data

Before developing SARs for skin corrosion, a data set of 277 organic chemicals (Table 7.1) was constructed from a variety of literature sources (Barratt, 1995b, 1996a & 1996b; ECETOC, 1995; NIH, 1999; Whittle *et al*., 1996). Chemicals taken from the ECETOC data bank (ECETOC, 1995) were classified for skin corrosion potential according to EU classification criteria (see Chapter 2); in the case of the chemicals taken from the other sources, the published classifications of corrosion potential were used.

Before developing SARs for skin irritation, a data set of 139 organic chemicals (Table 7.2) was compiled from the ECETOC data bank (ECETOC, 1995). These chemicals, which form a subset of the chemicals in Table 7.1, were classified for skin irritation potential in accordance with EU classification criteria (see Chapter 2). For 21 of these chemicals, there was more than one data sheet in the ECETOC data bank, and therefore more than one primary irritation index (PII) value. In these cases, the PII with the smallest variability was taken, as judged by the size of the bootstrap interquartile range (see Chapter 5.2). In all cases, the PII was calculated by applying Equation 2.1 to the erythema and œdema scores obtained at three time-points (see Chapter 2.2).

7.2.2 Calculation of physicochemical properties

The following physicochemical properties, which were considered to be possible predictors of acute skin toxicity, were calculated for the 277 chemicals in Table 7.1: molecular weight (MW), surface area (MSA) and volume (MV), logP, melting point (MP), surface tension (ST), pKa, the dermal permeability coefficient (K_p), dipole moment (DM), and the energies (E_{LUMO} and E_{HOMO}) of the lowest unoccupied molecular orbital (LUMO), and the highest occupied molecular orbital (HOMO). LogP, MP and K_p values were calculated with the Syracuse Research Corporation (SRC) KOWWIN, MPBPWIN and DERMWIN software packages (SRC, Syracuse, NY, USA), respectively, using the SMILES codes of the chemicals as the input. Values of MW and ST were calculated with the Advanced Chemistry Development (ACD) ChemSketch software, whereas pKa was calculated with the ACD pKa software (ACD, Toronto, Canada), using MOL files as the input. In the case of chemicals with more than one ionisable group, all pKa values were calculated.

To calculate the steric (MSA and MV) and molecular orbital properties (DM, E_{HOMO}, E_{LUMO}), chemical structures were entered as SMILES codes into the TSAR (version 3.2) molecular spreadsheet (Oxford Molecular Ltd [OML], Oxford, England). The CORINA (version 3.21) program (OML) converted the corresponding 2D structures into approximate 3D conformations. These 3D structures were then optimised in two stages: firstly, by using the COSMIC force field in TSAR, and secondly, by using the AM1 Hamiltonian in the VAMP (version 6.5) molecular orbital package (OML). The steric and molecular orbital properties were then calculated from the optimised 3D structures in VAMP. Predicted values of η and χ were calculated from E_{HOMO} and E_{LUMO} values by using Equations 3.36 and 3.37, respectively (see chapter 3).

7.2.3 Development of structure-activity relationships

A preliminary investigation of the skin corrosion data set was undertaken to detect any collinearities between potential predictor variables. Having eliminated redundant predictor variables, SARs were then derived by classification tree (CT) analysis, using the CART algorithm (Breiman *et al.*, 1984) in STATISTICA 5.5 for Windows (Statsoft Inc., Tulsa, OK, USA). Equal prior probabilities were set for the two classes (C/NC or

I/NI), the Gini index was used as the measure of node homogeneity, and a minimum node size of 5 observations was used as the stopping rule (i.e. a node would only be split if it contained more than 5 observations). The best-sized tree in the sequence of optimally-pruned trees was identified by applying 10-fold cross-validation pruning and selecting the tree with the minimum cross-validated cost (see Chapter 4 for an explanation of CT analysis) .

7.2.4 Development of prediction models

All PMs were derived by CT analysis, using the settings described in section 7.2.3. To develop a PM for skin corrosion potential based on measured pH values, a data set of 75 organic and inorganic chemicals (Table 7.3) was taken from Gordon *et al.* (1994). The PM obtained was validated by predicting the corrosivity classifications of a further 53 chemicals (Table 7.4), for which pH data had been provided by BIBRA International (Croydon, UK). The pH data in Table 7.4 were also used to develop a PM for skin irritation potential.

PMs based on *in vitro* endpoints were developed from the data obtained during the ECVAM Skin Corrosivity Validation Study (Barratt *et al.*, 1998; Fentem *et al.*, 1998) for the EPISKIN and rat skin transcutaneous electrical resistance (TER) assays. The EPISKIN data are cell viabilities, measured following treatment for 3 minutes, 1 hour and 4 hours; the TER data are electrical resistances measured across excised skin discs following their exposure for 2 and 24 hours (Table 7.4).

7.2.5 Assessment of the models

The two-group CMs were assessed in terms of their Cooper statistics (see § 4.3.4), which define an upper limit to predictive performance. In addition, cross-validated Cooper statistics, which provide a more realistic indication of a model's capacity to predict the classifications of independent data, were obtained by applying three-fold cross-validation to the best-sized CTs. In the three-fold cross-validation procedure, the data set is randomly divided into three approximately equal parts, the CT is reparameterised using two thirds of the data, and predicted classifications are made for the remaining third of the data. The cross-validated Cooper statistics are the mean

values of the usual Cooper statistics, taken over the three iterations of the cross-validation procedure.

Finally, confidence intervals were calculated for the Cooper statistics, using a purpose-written Minitab macro (Appendix A5). These calculations were performed to check whether the models derived had a predictive accuracy significantly greater than 50%. Confidence intervals were calculated for the standard Cooper statistics, but not for the cross-validated Cooper statistics, since the latter were obtained directly from 2x2 contingency tables generated in STATISTICA 5.5 for Windows, which were not associated with a set of known and predicted classifications for the given set of chemicals.

7.3 RESULTS

7.3.1 Structure-activity relationships

Preliminary data analysis

The variables MW, MSA and MV were found to be strongly and significantly correlated in the data set of 277 chemicals ($r_{MW\text{-}MSA}$ = 0.89; $r_{MW\text{-}MV}$ = 0.93; $r_{MV\text{-}MSA}$ = 0.99; all p values < 0.001). Therefore, due to the relative ease of its calculation, MW was used in preference to MSA and MV, to avoid redundancy among predictor variables. Similarly, η and χ were significantly correlated ($p<0.001$) with E_{LUMO} ($r = 0.79$ and -0.82 for η and χ, respectively). In this case, E_{LUMO} was used in preference to η and χ, since these properties are derived from E_{LUMO} (see chapter 3).

Structure-activity relationships for skin corrosion

A two-step decision rule was envisaged: in the first step, it was hypothesised that discrimination could be based on MP alone, on the grounds that chemicals existing as solids at skin temperature are not expected to be corrosive, whereas chemicals existing as liquids may or may not be, depending on other factors, which could be assessed in a second step. Separation of the 277 chemicals into two groups, one containing 88 chemicals having predicted MPs greater than 37°C, and the other containing 189 chemicals having predicted MPs less than or equal to 37°C, revealed that 74 of the 88

predicted solids (84%) are non-corrosive, as expected, whereas 14 of them (16%) are corrosive, contrary to expectation.

The best variable for discriminating between corrosive and non-corrosive liquids, as judged by the application of CT analysis to the values of MW, logP, logKp, ST, DM, E_{LUMO}, and E_{HOMO} for the 189 liquids, was found to be logP. However, the CT predicted liquids with logP values greater than 1.32 to be non-corrosive, and liquids with logP values less than or equal to 1.32 to be corrosive. The direction of this inequality is contrary to expectation, since corrosive chemicals are generally expected to be more hydrophobic than non-corrosive chemicals, and therefore have higher, not lower, values of logP. A possible explanation for this finding is that logP is significantly correlated with MW ($r=0.69$, $p<0.001$), i.e. it is the smaller chemicals that are more likely to be corrosive, not the less hydrophobic ones. Therefore, logP was removed from the set of input variables, and CT analysis was applied again. This time, CT analysis identified MW as the best discriminating variable, with an optimal cut-off value of 123 g/mol. On this basis of this finding, SAR 1 was formulated for predicting the corrosion potential of organic liquids, and the variable selection procedure was stopped:

If MW ≤ 123 g/mol, predict as C; otherwise predict as NC. (SAR 1)

Relationship between pKa and corrosion potential

The acid dissociation constant, pKa, was not used as an input variable when selecting for predictors of skin corrosion potential, because values were only available for 97 of the 189 liquids (i.e. those chemicals having ionisable groups). These chemicals (Table 7.5) can be divided according to chemical class, as follows: 27 carboxylic acids (26 C; 1 NC), 34 bases (23 C; 11 NC), 22 alcohols (3 C; 19 NC) and 14 phenols (8C; 6 NC). The subset of carboxylic acids, containing mostly corrosives, and the subset of alcohols, containing mostly non-corrosives, were considered too unbalanced for the derivation of CMs. However, since alcohols are very weak acids (with pKa values of around 14-15), it was decided to merge the carboxylic acids with the alcohols, to form a larger subset of acids (now defined in a broader sense). In addition, the three corrosive alcohols (propargyl alcohol, 2-butyn-1,4-diol, and 2-hydroxyethyl acrylate) were removed, so the new data set comprised 46 acids (26 C; 20 NC), which was sufficiently well balanced for model development. The application of CT analysis to the data set of 46 acids

produced SAR 2. Similarly, the application of CT analysis to the subset of bases produced SAR 3. The Cooper statistics of SARs 2 and 3 are reported in Table 7.6.

If the pKa of a carboxylic acid or alcohol ≤ 7.62, predict as C;
otherwise, predict as NC. (SAR 2)

If the pKa of a base > 9.65, predict as C; otherwise, predict as NC. (SAR 3)

In the case of multi-functional compounds, pKa is taken to be the minimal pKa value in SAR 2, and the maximal pKa value in SAR 3. This prevents underestimation of the acidity of acids and the basicity of bases.

Although CT analysis can also be used to determine a pKa cut-off value of 10.41 for phenols, above which they are predicted to be corrosive (with a sensitivity of 50%), and below which they are predicted to be non-corrosive (with a specificity of 100%), such a model was considered misleading on mechanistic grounds, since the phenols are weak acids with pKa values in the range 9.58 (3-methoxyphenol) to 11.34 (2-tert-butylphenol), and yet such a model is analogous to the model for bases, implying that the corrosive effects of phenols results from their basicity. At the pH of the skin (about 5.5), phenols exist predominantly in their undissociated form (Ar-OH), rather than in their phenoxide form ($Ar\text{-}O^-$). It therefore seems unlikely that the corrosivity of phenols is based on a pH-dependent mechanism.

Structure-activity relationships for skin irritation

A similar approach was adopted for the development of SARs for skin irritation potential. Initially, the data set of 139 chemicals (Table 7.2) was divided into two groups, one containing 41 chemicals predicted to be solids, and the other containing 98 chemicals predicted to be liquids. Of the 41 predicted solids, 28 (68%) are non-irritant, as expected, whereas 13 (32%) are irritants, contrary to expectation. The subsequent application of CT analysis to the data set of liquids, using MW, logP, logKp, ST, DM, E_{LUMO} and E_{HOMO} as input variables, identified ST as the best (and only) variable for

discriminating between I and NI chemicals, along which the optimal cut-off was found to be 35.8 dyne/cm. On this basis, the following SAR can be proposed:

If ST ≤ 35.8 dynes/cm, predict as I; otherwise, predict as NI. (SAR 4)

The dyne is a unit of force equal to 10^{-5} Newtons (N). The ST in dynes/cm can be converted into N/m by dividing by 10^3. The Cooper statistics obtained by applying SAR 4 are given in Table 7.6.

7.3.2 Prediction models

Prediction models based on pH data

The application of CT analysis to the pH data in Table 7.3 produced the CT shown in Figure 7.1. The CT is interpreted by reading from the root node (node 1) at the top of the tree to the terminal nodes (nodes 2, 4 and 5) at the bottom. Before the splitting process begins, all 75 observations are placed in node 1. According to the first decision rule, which is applied to all observations, observations with pH values less than 2.4 are placed in node 2 and predicted to be corrosive; otherwise they are placed in node 3 and subjected to the second decision rule. According to the second rule, observations are placed in node 4 and predicted to be non-corrosive if they have pH values less than or equal to 10.9; otherwise, they are placed in node 5 and predicted to be corrosive. The numbers above each node show how many observations (chemicals) are sent to each node, and the histograms illustrate the relative proportions of corrosive and non-corrosive chemicals in each node. The CT for skin corrosion potential can be summarised in the form of PM 1. Similarly, the CT for skin irritation potential (not shown) can be expressed as PM 2.

If pH < 2.4 or if pH > 10.9, then predict as C; otherwise, predict NC. (PM 1)
If pH < 4.4 or if pH > 9.2, then predict as I; otherwise, predict NI. (PM 2)

where pH is measured for a 10% solution (w/v in the case of liquids, and w/w in the case of solids. It is suggested that PM 1 is applied first, and that PM 2 is applied to those chemicals predicted to be NC by PM 1. The Cooper statistics obtained by applying PMs 1 and 2, including the application of PM 1 to an independent test set, are given in Table

7.7. The independent data set consisted of 53 chemicals from the ECVAM Skin Corrosivity Study for which pH data had been obtained for a 10% solution (Table 7.4).

Prediction models based on in vitro data

The application of CT analysis to the EPISKIN data for 59 chemicals (Table 7.4) produced the following PMs:

If the EPISKIN viability after 4h exposure < 36%, predict C;
otherwise, predict NC. (PM 3)

If the EPISKIN viability after 4h exposure < 67%, predict I;
otherwise, predict NI. (PM 4)

It is suggested that PM 4 is applied to chemicals predicted to be NC by PM 3.

A PM for predicting skin corrosion potential was also derived from *in vitro* data obtained by the rat skin transcutaneous electrical resistance (TER) assay:

If TER after 2h exposure < 8.8 kΩ, predict C; otherwise, predict NC. (PM 5)

The 2-hour endpoint was selected by the CT algorithm in preference to the 24-hour endpoint. A scatter plot of 2-hour vs 24-hour TER data showed no discrimination between I and NI chemicals, but this was not expected, since the TER assay is designed to detect chemicals with the potential to lyse the stratum corneum, but not chemicals that exert milder, inflammatory effects. The Cooper statistics obtained by applying PMs 3, 4 and 5 are given in Table 7.7.

7.4 DISCUSSION

In this chapter, a number of CMs for predicting the skin irritation and corrosion potentials of organic and inorganic chemicals are reported. These models either take the form of SARs based on easily obtained or calculated physicochemical data, or the form of PMs based on physicochemical or *in vitro* measurements. Since PMs are based on experimental measurements, it is important that the data used to develop the PMs have

been obtained by a standardised protocol. The pH measurements analysed in this study are standardised in the sense that they refer to measurements obtained in a 10% aqueous solution. The *in vitro* data (TER and EPISKIN) were taken from a validation study in which the use of standardised protocols was a requirement.

7.4.1 Contribution to existing knowledge

An important consideration of this study was that the models developed should take the form of explicit and objective algorithms for predicting animal-based classifications of acute local toxicity. The value of such decision rules is that they can be applied by a range of users (e.g. industrial toxicologists who need to perform in-house risk assessments) without the need for specialised software. This kind of model is lacking in many of the previous studies described in Chapter 2. In one of these studies, linear discriminant analysis was used (Barratt *et al.*, 1996a), but no attempt was made to define a linear model. In another study (e.g. Barratt *et al.*, 1996b), models with excellent performance statistics were reported, but the models were derived by using neural network analysis, from which explicit decision rules could not be formulated. An argument that can be made against the development of explicit models is that they are necessarily based on the assumption of a particular mathematical form, which may not be correct; for example, in this chapter, all of the models are based on simple cut-off values along one or more predictor variables. Against this argument, one of the contentions of this thesis is that the 'true' form of a model may never be known (indeed, there may be no such thing as a true model). Therefore, a pragmatic choice must be made according to the following considerations: a) the model should be no more complicated than necessary; b) it should make sense in mechanistic terms; and c) it should meet one or more criteria of predictive performance.

A similar approach has recently been reported by Gerner and colleagues, who developed decision rules on the basis of confidential industrial data (Gerner *et al.*, 2000; Zinke *et al.*, 2000). The main difference between the work of Gerner and colleagues and the work presented here is that the former is based on a series of exclusion rules to detect the *absence* of toxic potential (e.g. if MW > 1200 g/mol, then the substance has no local toxic effects). The emphasis of those papers was to describe the successful application of the decision rules, but very few rules were cited.

The results of this study confirm the importance of certain physicochemical properties previously reported to be useful predictors of skin corrosion potential (MP, MW, pKa and pH). In addition, a statistically significant relationship between ST and skin irritation potential was found. CMs developed on the basis of these properties were found to satisfy a minimal set of predictivity criteria, with the exception of the CM based on ST.

LogP was also found to discriminate between corrosive and non-corrosive chemicals, but the direction of the separation was contrary to expectation: low logP values were associated with the presence of corrosion, whereas high logP values were associated with the absence of corrosion. An inverse relationship between corrosion potential and logP also emerged (but was not commented upon) in several PCA studies (Barratt, 1995b; Barratt *et al.*, 1998). It was decided that the apparent importance of logP may be a reflection of the importance of MW, resulting from the collinearity between logP and MW. It has been argued (e.g. Barratt, 1995b) that logP plays a role in skin corrosion on the basis that hydrophobic chemicals are more likely than hydrophilic ones to diffuse across the stratum corneum. However, if it is assumed that the rate-limiting step in the production of a corrosive response is the transfer of the applied chemical from its bulk phase (solid or liquid) into the skin, one could question the importance of logP on the grounds that this provides a measure of the ability of a chemical to partition between octanol and water, rather than between the neat substance and the stratum corneum, which would be the more appropriate partitioning process to model, given that in the Draize skin test, most liquids and solids are applied neat, rather than as aqueous solutions. In other words, the octanol-water partition coefficient may be a poor substitute for the liquid-stratum corneum partition coefficient.

As far as the author is aware, the usefulness of the dermal permeability coefficient (Kp) for predicting acute skin toxicity has not been reported. In this study, no evidence was found for a relationship between Kp and any of the following: skin corrosion potential, skin irritation potential, or skin irritation potency (PII). A possible explanation for this is that corrosive chemicals do not need to diffuse through the stratum corneum to exert their effects in the viable epidermis - they can simply penetrate by lysis or solubilisation. Irritant chemicals, on the other hand, are expected to penetrate by

diffusion, so some relationship might be expected between the extent of irritancy and the degree of penetration. However, since irritant chemicals can initiate the inflammatory process by acting upon corneocytes in the lower stratum corneum (chapter 2), they do not have far to penetrate before triggering an irritant response. Thus, it is likely that all chemicals (irritant and non-irritant) will traverse the stratum corneum during the time-course of the animal test (1-3 days).

PMs 1 and 2 were based on measured pH values. A PM for corrosion potential is already cited in the OECD Test Guideline for skin corrosion (see chapter 2). However, this does not provide any guidance on the concentration at which the pH measurements should be made. In this work, it is clearly stated that PMs 1 and 2 are based on pH measurements in a 10% solution (w/w for solids, w/v for liquids), and are being used to predict the corrosion potential (PM 1) or irritation potential (PM 2) of the neat substance. It is therefore possible that these PMs will over-predict the effects of substances applied to the skin as dilute solutions. As far as the author is aware, no model similar to PM 2 for the prediction of skin irritation potential has been published, although a PM based on the combined use of pH and acid-alkaline reserve measurements has been proposed (see Chapter 2). An advantage of the PMs based on pH over SAR based on pKa measurements is that the former are applicable to both inorganic and organic substances, and to mixtures, whereas the latter are only applicable to certain groups of pure organic chemicals.

The other PMs cited in this chapter are based on *in vitro* data. The PM for skin corrosion based on EPISKIN measurements (PM 3) is similar to the one used in the ECVAM Validation Study (in which a cut-off value of 35% viability was adopted). The PM for skin irritation potential (PM 4) based on EPISKIN data is novel in that no such model has been published in the literature, as far as the author is aware. The PM for skin corrosion based on TER data (PM 5) is different from the one used in the ECVAM Skin Corrosivity Validation Study (Fentem *et al*., 1998) in that a cut-off value of 8.8 kΩ is proposed instead of 5 kΩ. In addition, PM 5 is less restrictive in that neutral organics and surfactants are included in the domain of the model, whereas the PM used in the validation study introduces an exception clause for these chemicals, to reduce the number of false positives.

7.4.2 Interpretation and assessment of the models

The importance of MP can be related to the physical state of the substance under the conditions of Draize test. In this study, it was assumed that chemicals with a MP less than or equal to 37°C would exist as liquids in the test procedure, and that, in general, liquids would be more likely than solids to cause corrosion and irritation. The results confirm that there is indeed a relationship between physical state and the potential for acute skin toxicity. The fact that some solids are corrosive or irritant may relate to the fact that their MPs are not much higher than 37°C and that they exist as wax-like substances, which are more capable of penetrating into the skin than are solids with higher MPs. For example, carvacrol and thymol, which are both irritant and corrosive, have predicted MPs of 38°C and 38.1°C, respectively. In the case of other solids, such as benzene sulphonyl chloride (MP = 61°C), the corrosive response may be due to a more toxic derivative (e.g. benzene sulphonic acid).

The importance of MW is probably related to the fact that small molecules are more likely to penetrate into the skin and cause corrosion than are larger chemicals. An alternative explanation could be that chemicals with lower MWs are applied in greater molar amounts than chemicals with higher MWs, since a fixed volume (or weight) of test substance is applied in the Draize test (see Chapter 2). Indeed, this could be regarded as a limitation in the protocol of the Draize test, which could be improved by adopting a fixed molar dose of the test substance.

The acid dissociation descriptor, pKa, was found to be useful for predicting the corrosion potentials of organic acids (carboxylic acids and alcohols) and bases. Three alcohols (propargyl alcohol, 2-butyn-1,4-diol, 2-hydroxyethyl acrylate) were excluded from the training set of acids, since they were considered to be unrepresenatative of alcohols in general. The corrosivity of the first two alcohols, which are both α,β-hydroxy alkynes, has been attributed to their oxidation to conjugated aldehydes, which then act as electrophiles (Barratt *et al.*, 1998). The corrosivity of 2-hydroxyethyl acrylate can be associated with the presence of the acrylate group, rather than the hydroxy group, since acrylic acid is also corrosive.

It was hypothesised that ST might play a role in acute local toxicity, on the grounds that liquids with a low surface tension will have a greater tendency to spread across the skin surface and therefore be more toxic. While the results suggest that there is indeed an inverse relationship between ST and skin irritation potential, the CM based on ST did not have a satisfactory predictivity. Furthermore, no meaningful relationship between ST and skin corrosion potential could be derived, since the mean ST of the corrosive chemicals (38.6 dyne/cm) was not significantly different from the mean for non-corrosives (38.1 dyne/cm)

The acidicity/basicity descriptor pH provides a useful means of identifying substances that are irritant or corrosive to the skin by disrupting its pH balance (away from a value of about 5.5). It is noticeable that the cut-offs for PM 1 are further toward the extremes of the pH scale than are the corresponding cut-offs for PM 2, which is consistent with the fact that corrosion is a stronger effect than irritation.

The performance of each CM is summarised in terms of its standard and cross-validated Cooper statistics (Table 7.6 for the SARs; Table 7.7 for the PMs). Strictly speaking, Cooper statistics should not be compared between different CMs, unless these models have been applied to the same set of data (i.e. if predicted classifications have been made for the same set of chemicals). However, if Cooper statistics are accompanied (as they are in Tables 7.6 and 7.7) by 95% confidence intervals (CIs) that reflect chemical variability, it is acceptable to compare the statistics between models. Thus, in principle, the CMs presented in Tables 7.6 and 7.7 could be compared. However, it was not the purpose of this study to compare the CMs with each another, but simply to determine whether each model is ‘acceptable’ for incorporation into a tiered testing strategy. It is therefore necessary to define what ‘acceptable’ means in this context. Firstly, it is necessary (but not sufficient) that the contingency table for each model should be statistically significant. Indeed, the χ^2 values for all of the models reported in this chapter are significant. However, it is clear from the Cooper statistics that the models are not equally predictive, so additional criteria are required. A minimal requirement for any classification model is that its concordance, sensitivity and specificity should be significantly greater than 50%, i.e. the predicted classifications generated by the model should be more reliable than predictions obtained by chance alone. It can be seen that SAR 4, which predicts the skin irritation potential of liquids, has a specificity of 32%

and therefore an unacceptable false positive rate of 68%. However, SARs 1-3, and PMs 1-5 are all acceptable, according to the above-mentioned criteria.

7.4.3 Strategic use of the models

To integrate the use of acceptable CMs, a stepwise assessment strategy can be formulated on the basis of two premises: a) corrosion potential should be identified before irritation potential, since corrosive chemicals are necessarily irritant, but non-corrosives may be irritant or non-irritant; and b) the order of the steps should reflect the relative ease and cost of applying the CMs. Thus, the first step should be based on SARs, since these are the easiest CMs to apply, not being based on experimental data, and subsequent steps should be based on PMs, using physicochemical (pH) data before *in vitro* (EPISKIN) data. A general outline of this scheme is given in Figure 7.2, and a specific implementation is evaluated in Chapter 10.

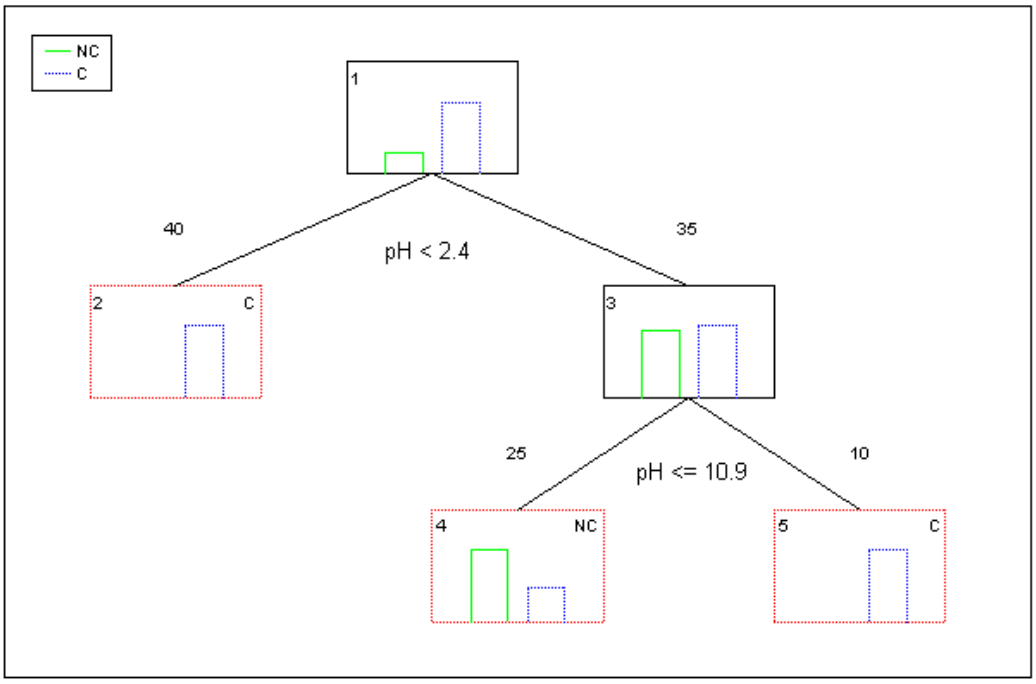
7.5 CONCLUSIONS

From the results of this study, it is concluded that:

1. The skin corrosion and irritation potentials of a diverse range of chemicals can be related to a few easily-obtained physicochemical properties (MP, MW, pKa, pH and ST). In particular, there is a tendency for solids to be non-corrosive and non-irritant. Liquids are corrosive if they have sufficiently low MWs, or if they are acids or bases with sufficiently extreme values of pKa. In addition, there is a tendency for liquids to be irritant if they have a sufficiently low ST. However, an SAR based on ST was not sufficiently predictive, according to the minimal set of criteria adopted in this chapter.

2. Acceptable PMs for skin corrosion and irritation potential can be derived from measured pH values, and from *in vitro* data obtained with the EPISKIN test.

Figure 7.1 Classification tree for distinguishing between corrosive and non-corrosive chemicals on the basis of pH measurements



Footnote to Figure 7.1

C = corrosive; NC = non-corrosive.

Figure 7.2 Outline of a tiered assessment strategy for acute local toxicity

Step 1: Apply SARs for corrosion ⇒ predict as C or NC
⇓
NC or no prediction
⇓
Step 2: Apply PM for corrosion based on pH data ⇒ predict as C or NC
⇓
NC or no prediction
⇓
Step 3: Apply PM for corrosion based on EPISKIN data ⇒ predict as C or NC
⇓
NC
⇓
Step 4: Apply PM for irritation based on pH data ⇒ predict I or NI
⇓
NI or no prediction
⇓
Step 5: Apply PM for irritation based on EPISKIN data ⇒ predict as I or NI
⇓
NI
⇓
Step 6: Perform Draize test ⇒ classify as C, I or NI

Table 7.1 Skin corrosion data set of 277 organic chemicals

	Chemical	Source	C/NC	MP	MW	ST	pKa1	pKa2	pKa3
1	1-Naphthoic acid	Barratt, 1996a	NC	106.7	172.2	56.5	3.68		
2	1-Naphthol	Barratt, 1996a	NC	67.7	144.2	51.0	9.40		
3	2,3-Lutidine	Barratt, 1996a	NC	-7.6	107.2	33.2	6.24		
4	2,3-Xylenol	Barratt, 1996a	C	25.4	122.2	37.2	10.42		
5	2,4,6-Trichlorophenol	Barratt, 1996a	NC	63.8	197.5	50.5	6.59		
6	2,4-Dichlorophenol	Barratt, 1996a	NC	46.8	163.0	47.8	8.05		
7	2,4-Dinitrophenol	Barratt, 1996a	NC	118.5	184.1	79.6	4.04		
8	2,4-Xylenol	Barratt, 1996a	C	25.4	122.2	37.2	10.61		
9	2,5-Dinitrophenol	Barratt, 1996a	NC	118.5	184.1	79.6	5.35		
10	2,5-Xylenol	Barratt, 1996a	C	25.4	122.2	37.2	10.42		
11	2,6-Xylenol	Barratt, 1996a	C	25.4	122.2	37.2	10.66		
12	2-Bromobenzoic acid	Barratt, 1995b	NC	81.6	201.0	53.7	2.85		
13	2-Butyn-1,4-diol	Barratt, 1996b	C	29.0	86.1	60.0	12.72	13.35	
14	2-Chlorobenzaldehyde	Barratt, 1996b	C	8.7	140.6	42.2			
15	2-Chloropropanoic acid	Barratt, 1996a	C	8.1	108.5	37.6	2.96		
16	2-Ethylphenol	Barratt, 1996a	NC	27.1	122.2	37.6	10.27		
17	2-Hydroxyethyl acrylate	Barratt, 1996b	C	-15.9	116.1	35.4	13.85		
18	2-Mercaptoethanoic acid	Barratt, 1996a	C	18.8	92.1	48.5	3.73	10.60	
19	2-Naphthoic acid	Barratt, 1996a	NC	106.7	172.2	56.5	4.20		
20	2-Naphthol	Barratt, 1996a	NC	67.7	144.2	51.0	9.57		
21	2-Nitrophenol	Barratt, 1996a	NC	70.8	139.1	60.2	7.14		
22	2-Phenylphenol	Barratt, 1996a	NC	86.6	170.2	44.5	9.99		
23	3-Methylbutanal	Barratt, 1996b	NC	-79.3	86.1	22.6			
24	3-Nitrophenol	Barratt, 1996a	NC	70.8	139.1	60.2	8.33		
25	3-Picoline	Barratt, 1996a	NC	-25.9	93.1	34.0	5.52		
26	3-Toluidine	Barratt, 1995b	NC	11.6	107.2	39.5	4.72		
27	4-Ethylbenzoic acid	Barratt, 1996a	NC	73.5	150.2	43.4	4.35		
28	4-Methoxyphenol	Barratt, 1996a	NC	25.2	124.1	38.6	10.40		

Table 7.1 Skin corrosion data set of 277 organic chemicals

	Chemical	Source	C/NC	MP	MW	ST	pKa1	pKa2	pKa3
29	4-Nitrophenol	Barratt, 1996a	NC	70.8	139.1	60.2	7.23		
30	4-Nitrophenylacetic acid	Barratt, 1996a	NC	124.3	181.2	61.9	3.86		
31	4-Picoline	Barratt, 1996a	NC	-25.9	93.1	34.0	5.94		
32	Acridine	Barratt, 1995b	NC	100.3	179.2	54.0	5.50		
33	Acrolein	Barratt, 1996b	C	-94.6	56.1	20.1			
34	Acrylic acid	Barratt, 1995b	C	-36.5	74.1	39.0	4.25		
35	Aminotris(methylphosphonic acid)	Barratt, 1996a	C	90.3	299.1	143.1	0.42	0.82	1.13
36	Barbituric acid	Barratt, 1996a	NC	199.0	128.1	48.7	4.00	11.46	17.35
37	Benzoic acid	Barratt, 1996a	NC	48.9	122.1	48.7	4.20		
38	Benzylamine	Barratt, 1996a	C	-6.2	93.1	41.7	4.61		
39	Butyric acid	Barratt, 1996a	C	3.0	88.1	32.5	4.76		
40	Catechol	Barratt, 1996a	NC	45.7	110.1	57.1	9.50	12.84	
41	Citric acid	Barratt, 1995b	NC	169.2	192.1	103.9	2.93	4.23	5.09
42	Cocoamine (dodecylamine)	Barratt, 1995b	C	35.1	185.4	29.9	10.67		
43	Cyanoacetic acid	Barratt, 1996a	C	38.0	85.1	57.0	2.47		
44	Cyclopropane carboxylic acid	Barratt, 1996a	C	13.0	86.1	57.9	4.78		
45	Decanoic acid	Barratt, 1995b	NC	62.7	172.3	33.1	4.79		
46	Formaldehyde	Barratt, 1996b	C	-110.9	30.0	12.6			
47	Fumaric acid	Barratt, 1996a	NC	84.1	116.1	67.6	3.15	4.79	
48	Glycolic acid	Barratt, 1996a	NC	23.3	76.1	61.3	3.74	16.46	
49	Glyoxylic acid	Barratt, 1996a	C	16.1	74.0	50.2	2.61		
50	Hexylcinnamic aldehyde	Barratt, 1996b	NC	44.4	216.3	35.7			
51	Hydrogenated tallow amine (hexadecylamine)	Barratt, 1996a	NC	75.6	241.5	30.7	10.67		
52	Hydroquinone	Barratt, 1996a	NC	45.7	110.1	57.1	10.33	11.86	
53	Imidazole	Barratt, 1995b	NC	18.5	68.1	48.6	6.95	14.52	
54	Iodoacetic acid	Barratt, 1996a	C	29.6	186.0	58.5	3.17		
55	Isobutanal	Barratt, 1996b	NC	-80.2	72.1	22.5			
56	Isobutyric acid	Barratt, 1996a	C	-8.3	88.1	30.4	4.85		

Table 7.1 Skin corrosion data set of 277 organic chemicals

	Chemical	Source	C/NC	MP	MW	ST	pKa1	pKa2	pKa3
57	Isoeugenol	Barratt, 1996a	NC	61.9	164.2	38.9	10.15		
58	Isoquinoline	Barratt, 1995b	NC	37.6	129.2	46.6	5.37		
59	Kojic acid	Barratt, 1996a	NC	96.2	142.1	71.5	8.05	13.78	
60	Lactic acid	Barratt, 1995b	C	22.7	90.1	49.7	3.90	16.52	
61	Malic acid	Barratt, 1996a	NC	112.7	134.1	86.2	3.61	4.82	16.29
62	Malonic (propanedioic) acid	Barratt, 1996a	NC	73.3	104.1	70.5	2.92	5.61	
63	3-Cresol	Barratt, 1995b	C	15.7	108.1	38.8	10.07		
64	Methoxyacetic acid	Barratt, 1996a	C	8.7	90.1	35.5	3.54		
65	Methyl isothiocyanate	Barratt, 1996b	C	-63.3	73.1	26.3			
66	Morpholine	Barratt, 1995b	C	-15.2	87.1	29.2	8.97		
67	Myristic (tetradecanoic) acid	Barratt, 1995b	NC	99.7	228.4	33.3	4.78		
68	2-Cresol	Barratt, 1995b	C	15.7	108.1	38.8	10.31		
69	Oxalic (ethanedioic) acid	Barratt, 1995b	C	63.0	90.0	87.3	1.38	4.28	
70	4-Cresol	Barratt, 1995b	C	15.7	108.1	38.8	10.21		
71	Propargyl alcohol	Barratt, 1996b	C	-49.0	56.1	37.9	13.21		
72	Propylphosphonic acid	Barratt, 1996a	C	28.3	124.1	48.9	2.43	8.10	
73	Pyridine	Barratt, 1995b	NC	-44.5	79.1	35.2	5.32		
74	Pyruvic acid	Barratt, 1996a	C	28.2	88.1	42.6	2.65		
75	Quinoline	Barratt, 1995b	NC	37.6	129.2	46.6	4.97		
76	Salicylic acid	Barratt, 1995b	NC	93.8	138.1	64.4	3.01	13.70	
77	Succinic acid	Whittle, 1996	NC	83.3	118.1	61.6	4.24	5.52	
78	Thymol	Barratt, 1996a	C	38.1	150.2	34.9	10.59		
79	*trans*-Cinnamic acid	Barratt, 1995b	NC	69.5	148.2	49.7	3.88		
80	3-Methoxyphenol	Barratt, 1996a	NC	25.2	124.1	38.6	9.58		
81	4-Ethylphenol	Barratt, 1996a	NC	27.1	122.2	37.6	10.25		
82	Phenol	Barratt, 1995b	C	-2.3	94.1	40.9	9.86		
83	1,1,1-Trichloroethane	ECETOC, 1995	NC	-72.0	133.4	28.9			
84	1,13-Tetradecadiene	ECETOC, 1995	NC	-1.2	194.4	26.5			

Table 7.1 Skin corrosion data set of 277 organic chemicals

	Chemical	Source	C/NC	MP	MW	ST	pKa1	pKa2	pKa3
85	1,3-Dibromopropane	ECETOC, 1995	NC	-27.0	201.9	36.0			
86	1,5-Hexadiene	ECETOC, 1995	NC	-96.7	82.2	19.9			
87	1,6-Dibromohexane	ECETOC, 1995	NC	7.9	244.0	35.2			
88	1,9-Decadiene	ECETOC, 1995	NC	-46.8	138.3	24.2			
89	10-Undecenoic Acid	ECETOC, 1995	NC	71.5	184.3	33.4	4.78		
90	1-Bromo-2-chloroethane	ECETOC, 1995	NC	-58.0	143.4	30.4			
91	1-Bromo-4-chlorobutane	ECETOC, 1995	NC	-33.6	171.5	31.3			
92	1-Bromo-4-fluorobenzene	ECETOC, 1995	NC	-19.1	175.0	33.8			
93	1-Bromohexane	ECETOC, 1995	NC	-41.6	165.1	27.7			
94	1-Bromopentane	ECETOC, 1995	NC	-53.8	151.1	26.9			
95	1-Decanol	ECETOC, 1995	NC	7.9	158.3	29.8	15.21		
96	1-Formyl-1-methyl-4(4-methyl-3-penten-1-yl)-3-cyclohexane	ECETOC, 1995	NC	46.5	208.4	34.2			
97	2,3-Dichloroproprionitrile	ECETOC, 1995	NC	-21.2	124.0	37.5			
98	2,4-Decadienal	ECETOC, 1995	NC	6.0	154.3	31.1			
99	2,4-Dimethyl-3-cyclohexene-1-carboxaldehyde	ECETOC, 1995	NC	-10.1	138.2	34.4			
100	2,4-Dimethyltetrahydrobenzaldehyde	ECETOC, 1995	NC	-10.1	138.2	34.4			
101	2,4-Dinitromethylaniline	ECETOC, 1995	NC	108.9	197.2	65.4	-3.46		
102	2,4-Hexadienal	ECETOC, 1995	NC	-56.2	96.1	26.4			
103	2,4-Xylidine	ECETOC, 1995	NC	34.7	135.2	36.2	9.57		
104	2,5-Methylene-6-propyl-3-cyclo-hexen-carbaldehyde	ECETOC, 1995	NC	15.2	164.3	39.5			
105	2,6-Dimethyl-2,4,6-octatriene	ECETOC, 1995	NC	-21.2	134.2	26.0			
106	2,6-Dimethyl-4-heptanol	ECETOC, 1995	NC	-38.1	144.3	26.6	15.31		
107	2-Bromobutane	ECETOC, 1995	NC	-78.1	137.0	24.6			
108	2-Bromopropane	ECETOC, 1995	NC	-91.0	123.0	23.1			
109	2-Chloronitrobenzene	ECETOC, 1995	NC	48.8	157.6	48.3			
110	2-Ethoxyethyl methacrylate	ECETOC, 1995	NC	-25.2	158.2	27.9			
111	2-Ethylhexanal	ECETOC, 1995	NC	-42.3	128.2	25.7			

Table 7.1 Skin corrosion data set of 277 organic chemicals

	Chemical	Source	C/NC	MP	MW	ST	pKa1	pKa2	pKa3
112	2-Ethylhexylpalmitate	ECETOC, 1995	NC	117.2	368.7	30.7			
113	2-Fluorotoluene	ECETOC, 1995	NC	-54.2	110.1	27.6			
114	2-Methoxyethyl acrylate	ECETOC, 1995	C	-56.2	128.2	23.7			
115	2-Methoxyphenol (guaiacol)	Barratt, 1996a	NC	25.2	124.1	38.6	9.97		
116	2-Methyl-4-phenyl-2-butanol	ECETOC, 1995	NC	30.4	164.3	36.1	15.17		
117	2-Methylbutyric acid	ECETOC, 1995	C	3.6	102.1	30.9	4.80		
118	2-Phenylethanol (phenylethylalcohol)	ECETOC, 1995	NC	5.8	122.2	51.4	13.61	15.04	
119	2-Phenylpropanal (2-phenylpropionaldehyde)	ECETOC, 1995	NC	-10.0	134.2	34.3			
120	2-*tert*-Butylphenol	ECETOC, 1995	C	36.9	150.2	32.9	11.34		
121	3,3'-Dithiopropionic acid	ECETOC, 1995	NC	141.5	210.3	68.0	3.94	4.56	
122	3,7-Dimethyl-2,6-nonadienal	ECETOC, 1995	NC	-3.9	180.3	27.3			
123	3-Chloro-4-fluoronitrobenzene	ECETOC, 1995	NC	44.2	175.6	45.9			
124	3-Diethylaminopropionitrile	ECETOC, 1995	NC	-0.4	126.2	32.1	9.16		
125	3-Mercapto-1-propanol	ECETOC, 1995	NC	-33.6	92.2	35.9	10.39	15.17	
126	3-Methoxypropylamine	ECETOC, 1995	NC	-40.4	89.1		9.73		
127	3-Methylphenol	ECETOC, 1995	NC	15.7	108.1	38.8	10.07		
128	3-Methylbutyraldehyde	ECETOC, 1995	NC	-79.3	86.1	22.6			
129	4-(Methylthio)-benzaldehyde	ECETOC, 1995	NC	28.6	152.2	43.3			
130	4,4'-Methylene-bis-(2,6-ditert-butylphenol)	ECETOC, 1995	NC	208.5	424.7	33.5	12.39	13.10	
131	4-Amino-1,2,4-triazole	ECETOC, 1995	NC	31.0	84.1	80.6	3.49		
132	4-Tricyclo-decylindene-8-butanal	ECETOC, 1995	NC	233.9	494.9	37.1			
133	6-Butyl-2,4-dimethyldihydropyrane	ECETOC, 1995	NC	-2.3	168.3	35.0			
134	α-Hexyl cinnamic aldehyde	ECETOC, 1995	NC	44.4	216.3	35.7			
135	α-Ionol	ECETOC, 1995	NC	45.2	194.3	34.9	14.52		
136	Allyl bromide	ECETOC, 1995	C	-80.5	121.0	24.7			
137	Allyl heptanoate	ECETOC, 1995	NC	-10.8	170.3	28.4			
138	Allyl phenoxyacetate	ECETOC, 1995	NC	36.5	192.2	36.2			
139	α-Terpineol	ECETOC, 1995	NC	12.4	154.3	33.2	15.09		
140	α-Terpinyl acetate	ECETOC, 1995	NC	21.5	196.3	31.8			

Table 7.1 Skin corrosion data set of 277 organic chemicals

Chemical	Source	C/NC	MP	MW	ST	pKa1	pKa2	pKa3
141 Benzyl acetate	ECETOC, 1995	NC	-0.5	150.2	35.9			
142 Benzyl acetone	ECETOC, 1995	NC	12.8	148.2	34.1			
143 Benzyl alcohol	ECETOC, 1995	NC	-5.4	108.1	40.7	14.36		
144 Benzyl benzoate	ECETOC, 1995	NC	70.8	212.3	44.0			
145 Benzyl salicylate	ECETOC, 1995	NC	115.5	228.3	51.4	8.11		
146 β-Ionol	ECETOC, 1995	NC	54.5	194.3	35.8	14.41		
147 Butyl propanoate	ECETOC, 1995	NC	-44.6	130.2	25.8			
148 Carvacrol	ECETOC, 1995	C	38.1	150.2	34.9	10.37		
149 Cinnamaldehyde	ECETOC, 1995	NC	0.0	132.2	38.9			
150 Cinnamyl alcohol	ECETOC, 1995	NC	15.8	134.2	42.6	14.61		
151 *cis*-Cyclooctene	ECETOC, 1995	NC	-58.8	110.2	28.2			
152 *cis*-Jasmone	ECETOC, 1995	NC	40.2	164.3	31.1			
153 Citrathal	ECETOC, 1995	NC	4.8	226.4	27.1			
154 Cyclamen aldehyde	ECETOC, 1995	NC	29.1	190.3	32.4			
155 Diacetyl	ECETOC, 1995	NC	-41.7	86.1	26.4			
156 Dichloromethane	ECETOC, 1995	NC	-89.5	84.9	23.1			
157 Diethyl phthalate	ECETOC, 1995	NC	-1.7	222.2	39.3			
158 Diethylaminopropylamine	ECETOC, 1995	C	0.7	130.2	30.8	8.94	10.54	
159 Dihydromercenol	ECETOC, 1995	NC	-10.6	156.3	27.7	15.29		
160 Dimethyl disulphide	ECETOC, 1995	NC	-69.7	94.2	32.1			
161 Dimethylbenzylcarbinyl acetate	ECETOC, 1995	NC	28.3	192.3	33.9			
162 Dimethyldipropylenetriamine	ECETOC, 1995	C	40.4	159.3	33.2	7.36	9.56	10.30
163 Dimethylisopropylamine	ECETOC, 1995	C	-95.4	87.2	21.0	9.91		
164 Dimethyl butylamine	ECETOC, 1995	C	-70.6	101.2	23.4	9.83		
165 Dipropyl disulphide	ECETOC, 1995	NC	-21.8	150.3	32.8			
166 Dipropylene glycol	ECETOC, 1995	NC	6.1	134.2	36.4	14.18	14.80	
167 dl-Citronellol	ECETOC, 1995	NC	-12.2	156.3	28.5	15.13		
168 d-Limonene	ECETOC, 1995	NC	-40.8	136.2	25.8			

Table 7.1 Skin corrosion data set of 277 organic chemicals

Chemical	Source	C/NC	MP	MW	ST	pKa1	pKa2	pKa3
169 Dodecanoic (lauric) acid	ECETOC, 1995	NC	81.9	200.3	33.2	4.78		
170 Erucamide	ECETOC, 1995	NC	183.4	337.6	33.4	16.61		
171 Ethyl thioethyl methacrylate	ECETOC, 1995	NC	-8.5	174.3	31.9			
172 Ethyl tiglate	ECETOC, 1995	NC	-53.9	128.2	26.1			
173 Ethyl triglycol methacrylate	ECETOC, 1995	NC	51.3	246.3	31.5			
174 Ethyl trimethyl acetate	ECETOC, 1995	NC	-68.4	116.2	24.6			
175 Eucalyptol	ECETOC, 1995	NC	8.1	154.3	32.4			
176 Eugenol	ECETOC, 1995	NC	60.6	164.2	36.5	10.29		
177 Fluorobenzene	ECETOC, 1995	NC	-73.0	96.1	27.4			
178 Geraniol	ECETOC, 1995	NC	-10.8	154.3	29.1	14.57		
179 Geranyl dihydrolinalol	ECETOC, 1995	NC	60.0	292.5	30.3	15.23		
180 Geranyl linalol	ECETOC, 1995	NC	58.5	290.5	30.4	14.41		
181 Glycol bromoacetate	ECETOC, 1995	C	1.2	303.9	48.4			
182 Heptanal	ECETOC, 1995	NC	-43.0	114.2	26.0			
183 Heptyl butyrate	ECETOC, 1995	NC	1.7	186.3	28.7			
184 Heptylamine	ECETOC, 1995	C	-21.6	115.2	26.0	10.69		
185 Hexyl salicylate	ECETOC, 1995	NC	99.7	222.3	40.6	8.17		
186 Hydroxycitronellal	ECETOC, 1995	NC	23.4	172.3	31.6			
187 Isobornyl acetate	ECETOC, 1995	NC	34.1	196.3	32.6			
188 Isobutyraldehyde	ECETOC, 1995	NC	-92.1	72.1	21.0			
189 Isopropanol	ECETOC, 1995	NC	-89.2	60.1	22.6	15.31		
190 Isopropyl isostearate	ECETOC, 1995	NC	80.6	326.6	29.8			
191 Isopropyl myristate	ECETOC, 1995	NC	44.4	270.5	29.6			
192 Isopropyl palmitate	ECETOC, 1995	NC	72.0	298.5	30.0			
193 Isostearic acid	ECETOC, 1995	NC	125.2	284.5	32.8	4.78		
194 Isostearyl alcohol	ECETOC, 1995	NC	77.3	270.5	30.8	15.19		
195 Lilestralis/lilial	ECETOC, 1995	NC	46.3	204.3	30.8			

Table 7.1 Skin corrosion data set of 277 organic chemicals

Chemical	Source	C/NC	MP	MW	ST	pKa1	pKa2	pKa3
196 Linalol	ECETOC, 1995	NC	-11.4	154.3	28.2	14.51		
197 Linalol oxide	ECETOC, 1995	NC	31.1	170.3	39.9	14.50		
198 Linalyl acetate	ECETOC, 1995	NC	-2.1	196.3	27.7			
199 Methacrolein	ECETOC, 1995	C	-90.6	70.1	20.8			
200 Methyl 2-methylbutyrate	ECETOC, 1995	NC	-68.4	116.2	24.6			
201 Methyl caproate	ECETOC, 1995	NC	-44.6	130.2	26.6			
202 Methyl laurate	ECETOC, 1995	NC	23.2	214.4	29.3			
203 Methyl lavender ketone (1-hydroxy-3-decanone)	ECETOC, 1995	NC	42.7	172.3	33.1	14.36		
204 Methyl linoleate	ECETOC, 1995	NC	70.8	294.5	31.2			
205 Methyl palmitate	ECETOC, 1995	NC	63.2	270.5	30.2			
206 Methyl stearate	ECETOC, 1995	NC	81.6	298.5	30.5			
207 Methyl trimethyl acetate	ECETOC, 1995	NC	-62.5	116.2	24.1			
208 Decylidene methyl anthranilate	ECETOC, 1995	NC	99.9	289.4	33.6	4.28		
209 N,N-Dimethylbenzylamine	ECETOC, 1995	NC	-12.8	135.2	32.7	8.80		
210 Nonanal	ECETOC, 1995	NC	-19.5	142.2	27.4			
211 Octanoic acid	ECETOC, 1995	C	48.4	144.2	33.0	4.78		
212 Oleyl propylene diamine dioleate	ECETOC, 1995	NC	142.1	324.6	32.7	8.73	10.79	
213 Phenethyl bromide	ECETOC, 1995	NC	2.5	185.1	39.6			
214 p-Isopropylphenylacetaldehyde	ECETOC, 1995	NC	18.4	162.2	33.3			
215 p-Mentha-1,8-dien-7-ol	ECETOC, 1995	NC	11.1	152.2	32.1	14.51		
216 p-*tert*-Butyl dihydrocinnamaldehyde	ECETOC, 1995	NC	46.3	190.3	31.5			
217 Salicylaldehyde	ECETOC, 1995	NC	42.6	122.1	52.0			
218 Tetrachloroethylene	ECETOC, 1995	NC	-60.6	165.8	35.6			
219 Tetrahydrogeranial	ECETOC, 1995	NC	-30.0	156.3	26.1			
220 Tonalid	ECETOC, 1995	NC	98.7	244.4	31.1			
221 Trichloroethylene	ECETOC, 1995	NC	-60.6	165.8	31.0			
222 1-(2-Aminoethyl)piperazine	NIH, 1999	C	53.7	129.2	33.7	4.35	8.56	10.11

Table 7.1 Skin corrosion data set of 277 organic chemicals

Chemical	Source	C/NC	MP	MW	ST	pKa1	pKa2	pKa3
223 1,2-Diaminopropane	NIH, 1999	C	-22.9	74.1	33.8	6.80	9.92	
224 1,4-Diaminobutane	NIH, 1999	C	0.9	88.2	35.8	9.26	10.68	
225 2,3-Dimethylcyclohexylamine	NIH, 1999	C	-11.1	127.2	26.3	10.70		
226 2-Ethylhexylamine	NIH, 1999	C	-21.0	129.3	27.4	10.75		
227 2-Mercaptoethanol	NIH, 1999	C	-45.6	78.1	37.0	9.74	15.06	
228 3-Diethylaminopropylamine	NIH, 1999	C	0.7	130.2	30.8	8.94	10.54	
229 Acetic acid	NIH, 1999	C	-21.3	60.1	31.9	4.79		
230 Acetic anhydride	NIH, 1999	C	-95.1	102.1	29.1			
231 Acetyl bromide	NIH, 1999	C	-53.0	123.0	30.5			
232 Benzene sulphonyl chloride	NIH, 1999	C	61.2	176.6	42.6			
233 Benzyl chloroformate	NIH, 1999	C	11.6	170.6	41.0			
234 Bromoacetic acid	NIH, 1999	C	29.2	139.0	51.3	2.73		
235 Bromoacetyl bromide	NIH, 1999	C	-1.7	201.9	45.6			
236 Butanoic acid	NIH, 1999	C	3.0	88.1	32.5	4.76		
237 Butylamine	NIH, 1999	C	-58.8	73.1	25.3	10.69		
238 Butylbenzene	NIH, 1999	NC	-23.3	134.2	30.0			
239 Butyric anhydride	NIH, 1999	C	-44.6	158.2	30.9			
240 Chloroacetic acid	NIH, 1999	C	10.9	94.5	41.9	2.65		
241 Crotonic acid	NIH, 1999	C	2.4	86.1	34.2	4.80		
242 Cyanuric chloride	NIH, 1999	C	68.8	184.4	64.3	-2.92		
243 Cyclohexylamine	NIH, 1999	C	-27.1	99.2	31.4	10.57		
244 Dichloroacetic acid	NIH, 1999	C	24.2	128.9	47.0	1.37		
245 Dichloroacetyl chloride	NIH, 1999	C	-32.5	147.4	36.7			
246 Dichlorophenyl phosphine	NIH, 1999	C	-4.9	179.0	*			
247 Dicyclohexylamine	NIH, 1999	C	27.7	181.3	33.1	11.43		
248 Diethylamine	NIH, 1999	C	-79.7	73.1	20.2	10.76		
249 Diethylene triamine	NIH, 1999	C	17.8	103.2	39.3	3.60	8.98	10.16
250 Dimethylcarbamyl chloride	NIH, 1999	C	-15.9	107.5	30.6	-1.85		

Table 7.1 Skin corrosion data set of 277 organic chemicals

Chemical	Source	C/NC	MP	MW	ST	pKa1	pKa2	pKa3
251 Dodecyl trichlorosilane	NIH, 1999	C	51.0	303.8	28.3			
252 Ethanolamine	NIH, 1999	C	-27.6	61.1	39.7	9.16	12.87	
253 Ethylene diamine	NIH, 1999	C	-23.8	60.1	36.8	6.77	9.89	
254 Formic acid	NIH, 1999	C	-25.0	46.0	35.8	3.74		
255 Fumaryl chloride	NIH, 1999	C	6.8	153.0	40.6			
256 Hexanoic acid	NIH, 1999	C	26.2	116.2	32.8	4.78		
257 Hexanol	NIH, 1999	NC	-37.9	102.2	27.9	15.37		
258 Maleic acid	NIH, 1999	NC	84.1	116.1	67.6	3.15	4.79	
259 Maleic anhydride	NIH, 1999	C	-51.6	98.1	53.7			
260 Mercaptoacetic acid	NIH, 1999	C	18.8	92.1	48.5	3.73	10.60	
261 Nonanol	NIH, 1999	NC	-3.2	144.3	29.5	15.22		
262 2-Anisoyl chloride	NIH, 1999	C	36.7	170.6	38.2			
263 Octadecyl trichlorosilane	NIH, 1999	C	107.7	387.9	29.6			
264 Octyl trichlorosilane	NIH, 1999	C	8.1	247.7	26.9			
265 Pentanoyl (valeryl) chloride	NIH, 1999	C	-42.4	120.6	27.8			
266 Phenyl acetyl chloride	NIH, 1999	C	13.7	154.6	39.2			
267 Phenyl trichlorosilane	NIH, 1999	C	5.8	211.6	31.6			
268 Propanoic acid	NIH, 1999	C	-9.0	74.1	32.3	4.79		
269 Pyrrolidine	NIH, 1999	C	-36.0	71.1	36.8	11.26		
270 Tetraethylenepentamine	NIH, 1999	C	112.7	189.3	41.1	3.22	3.88	9.13
271 Tributylamine	NIH, 1999	NC	0.8	185.4	27.4	9.99		
272 Trichloroacetic acid	NIH, 1999	C	26.7	163.4	53.0	1.10		
273 Trichlorotoluene	NIH, 1999	NC	10.4	195.5	38.7			

Table 7.1 Skin corrosion data set of 277 organic chemicals

Chemical	Source	C/NC	MP	MW	ST	pKa1	pKa2	pKa3
274 Triethanolamine	NIH, 1999	NC	83.3	149.2	54.9	7.77	14.17	14.73
275 Triethylene tetramine	NIH, 1999	C	68.2	146.2	40.4	3.14	6.60	9.53
276 Trifluoroacetic acid	NIH, 1999	C	-24.0	114.0	21.9	0.67		
277 Undecanol	NIH, 1999	NC	18.7	172.3	30.1	15.20		

Footnote to Table 7.1

C = corrosive; MP = melting point (°C); MW = molecular weight (g/mol); NC = non-corrosive;
pKa(n) = the nth value of the negative logarithm of the acid dissociation constant; ST = surface tension.

Table 7.2 Skin irritation data set of 139 organic chemicals

	Chemical[a]	C/NC	I/NI	Erythema	Oedema	PII
1	2-Methylbutyric acid	NC	I	2.3	2.7	5.0
2	10-Undecenoic Acid	NC	NI	1.8	0.6	2.4
3	Ethyltriglycol methacrylate	NC	NI	0.2	0.0	0.2
4	Ethylthioethyl methacrylate	NC	NI	0.6	0.0	0.6
5	2-Ethoxyethyl methacrylate	NC	NI	1.6	0.0	1.6
6	2-Methoxyethyl acrylate	C	I	3.6	3.5	7.1
7	Benzyl alcohol [2]	NC	NI	1.4	0.4	1.8
8	dl-Citronellol [1]	NC	I	2.0	2.2	4.2
9	1-Decanol	NC	I	2.2	1.2	3.3
10	Dihydromercenol [1]	NC	NI	1.3	0.2	1.6
11	2,6-Dimethyl-4-heptanol	NC	NI	0.0	0.0	0.0
12	Dipropylene glycol [2]	NC	NI	0.0	0.0	0.0
13	Geraniol [3]	NC	NI	1.9	1.0	2.9
14	Geranyl dihydrolinalol	NC	NI	1.4	0.8	2.3
15	Geranyl linalol	NC	I	2.2	2.1	4.3
16	α-Ionol	NC	NI	0.9	0.4	1.3
17	β-Ionol	NC	NI	1.2	0.7	1.9
18	Linalol [2]	NC	I	2.0	1.4	3.4
19	p-Mentha-1,8-dien-7-ol	NC	NI	2.0	1.3	3.3
20	2-Methyl-4-phenyl-2-butanol	NC	NI	1.3	0.3	1.6
21	Phenylethyl alcohol [2]	NC	NI	0.7	0.2	0.9
22	Isopropanol	NC	NI	0.8	0.0	0.8
23	Isostearyl alcohol	NC	NI	1.6	1.0	2.6
24	α-Terpineol [3]	NC	I	1.9	2.1	4.0
25	p-tert-Butyl dihydrocinnamaldehyde	NC	NI	1.6	0.8	2.4
26	Isobutyraldehyde	NC	NI	0.1	0.0	0.1
27	Cinnamaldehyde	NC	I	2.0	1.7	3.7
28	Citrathal	NC	NI	2.0	1.7	3.6
29	Cyclamen aldehyde [4]	NC	I	2.0	1.4	3.4
30	2,4-Decadienal	NC	I	2.3	2.5	4.8
31	2,4-Dimethyl-3-cyclohexen-1-carboxaldehyde	NC	I	2.0	1.2	3.2
32	3,7-Dimethyl-2,6-nondien-1-al (Ethyl citral)	NC	I	2.0	1.8	3.8
33	2,4-Dimethyltetrahydrobenzaldehyde	NC	NI	1.9	0.9	2.8
34	2-Ethylhexanal	NC	I	2.0	1.9	3.9
35	Heptanal	NC	I	3.0	2.0	5.0
36	2,4-Hexadienal	NC	I	3.6	3.5	7.1
37	α-Hexyl cinnamic aldehyde [2]	NC	I	2.0	2.0	4.0
38	Hydroxycitronellal [1]	NC	NI	0.9	0.2	1.1
39	Lilestralis/lilial [2]	NC	NI	1.9	1.7	3.6
40	3-Methylbutyraldehyde	NC	NI	1.6	1.3	2.8
41	2,5-Methylene-6-propyl-3-cyclo-hexene-carbaldehyde	NC	NI	1.8	0.6	2.4
42	Nonanal	NC	I	2.0	1.4	3.4
43	2-Phenylpropionaldehyde	NC	I	2.0	0.9	2.9
44	p-Isopropylphenylacetaldehyde	NC	NI	1.3	0.9	2.3
45	Salicylaldehyde	NC	NI	1.5	1.0	2.5
46	Tetrahydrogeranial	NC	NI	1.9	0.7	2.6
47	4-Tricyclo-decylindene-8-butanal	NC	I	2.1	1.2	3.3
48	Methacrolein	C	I	2.0	2.1	4.1
49	Erucamide	NC	NI	0.0	0.0	0.0

Table 7.2 Skin irritation data set of 139 organic chemicals

	Chemical[a]	C/NC	I/NI	Erythema	Oedema	PII
50	Diethylaminopropylamine	C	I	b	b	b
51	N,N-Dimethylbenzylamine	NC	I	b	b	b
52	Dimethyl-n-butylamine	C	I	4.0	1.1	5.1
53	Dimethylisopropylamine	C	I	4.0	1.6	5.6
54	Dimethyldipropylenetriamine	C	I	b	b	b
55	Heptylamine	C	I	4.0	2.7	6.7
56	3-Methoxy propylamine	C	I	4.0	2.7	6.7
57	Oleyl propylene diamine dioleate	NC	I	2.1	1.5	3.7
58	2,4-Xylidine	NC	NI	1.0	0.4	1.4
59	Allyl bromide	C	I	4.0	3.2	7.2
60	2-Bromobutane	NC	NI	1.8	0.7	2.5
61	1-Bromo-4-chlorobutane	NC	NI	0.0	0.0	0.0
62	1-Bromo-2-chloroethane	NC	NI	1.8	0.5	2.3
63	1-Bromo-4-fluorobenzene	NC	NI	0.3	0.0	0.3
64	1-Bromohexane	NC	I	2.4	2.4	4.0
65	1-Bromopentane	NC	I	2.7	1.8	4.5
66	2-Bromopropane	NC	NI	1.4	0.0	1.4
67	1,6-Dibromohexane	NC	NI	0.9	0.0	0.9
68	1,3-Dibromopropane	NC	NI	1.9	0.0	1.9
69	Phenethyl bromide	NC	NI	0.0	0.0	0.0
70	Dichloromethane	NC	I	4.0	1.7	5.7
71	Tetrachloroethylene	NC	I	4.0	1.7	5.7
72	1,1,1-Trichloroethane	NC	I	3.7	1.5	5.2
73	Trichloroethylene	NC	I	4.0	1.4	5.4
74	Allyl heptanoate	NC	NI	1.7	0.4	2.1
75	Allyl phenoxyacetate	NC	NI	0.3	0.1	0.4
76	Benzyl acetate [2]	NC	NI	0.8	0.0	0.8
77	Benzyl benzoate [1]	NC	NI	0.0	0.0	0.0
78	Benzyl salicylate [1]	NC	NI	0.3	0.0	0.3
79	Isobornyl acetate [1]	NC	I	2.0	2.0	4.0
80	Butyl propanoate	NC	NI	1.1	0.0	1.1
81	Diethyl phthalate [1]	NC	NI	0.0	0.0	0.0
82	2-Ethylhexylpalmitate	NC	NI	0.6	0.0	0.6
83	Ethyl tiglate	NC	NI	1.0	0.1	1.2
84	Ethyl trimethyl acetate	NC	NI	0.4	0.0	0.4
85	Glycolbromoacetate	C	I	b	b	b
86	Heptyl butyrate	NC	NI	1.5	0.2	1.7
87	Hexyl salicylate [4]	NC	I	2.0	1.3	3.3
88	Linalyl acetate [1]	NC	NI	1.9	1.8	3.7
89	Methyl caproate	NC	NI	1.7	1.1	2.8
90	Methyl laurate	NC	I	2.0	1.9	3.9
91	Methyl linoleate	NC	NI	1.9	1.2	3.1
92	Methyl 2-methylbutyrate	NC	NI	0.6	0.0	0.7
93	Methyl palmitate	NC	I	2.6	2.0	4.6
94	Methyl stearate	NC	NI	1.5	0.7	2.1
95	Methyl trimethyl acetate	NC	NI	0.0	0.0	0.0
96	Isopropyl myristate	NC	NI	1.0	0.2	1.2
97	Isopropyl palmitate	NC	NI	1.2	0.2	1.4
98	Isopropyl isostearate	NC	NI	0.1	0.0	0.1

Table 7.2 Skin irritation data set of 139 organic chemicals

	Chemical[a]	C/NC	I/NI	Erythema	Oedema	PII
99	α-Terpinyl acetate [2]	NC	I	2.0	2.3	4.3
100	Eucalyptol	NC	NI	1.9	0.4	2.2
101	Octanoic acid	C	I	3.0	1.4	4.5
102	Isostearic acid	NC	I	1.9	2.4	4.3
103	3-Chloro-4-fluoronitrobenzene	NC	NI	1.0	0.2	1.2
104	Fluorobenzene	NC	NI	0.1	0.0	0.1
105	2-Fluorotoluene	NC	NI	0.1	0.0	0.1
106	*cis*-Cyclooctene	NC	NI	1.7	0.4	2.1
107	1,9-Decadiene	NC	I	2.0	1.0	3.0
108	1,5-Hexadiene	NC	NI	0.0	0.0	0.0
109	1,13-Tetradecadiene	NC	NI	1.3	0.2	1.6
110	Benzyl acetone	NC	NI	0.8	0.2	1.0
111	Diacetyl	NC	NI	0.6	0.0	0.6
112	*cis*-Jasmone	NC	NI	1.7	1.7	2.6
113	Methyl lavender ketone (1-hydroxy-3-decanone)	NC	I	2.2	1.6	3.8
114	2,3-Dichloropropionitrile	NC	I	2.0	0.0	2.0
115	3-Diethylaminopropionitrile	NC	NI	0.0	0.0	0.0
116	Carvacrol	C	I	4.0	1.7	5.7
117	2-*tert*-Butylphenol	C	I	4.0	1.7	5.7
118	Eugenol	NC	NI	1.9	1.0	2.9
119	Guaiacol (2-methoxyphenol)	NC	NI	1.8	0.6	2.4
120	4,4'-Methylene bis (2,6-di-*tert*-butyl phenol)	NC	NI	0.0	0.0	0.0
121	Dimethyl disulphide	NC	NI	1.8	1.2	3.0
122	Dipropyl disulphide	NC	I	2.6	0.0	2.6
123	3-Mercapto-1-propanol	NC	NI	0.8	0.3	1.1
124	4-(Methylthio)benzaldehyde	NC	NI	0.9	0.0	0.9
125	6-Butyl-2,4-dimethyldihydropyrane	NC	NI	1.7	0.3	2.0
126	Decylidene methyl anthranilate	NC	NI	1.5	0.5	2.1
127	2,6-Dimethyl-2,4,6-octatriene	NC	I	2.0	1.0	3.0
128	1-Formyl-1-methyl-4(4-methyl-3-penten-1-yl)-3-cyclohexane	NC	I	2.0	1.3	3.3
129	d-Limonene [2]	NC	NI	1.9	1.3	3.2
130	Linalol oxide	NC	NI	2.0	0.6	2.6
131	Cinnamyl alcohol [1]	NC	NI	0.0	0.0	0.0
132	2,4-Dinitromethylaniline	NC	NI	0.0	0.0	0.0
133	Dimethylbenzylcarbinyl acetate [1]	NC	NI	1.0	0.2	1.2
134	Dodecanoic (lauric) acid	NC	NI	0.4	0.0	0.4
135	2-Chloronitrobenzene	NC	NI	0.0	0.0	0.0
136	3-Methyl phenol	NC	NI	0.0	0.0	0.0
137	3,3'-Dithiopropionic acid	NC	NI	0.0	0.0	0.0
138	4-Amino-1,2,4-triazole	NC	NI	0.0	0.0	0.0
139	Tonalid	NC	NI	0.0	0.0	0.0

Footnote to Table 7.2

I = irritant; NC = non-corrosive; NI = non-irritant.

[a]The numbers in square brackets refer to the data sheet taken from the ECETOC Skin Irritation Databank in cases where there was more than one data sheet per chemical; [b]Not possible to calculate because some or all tissue scores were missing.

Table 7.3 pH data for 75 organic and inorganic chemicals[a]

Chemical		C/NC	EU CLASS	pH[b]
1	Benzyl chloroformate	NC	NC	2.5
2	Fluorosulphonic acid	C	R34	0
3	Nitric acid	C	R34	0
4	Selenic acid	C	R34	0
5	Sulphur monochloride	C	R34	5.2
6	Trifluoroacetic acid	C	R34	0.8
7	Acetic acid	C	R34	2.3
8	Acetic anhydride	C	R34	2
9	Ethylhexylamine	C	R35	12
10	Acrylic acid	C	R34	2.1
11	Aluminium bromide	C	R34	1.2
12	Aluminium chloride	C	R34	3
13	Ammonium hydrogen difluoride	C	R34	5.2
14	Ammonium hydrogen sulphate	C	R34	0.8
15	2-Anisoyl chloride	C	R34	0.7
16	Antimony tribromide	C	R34	0.4
17	Antimony trichloride	C	R34	0.3
18	Dicyclohexylamine	C	R35	9.6
19	Boron trifluoride-acetic acid	C	R34	1
20	Bromoacetic acid	C	R34	1.4
21	Butanoic anhydride	C	R35	3.1
22	Chloroacetic acid	C	R34	1.4
23	Cyanuric chloride	C	R35	1.7
24	Cyclohexylamine	C	R34	12.3
25	Dichloroacetic acid	C	R34	0.6
26	Dichloroacetyl chloride	C	R34	0.5
27	Dichlorophenyl phosphine	C	R34	0
28	Diethylene triamine	C	R34	12
29	Benzene sulphonyl chloride	C	R35	1.8
30	Dimethyl carbamyl chloride	C	R34	2.3
31	Crotonic acid	C	R35	2.3
32	Dodecyl trichlorosilane	C	R34	0.5
33	Ethylene diamine	C	R34	12.1
34	Ferrous chloride	C	R34	2.1
35	Fluroboric acid	C	R34	1.3
36	Formic acid	C	R34	1.6
37	Fumaryl chloride	C	R34	0.1
38	Hydrobenzene sulphonic acid	C	R34	0.6
39	Hydrogen bromide	C	R34	0.3
40	Iodine monochloride	C	R34	0.8
41	Lithium hydroxide	C	R34	11.8
42	Mercaptoacetic acid	C	R34	0.3
43	Octadecyl trichlorosilane	C	R34	0.3
44	Octyl trichlorosilane	C	R34	0.1
45	Phenylacetyl chloride	C	R34	0.9
46	Phenyl trichlorosilane	C	R34	0
47	Tetraethylene pentamine	C	R35	11.9
48	Hydroxylamine sulphate	C	R35	3.6
49	Potassium hydrogen sulphate	C	R34	0.9

Table 7.3 cont'd pH data for 75 organic and inorganic chemicals[a]

Chemical		C/NC	EU CLASS	pH[b]
50	Potassium hydroxide	C	R34	14
51	Sodium hydrogen fluoride	C	R35	5.2
52	Sodium hydroxide	C	R34	13.8
53	Sulphuric acid	C	R34	0
54	Sulphurous acid	C	R34	1.8
55	Tetramethyl ammonium hydroxide	C	R34	13.6
56	Thiophosphoryl chloride	C	R34	5.8
57	Trichloroacetic acid	C	R34	0.7
58	Trichlorotoluene	NC	NC	3.3
59	Triethylene tetramine	C	R34	11.9
60	Valeryl chloride	C	R34	0.5
61	Ethyl triglycol methacrylate	NC	NC	4.5
62	Nonyl acrylate	NC	NC	6.9
63	Fuel additive	NC	NC	4.1
64	Amphoteric surfactant	NC	NC	7.3
65	Anionic surfactant	NC	NC	8.3
66	Anionic surfactant	NC	NC	7.9
67	Liquid bleach	NC	NC	4.5
68	Liquid bleach	NC	NC	9.5
69	Degreaser	NC	NC	7.3
70	Dilutable cleaner	NC	NC	9.9
71	Dilutable cleaner	NC	NC	3.8
72	Concentrated organic cleaner	NC	NC	7.7
73	3,3'-Dithiopropanoic acid	NC	NC	3.3
74	Mercaptopropanol	NC	NC	6.4
75	Benzalkonium chloride	NC	NC	4.9

Footnote to Table 7.3

C = corrosive (R34 = moderate corrosive; R35 = severe corrosive); NC = non-corrosive.

[a]Data are taken from Gordon (1994).
[b]pH of 10% solution.

Table 7.4 Data for the 60 chemicals tested in the ECVAM Skin Corrosivity Study

Chemical		C/NC	PII	Erythema	Oedema	I/NI	pH[c]	EPISKIN 3m	EPISKIN 1h	EPISKIN 4h	TER 2h	TER 24h
1	Hexanoic acid	C	a	a	a	I	2.57	15.49	2.30	2.64	1.15	0.92
2	1,2-Diaminopropane	C	a	a	a	I	12.02	63.78	49.76	25.36	1.01	0.91
3	Carvacrol	C	5.7	4.0	1.7	I	4.91	126.04	16.17	16.28	6.50	2.91
4	Boron trifluoride dihydrate	C	a	a	a	I	0.57	2.26	3.79	3.43	0.58	0.54
5	Methacrolein	C	4.1	2.0	2.1	I	4.18	105.75	36.21	21.07	8.64	9.45
6	Phenethyl bromide	NC	0.0	0.0	0.0	NI	5.40	140.42	148.54	156.59	15.22	12.14
7	3,3'-Dithiodipropionic acid	NC	0.0	0.0	0.0	NI	3.22	98.31	108.10	99.79	14.41	16.35
8	Isopropanol	NC	0.8	0.8	0.0	NI	5.86	96.05	99.27	89.67	9.39	5.83
9	2-Methoxyphenol (Guaiacol)	NC	2.4	1.8	0.6	NI	4.86	143.83	16.27	8.87	11.73	9.40
10	2,4-Xylidine (2,4-Dimethylaniline)	NC	1.4	1.0	0.4	NI	8.73	117.38	54.50	39.40	9.18	3.32
11	2-Phenylethanol (phenylethylalcohol)	NC	0.9	0.7	0.2	NI	5.31	115.65	106.72	69.14	12.62	11.73
12	Dodecanoic (lauric) acid	NC	0.4	0.4	0.0	NI	4.72	106.41	121.01	112.69	18.13	16.11
13	3-Methoxypropylamine	C	6.7	4.0	2.7	I	11.78	50.87	41.83	17.60	1.34	1.04
14	Allyl bromide	C	7.2	4.0	3.2	I	3.15	117.84	28.23	22.46	4.61	4.58
15	Dimethyldipropylenetriamine	C	b	b	b	I	11.38	79.23	38.39	17.07	1.21	1.29
16	Methyl trimethylacetate	NC	0.0	0.0	0.0	NI	4.96	112.89	96.97	70.56	8.78	4.89
17	Dimethylisopropylamine	C	5.6	4.0	1.6	I	11.81	78.22	22.30	13.99	1.19	0.93
18	Potassium hydroxide (10% aq.)	C	b	b	b	I	13.76	64.19	15.30	11.11	1.14	1.30
19	Tetrachloroethylene	NC	5.7	4.0	1.7	I	7.13	104.55	91.00	55.96	15.42	3.23
20	Ferric [iron (III)] chloride	C	a	a	a	I	1.11	91.04	66.28	33.24	3.78	0.98
21	Potassium hydroxide (5% aq.)	NC	5.2	3.2	2.0	I	13.67	44.42	12.11	9.85	1.64	1.76
22	Butyl propanoate	NC	1.1	1.1	0.0	NI	4.57	102.89	117.33	77.09	4.63	3.62
23	2-tert-Butylphenol	C	5.7	4.0	1.7	I	8.17	93.16	8.09	8.17	4.61	9.11
24	Sodium carbonate (50% aq.)	NC	2.3	1.7	0.7	I	10.95	115.11	112.61	88.39	8.60	3.62
25	Sulphuric acid (10% wt.)	C	a	a	a	I	0.33	94.95	22.08	2.12	9.13	4.14
26	Isostearic acid	NC	4.3	1.9	2.4	I	4.78	103.39	94.21	108.30	21.27	18.22
27	Methyl palmitate	NC	4.6	2.6	2.0	I	5.69	104.63	104.59	99.06	21.58	22.50

Table 7.4 cont'd Data for the 60 chemicals tested in the ECVAM Skin Corrosivity Study

Chemical		C/NC	PII	Erythema	Oedema	I/NI	pH[c]	EPISKIN 3m	EPISKIN 1h	EPISKIN 4h	TER 2h	TER 24h
28	Phosphorus tribromide	C	a	a	a	I	0.12	3.55	14.34	11.87	0.60	0.72
29	65/35 Octanoic/decanoic acids	C	a	a	a	I	3.72	108.83	6.16	4.25	3.78	1.74
30	4,4'-Methylene-bis-(2,6-ditert-butylphenol)	NC	0.0	0.0	0.0	NI	6.44	120.08	99.06	103.29	14.09	13.84
31	2-Bromobutane	NC	2.5	1.8	0.7	NI	3.89	132.81	121.21	55.02	5.09	4.67
32	Phosphorus pentachloride	C	a	a	a	I	0.80	43.34	15.25	3.28	0.69	0.67
33	4-(Methylthio)-benzaldehyde	NC	0.9	0.9	0.0	NI	6.38	134.39	176.61	161.64	12.51	13.16
34	70/30 Oleine/octanoic acid	NC	a	a	a	I	3.26	107.55	104.67	59.56	12.13	8.23
35	Hydrogenated tallow amine	NC	3.6	1.8	1.8	I	10.23	98.63	101.41	104.20	13.98	14.26
36	2-Methylbutyric acid	C	5.0	2.3	2.7	I	2.81	44.05	3.58	3.82	2.28	1.76
37	Sodium undecylenate (34% aq.)	NC	1.7	1.7	0.0	NI	8.98	122.64	34.37	12.22	2.64	1.24
38	Tallow amine	C	b	b	b	I	10.34	94.44	110.84	113.97	14.11	7.70
39	2-Ethoxyethyl methacrylate	NC	1.6	1.6	0.0	NI	9.52	135.96	200.93	178.77	9.39	8.76
40	Octanoic acid (caprylic acid)	C	4.5	3.0	1.4	I	3.67	80.19	4.20	3.17	14.75	5.92
41	20/80 Coconut/palm soap	NC	2.7	2.0	0.7	I	10.72	112.49	116.04	136.18	11.13	1.50
42	2-Mercaptoethanol, Na salt (46% aq.)	C	4.7	2.0	2.7	I	12.29	101.39	104.44	101.49	1.58	1.37
43	Hydrochloric acid (14.4% wt.)	C	a	a	a	I	-0.61	73.57	3.29	3.55	4.67	1.18
44	Benzyl acetone	NC	1.0	0.8	0.2	NI	4.81	139.35	158.09	151.71	13.78	12.39
45	Heptylamine	C	6.7	4.0	2.7	I	11.88	97.22	355.39	348.54	0.87	0.98
46	Cinnamaldehyde	NC	3.7	2.0	1.7	I	4.03	129.99	118.92	64.14	14.79	11.32
47	60/40 Octanoic/decanoic acids	C	7.1	3.3	3.8	I	3.77	108.75	7.01	5.12	6.28	1.61
48	Glycol bromoacetate (85%)	C	b	b	b	I	1.98	95.55	3.80	3.98	8.67	5.67
49	Eugenol	NC	2.9	1.9	1.0	NI	3.68	136.42	91.06	49.02	11.11	9.56
50	55/45 Octanoic/decanoic acids	C	5.2	3.3	1.9	I	3.80	122.38	7.72	4.67	4.98	1.44
51	Methyl laurate	NC	3.9	2.0	1.9	I	5.67	110.30	101.50	115.60	16.20	8.08
52	Sodium bicarbonate	NC	0.1	0.1	0.0	NI	7.89	101.19	109.76	102.33	12.97	4.16
53	Sulphamic acid	NC	a	a	a	NI	0.70	101.44	25.02	2.53	8.86	4.07

Table 7.4 cont'd Data for the 60 chemicals tested in the ECVAM Skin Corrosivity Study

Chemical		C/NC	PII	Erythema	Oedema	I/NI	pH[c]	EPISKIN 3m	EPISKIN 1h	EPISKIN 4h	TER 2h	TER 24h
54	Sodium bisulphite	NC	1.0	1.0	0.0	NI	3.85	116.31	103.18	107.76	13.14	6.42
55	1-(2-Aminoethyl)piperazine	C	a	a	a	I	11.67	117.32	88.34	53.89	1.06	1.15
56	1,9-Decadiene	NC	3.0	2.0	1.0	I	4.15	116.16	116.28	129.68	14.76	3.39
57	Phosphoric acid	C	a	a	a	I	1.63	87.02	36.78	3.98	0.86	0.92
58	10-Undecenoic acid	NC	2.4	1.8	0.6	NI	3.88	104.41	72.22	53.90	21.49	21.68
59	4-Amino-1,2,4-triazole	NC	0.0	0.0	0.0	NI	5.92	107.66	106.11	107.30	13.27	14.31
60	Sodium lauryl sulphate (20% aq.)	NC	6.8	3.8	3.0	I	3.78	114.14	109.82	71.59	3.61	1.10

Footnote to Table 7.4

[a]Data supplied to ECVAM in confidence by an industrial company.
[b]Not possible to calculate because some or all of the tissue scores were missing.
[c]pH measurements were obtained by BIBRA International (Croydon, UK) and refer to a 10% solution, unless specified otherwise.

Table 7.5 List of liquids containing ionisable groups

Chemical		Group	C/NC	Chemical		Group	C/NC
15	2-Chloropropanoic acid	acid	C	3	2,3-Lutidine	base	NC
18	2-Mercaptoethanoic acid	acid	C	4	2,3-Xylenol	base	C
34	Acrylic acid	acid	C	25	3-Picoline	base	NC
39	Butyric acid	acid	C	26	3-Toluidine	base	NC
44	Cyclopropane COOH	acid	C	31	4-Picoline	base	NC
48	Glycolic acid	acid	NC	38	Benzylamine	base	C
49	Glyoxylic acid	acid	C	42	Cocoamine (dodecylamine)	base	C
54	Iodoacetic acid	acid	C	53	Imidazole	base	NC
56	Isobutyric acid	acid	C	66	Morpholine	base	C
60	Lactic acid	acid	C	73	Pyridine	base	NC
64	Methoxyacetic acid	acid	C	103	2,4-Xylidine	base	NC
72	Propylphosphonic acid	acid	C	124	3-Diethylaminopropionitrile	base	NC
74	Pyruvic acid	acid	C	126	3-Methoxypropylamine	base	C
117	2-Methylbutyric acid	acid	C	131	4-Amino-1,2,4-triazole	base	NC
229	Acetic acid	acid	C	158	Diethylaminopropylamine	base	C
234	Bromoacetic acid	acid	C	163	Dimethylisopropylamine	base	C
236	Butanoic acid	acid	C	164	Dimethyl butylamine	base	C
240	Chloroacetic acid	acid	C	184	Heptylamine	base	C
241	Crotonic acid	acid	C	209	N,N-Dimethylbenzylamine	base	NC
244	Dichloroacetic acid	acid	C	223	1,2-Diaminopropane	base	C
250	Dimethylcarbamyl chloride	acid	C	224	1,4-Diaminobutane	base	C
254	Formic acid	acid	C	225	2,3-Dimethylcyclohexylamine	base	C
256	Hexanoic acid	acid	C	226	2-Ethylhexylamine	base	C
260	Mercaptoacetic acid	acid	C	227	2-Mercaptoethanol	base	C
268	Propanoic acid	acid	C	228	3-Diethylaminopropylamine	base	C
272	Trichloroacetic acid	acid	C	237	Butylamine	base	C
276	Trifluoroacetic acid	acid	C	243	Cyclohexylamine	base	C
13	2-Butyn-1,4-diol	alcohol	C	247	Dicyclohexylamine	base	C
17	2-Hydroxyethyl acrylate	alcohol	C	248	Diethylamine	base	C
71	Propargyl alcohol	alcohol	C	249	Diethylene triamine	base	C
95	Decanol	alcohol	NC	252	Ethanolamine	base	C
106	2,6-Dimethyl-4-heptanol	alcohol	NC	253	Ethylene diamine	base	C
116	2-Methyl-4-phenyl-2-butanol	alcohol	NC	269	Pyrrolidine	base	C
118	2-Phenylethanol	alcohol	NC	271	Tributylamine	base	NC
125	3-Mercapto-1-propanol	alcohol	NC	8	2,4-Xylenol	phenol	C
139	α-Terpineol	alcohol	NC	10	2,5-Xylenol	phenol	C
143	Benzyl alcohol	alcohol	NC	11	2,6-Xylenol	phenol	C
150	Cinnamyl alcohol	alcohol	NC	16	2-Ethylphenol	phenol	NC
159	Dihydromercenol	alcohol	NC	28	4-Methoxyphenol	phenol	NC
166	Dipropylene glycol	alcohol	NC	63	3-Cresol	phenol	C
167	dl-Citronellol	alcohol	NC	68	2-Cresol	phenol	C
178	Geraniol	alcohol	NC	70	4-Cresol	phenol	C
189	Isopropanol	alcohol	NC	80	3-Methoxyphenol	phenol	NC
196	Linalol	alcohol	NC	81	4-Ethylphenol	phenol	NC
197	Linalol oxide	alcohol	NC	82	Phenol	phenol	C
215	p-Mentha-1,8-dien-7-ol	alcohol	NC	115	2-Methoxyphenol (guaiacol)	phenol	NC
257	Hexanol	alcohol	NC	120	2-*tert*-Butylphenol	phenol	C
264	Nonanol	alcohol	NC	127	3-Methyl phenol	phenol	NC
277	Undecanol	alcohol	NC				

Table 7.6 Performance of the SARs for skin irritation and corrosion

Model	Sensitivity (95% CI)	Specificity (95% CI)	Concordance (95% CI)	False positives (95% CI)	False negatives (95% CI)
SAR 1[a]	70	68	69	32	30
	(60-79)	(58-77)	(62-75)	(23-42)	21-40
SAR 1[b]	66	62	64	38	34
SAR 2[c]	100	95	98	5	0
	(100-100)	(86-100)	(94-100)	(0-14)	(0-0)
SAR 2[d]	100	95	98	5	0
SAR 3[e]	91	82	88	18	9
	(80-100)	(59-100)	(78-99)	(0-41)	(0-20)
SAR 3[f]	91	73	85	27	9
SAR 4[g]	95	32	56	68	5
	(87 -100)	(22-44)	(46-66)	(56-78)	(0-13)
SAR 4[h]	89	35	56	65	11

Footnote to Table 7.6

[a]Statistics based on the application of SAR 1 to its training set of 189 organic liquids; [b]Cross-validated statistics based on the 3-fold cross-validation of SAR 1; [c]Statistics based on the application of SAR 2 to its training set of 27 carboxylic acids and 19 alcohols; [d]Cross-validated statistics based on the 3-fold cross-validation of SAR 2; [e]Statistics based on the application of SAR 3 to its training set of 34 organic bases; [f]Cross-validated statistics based on the 3-fold cross-validation of SAR 3; [g]Statistics based on the application of SAR 4 to its training set of 98 liquids; [h]Cross-validated statistics based on the 3-fold cross-validation of SAR 4.

The 95% confidence intervals were derived by bootstrap resampling the appropriate data set 1000 times, except in the cases of SAR 1, for which only 100 bootstrap samples were obtained, due to the relatively large training set for this model.

Shaded cells indicate ‘unsatisfactory’ Cooper statistics.

Table 7.7 Performance of the PMs for skin irritation and corrosion

Model	Sensitivity (95% CI)	Specificity (95% CI)	Concordance (95% CI)[s]	False positives (95% CI)	False negatives (95% CI)
PM 1[a]	86	100	89	0	14
	(77-95)	(100-100)	(82-97)	(0-0)	(5-23)
PM 1[b]	54	97	77	3	46
PM 2[c]	82	86	83	14	18
	(70-94)	(68-100)	(73-93)	(0-32)	(6-30)
PM 2[d]	79	79	79	21	20
PM 3[e]	88	85	86	15	12
	(75-100)	(73-97)	(78-95)	(3-27)	(0-25)
PM 3[f]	88	79	83	21	12
PM 4[g]	76	93	80	7	24
	(62-88)	(78-100)	(69-90)	(0-22)	(12-38)
PM 4[h]	67	79	69	21	33
PM 5[i]	88	77	82	23	12
	(64-100)	(60-89)	(69-90)	(11-40)	(0-36)
PM 5[j]	88	74	80	26	12

Footnote to Table 7.7

[a]Statistics based on the application of PM 1 to its training set of 75 chemicals; [b]Statistics based on the application of PM 1 to a test set of 53 chemicals; [c]Statistics are based on the application of PM 2 to its training set of 53 chemicals; [d]Cross-validated statistics based on the 3-fold cross-validation of PM 2; [e]Statistics based on the application of PM 3 to its training set of 59 chemicals; [f]Cross-validated statistics based on the 3-fold cross-validation of PM 3; [g]Statistics based on the application of PM 4 to its training set of 59 chemicals; [h]Cross-validated statistics based on the 3-fold cross-validation of PM 4; [i]Statistics based on the application of PM 5 to its training set of 60 chemicals; [j]Cross-validated statistics based on the 3-fold cross-validation of PM 5.

The 95% confidence intervals were derived by bootstrap resampling the appropriate data set 1000 times.

CHAPTER 8

DEVELOPMENT OF NEW MODELS FOR EYE PENETRATION

8.1 INTRODUCTION

The ability of a chemical to penetrate across the tissue barriers of the eye, such as the conjunctiva, cornea and sclera, is an important consideration, not only in drug development, but also in chemical risk assessment. Various efforts have been made to model the permeabilities of ocular tissues by using *in vitro* methods (e.g. Goskonda *et al.*, 1999) or computer-based methods (e.g. Yoshida & Topliss, 1996). The work described in this chapter, which is based on Worth & Cronin (2000b), had the following objectives:

1) to determine whether a statistically-significant structure-permeability relationship (SPR) for predicting the transport of structurally diverse chemicals across the rabbit cornea could be developed on the basis of just two physicochemical properties - the logarithm of the octanol-water partition coefficient (logP) and molecular weight (MW). Such a model was sought because these physicochemical properties had been reported as significant predictors of skin permeability (Potts & Guy, 1992; Cronin *et al.*, 1998).
2) to identify additional physicochemical properties that are predictive of corneal permeability, and to use these as predictor variables in a SPR.
3) to see if any relationships exist between the corneal permeability coefficients of chemicals and their abilities to cause eye lesions.

8.2 METHOD

8.2.1 *In vitro* permeability measurements

Corneal permeability coefficients for 112 chemicals (Table 8.1) were taken from Prausnitz & Noonan (1998). In compiling the data set, Prausnitz & Noonan excluded any chemicals that had been reported to undergo active transport or to be chemically altered during transport. The permeability data for two macromolecules (cyclosporin A and inulin) were excluded from these analyses, because their large sizes (radii of 10Å and 14Å, respectively) were unrepresentative, and chemicals with radii greater than about 10Å cannot usually cross the cornea at measurable rates (Hamalainen *et al.*, 1997a). Another reason for excluding cyclosporin A in the development of SPRs is that

this macromolecule is extruded from the outer epithelial cells of the cornea by an efflux pump (Kawazu *et al.*, 1999). The permeability data were collated from *in vitro* studies performed in a number of laboratories, so a certain amount of inter-laboratory variability in the data was expected.

8.2.2 Calculation of physicochemical properties

The logarithm of the octanol-water partition coefficient (logP) and the logarithm of aqueous solubility (in mol/l) were calculated by using the KOWWIN software (Syracuse Research Corporation, Syracuse, NY, USA). The Molconn-Z software (Hall Associates Consulting, Quincy, MA, USA) was used to calculate molecular weight (MW), the number of hydrogen-bond donors (n_{HD}), the number of hydrogen bond acceptors (n_{HA}), the number of potential hydrogen bonds (n_H; $n_H = n_{HA} + n_{HD}$), and the following topological indices: simple and valence path molecular connectivity of orders 0 to 4, simple and valence difference connectivity indices of orders 0 to 4, and kappa indices of orders 0 to 3 ($\kappa 0$–$\kappa 3$).

To calculate steric and molecular orbital descriptors, chemical structures were input as SMILES codes into the TSAR (version 3.2) molecular spreadsheet (Oxford Molecular Ltd. Oxford, England). The CORINA (version 3.21) program (Oxford Molecular Ltd., Oxford, England) converted the corresponding 2D structures into reasonable 3D conformations. These 3D structures were energy-minimised by using the COSMIC force field in TSAR. The 3D structure of each molecule was subsequently optimised by using the AM1 Hamiltonian (Dewar *et al.*, 1985) in the VAMP (version 6.5) molecular orbital package (Oxford Molecular Ltd, Oxford, England). For each energy-optimised molecule, the following properties were calculated: the energies of the lowest unoccupied molecular orbital (LUMO) and the highest occupied molecular orbital (HOMO); molecular surface area and volume; lengths of the first, second and third inertial axes; dipole moment; and heat of formation.

8.2.3 Statistical analyses

All statistical analyses were performed with Minitab 11 for Windows (Minitab Inc., State College, PA, USA). Multiple linear regression (MLR) was applied to the data set

of 112 chemicals to derive two SPR models, one based on MW and logP, and the other based on logP, n_H and $\kappa 3$. The last three variables were identified by stepwise linear regression as the most significant predictors of the logarithm of the permeability coefficient (logKp) across the rabbit cornea.

Before deriving the SPRs for corneal permeability, certain chemicals were removed from the training set, because they were outliers along one or more predictor variables. Removal of extreme observations from the independent variables prior to model development allows the domains of the eventual models to be defined with greater certainty. Outliers were defined as observations that were greater than $Q_3+1.5(Q_3-Q_1)$, or smaller than $Q_1-1.5(Q_3-Q_1)$, where Q_1 and Q_3 are the first and third quartiles (25^{th} and 75^{th} percentiles), respectively.

Cross-validated coefficients of determination [r^2(cv)] for the SPRs were calculated by the leave-one-out cross-validation procedure, using a purpose-written Minitab macro (See Appendix A2). Cross-validated r^2 values are considered to provide a more realistic indication of model predictivity than are conventional r^2 values. The Minitab macro was also used to generate predicted values of corneal permeability, and a cross-validated standard error of the estimate (i.e. the mean standard error of the estimate, taken over all iterations of the leave-one-out procedure).

The stabilities of the regression coefficients were assessed by performing a resampling routine, coded in the form of another Minitab macro (see Appendix A10). In each iteration of the routine, 20 chemicals were selected, at random and without replacement, from the training set, and MLR was performed with the physicochemical and permeability data for these 20 chemicals. This procedure was carried out 1000 times, each time storing the coefficients of the regression equation. The distribution of each parameter, taken over 1000 iterations, was then used to determine the variability of each parameter. Typically, the sampling distributions are expected to be approximately normal, or to at least consist of a single node. If a parameter distribution is observed to be multi-modal, this indicates that a single regression model is insufficient to fit the data.

To explore the possibility of relationships occurring between the corneal permeabilities of chemicals and their abilities to cause eye lesions, the predicted corneal permeability coefficients for 66 chemicals were compared with various eye irritation endpoints by using matrix and scatter plots. The endpoints examined were the bootstrap mean values of the WDS, conjunctival erythema, œdema and discharge, corneal opacity and area, the weighted corneal score (= 5 x opacity x area), and iritis (see chapter 5).

8.3 RESULTS

8.3.1 Prediction of corneal permeability from molecular weight and lipophilicity

Before carrying out MLR on logP and MW, five chemicals (EDTA, mannitol, sucrose, the quaternary ammonium derivative of sulphonamide, and tobramycin) were removed from the training set, since these were outliers along logP. Application of MLR to the remaining 107 chemicals produced a two-variable SPR (Equation 8.1), which is significant at the 99.9% level. The domain of this SPR is defined by logP values in the range –2.2 to 4.2, and MW values from 17 to 468 g/mol:

$\log K_p = 0.344 \log P - 0.00321 MW - 4.73$ (Equation 8.1)

$n = 107$, $s = 0.50$, $s(cv) = 0.50$, $r = 0.65$, $r^2 = 0.42$, $r^2(cv) = 0.38$
$F = 37.6$ ($p < 0.001$), $t(\log P) = 8.6$ ($p < 0.001$), $t(MW) = -5.2$ ($p < 0.001$)

Resampling the data set of 107 chemicals (1000 times, 20 chemicals a time), and taking the median values of the parameter distributions, produced the following equation:

$\log K_p = 0.355 \log P – 0.00326 MW – 4.74$

The sampling distributions of the regression parameters are all approximately normal, with standard deviations of 0.106, 0.0015 and 0.38 for logP, MW and the constant, respectively. The estimates for the parameters are therefore considered to be stable. The two variables (logP and MW) are considered to be independent since they are not significantly correlated ($r = 0.45$).

As with any regression model, the statistical quality of the two-variable SPR can be improved by removing observations that are poorly predicted or exert a large influence on the regression parameters. For example, by removing two chemicals (ibuprofen and tetracaine) that are poorly predicted (i.e. have standardised residuals greater than 2), the r^2 value increases from 0.42 to 0.51:

$\log K_p = 0.390 \log P - 0.00327 MW - 4.77$

$n = 105$, $s = 0.46$, $s(cv) = 0.46$, $r = 0.71$, $r^2 = 0.51$, $r^2(cv) = 0.47$
$F = 52.6$ ($p < 0.001$), $t(\log P) = 10.1$ ($p < 0.001$), $t(MW) = -6.2$ ($p < 0.001$)

8.3.2 Prediction of corneal permeability from lipophilicity, hydrogen bonding and molecular shape

Stepwise regression analysis identified the three most significant variables (in order of decreasing significance) to be logP, n_H and $\kappa 3$. Before carrying out MLR on these variables, 17 chemicals were removed from the data set, because they were outliers along one or more of the three variables: acebutolol, alprenolol, atenolol, betaxolol, bevantolol, cromolyn, EDTA, labetalol, mannitol, metoprolol, nadolol diacetate, oxprenolol, sotalol, sucrose, the quaternary ammonium derivative of sulphonamide, tetracaine, and tobramycin.

The regression of $\log K_p$ against logP, n_H and $\kappa 3$ produced a three-variable SPR (Equation 8.2), which is significant at the 99.9 % level. The domain of this SPR is defined by logP values in the range –2.2 to 4.2, n_H values from 2 to 13, and $\kappa 3$ values from 0 to 6.

$\log K_p = 0.230 \log P - 0.0679 n_H - 0.168 \kappa 3 - 4.35$ (Equation 8.2)

$n = 95$, $s = 0.46$, $s(cv) = 0.46$, $r = 0.72$, $r^2 = 0.52$, $r^2(cv) = 0.47$
$F = 33.0$ ($p < 0.001$), $t(\log P) = 6.1$ ($p < 0.001$), $t(n_H) = -3.9$ ($p < 0.001$),
$t(\kappa 3) = -3.5$ ($p = 0.001$)

To assess the stability of the parameters in this model, the subset of 95 chemicals was resampled (1000 times, 20 chemicals a time). Taking the median values of the parameter distributions produced the following equation:

$$\log K_p = 0.229 \log P - 0.0684 n_H - 0.167 \kappa 3 - 4.32$$

The sampling distributions of the regression parameters are approximately normal, with standard deviations of 0.086, 0.044, 0.076 and 0.12 for logP, n_H, $\kappa 3$ and the constant, respectively. The estimates are therefore considered to be stable. In addition, there are no significant correlations between the three descriptor variables ($r_{nH\text{-}logP}$ =-0.31; $r_{\kappa 3\text{-}logP}$ = 0.20; $r_{nH\text{-}\kappa 3}$ = 0.30), which provides further support for the stability of the model.

8.3.3 Relationships between corneal permeability and eye irritation endpoints

A subset of 66 chemicals (Table 8.2) was selected from an eye irritation data set of 92 chemicals (Table 5.7) on the basis that the logP, $\kappa 3$ and n_H values of these chemicals fall within the domain of the SPR defined by Equation 8.2, which was used to predict their corneal permeability coefficients. Examination of a matrix plot (Figure 8.1) revealed no apparent relationship between predicted corneal permeability and the three conjunctival endpoints. The other eye irritation endpoints (corneal opacity and area, the weighted corneal score, the WDS, and iritis) showed a tendency for irritants to be located in the mid-range of corneal permeabilities. However, the matrix plot shows that the non-irritants also have intermediate values of corneal permeability, so it would appear difficult to distinguish between irritants and non-irritants on the basis of corneal permeability. These effects are most noticeable in the scatter plot of iritis against corneal permeability (Figure 8.2).

8.4 DISCUSSION

In this study, two mechanistically-based SPRs (Equations 8.1 and 8.2) were derived for predicting the corneal permeability coefficients of a diverse range of organic chemicals. The two models have statistically significant predictor variables, stable coefficients, and well-defined domains of application. While the two-variable SPR is more economical and user-friendly in terms of its predictor variables, the three-variable SPR produces a

better fit to the data, and provides more mechanistic insight into the transcorneal penetration process.

The mathematical form of both SPRs was chosen to be linear, because in the absence of prior information, this is the simplest mathematical form to assume for any model. In support of this choice, scatter plots of $\log K_p$ against each predictor variable did not indicate any non-linear dependencies. Furthermore, an SPR based only on logP and $\log P^2$ produced a poor fit ($r^2 = 0.27$) to the reduced data set of 107 chemicals, and the inclusion of $\log P^2$ as a predictor variable (in addition to, or instead of, logP) in the two-variable and three-variable SPRs did not improve their statistical quality.

According to the two-variable model (Equation 8.1), corneal permeability increases with increasing lipophilicity (logP) and with decreasing molecular weight (MW), an approximate measure of molecular size. The importance of the latter variable is consistent with the observation that the diffusion parameters of hydrophilic drugs increase with decreasing molecular weight (Sasaki *et al.*, 1997).

The three-variable model (Equation 8.2) indicates that corneal permeability increases with increasing logP (lipophilicity), and with decreasing n_H (hydrogen-bonding potential) and $\kappa 3$. The last variable can be interpreted as a measure of the centrality of branching: the more linear a molecule, the greater its value of $\kappa 3$, whereas the greater the degree of branching at the centre of the molecule, the smaller the value of $\kappa 3$ (Hall & Kier, 1991). Consistent with the importance of $\kappa 3$ is the observation that the rabbit corneal permeability of a series of local anæsthetics decreased with increasing molecular length (Igarashi *et al.*, 1984). The lipophilic barrier is likely to be the corneal epithelium on the outer surface of the cornea, since this is a stratified epithelium comprising 5-7 layers of cells, inter-connected by tight junctions. The importance of n_H could be a reflection of the barrier to diffusion provided by the corneal stroma, which accounts for about 90% of the cornea's thickness (450 μm), and consists of a highly ordered array of collagen bundles. Thus, molecules that are sufficiently lipophilic to cross the epithelium will be further impeded if they engage in hydrogen bonding with the collagen matrix. The importance of the third variable ($\kappa 3$) could also be a reflection that the stroma is a highly ordered matrix, the dimensions of which favour the transport

of non-linear molecules over linear ones, but it could also relate to the barrier properties of the epithelia on both sides of the cornea.

In comparison with previous attempts to model corneal permeability (e.g. Yoshida & Topliss, 1986), the SPRs presented here have been based on a larger number of chemicals covering a broader range of chemical structures. For example, Yoshida & Topliss presented an SPR with a high r value (0.9), but this was derived from just 26 chemicals, mostly steroids and β-blockers. The Yoshida & Topliss model is based on two predictors: logD and ΔlogP (where logD = the logarithm of the distribution coefficient at pH 7.65, and $\Delta\log P = \log P_{octanol} - \log P_{alkane}$). Data for logD can be found in the literature (e.g. Prausnitz & Noonan, 1998), but the measurements have generally been obtained at different pH values and are therefore incomparable. Data for ΔlogP are not so readily available, so in cases where these data were missing, Yoshida & Topliss used a solvation equation to predict ΔlogP from five other variables (R_2, a term representing molar refraction; π_2, a descriptor for dipolarity/polarisability; $\Sigma\alpha_2^H$ and $\Sigma\beta_2^H$, descriptors for hydrogen bond acidity and basicity; and V_x, the characteristic volume of McGowan). Interestingly, according to Yoshida & Topliss, ΔlogP expresses mainly hydrogen-bonding acidity, which, in mechanistic terms, is related to the hydrogen-bonding predictor (n_H) identified as significant by stepwise regression in this study.

At first sight, the SPRs for corneal permeability do not appear to be as predictive as the SPRs reported above for skin permeability (Cronin *et al.*, 1998; Potts & Guy, 1992). It should be noted, however, that the r^2 values of the two SPRs for corneal permeability can be improved by removing additional chemicals from the training set (i.e. chemicals that are poorly predicted or exert an unduly large influence on the regression parameters). For example, the r^2 value of the two-variable SPR can be increased from 0.42 to 0.51 by removing just two chemicals (ibuprofen and tetracaine). Since the SPRs for skin permeability and for corneal permeability were based on different sets of chemicals, and possibly on data sets of differing quality, the statistical quality of the different models is not strictly comparable.

To better define the domains of the SPRs, a number of chemicals were removed from the training set before developing the models. The five chemicals removed before developing the two-variable SPR had extreme logP values. In the cases of EDTA, mannitol, sucrose, and tobramycin, these extreme values could be a result of the large numbers of hydrogen bonds formed, whereas the quaternary ammonium derivative of sulphonamide is an ionic species that does not exist in its non-ionised (permeable) form at the experimental pH of 7.0 (Prausnitz & Noonan, 1998). The same chemicals were also removed before developing the three-variable SPR, since this is also based on logP. In addition, a further 12 chemicals (acebutolol, alprenolol, atenolol, betaxolol, bevantolol, cromolyn, labetalol, metoprolol, nadolol diacetate, oxprenolol, sotalol and tetracaine) were removed before developing the three-variable SPR, since these were outliers along $\kappa 3$. It is noteworthy that, with the exception of cromolyn and tetracaine, these chemicals are all β-adrenoceptor antagonists, 10 of which contain the following substructure: Ar-O-CH_2-CH(OH)-CH_2-NH-R, where Ar is an aromatic ring structure and R is an isopropyl or *tert*-butyl group. The other two chemicals (bevantolol and labetalol) have a very similar structure (in bevantolol, the R group is an ethylbenzene derivative, whereas in labetalol, there is no methoxy group). It therefore appears that the domain of the three-variable SPR excludes the family of β-blocking drugs.

A number of chemical types were consistently under-predicted or over-predicted by the SPRs. For example, the glucose derivatives showed a tendency to be underpredicted, perhaps because they can traverse the cornea by facilitated transport in addition to passive diffusion. Evidence for a facilitated glucose transporter in the corneal epithelium was presented by Takahashi *et al.* (1996). The steroids were anomalous in that they are under-predicted by the two-variable SPR, and over-predicted by the three-variable SPR. Since there are 10 steroids in the data set, it is acceptable, in principle, to derive a separate two-variable SPR for these substances (according to Topliss & Costello [1972], the ratio of observations to variables should be at least 5 to avoid the derivation of chance relationships). However, when MLR of $\log K_p$ was performed against logP and MW, the residuals showed a significant degree of autocorrelation, and were not normally distributed, suggesting that the model could be unreliable, despite its reasonable r^2 value of 0.68. It is noteworthy that the corneal permeability for the 10

steroids were all derived from a single reference (Schoenwald & Ward, 1978), which suggests that a systematic laboratory bias might be the reason from their anomalous behaviour. The corneal permeability coefficient of water was under-predicted by both SPRs. The under-prediction of water could be an artefact, possibly because no account had been taken of its high diffusion coefficient in the experimental set-up, with the result that the measured permeability coefficient was overestimated. Alternatively, there could be a mechanistic reason, related to the presence of water channels or small pores in the corneal membrane. Evidence for the presence of a water channel (aquaporin 5) in the corneal epithelium has been reported by Kang *et al.* (1999), and evidence for the presence of pores has been presented by Hamalainen *et al.* (1997a, 1997b). The permeabilities of two other small polar molecules were also under-predicted (butanol and glycerol), but the permeability of ethanolamine was over-predicted.

Perhaps the most significant reason for poor predictions is that the permeability data (Table 8.1) were pooled from 20 primary literature sources, which almost certainly means that there were significant inter-laboratory differences in the experimental protocols used to obtain the permeability coefficients. The inherent variability of the experimental data will place a ceiling on the r^2 value that can be obtained with any model, however well-chosen the predictor variables. An inspection of the original data showed that permeability measurements for the same chemical can vary by up to one order of magnitude, i.e. log Kp values were generally accurate within one log unit. The SPRs derived in this study have a standard error of the estimate (s) of about 0.5 log units, i.e. there is an uncertainty of about ± 1 log unit (the 95% confidence interval is approximately 2.s). Thus, the uncertainty in the predicted values is about the same as the uncertainty in the original data, which means that the SPRs are not overfitting the data. If the corneal permeability measurements had been obtained according to a standardised protocol, the SPRs would probably have been more accurate.

In the pharmaceutical industry, SPRs for corneal permeability could be used to determine the delivery of potential drugs to the appropriate tissue in the eye. It has been shown that the corneal epithelium contains a greater proportion of paracellular space than either the conjunctiva or the sclera (Hamalainen *et al.*, 1997a; 1997b). This

suggests that transcorneal transport is generally the rate-limiting step in drug delivery, and therefore the most appropriate step to model.

From the results of this study, no obvious relationship could be identified between corneal permeability and several eye irritation endpoints (Figure 8.1). A relationship between corneal permeability and the corneal and iris endpoints was expected on the basis that chemicals with greater corneal permeabilities should be more likely to cause lesions in the cornea and iris than chemicals with smaller permeabilities. Ideally, this hypothesis would have been tested by comparing measured permeabilities against measured ocular effects. However, eye irritation data were not available for the Prausnitiz & Noonan permeability data set, and corneal permeabilities were not available for the eye irritation data set. Therefore, the hypothesis was tested by predicting the corneal permeabilities for those chemicals in the eye irritation data set that fell within the domain of the better SPR (Equation 8.2). The absence of the expected relationship could be due to the limited predictive quality of the SPR (r^2 = 52%), or there could be mechanistic reasons. In particular, the fact that non-irritating chemicals span the full range of corneal permeabilities (Figures 8.1 and 8.2) is not surprising, since not all chemicals that reach the cornea and iris are expected to exert toxic effects. In addition, the observation that chemicals with low predicted permeabilities are non-irritant is expected, on the basis that these chemicals are unable to reach the cornea and iris.

8.5 CONCLUSIONS

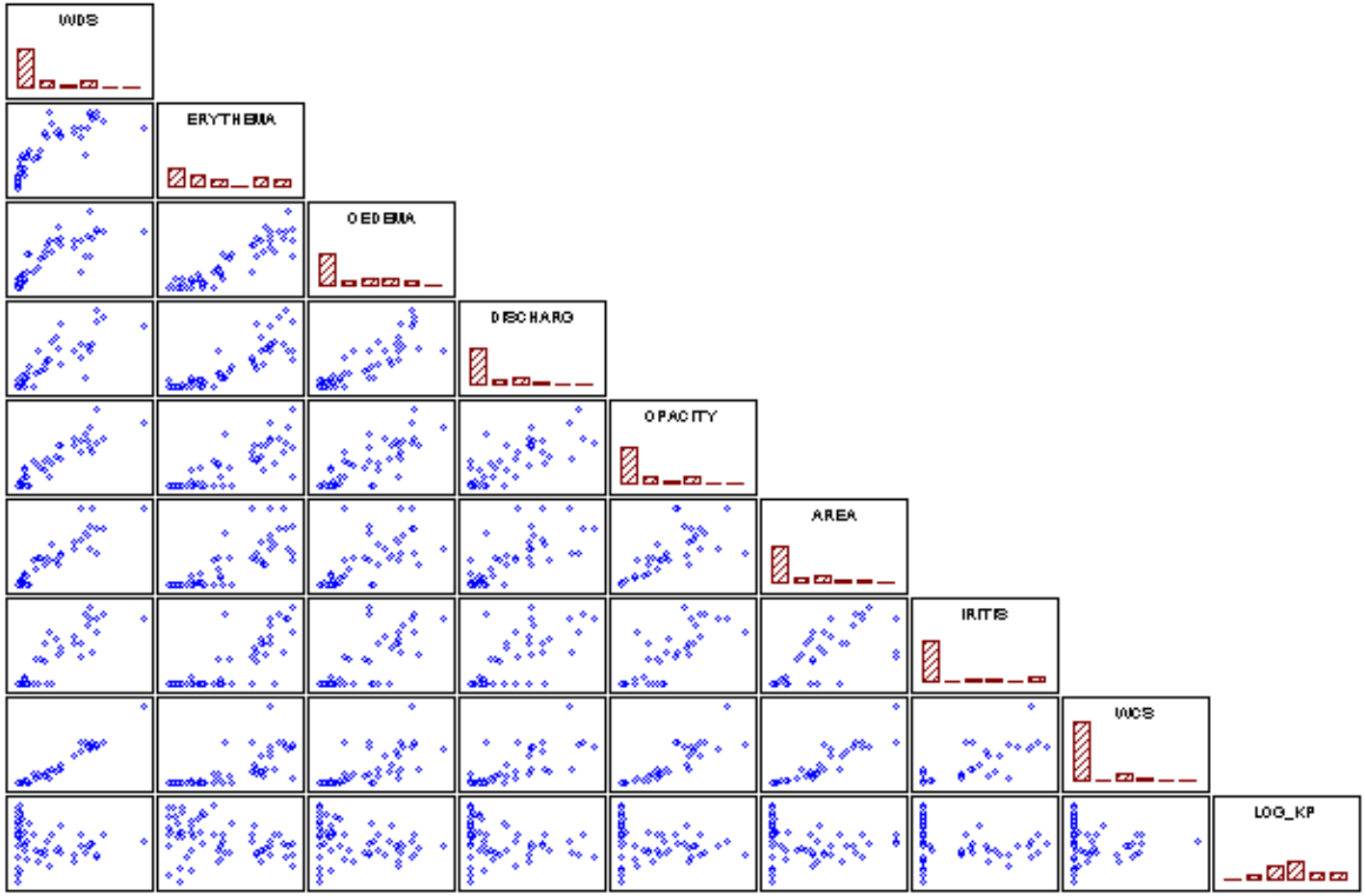
The analyses carried out in this study suggest that the corneal permeability coefficients of a diverse range of organic chemicals can be predicted by using two (logP and MW) or three (logP, n_H and $\kappa 3$) predictor variables. The SPRs are applicable to a diverse range of chemical classes, including glucose derivatives, steroids, sulphonamides, nitrogen-containing heterocycles, and anilines. Chemicals known to undergo metabolism (e.g. pro-drugs) or extrusion at the outer corneal surface (e.g. cyclosporin A) are excluded from the domains of the two SPRs. In addition, the family of β-adrenoceptor antagonists is excluded from the domain of the three-variable SPR. The

two SPRs are probably modelling passive diffusion, and outliers are likely to be chemicals that permeate by means of specific transport mechanisms.

The models presented in this study should be regarded as provisional, since their predictive capacities could probably be improved if they were reparameterised by using a sufficient amount of high-quality permeability data, obtained according to a single, standardised *in vitro* protocol. It remains to be seen, however, whether a single SPR will be capable of predicting the corneal permeabilities of a broad range of chemical structures with greater accuracy, or whether different class-specific SPRs will need to be developed. Depending on the availability of mechanistic information, it might also be possible to develop different SPRs to account for different modes of transcorneal penetration (passive transcellular diffusion, passive paracellular diffusion, passive carrier-mediated transport, and active transport). However, since it is unlikely that SPRs will be able to model the transcorneal transport of all types of substances (e.g. pro-drugs and formulations), it is suggested that SPRs are used in combination with *in vitro* tests, so that substances falling outside the domain of the SPRs are the only ones that would need to be tested *in vitro*.

Finally, there appears to be some relationship between corneal permeability and the severity of lesions in the cornea and iris. However, the relationship does not appear to be sufficiently well-defined for reliable predictions to be made of eye irritation on the basis of corneal permeability alone.

Figure 8.1 Relationships between corneal permeability and eye irritation endpoints

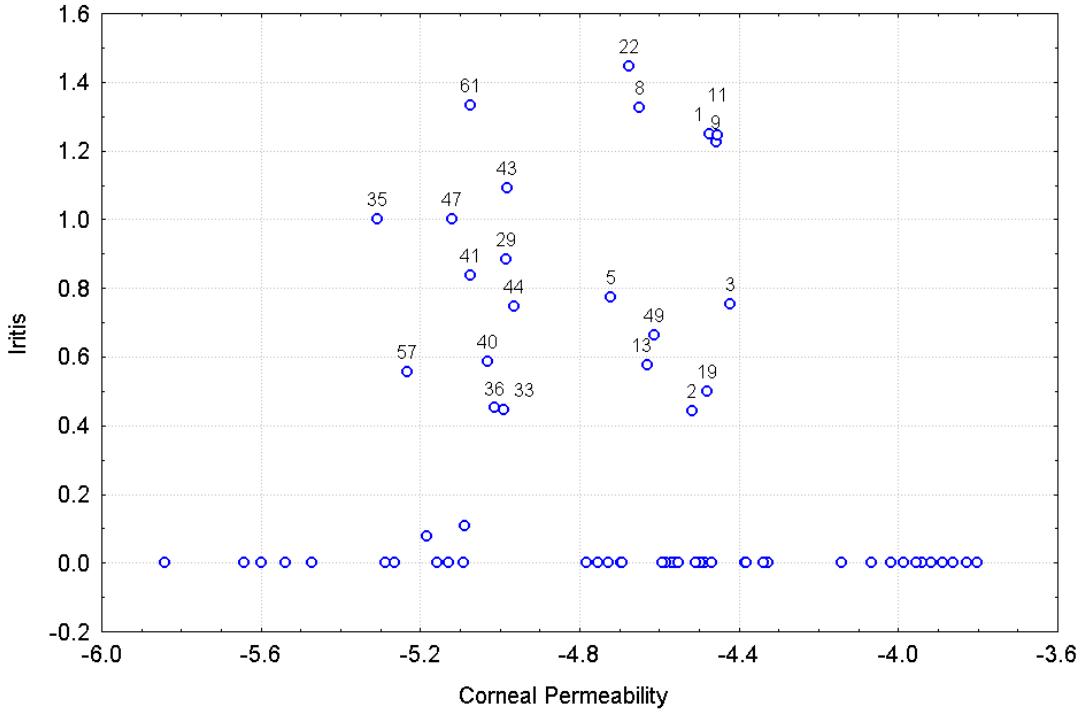


Footnote to Figure 8.1

WDS = weighted Draize score; ERYTHEMA = conjunctival erythema (redness); OEDEMA = conjunctival œdema (chemosis); DISCHARG = conjunctival discharge; OPACITY = corneal opacity; AREA = corneal area; IRITIS = iris lesion; WCS = weighted corneal score; log KP = logarithm of the corneal permeability coefficient.

Each tissue score is a mean value, taken over three time-points (24h, 48h & 72h) and all rabbits in the experimental group.

Figure 8.2 Relationship between corneal permeability and iritis



Footnote to Figure 8.2

1 = acetone; 3= isopropanol; 5 = γ-butyrolactone; 8 = furfuryl hexanol; 9 = pyridine; 11 = cyclohexene; 13 = 4-fluoroaniline; 19 = 2,2-dimethylbutanoic acid; 22 = 4-carboxybenzaldehyde; 29 = methylthioglycolate; 33= glycidyl methacrylate; 35 = methyl cyanoacetate; 36 = ethylthioglycolate; 40 = butanone; 41 = methyl acetate; 43 = isobutanol; 44 = butanol; 47 = diethylaminopropionitrile; 49 = 2-ethyl-1-hexanol; 57 = 2-methoxyethyl acrylate; 61 = hexanol

Table 8.1 Corneal permeability and physicochemical data for 112 chemicals

No	Chemical	logKp	logP	MW	κ3	n_H
1	Acebutolol	-6.07	1.19	336.4	8.7924	9
2	Acetazolamide	-6.02	-0.72	222.2	3.3241	11
3	2-Benzoylamino-1,3,4-thiadiazole 5-sulphonamide	-6.22	0.88	284.3	3.9958	11
4	2-Isopentenyl amino 1,3,4-thiadiazole-5-sulphonamide	-6.25	1.00	276.3	4.0000	11
5	N-Methylacetazolamide	-5.64	-0.25	236.3	3.2727	10
6	Alprenolol	-4.54	2.81	249.4	6.6667	5
7	Aniline	-4.44	1.08	93.1	1.5000	3
8	4-Aminoacetanilide	-5.44	0.18	150.2	3.2653	6
9	4-Aminoacetophenone	-4.49	0.76	135.2	2.2857	4
10	4-Aminobenzamide	-5.77	-0.33	136.2	2.6509	6
11	4-Aminobenzyl alcohol	-4.77	0.16	123.2	2.0000	5
12	4-Aminophenylethanol	-4.70	0.65	137.2	2.6509	5
13	4-Chloroaniline	-4.46	1.72	127.6	1.8000	4
14	4-Ethoxyaniline	-4.51	1.65	137.2	2.6509	4
15	4-Ethylaniline	-4.57	2.11	121.2	2.0000	3
16	4-Isopropylaniline	-4.62	2.53	135.2	2.2857	3
17	4-Methoxyaniline	-4.46	1.16	123.2	2.0000	4
18	4-Toluidine	-4.52	1.62	107.2	1.8000	3
19	Atenolol	-5.96	-0.03	266.3	8.0000	9
20	Benzolamide	-6.85	0.12	320.4	3.9862	12
21	Betaxolol	-4.57	2.98	307.4	9.0369	6
22	Bevantolol	-4.27	2.68	345.4	7.6371	7
23	Bromacetazolamide	-6.17	-0.05	301.1	3.5918	12
24	Bufuralol	-4.14	3.43	261.4	3.9862	5
25	Butanol	-4.12	0.84	74.1	4.0000	2
26	Chloramphenicol	-5.17	0.92	323.1	5.3789	12
27	Chlorzolamide	-4.74	0.94	275.7	2.8311	9
28	Cimetidine	-6.15	0.57	252.3	5.9282	9
29	Clonidine	-4.36	1.89	230.1	2.5344	7
30	Cocaine	-5.21	2.17	303.4	3.4405	5
31	Cortexolone	-4.52	3.15	346.5	2.3706	6
32	Corynanthine	-4.96	2.11	354.4	2.9512	6
33	Cromolyn	-5.96	1.55	468.4	6.4784	14
34	Cyclophosphamide	-4.96	0.97	261.1	3.2727	9
35	Deoxycorticosterone	-4.40	3.12	330.5	2.4814	4
36	2-Deoxyglucose	-5.13	-2.17	164.2	1.7729	9
37	1-Cyclopropyl 2-deoxyglucose	-5.66	-0.95	204.2	2.0204	9
38	1-Ethyl 2-deoxyglucose	-5.54	-1.26	192.2	2.0833	9
39	1-Isopropyl 2-deoxyglucose	-5.74	-0.84	206.2	2.3432	9
40	1-Methyl 2-deoxyglucose	-5.34	-1.75	178.2	1.8595	9
41	Dexamethasone	-5.30	1.72	392.5	2.3951	9
42	Dexamethasone-17-acetate	-4.43	2.46	434.5	3.1074	9
43	2,5-Dimethoxyaniline	-4.48	1.24	153.2	2.2145	5
44	Edetic acid (EDTA)	-5.68	-3.86	292.2	12.4898	14
45	Ethanolamine	-6.30	-1.61	61.1	4.0000	5
46	Ethoxyzolamide	-4.25	2.08	258.3	2.8311	8
47	2-Benzothiazole-sulphonamide	-4.44	1.50	214.3	1.9200	7
48	6-Hydroxy 2-benzothiazole-sulphonamide	-5.25	1.02	230.3	2.1728	9
49	6-Chloro 2-benzothiazole-sulphonamide	-4.37	2.15	248.7	2.1728	8
50	4,6-Dichloro 2-benzothiazole-sulphonamide	-4.41	2.79	283.1	2.2400	9

Table 8.1 continued

No	Chemical	logKp	logP	MW	κ3	n_H
51	6-Amino 2-benzothiazole-sulphonamide	-5.17	0.59	229.3	2.1728	10
52	6-Nitro 2-benzothiazole-sulphonamide	-5.18	1.32	259.3	2.6514	9
53	6-Hydroxyethoxy 2-benzothiazole-sulphonamide	-5.82	0.61	274.3	3.2633	10
54	6-Benzyloxy 2-benzothiazole-sulphonamide	-4.33	3.29	320.4	4.0500	8
55	6-Acetamido 2-benzothiazole-sulphonamide	-5.33	0.61	271.3	3.2633	10
56	Fluorometholone	-4.77	2.06	376.5	2.2272	7
57	Flurbiprofen	-4.68	3.81	244.3	3.5262	4
58	Flurbiprofen amide	-4.66	2.93	243.3	3.5262	5
59	Glycerol	-5.35	-1.65	92.1	3.0000	6
60	Hydrocortisone	-5.07	1.62	362.5	2.4860	8
61	Ibuprofen	-5.66	3.79	206.3	4.1653	3
62	Indomethacin	-4.16	4.23	357.8	4.1353	6
63	Labetalol	-4.85	2.41	328.4	6.6824	10
64	Levobunolol	-4.77	2.48	291.4	5.2898	6
65	Mannitol	-5.62	-3.01	182.2	3.5156	12
66	Methazolamide	-5.59	-0.65	236.3	3.2727	10
67	5-Imino-4-methyl-1,3,4 thiadiazoline-2-sulphonamide	-6.11	0.36	194.2	1.9753	10
68	Methylenedianiline	-4.60	2.18	198.3	3.5000	6
69	Metoprolol	-4.55	1.69	267.4	8.0000	6
70	MK-927	-5.34	0.86	338.5	3.6213	10
71	Nadolol	-5.85	1.17	309.4	5.5515	9
72	Nadolol diacetate	-5.32	2.77	393.5	8.3457	9
73	Oxprenolol	-4.55	1.83	265.4	7.3728	6
74	Penbutolol	-4.22	4.20	291.4	5.9504	5
75	Phenylephrine	-6.03	-0.21	167.2	3.1142	6
76	Pilocarpine	-4.77	0.24	208.3	2.3971	4
77	Pindolol	-5.00	1.48	248.3	4.2667	6
78	Prednisolone	-5.43	1.40	360.4	2.4860	8
79	Prednisolone acetate	-4.48	2.14	402.5	3.2770	8
80	Procaine	-5.38	1.99	236.3	5.4444	6
81	Progesterone	-4.70	3.67	314.5	2.2893	2
82	Propanolol	-4.47	2.60	259.3	4.7950	5
83	Rauwolfine	-5.04	1.88	326.4	1.4405	6
84	SKF 72223	-4.31	1.91	193.2	2.0204	4
85	SKF 86466	-4.15	2.87	195.7	2.0833	2
86	SKF 86607	-4.10	1.10	191.3	1.3333	5
87	Sotalol	-6.15	0.37	272.4	6.6667	8
88	Sucrose	-5.37	-4.27	342.3	3.3833	19
89	Sulphacetamide	-6.00	-0.60	214.2	3.5918	8
90	Quaternary ammonium derivative of sulphonamide	-5.44	-2.85	229.3	4.5714	5
91	2-Benzimidazole derivative of sulphonamide	-5.52	1.38	273.3	2.8800	7
92	4-Chloro-N-methylbenzene derivative of sulphonamide	-5.19	3.24	281.8	3.5262	5
93	4-Chlorobenzene derivative of sulphonamide	-5.26	2.78	267.7	3.4844	6
94	Sulphanilamide	-6.30	-0.55	172.2	2.5000	8
95	Testosterone	-4.38	3.27	288.4	1.9263	3
96	Tetracaine	-5.82	3.02	264.4	7.3728	5
97	6-Sulphonamido-3-substituted-3-H-1,3,4-thiadiazolo[2,3-C]-1,2,4-thiadiazole	-5.10	1.61	313.4	3.0296	11
98	3-Chloro derivative of chemical #97	-4.89	2.25	347.8	3.2766	12
99	4-Chloro derivative of chemical #97	-5.08	2.25	347.8	3.2766	12
100	4-Methoxy derivative of chemical #97	-5.35	1.69	343.4	3.5046	12

Table 8.1 continued

No	Chemical	logKp	logP	MW	κ3	n_H
101	3-Methoxy derivative of chemical #97	-5.28	1.69	343.4	6.2457	12
102	4-Hydroxy derivative of chemical #97	-6.46	1.13	329.4	5.6529	13
103	3-Fluoro derivative of chemical #97	-5.19	1.81	331.4	5.6529	12
104	4-Fluoro derivative of chemical #97	-5.39	1.81	331.4	5.6529	12
105	4-Dimethylamino derivative of chemical #97	-5.24	1.79	356.4	6.4815	12
106	Timolol	-4.92	1.75	316.4	8.0222	10
107	Tobramycin	-6.28	-5.76	467.5	11.1600	29
108	Triamcinolone acetonide	-4.92	2.69	434.5	6.7804	9
109	Trichlormethazolamide	-4.62	0.70	339.6	4.9383	13
110	Trifluormethazolamide	-5.41	-0.24	290.2	4.9383	13
111	Water	-3.82	-0.57	17.0	2.0000	2
112	Yohimbine	-4.74	2.11	354.4	7.1111	6

Footnote to Table 8.1

κ3 = third-order kappa index; log Kp = logarithm of the corneal permeability coefficient; logP = logarithm of the octanol-water partition coefficient; MW = molecular weight; n_H = number of potential hydrogen bonds.

Table 8.2 Data set of 66 chemicals for which predicted corneal permeability was compared with eye irritation

	Chemical	Redness	Chemosis	Discharge	Opacity	Area	Iritis	WCS	WDS	MW	logP	κ3	n_H	logKp
1	Acetone	2.8	2.5	2.8	1.9	2.9	1.2	27.9	50.4	58.1	-0.24	0.000	1	-4.47
2	Ethanol	2.1	1.3	0.7	1.1	1.3	0.4	7.4	18.0	46.1	-0.14	0.000	2	-4.52
3	Isopropanol (propan-2-ol)	2.3	1.6	0.8	0.7	1.1	0.8	4.0	17.6	60.1	0.28	0.000	2	-4.42
4	Isopropyl bromide	0.3	0.2	0.0	0.0	0.0	0.0	0.0	1.1	123.0	2.08	0.000	1	-3.94
5	γ-Butyrolactone	2.2	2.1	1.3	1.9	2.2	0.8	21.0	36.2	86.1	-0.31	0.980	2	-4.72
6	Cyclopentanol	2.0	1.9	1.0	1.2	1.4	0.0	8.9	18.7	86.1	1.15	0.980	2	-4.39
7	Methylcyclopentane	0.8	0.1	0.2	0.0	0.0	0.0	0.0	2.3	84.2	3.1	0.980	0	-3.80
8	Furfuryl hexanol	1.3	1.3	0.3	2.3	2.7	1.3	31.1	43.5	98.1	0.45	1.185	3	-4.65
9	Pyridine	2.9	3.3	1.3	2.7	1.9	1.2	25.2	46.1	79.1	0.8	1.333	1	-4.46
10	2-xylene	0.2	0.4	0.0	0.0	0.0	0.0	0.0	1.2	106.2	3.09	1.488	0	-3.89
11	Cyclohexanol	2.4	2.5	2.2	2.9	4.0	1.2	58.2	78.6	100.2	1.23	1.500	2	-4.45
12	Toluene	1.5	1.4	0.7	0.0	0.0	0.0	0.0	7.3	92.1	2.54	1.500	0	-4.02
13	4-Fluoroaniline	2.7	2.2	1.6	1.6	4.0	0.6	31.5	47.4	111.1	1.28	1.800	4	-4.63
14	Styrene	0.7	0.6	0.3	0.1	0.1	0.0	0.0	3.2	104.2	2.89	1.800	0	-3.99
15	2,6-Dichlorobenzoyl chloride	2.6	1.6	1.7	1.0	2.2	0.0	10.8	22.3	209.5	2.73	1.975	4	-4.33
16	3-Ethyltoluene	0.8	0.0	0.1	0.0	0.0	0.0	0.0	1.8	120.2	3.58	2.000	0	-3.86
17	2,4-Difluoronitrobenzene	1.3	0.4	0.4	0.0	0.0	0.0	0.0	4.2	159.1	2.21	2.215	4	-4.49
18	3-Chloro-4-fluoronitrobenzene	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.8	175.5	2.66	2.215	4	-4.38
19	2,2-Dimethylbutanoic acid	2.0	0.7	1.1	1.5	4.0	0.5	30.0	40.0	116.2	1.94	2.222	3	-4.48
20	4-Methylthiobenzaldehyde	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.4	152.2	2.31	2.286	2	-4.34
21	1-Methylpropylbenzene	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.7	134.2	3.94	2.286	0	-3.83
22	4-Carboxybenzaldehyde	3.0	1.3	1.3	1.8	3.1	1.4	27.6	46.1	150.1	1.59	2.500	4	-4.68
23	4-Bromophenetole	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.4	201.1	3.46	2.651	1	-4.07
24	3,3-Dimethylpentane	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.2	3.67	2.667	0	-3.95
25	Methyl trimethyl acetate	0.6	0.3	0.1	0.0	0.0	0.0	0.0	1.9	116.2	1.74	2.813	2	-4.56

Table 8.2 continued

	Chemical	Redness	Chemosis	Discharge	Opacity	Area	Iritis	WCS	WDS	MW	logP	κ3	logP	n_H	logKp
26	2,2-Dimethyl-3-pentanol	0.8	0.3	0.0	0.8	0.6	0.0	2.1	4.5	116.2	-0.24	2.813	2.13	2	-4.47
27	*cis*-Cyclooctene	1.1	0.0	0.0	0.0	0.0	0.0	0.0	2.2	110.2	-0.14	2.813	3.94	0	-3.92
28	Glycerol	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.6	92.1	0.28	3.000	-1.65	6	-5.64
29	Methylthioglycolate	2.4	1.8	1.3	2.2	1.9	0.9	20.7	36.4	106.1	2.08	3.000	0.32	3	-4.98
30	1,2,3-Trimercaptopropane	1.3	1.5	0.3	0.0	0.0	0.0	0.0	6.2	140.3	-0.31	3.000	1.57	3	-4.70
31	Ethyl trimethyl acetate	0.6	0.4	0.1	0.2	0.2	0.0	0.1	2.3	130.2	1.15	3.556	2.23	2	-4.57
32	Ethyl-2-methyl acetoacetate	1.2	0.9	1.3	0.9	1.4	0.0	6.4	13.3	144.2	3.1	3.703	0.21	3	-5.13
33	Glycidyl methacrylate	1.5	1.2	0.6	0.8	1.4	0.4	5.3	14.4	142.2	0.45	3.703	0.81	3	-4.99
34	3-Methoxy-1,2-propanediol	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	106.1	0.8	3.840	-1.15	5	-5.60
35	Methyl cyanoacetate	3.0	2.1	1.0	0.4	1.3	1.0	2.9	20.6	99.1	3.09	3.840	-0.47	3	-5.31
36	Ethylthioglycolate	1.2	0.8	0.4	1.0	0.8	0.5	4.0	11.6	120.2	1.23	3.840	0.81	3	-5.01
37	2-Methyl-1-pentanol	1.1	0.7	0.0	1.1	1.2	0.0	6.9	10.7	102.2	2.54	3.840	1.75	2	-4.73
38	3-Methylhexane	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.2	100.2	1.28	3.840	3.71	0	-4.14
39	Propylene glycol	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.6	76.1	2.89	4.000	-0.78	4	-5.47
40	Butanone	2.1	2.2	2.3	1.3	1.7	0.6	10.6	27.0	72.1	2.73	4.000	0.26	1	-5.03
41	Methyl acetate	2.2	2.1	1.8	1.3	1.3	0.8	8.4	25.1	74.1	3.58	4.000	0.37	2	-5.07
42	3-Chloropropionitrile	0.6	0.3	0.2	0.7	0.4	0.1	1.5	4.5	89.5	2.21	4.000	0.6	3	-5.09
43	Isobutanol	2.4	2.3	1.4	2.1	2.6	1.1	26.9	44.6	74.1	2.66	4.000	0.77	2	-4.98
44	Butanol	2.4	2.1	1.6	2.0	2.4	0.7	24.2	40.3	74.1	1.94	4.000	0.84	2	-4.96
45	1,3-Dibromopropane	0.3	0.3	0.0	0.0	0.0	0.0	0.0	1.3	201.9	2.31	4.000	2.5	2	-4.58
46	2-Bromobutane	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	137.0	3.94	4.000	2.58	1	-4.50
47	Diethylaminopropionitrile	3.0	2.6	2.4	3.5	1.7	1.0	29.0	50.1	126.2	1.59	4.500	0.53	2	-5.12
48	Allyl methacrylate	0.8	0.3	0.2	0.0	0.1	0.0	0.0	2.7	126.2	3.46	4.500	2.12	2	-4.75
49	2-Ethyl-1-hexanol	2.1	1.9	0.8	1.8	1.5	0.7	13.9	27.2	130.2	3.67	4.500	2.73	2	-4.61
50	Ethyl acetate	1.2	0.6	0.3	0.4	0.3	0.1	0.7	5.5	88.1	1.74	5.333	0.86	2	-5.18

Table 8.2 continued

	Chemical	Redness	Chemosis	Discharge	Opacity	Area	Iritis	WCS	WDS	MW	logP	κ3	n_H	logKp
51	1-Bromo-4-chlorobutane	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	171.5	2.9	5.333	3	-4.78
52	1,4-Dibromobutane	0.3	0.3	0.0	0.0	0.0	0.0	0.0	1.4	215.9	2.99	5.333	2	-4.69
53	1,5-Hexadiene	0.9	0.2	0.1	0.0	0.0	0.0	0.0	2.4	82.1	3.02	5.333	0	-4.55
54	Bromopentane	0.6	0.0	0.0	0.0	0.0	0.0	0.0	1.1	151.0	3.14	5.333	1	-4.59
55	2-Methylpentane	0.3	0.0	0.1	0.0	0.0	0.0	0.0	0.8	86.2	3.21	5.333	0	-4.51
56	Cellosolve acetate	1.2	0.7	0.5	0.4	0.5	0.0	1.0	6.1	132.2	0.31	5.878	4	-5.54
57	2-Methoxyethyl acrylate	2.3	2.7	1.4	1.9	1.0	0.6	9.4	25.5	128.2	1.05	5.878	2	-5.23
58	Thiodiglycol	0.5	0.3	0.3	0.0	0.0	0.0	0.0	2.4	122.2	-0.62	6.000	5	-5.84
59	2,4-Pentanediol	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.9	104.1	0.13	6.000	4	-5.60
60	Methyl isobutyl ketone (4-methylpentan-2-one)	0.8	0.2	0.0	0.1	0.1	0.0	0.0	2.1	100.2	1.16	6.000	1	-5.16
61	Hexanol	2.7	2.5	2.5	2.2	2.9	1.3	31.7	53.7	102.2	1.82	6.000	2	-5.08
62	1,5-Dibromopentane	0.6	0.3	0.0	0.0	0.0	0.0	0.0	1.8	229.9	3.48	6.000	2	-4.69
63	Bromohexane	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.4	165.1	3.63	6.000	1	-4.59
64	2-Ethoxyethyl methacrylate	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	158.2	1.49	6.400	3	-5.29
65	Trifluoroethyl methacrylate	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.4	168.1	2.18	6.400	5	-5.26
66	Ethyl thioethyl methacrylate	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	174.3	2.34	6.400	3	-5.09

Footnote to Table 8.2

κ3= third-order kappa index; logKp = logarithm of the corneal permeability, predicted with Equation 8.2; logP = logarithm of the octanol-water partition coefficient; MW= molecular weight; n_H = number of hydrogen bonds; WCS = weighted corneal score; WDS = weighted Draize score.

CHAPTER 9

DEVELOPMENT OF NEW MODELS FOR ACUTE EYE TOXICITY

9.1 OBJECTIVES

The aim of the work presented in this chapter was to develop new models for predicting the eye irritation potential of organic chemicals. In particular, the following types of model were sought: a) SARs based on physicochemical properties that are easily calculated; b) PMs based on measured pH values; and c) PMs based on *in vitro* data. While the emphasis of this chapter is to describe the development and assessment of individual models, the integrated use of these models is discussed in Chapter 10.

9.2 METHOD

9.2.1 Collection and treatment of data

For the development of SARs, a data set of 119 organic chemicals (Tables 9.1 & 9.2) was compiled by taking 82 chemicals from the ECETOC reference chemicals data bank for eye irritation (ECETOC, 1998a) and 37 chemicals from Spielmann *et al.* (1996). Chemicals were only included if they were predicted to be liquids at physiological temperature (37 °C). Chemicals predicted to be solids were excluded from the analyses, partly because the exposure of the rabbit eye to solids is not comparable to that of liquids (Balls *et al.*, 1999), and partly because the irritant effects of solids may be caused by physical abrasion, rather than by chemical or biochemical effects. All chemicals were classified as eye irritants or non-irritants by applying EU and OECD classification criteria to the animal data (see chapter 2). The weighted Draize score (WDS) was calculated for the 82 ECETOC chemicals, but could not be calculated for the remaining chemicals since not all of the necessary tissue scores were reported by Spielmann *et al.* (1996).

For the development of a PM based on pH data, a training set of 165 chemicals (Table 9.3) was compiled from Régnier & Imbert (1992), and a test set of 49 chemicals (Table 9.4) was taken from Balls *et al.* (1995). The pH data for these 49 chemicals were provided by BIBRA International (Croydon, UK). The pH values in the two data sets are based on measurements carried out in a 10% aqueous solution (w/v for solids and v/v for liquids).

For the development of PMs based on *in vitro* data, two validation study data sets were investigated. The first *in vitro* data set (Table 9.5) contains 59 of the 60 chemicals tested in the European Commission/British Home Office (EC/HO) validation study (Balls *et al.*, 1995). One of the 60 chemicals, thiourea, was omitted from the data set because it is acutely toxic. In the validation study, 18 endpoints obtained with 9 test systems were examined. The analyses reported here are based on 15 of the *in vitro* endpoints, obtained with 7 test systems (EC/HO data for the HETCAM and NRU tests were omitted). The data analysed for each endpoint were the average values of measurements made in three or more laboratories, the raw data being taken from the ECVAM archive. The second *in vitro* data set (Table 9.6) contains 143 chemicals taken from a validation study carried out in Germany (Spielmann *et al.*, 1996). Seven *in vitro* endpoints were investigated, one being the concentration of test chemical that causes 50% inhibition of neutral red uptake into 3T3 cells (IC50), and the other six being endpoints determined in the HETCAM test: the mean time for hæmorrhage (TH), lysis (TL) and coagulation (TC), using either the undiluted substance (TH100, TL100, TC100) or a 10% solution (TH10, TL10, TC10). These data were taken directly from Spielmann *et al.* (1996), and represent the mean values of measurements made in two or more laboratories.

9.2.2 Calculation of physicochemical properties

Physicochemical properties considered to be possible predictors of eye irritation potential were calculated from the molecular structures of the 119 chemicals, using the methods described in Chapters 6, 7 and 8. Briefly, logP and logS were calculated with the Syracuse Research Corporation KOWWIN and WS-KOWWIN software, respectively; molecular weight (MW) and surface tension (ST) were calculated with the Advanced Chemistry Development (ACD) ChemSketch software; and pKa values with the ACD pKa software. In the case of chemicals with more than one ionisable group, all pKa values were calculated. Calculations of molecular surface area (MSA), molecular volume (MV), dipole moment (DM), and of the energies (E_{LUMO} and E_{HOMO}) of the lowest unoccupied molecular orbital (LUMO) and the highest occupied molecular orbital (HOMO), were performed on energy-minimised structures in HyperChem 5.1 for Windows, as described in chapter 6. The molecular orbital energies were used to calculate two additional energy variables, hardness (η) and electronegativity (χ), by applying Equations 3.36 and 3.37, respectively (see Chapter 3). Finally, the first-order

difference molecular connectivity valence index (dV1) and the third-order kappa index (κ3) were calculated with Molconn-Z. The first of these connectivity indices, dV1, was found to discriminate between irritants and non-irritants (defined on the basis of a MMAS cut-off) in chapter 6, whereas κ3 was found to be a predictor of corneal permeability in chapter 8, and a predictor of the molar eye score by Kulkarni & Hopfinger (1999).

9.2.3 Development of structure-activity relationships

A preliminary inspection of the physicochemical data, using dot plots, scatter plots, and matrix plots, was performed to visually identify the best variables or pairwise combinations of variables capable of distinguishing between irritant and non-irritant chemicals. Any variable seen to separate irritants from non-irritants in a linear manner was subjected to one of the following statistical methods: linear discriminant analysis (LDA), binary logistic regression (BLR) or classification tree (CT) analysis. As described in chapter 4, the use of LDA is appropriate when the data in the two groups (I and NI) are normally distributed and have approximately equal variance-covariance matrices. If the data in one or both groups are not normally distributed, BLR can be used, provided that the residuals associated with the logistic model are normally distributed. If the assumptions of both LDA and BLR are violated, CT analysis can be used as a last resort. Any pairwise combination of variables seen to separate irritants from non-irritants by the formation of an embedded cluster was analysed by the sequential application of CSA and ECM, as described in chapter 6. LDA, BLR and CT analyses were performed with STATISTICA, whereas CSA and ECM were performed with purpose-written Minitab macros (Appendices A8 and A10).

9.2.4 Development of prediction models

A PM based on pH data was derived from a training set of 165 chemicals (Table 9.3) by CT analysis, as described in chapter 7. The performance of the resulting CM was assessed by applying the model both to its training set and to a test set of 49 chemicals (Table 9.4).

A preliminary inspection of the two *in vitro* data sets was performed to identify the best variables or pairwise combinations of variables capable of distinguishing between irritant and non-irritant chemicals, and PMs were developed by BLR or CT analysis.

9.2.5 Assessment of classification models

The predictive capacities of the two-group CMs were assessed in terms of their Cooper statistics, as described in chapter 7.

9.3 RESULTS

9.3.1 Comparison of EU and OECD classifications

Table 9.1 shows that the application of EU and OECD classification criteria to the Draize rabbit data produces the same classifications of eye irritation potential for 104 of the 119 chemicals (i.e. 87%), with the EU risk phrase R36 corresponding to the OECD class B, and the EU risk phrase R41 corresponding to the OECD class A. However, due to the lower thresholds of the OECD system (see chapter 2), 15 of the 119 chemicals are classified as non-irritants by the EU system and as class B irritants by the OECD system. These chemicals are: 2-methyl-1-pentanol, cyclopentanol, ethanol, isopropanol, 2-ethyl-1-hexanol, 2,2-dimethylbutanoic acid, ethylthioglycolate, butanal, ethylbutanal, isobutanal, paraformaldehyde, 3-cyclohexene-1-methanol, 2,5-dimethylhexanediol, 3,6-dimethyloctanol, and 1-chloroctan-8-ol. In this study, the EU classifications (30 I and 89 NI) were used for model development.

9.3.2 Structure-activity relationships

A preliminary inspection of the data revealed that MW was highly correlated with both MV and MSA in the data set of 119 chemicals ($r_{MW-MV} = 0.83$ and $r_{MW-MSA} = 0.83$), so the latter two variables were not considered further as predictor variables. Similarly, logS was not considered as a predictor variable, due to its high collinearity with logP ($r = -0.98$, $p<0.01$). A visual inspection of the data for the remaining variables indicated that a cut-off between irritant and non-irritant chemicals could be defined along MW

(Figure 9.1), whereas an elliptic boundary could be drawn in the two-dimensional space formed by dV1 and logP (Figure 9.2).

Linear model based on molecular weight

To illustrate that different classification models can be produced by using LDA, BLR and CT analysis, the results obtained by applying each of these methods to the MW data for 119 chemicals are briefly reported. Strictly, LDA is not appropriate, since the irritants are not normally distributed (as judged by the Lilliefors test). However, if the normality condition is ignored, LDA produces a statistically significant ($p<0.01$) discriminant model with a cut-off value of 121 g/mol. The application of BLR produces the logistic model described by Equation 9.1, which is also statistically significant ($p<0.01$), and corresponds to a cut-off value of 77 g/mol (=1.996/0.026).

$$p_I = \exp(1.996 - 0.026\ MW) / \{1+\exp(1.996 - 0.026\ MW)\} \qquad \text{(Equation 9.1)}$$

However, the residuals of this model are not normally distributed (they are bimodal), suggesting that the use of a logistic model is not appropriate. Using CT analysis as a last resort produces a cut-off of 137 g/mol, from which the following SAR can be formulated:

MW ≤ 137 g/mol, then predict I; otherwise, predict NI. (SAR 1)

The cut-off value 137 g/mol is illustrated in Figure 9.1, and the performance of SAR 1 is summarised in Table 9.7.

Elliptic model based on lipophilicity and molecular shape

The application of CSA to the data set of 119 chemicals confirmed that embedded clustering is statistically significant ($p < 0.01$) along both logP and dV1. Following the removal of two irritant outliers (octanol [#29] and camphen [#114]), the application of ECM generated the elliptic model described by SAR 2. The 2D ellipse described by SAR 2 is illustrated in Figure 9.2, and the performance of the model is summarised in Table 9.7.

Classify an undiluted, organic liquid as an eye irritant if:

$(\log P\text{-}0.66)^2 / 1.21^2 + (dV1 + 1.22)^2 / 1.12^2 \leq 1$ (SAR 2)

Relationship between pKa and irritation potential

Of the 119 organic liquids in Table 9.1, 47 chemicals (Table 9.2) are ionisable, and can be classified into the following groups: 3 acids (2 I, 1 NI), 33 alcohols (13 I, 20 NI), and 12 bases (6 I, 6NI). Ten of the 47 chemicals have more than one ionisation centre and therefore more than one pKa value. Dot plots of the minimal pKa values for the 36 acids and alcohols, and of the maximal pKa values for the 12 bases, provided no evidence for a separation between irritants and non-irritants along pKa.

9.3.3 Prediction models

Prediction models based on pH data

The distribution of pH values for irritants and non-irritants in the data set of 165 chemicals is shown in Figure 9.3. It can be seen that irritants span a wide range of pH values from 0 to about 12, whereas the non-irritants span a much narrower range from about 3 to 9. Using the cut-off values generated by CT analysis, the following PM can be formulated:

If pH < 3.2 or if pH > 8.6, then predict I; otherwise, predict NI. (PM 1)

The Cooper statistics obtained by applying PM 1 to its training set and to an independent test set are summarised in Table 9.7.

Prediction models based on in vitro data

A preliminary inspection of the *in vitro* data from the EC/HO study (Balls *et al.*, 1995) suggested that the four endpoints measured in the isolated rabbit eye (IRE) test (opacity and swelling at 1h and 4h) provide the best discrimination between irritants and non-irritants (Figure 9.4). Following the removal of two outliers (2,2-dimethylbutanoic acid [#27] and ethanol [#41]), CT analysis produced the following PMs 2-5 for the IRE test. It is interesting to note that the two outliers are both classified differently under the EU and OECD systems, which is indicative of their borderline nature.

If opacity score at 1h > 0.9, then predict I; otherwise, predict NI. (PM 2)

If opacity score at 4h > 1.7, then predict I; otherwise, predict NI. (PM 3)

If swelling at 1h > 18%, then predict I; otherwise, predict NI. (PM 4)

If swelling at 4h > 23%, then predict I; otherwise, predict NI. (PM 5)

A preliminary inspection of the data from the German validation study (Spielmann *et al.*, 1996) indicated that the separation between irritants and non-irritants was more apparent if the seven endpoints were expressed in logarithmic form (Figure 9.5). However, it was difficult to judge by visual inspection which endpoint or combination of endpoints provides the best discrimination. Therefore, forward stepwise BLR was applied to the data, using the generalised linear modelling module in STATISTICA 5.5 for Windows. The seven endpoints entered the model in the following order: log(TH10), log(IC50), log(TC10), log(TC100), log(TH100), log(TL10), log(TL100). Only the first three variables are significant at the 99.9% level. The application of BLR to the most significant variable, log(TH10), produced a model whose residuals were bimodal. However, the application of BLR to the two most significant variables, log(TH10) and log(IC50), produced a model whose residuals were approximately normal. The two-variable logistic model is given by Equation 9.2, and is illustrated in Figure 9.6:

$$p_I = \exp(Z) / \{1+\exp(Z)\} \qquad \text{(Equation 9.2)}$$

where $Z = 5.25 - 3.04 \log(TH10) - 0.79 \log(IC50)$

Figure 9.6 shows that a chemical is more likely to be an eye irritant if its log(IC50) value is low (i.e. if the chemical is cytotoxic to 3T3 cells) and if its log(TH10) value is low (i.e. if a 10% solution of the chemical produces rapid hæmorrhaging of the chorioallantoic membrane). From Equation 9.2, it follows that if $Z > 0$, then $p_I > 0.5$ (i.e. the chemical is more likely to be irritant than non-irritant). Thus, the logistic model can be rewritten in the form of the following PM:

If $3.04 \log(TH10) + 0.79 \log(IC50) < 5.25$, predict I; otherwise, predict NI. (PM 6)

PM 6 was developed from and applied to the data for 129 chemicals, since data were missing for 14 of the 143 chemicals. The performance of PMs 2-6 are summarised in Table 9.7.

9.4 DISCUSSION

In this chapter, several CMs for predicting eye irritation potential are reported. The classification models take the form of either SARs based on easily obtained or calculated physicochemical data, or PMs based on physicochemical or *in vitro* data. Since PMs are based on experimental measurements, it is important that the data used to develop the PMs have been obtained by application of a standardised protocol. The pH measurements analysed in this study are standardised in the sense that they refer to measurements obtained in a 10% aqueous solution. The *in vitro* data were taken from two validation studies in which the use of standardised protocols was a requirement for assessment of the *in vitro* methods. In addition, the data for the 3T3 NRU and HETCAM tests were subjected to independent quality control before being used in the German validation study.

9.4.1 Contribution to existing knowledge

The results obtained in this study are consistent with several findings made in previous SAR studies (see chapter 2). For example, the observation of a relationship between eye irritation potential and MW is consistent with the finding by Rosenkranz *et al.* (1998) that the mean MW of irritants is lower than the mean MW of non-irritants. Rosenkranz and colleagues did not, however, publish a PM based on MW. In addition, the lack of an apparent relationship between eye irritation potential and the orbital energy descriptors (E_{HOMO}, E_{LUMO}, electronegativity and hardness) is consistent with the conclusion made by Rosenkranz *et al.* (1998) that chemical reactivity is not necessary for eye irritation. The embedded clustering of irritants along logP has been reported previously by Cronin (1996), and also by the author (Worth & Fentem, 1999). The use of the molecular connectivity index dV1 as a predictor of eye irritation potential does not appear to have been reported before, other than by the author (Worth & Cronin, 2000a).

The possibility of a relationship between pKa and eye irritation potential was explored, because a relationship between pKa and the weighted corneal score was reported by Sugai *et al.* (1991) for a homogeneous set of salicylates. However, since no relationship could be found, it appears that pKa is not a useful predictor of eye irritation potential of heterogeneous sets of chemicals.

The importance of DM (e.g. Barratt, 1997) was not confirmed for the heterogeneous set of chemicals investigated in this study. Again, this may be a property that is only predictive for relatively homogeneous groups of chemicals, such as the 'neutral organics' investigated by Barratt. In any case, DM is not an ideal predictor variable, since its calculated value is dependent on the conformation and electron distribution of the molecular structure, both of which vary according to the molecular modelling methods employed. Thus, the use of more 'robust' physicochemical properties variables, such as MW and logP, is preferable for the development of SARs.

Contrary to expectation, the kappa index (κ3) was not found to be a useful predictor of eye irritation potential. The mean value for non-irritants (5.07) was found to be higher than the mean value for irritants (3.76), but there was significant overlap between the distributions of irritant an non-irritant κ3 values. Using the data for the 82 ECETOC chemicals, κ3 was found to be weakly, but significantly, correlated with the weighted Draize score ($r = -0.3$, $p<0.01$). The observation of an inverse relationship between κ3 and the WDS is consistent with the inverse relationship found between the corneal permeability coefficient and κ3 (see chapter 8), but is inconsistent with the positive correlation reported by Kulkarni & Hopfinger (1999; see chapter 3).

As far as the author is aware, a PM for eye irritation potential based on pH data alone has not been published (other than by Worth & Fentem, 1999), although Régnier & Imbert (1992) reported the combined use of pH and acid-alkaline reserve. In addition, a PM based on the isolated rabbit eye (IRE) test does not appear to have been published. In the EC/HO validation study, the IRE test was assessed entirely on the basis of its reproducibility and on the correlations between its four endpoints and the MMAS. At the time of the EC/HO study, the concept of the PM had not been developed, so the assessment of PMs did not form an integral part of interlaboratory validation studies. Any PM based on IRE data should be of regulatory interest, given that the IRE test is

accepted by the regulatory authorities in the United Kingdom as an alternative to the Draize eye test.

The models developed on the basis of 3T3 NRU and HETCAM data represent a re-analysis of the German validation study (Spielmann *et al.*, 1996). In the validation study, stepwise LDA was used to identify the best (most discriminating) three endpoints for distinguishing between severe (R41) and non-severe (NI, R36) irritants as TC10, log(IC50) and TC100 (in decreasing order of statistical significance). In contrast, stepwise BLR was used in this study to identify the best three endpoints for distinguishing between irritants (R36 and R41) and non-irritants as log(TH10), log(IC50) and log(TC10). It should be noted that the *in vitro* data for the HETCAM and NRU tests are not normally distributed within groups, which may invalidate any conclusions drawn from the use of LDA to assess the statistical significance of the seven endpoints.

9.4.2 Interpretation and assessment of the models

The classification model based on MW (SAR 1) states that for an organic liquid to be irritating to the eye, it must have a molecular size below a certain threshold. The high sensitivity of 97% (Table 9.7) indicates that only a small proportion of irritant chemicals (3%) have a MW greater than the cut-off value of 137 g/mol. However, not all chemicals that are small enough to penetrate the eye are irritants, as reflected by the fact that 51% of the non-irritant chemicals also have a MW less than 137 g/mol (false positive rate = 51%). As a stand-alone model for predicting eye irritation potential, the performance of SAR is unsatisfactory - it is good at identifying known irritants, but this is at the expense of over-classifying known non-irritants. However, the model could be useful in the context of a tiered testing strategy (see § 9.4.3 below).

A mechanistic interpretation of the elliptic model based on logP and dV1 is offered in chapter 6. Given that the sensitivity, specificity and concordance for this model (Table 9.7) are all greater than 50%, SAR 2 could be regarded as an acceptable stand-alone model for predicting the eye irritation potential of organic liquids.

The PM based on pH (PM 1) has a clear mechanistic interpretation - chemicals that do not form acidic or basic solutions are unlikely to be irritating to the eye. This is reflected by the fact that the model correctly identifies 97% of the known non-irritants in its training set and 94% of the non-irritants in its test set. However, not all chemicals that have intermediate pH values in the range 3.2 to 8.6 are non-irritant, as reflected by the fact that 47% of the known irritants are under-classified as non-irritant. Thus, PM 1 would be unsatisfactory as a stand-alone model, but may be useful when used in combination with other models (see § 9.4.3 below).

The PMs based on IRE data (PMs 2-5) express the fact that eye irritation is associated with opacity and swelling in the cornea. These PMs identify 46-69% of the irritants, and 89-100% of the non-irritants, in the training set of 57 chemicals. Of the four models, only PMs 4 and 5 can be regarded as satisfactory stand-alone models, since PMs 2 and 3 have sensitivities that are not significantly greater than 50%.

The PM based on the combined use of HETCAM and NRU tests (PM 6) indicates that irritant chemicals are generally more cytotoxic than non-irritants, and supports the use of the chorioallantoic membrane as a model for eye irritation. Given that the sensitivity, specificity and concordance of PM 6 (Table 9.7) are all greater than 50%, this model could be regarded as an acceptable stand-alone model.

9.4.3 Strategic use of the models

In the preceding discussion, the various classification models are judged in terms of their Cooper statistics (sensitivity, specificity and concordance), which is the conventional way of assessing two-group classification models in the field of alternative methods. Although different criteria may be adopted for the acceptability of models, depending on the toxicological endpoint they are designed to predict, a minimum set of conditions would be that the sensitivity, specificity and concordance for a stand-alone method should all be significantly greater than 50%. This would lead to the rejection of SAR 1 and PM 1 as stand-alone models. However, by integrating these models into a tiered testing strategy, their strengths can be combined without introducing their weaknesses. Thus, SAR 1 could be rephrased as follows:

MW > 137 g/mol, then predict NI; otherwise, make no prediction (SAR 1b)

In other words, SAR 1b is being used to identify non-irritants, but not irritants. This can be justified on the basis that even though SAR 1 only identifies 49% of the known non-irritants in the data set, of those chemicals it does predict to be non-irritant, 98% of these predictions are correct. This is the same as saying that the negative predictivity of SAR 1 is 98%.

Similarly, PM 1 could be recast as follows:

If pH < 3.2 or if pH > 8.6, then predict I; otherwise, make no prediction (PM 1b)

Thus, PM 1b is being used to identify irritants, but not non-irritants. This can be justified on the basis that even though PM 1b only identifies 53% of the known irritants in its training set, and 44% of the known irritants in a test set, 93% of the irritant predictions made by PM1b are correct in both the data sets. In other words, PM 1b has a positive predictivity of 93%. This approach is mechanistically reasonable, in that while chemicals at the extremes of the pH range are expected to be irritant, not all irritant chemicals are expected to act by a pH-dependent mechanism. Therefore, even chemicals with intermediate pH values could be irritant, which is clearly illustrated in Figure 9.3.

The manner in which SAR 1b and PM1b should be integrated will depend on the underlying rationale of the tiered testing strategy. In the OECD testing strategies (see chapter 2), any chemical predicted to be toxic is classified without further testing, whereas chemicals predicted to be non-toxic undergo further testing until the absence of toxic potential is confirmed by conducting an animal experiment. However, tiered testing strategies could be developed according to a different rationale, as illustrated in Figure 9.7. In this approach, CMs with a high specificity (and high positive predictivity) are used to classify chemicals as irritants without further testing, whereas CMs with a high sensitivity (and high negative predictivity) are used to identify chemicals that are so unlikely to be irritant that they can be tested directly in the rabbit, without the need to conduct *in vitro* testing. Of course, if no model is available for any given step, or if the model is not applicable to the chemical in question, testing must proceed to the next

step. An evaluation of this approach, and of the OECD approach, is presented in chapter 10.

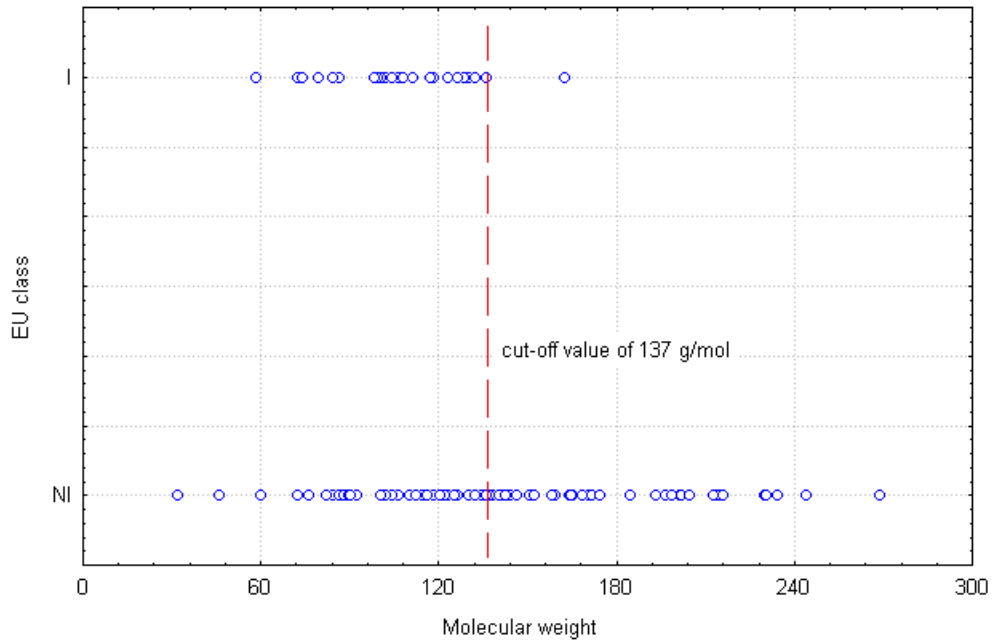
It follows from the preceding discussion that models designed for use in the context of a tiered testing strategy do not need to meet the same performance criteria as models intended for stand-alone use. In particular, a model with a sensitivity ≤ 50% may be useful, if it has a high specificity (and therefore a low false positive rate). Similarly, a model with a low specificity may be useful, as long as its sensitivity is high and its false negative rate is low. Bearing this in mind, a possible approach to the design of a tiered testing strategy would be to parameterise models based on suitable predictor variables in such a way that high sensitivities or specificities are ensured. In the case of 'simple' classification models, such as linear models based on a single predictor variable, cut-off values could be defined in terms of percentiles. For example, given the distributions of irritant and non-irritant chemicals along MW (Figure 9.1), one could choose the 95th percentile in the distribution of 'irritant' MWs to define the cut-off for the SAR as 134 g/mol (instead of 137 g/mol). This cut-off would ensure that 95% of the irritant chemicals are correctly identified (sensitivity = 95%) and that 5% of the irritant chemicals are under-predicted (false negative rate = 5%). This 'percentile approach' was adopted in Worth & Fentem (1999), in which a fixed sensitivity of 75% was used to derive cut-off values along 17 *in vitro* endpoints for eye irritation potential.

9.5 CONCLUSIONS

It is concluded that:

1) eye irritation potential can be predicted by using: a) SARs based on molecular size (MW), partitioning behaviour (logP), and molecular shape (dV1); and b) PMs based on pH, the degree of swelling and opacity in the IRE test, cytotoxicity (to 3T3 fibroblasts), and the onset time of hæmorrhaging in the hen's egg chorioallantoic membrane.
2) CMs designed for incorporation into a tiered testing strategy do not need to be subjected to the same performance criteria as CMs intended for stand-alone use, and the parameters of these models could be derived by a percentile method.

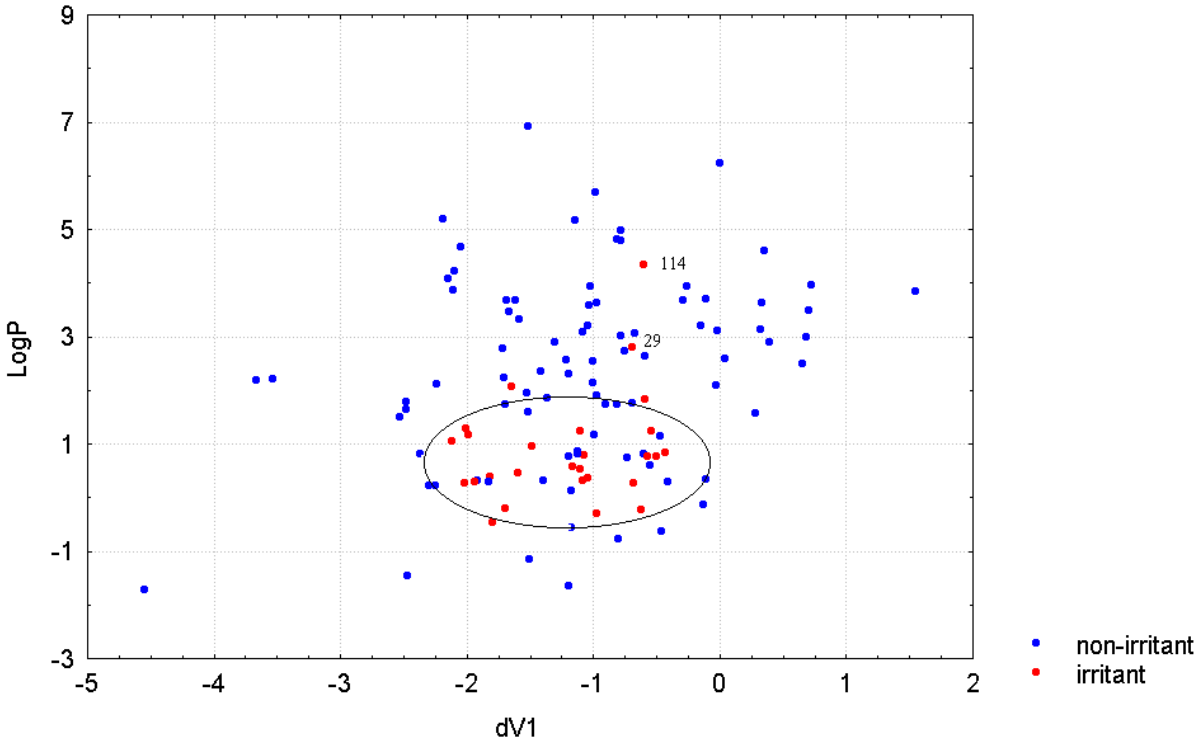
Figure 9.1 Distribution of eye irritants and non-irritants by molecular weight



Footnote to Figure 9.1

I = irritant (EU risk phrase R36 or R41); NI = non-irritant

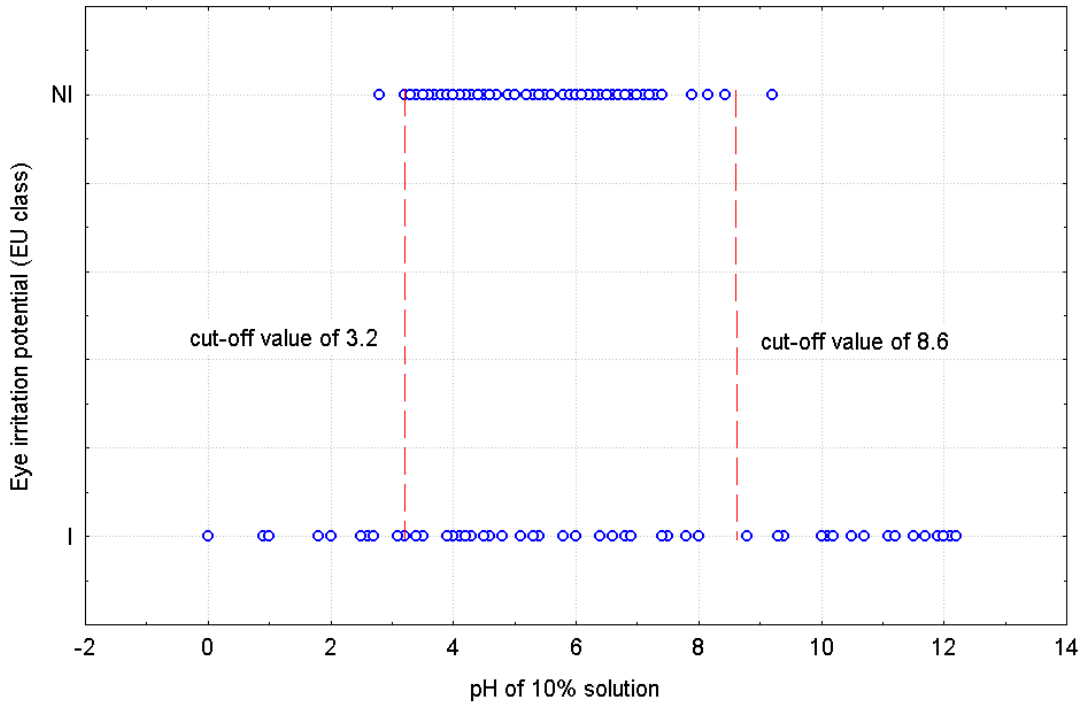
Figure 9.2 Embedded clustering of eye irritants



Footnote to Figure 9.2

Chemical # 29 = octanol; # 114 = camphen

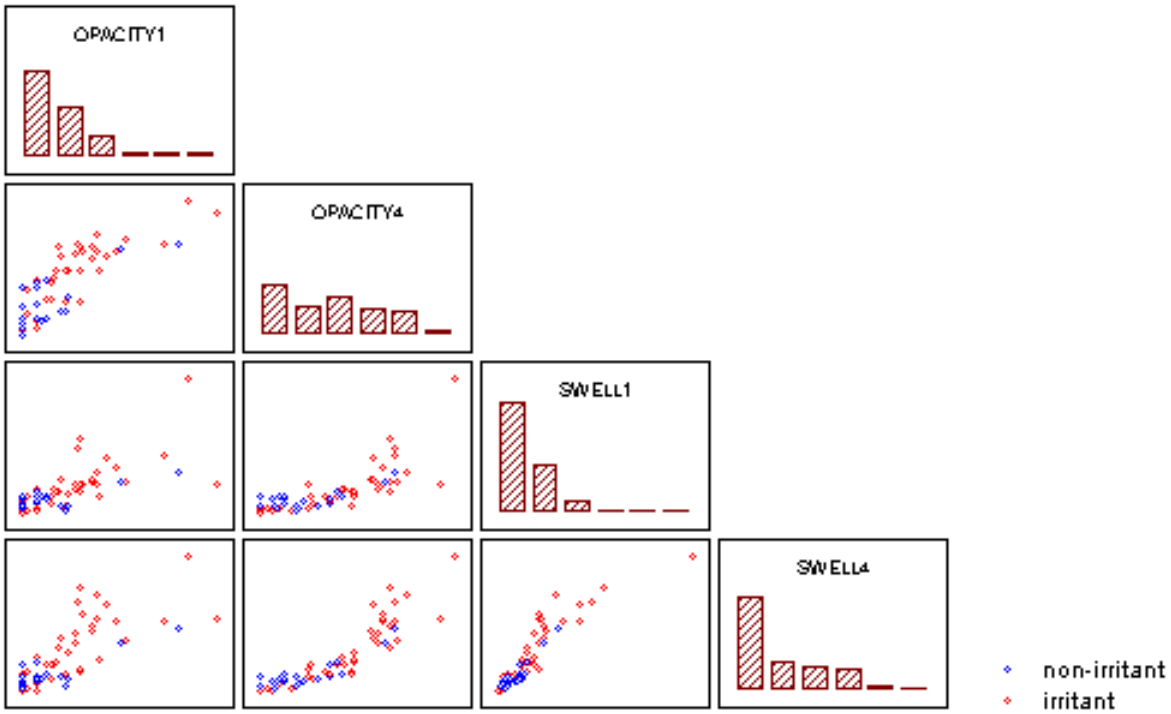
Figure 9.3 Distribution of eye irritants and non-irritants by pH



Footnote to Figure 9.3

I = irritant (EU risk phrase R36 or R41); NI = non-irritant

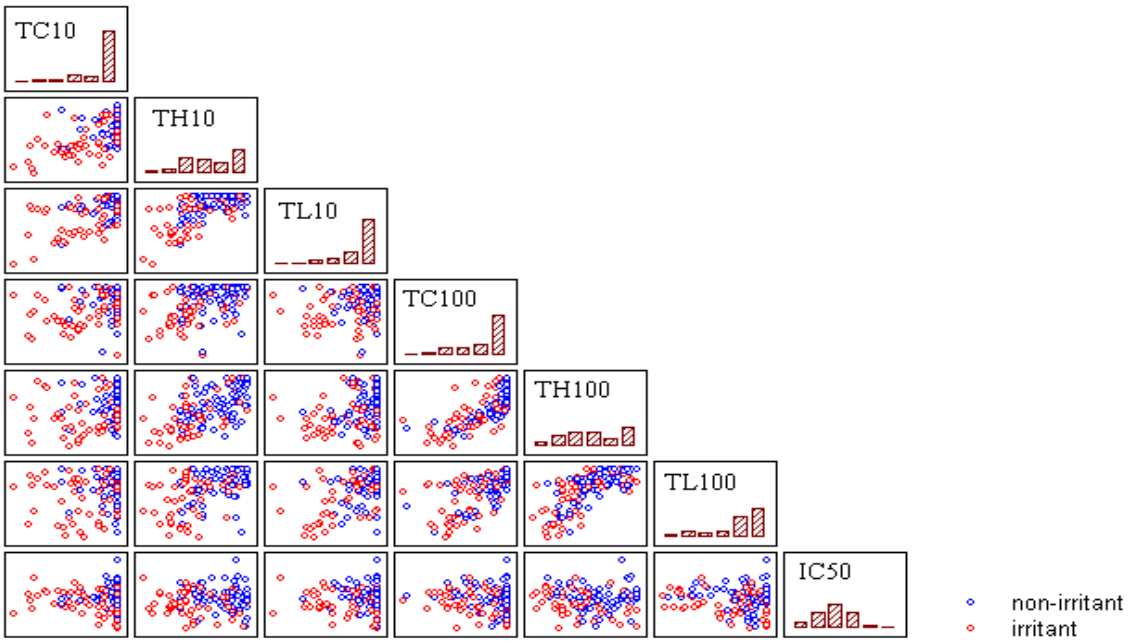
Figure 9.4 Ability of the isolated rabbit eye test to discriminate between eye irritants and non-irritants



Footnote to Figure 9.4

Endpoints of the isolated rabbit eye test: OPACITY1 = corneal opacity score at 1h; OPACITY4 = corneal opacity score at 4h; SWELL1 = corneal swelling (%) at 1h; SWELL4 = corneal swelling (%) at 4h.

Figure 9.5 Endpoints of the HETCAM and neutral red uptake tests



Footnote to Figure 9.5

Endpoints of the HETCAM test: TC10 = time for coagulation with a 10% solution; TH10 = time for hæmorrhage with a 10% solution; TL10 = time for lysis with a 10% solution; TC100 = time for coagulation with the undiluted substance; TH100 = time for hæmorrhage with the undiluted substance; TL100 = time for lysis with the undiluted substance.
Endpoint of the NRU test: IC50 = concentration causing a 50% reduction in neutral red uptake.

All endpoints are expressed in logarithmic form.

Figure 9.6 Logistic model for eye irritation potential based on the HETCAM and neutral red uptake tests

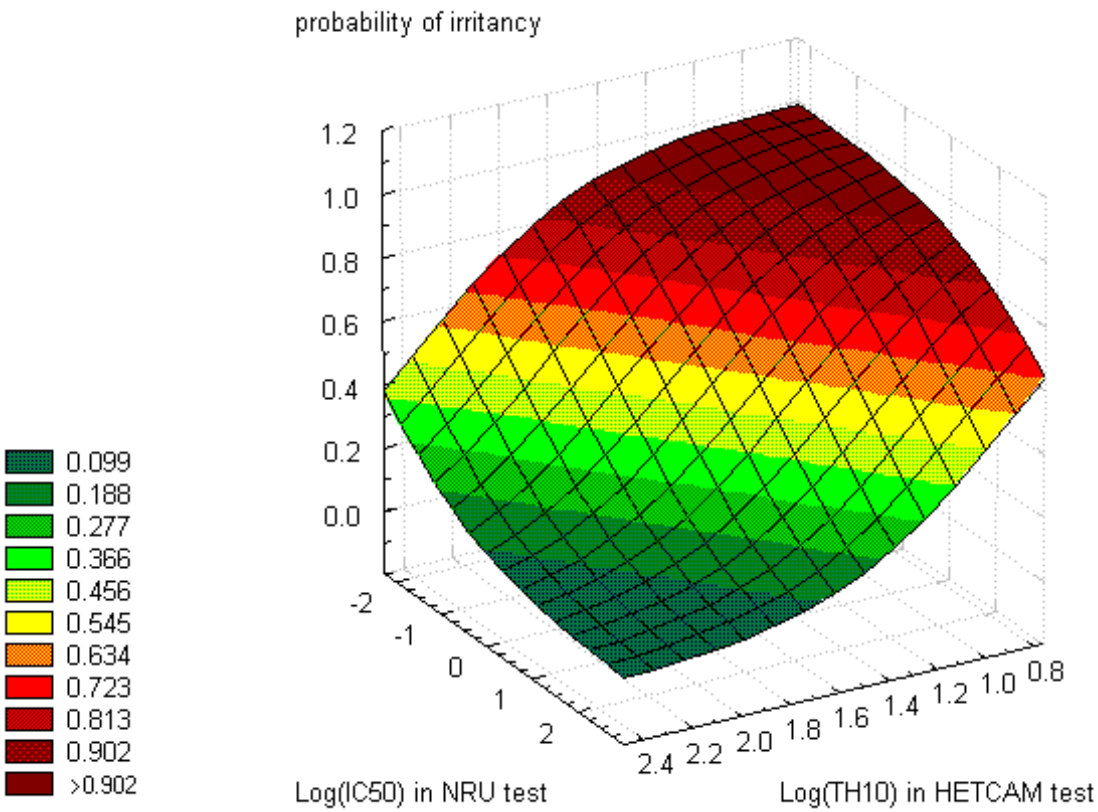


Figure 9.7 Outline of a tiered testing strategy for eye irritation

Step 1:	Apply a CM for I chemicals (e.g. PM 1b) ⇓ No prediction ⇓	⇒	Predict as I Stop testing
Step 2:	Apply a CM for NI chemicals (e.g. SAR 1b) ⇓ No prediction ⇓	⇒	Predict as NI Proceed to Step 4
Step 3:	Apply a PM based on *in vitro* data for I chemicals (e.g. PM 4b) ⇓ Predict as NI ⇓	⇒	Predict as I Stop testing
Step 4:	Perform Draize eye test	⇒	Classify as I or NI

Table 9.1 Draize rabbit eye irritation data set for 119 organic liquids

No	Chemical	Source[a]	Erythema	Chemosis	Opacity	Iritis	Classification	
							EU	OECD
1	3,3-Dimethylpentane	ECETOC	0.0	0.0	0.0	0.0	NI	NI
2	3-Methylhexane	ECETOC	0.1	0.0	0.0	0.0	NI	NI
3	2-Methylpentane	ECETOC	0.3	0.0	0.0	0.0	NI	NI
4	1,9-Decadiene	ECETOC	0.4	0.0	0.0	0.0	NI	NI
5	Dodecane	ECETOC	0.4	0.0	0.0	0.0	NI	NI
6	1,5-Dimethylcyclooctadiene	ECETOC	0.6	0.0	0.1	0.0	NI	NI
7	*cis*-Cyclooctene	ECETOC	1.1	0.0	0.0	0.0	NI	NI
8	Methylcyclopentane	ECETOC	0.8	0.1	0.0	0.0	NI	NI
9	1,5-Hexadiene	ECETOC	0.9	0.2	0.0	0.0	NI	NI
10	4-Bromophenetole	ECETOC	0.2	0.0	0.0	0.0	NI	NI
11	2-Xylene	ECETOC	1.2	1.7	0.0	0.0	NI	NI
12	1-Methylpropylbenzene	ECETOC	0.3	0.0	0.0	0.0	NI	NI
13	3-Ethyltoluene	ECETOC	0.8	0.0	0.0	0.0	NI	NI
14	2,4-Difluoronitrobenzene	ECETOC	1.3	0.4	0.0	0.0	NI	NI
15	Styrene	ECETOC	0.7	0.6	0.1	0.0	NI	NI
16	Toluene	ECETOC	1.5	1.4	0.0	0.0	NI	NI
17	4-Fluoroaniline	ECETOC	2.7	2.2	1.6	0.6	R36	B
18	Pyridine	ECETOC	2.9	3.3	2.7	1.2	R36	B
19	3-Methoxy-1,2-propanediol	ECETOC	0.0	0.0	0.0	0.0	NI	NI
20	2,4-Pentanediol	ECETOC	0.4	0.0	0.0	0.0	NI	NI
21	Propylene glycol	ECETOC	0.3	0.0	0.0	0.0	NI	NI
22	Glycerol	ECETOC	0.3	0.0	0.0	0.0	NI	NI
23	Isomyristyl alcohol	ECETOC	0.5	0.5	0.0	0.0	NI	NI
24	2,2-Dimethyl-3-pentanol	ECETOC	0.8	0.3	0.8	0.0	NI	NI
25	2-Methyl-1-pentanol	ECETOC	1.1	0.7	1.1	0.0	NI	B
26	Cyclopentanol	ECETOC	2.0	1.9	1.2	0.0	NI	B
27	Ethanol	ECETOC	2.1	1.3	1.1	0.4	NI	B
28	Isopropanol (propan-2-ol)	ECETOC	2.3	1.6	0.7	0.8	NI	B
29	Octanol	ECETOC	2.2	2.5	1.7	0.7	R36	B
30	Furfuryl hexanol	ECETOC	1.3	1.3	2.3	1.3	R36	B
31	2-Ethyl-1-hexanol	ECETOC	2.1	1.9	1.8	0.7	NI	B
32	Isobutanol	ECETOC	2.4	2.3	2.1	1.1	R36	B
33	Butanol	ECETOC	2.4	2.1	2.0	0.7	R36	B
34	Hexanol	ECETOC	2.7	2.5	2.2	1.3	R36	B
35	Butyl cellosolve(2-butoxyethanol)	ECETOC	2.3	2.8	2.3	1.3	R36	B
36	Cyclohexanol	ECETOC	2.4	2.5	2.9	1.2	R36	B
37	Methyl trimethyl acetate	ECETOC	0.6	0.3	0.0	0.0	NI	NI
38	Ethyl trimethyl acetate	ECETOC	0.6	0.4	0.2	0.0	NI	NI
39	Butyl acetate	ECETOC	1.0	0.3	0.2	0.0	NI	NI
40	Cellosolve (2-ethoxyethyl) acetate	ECETOC	1.2	0.7	0.4	0.0	NI	NI
41	Ethyl acetate	ECETOC	1.2	0.6	0.4	0.1	NI	NI
42	Ethyl-2-methyl acetoacetate	ECETOC	1.2	0.9	0.9	0.0	NI	NI
43	Methyl cyanoacetate	ECETOC	3.0	2.1	0.4	1.0	R36	B
44	Methyl acetate	ECETOC	2.2	2.1	1.3	0.8	R36	B
45	Ethyleneglycol diethylether	ECETOC	0.0	0.0	0.0	0.0	NI	NI
46	Di-isobutyl ketone (2,6-dimethyl-4-heptanone)	ECETOC	0.1	0.0	0.0	0.0	NI	NI
47	Methyl isobutyl ketone (4-methylpentan-2-one)	ECETOC	0.8	0.2	0.1	0.0	NI	NI

Table 9.1 continued

No	Chemical	Source[a]	Erythema	Chemosis	Opacity	Iritis	Classification	
							EU	OECD
48	Methyl amyl ketone (heptanone)	ECETOC	1.2	1.2	0.3	0.1	NI	NI
49	Methyl ethyl ketone (butanone)	ECETOC	2.1	2.2	1.3	0.6	R36	B
50	Acetone	ECETOC	2.8	2.5	1.9	1.2	R36	B
51	2-Bromobutane	ECETOC	0.0	0.0	0.0	0.0	NI	NI
52	1-Bromo-4-chlorobutane	ECETOC	0.0	0.0	0.0	0.0	NI	NI
53	Bromooctane	ECETOC	0.0	0.0	0.0	0.0	NI	NI
54	Bromohexane	ECETOC	0.0	0.0	0.0	0.0	NI	NI
55	Bromopentane	ECETOC	0.6	0.0	0.0	0.0	NI	NI
56	1,6-Dibromohexane	ECETOC	0.2	0.1	0.0	0.0	NI	NI
57	Isopropyl bromide	ECETOC	0.3	0.2	0.0	0.0	NI	NI
58	1,4-Dibromobutane	ECETOC	0.3	0.3	0.0	0.0	NI	NI
59	1,5-Dibromopentane	ECETOC	0.6	0.3	0.0	0.0	NI	NI
60	1,3-Dibromopropane	ECETOC	0.3	0.3	0.0	0.0	NI	NI
61	γ-Butyrolactone	ECETOC	2.2	2.1	1.9	0.8	R36	B
62	2,2-Dimethylbutanoic acid	ECETOC	2.0	0.7	1.5	0.5	NI	B
63	2-Ethoxyethyl methacrylate	ECETOC	0.0	0.0	0.0	0.0	NI	NI
64	Ethyl thioethyl methacrylate	ECETOC	0.0	0.0	0.0	0.0	NI	NI
65	Nonyl acrylate	ECETOC	0.0	0.0	0.0	0.0	NI	NI
66	Nonyl methacrylate	ECETOC	0.0	0.0	0.0	0.0	NI	NI
67	Iso-octyl acrylate	ECETOC	0.1	0.0	0.0	0.0	NI	NI
68	Heptyl methacrylate	ECETOC	0.2	0.0	0.0	0.0	NI	NI
69	Trifluoroethyl methacrylate	ECETOC	0.2	0.0	0.0	0.0	NI	NI
70	Allyl methacrylate	ECETOC	0.8	0.3	0.0	0.0	NI	NI
71	Glycidyl methacrylate	ECETOC	1.5	1.2	0.8	0.4	NI	NI
72	2-Methoxyethyl acrylate	ECETOC	2.3	2.7	1.9	0.6	R36	B
73	3-Chloropropionitrile	ECETOC	0.6	0.3	0.7	0.1	NI	NI
74	Diethylaminopropionitrile	ECETOC	3.0	2.6	3.5	1.0	R41	A
75	2-Ethylhexylthioglycolate	ECETOC	0.0	0.0	0.0	0.0	NI	NI
76	4-Methylthiobenzaldehyde	ECETOC	0.1	0.1	0.0	0.0	NI	NI
77	Iso-octylthioglycolate	ECETOC	0.1	0.0	0.0	0.0	NI	NI
78	Dipropyl disulphide	ECETOC	0.4	0.0	0.0	0.0	NI	NI
79	Thiodiglycol	ECETOC	0.5	0.3	0.0	0.0	NI	NI
80	1,2,3-Trimercaptopropane	ECETOC	1.3	1.5	0.0	0.0	NI	NI
81	Ethylthioglycolate	ECETOC	1.2	0.8	1.0	0.5	NI	B
82	Methylthioglycolate	ECETOC	2.4	1.8	2.2	0.9	R36	B
83	2,2,3-Trimethyl-3-cyclo-pentene-1-acetaldehyde	Spielmann	1.7	0.8	0.9	0.1	NI	NI
84	Silan 108 (trimethoxyoctyl silane)	Spielmann	0.7	0.0	0.0	0.0	NI	NI
85	Silan 103 (trimethoxypropyl silane)	Spielmann	1.6	0.3	0.3	0.1	NI	NI
86	Methyltriglycol	Spielmann	0.5	0.0	0.0	0.0	NI	NI
87	Ethylacrolein	Spielmann	1.7	3.3	3.0	nd	R41	A
88	Polyethylene glycol butyl ether	Spielmann	2.5	2.0	1.6	0.8	R36	B
89	1,2-Epoxycyclooctane	Spielmann	1.4	1.0	0.7	0.3	NI	NI
90	Methylacetate	Spielmann	2.7	1.8	1.3	1.0	R36	B
91	Polysolvan (butyl glycolate)	Spielmann	2.6	1.5	1.8	0.9	R36	B
92	α-Lactid	Spielmann	0.9	1.2	0.0	0.0	NI	NI
93	Isopropyl acetoacetate	Spielmann	1.5	0.7	0.2	0.2	NI	NI

Table 9.1 continued

No	Chemical	Source[a]	Erythema	Chemosis	Opacity	Iritis	Classification	
							EU	OECD
94	Butanal	Spielmann	2.3	1.3	0.9	0.4	NI	B
95	3-Methoxybenzaldehyde	Spielmann	0.1	0.0	0.0	0.0	NI	NI
96	Butanol	Spielmann	2.9	3.0	2.1	1.0	R36	B
97	Anisole	Spielmann	3.0	1.7	0.7	0.7	R36	B
98	4-Anisidine	Spielmann	2.6	2.1	1.8	1.0	R36	B
99	Ethylbutanal	Spielmann	2.0	1.0	0.0	0.3	NI	B
100	Isobutanal	Spielmann	2.0	1.3	1.1	0.4	NI	B
101	1,2-Epoxydodecane	Spielmann	0.8	0.0	0.3	0.0	NI	NI
102	Isononylaldehyde	Spielmann	0.6	0.3	0.0	0.0	NI	NI
103	2-Methyl-1-propanol	Spielmann	2.5	1.5	1.0	0.3	R36	B
104	2-Hydroxyisobutyric acid ethylester	Spielmann	2.3	3.3	3.0	1.0	R41	A
105	Paraformaldehyde (methanal)	Spielmann	3.0	3.6	1.7	1.0	NI	B
106	3-Cyclohexene-1-methanol	Spielmann	2.8	2.0	1.9	0.6	NI	B
107	4,4-Dimethyl-3-oxopentanenitrile	Spielmann	0.1	0.1	0.0	0.0	NI	NI
108	2-Hydroxyisobutyric acid	Spielmann	2.3	4.0	4.0	nd	R41	A
109	Hexahydrofarnesyl acetone	Spielmann	0.7	0.0	0.0	0.0	NI	NI
110	Trioxane	Spielmann	1.7	0.3	0.0	0.1	NI	NI
111	Butyl carbamate	Spielmann	2.9	2.0	1.3	1.0	R36	B
112	Isobornyl acetate	Spielmann	0.2	0.2	0.0	0.0	NI	NI
113	Acetoacetic acid glycolester	Spielmann	2.0	1.3	0.9	0.6	NI	NI
114	Camphen (2,2-dimethyl-3-methylene bicyclo 2.2.1 heptane)	Spielmann	2.5	1.5	0.9	0.6	R36	B
115	Methylpentynol	Spielmann	2.0	2.0	1.0	0.3	R36	B
116	2,5-Dimethylhexanediol	Spielmann	2.3	1.8	1.0	0.3	NI	B
117	3,6-Dimethyloctanol	Spielmann	1.6	0.5	1.0	0.0	NI	B
118	1-Chloroctan-8-ol	Spielmann	2.0	0.6	1.0	0.3	NI	B
119	Isotridecanal	Spielmann	0.2	0.0	0.0	0.0	NI	NI

Footnote to Table 9.1

[a]ECETOC (1998), Spielmann *et al.* (1996).
A = severe irritant (OECD system); B = mild irritant (OECD system); nd= no data; NI = non-irritant; R36 = mild irritant (EU system); R41 = severe irritant (EU system).

Differences in the EU and OECD classification systems are indicated by shading.

Table 9.2 Physicochemical properties for the data set of 119 organic liquids

No	Chemical	I/NI	MW	LogP	dV1	pKa1	pKa2	pKa3
1	3,3-Dimethylpentane	NI	100.2	3.67	-0.2929			
2	3-Methylhexane	NI	100.2	3.71	-0.1062			
3	2-Methylpentane	NI	86.2	3.21	-0.1442			
4	1,9-Decadiene	NI	138.3	4.98	-0.7812			
5	Dodecane	NI	170.3	6.23	0.0000			
6	1,5-Dimethylcyclooctadiene	NI	136.2	4.82	-0.8133			
7	*cis*-Cyclooctene	NI	110.2	3.94	-0.2644			
8	Methylcyclopentane	NI	84.2	3.1	-0.0204			
9	1,5-Hexadiene	NI	82.1	3.02	-0.7812			
10	4-Bromophenetole	NI	201.1	3.46	-1.6650			
11	2-Xylene	NI	106.2	3.09	-1.0869			
12	1-Methylpropylbenzene	NI	134.2	3.94	-1.0221			
13	3-Ethyltoluene	NI	120.2	3.58	-1.0322			
14	2,4-Difluoronitrobenzene	NI	159.1	2.21	-3.5352			
15	Styrene	NI	104.2	2.89	-1.3066			
16	Toluene	NI	92.1	2.54	-1.0035			
17	4-Fluoroaniline	I	111.1	1.28	-2.0044	4.7		
18	Pyridine	I	79.1	0.8	-1.0749	5.3		
19	3-Methoxy-1,2-propanediol	NI	106.1	-1.15	-1.5057	13.6	15.3	
20	2,4-Pentanediol	NI	104.1	0.13	-1.1686	14.7	15.6	
21	Propylene glycol	NI	76.1	-0.78	-0.7974	14.4	15.8	
22	Glycerol	NI	92.1	-1.65	-1.1962	13.5	14.9	16.1
23	Isomyristyl alcohol	NI	214.4	5.68	-0.9868	15.2		
24	2,2-Dimethyl-3-pentanol	NI	116.2	2.13	-1.0046	15.3		
25	2-Methyl-1-pentanol	NI	102.2	1.75	-0.6945	15.1		
26	Cyclopentanol	NI	86.1	1.15	-0.4673	15.3		
27	Ethanol	NI	46.1	-0.14	-0.1303	15.2		
28	Isopropanol (propan-2-ol)	NI	60.1	0.28	-0.4153	15.3		
29	Octanol	I	130.2	2.81	-0.6881	15.3		
30	Furfuryl hexanol	I	98.1	0.45	-1.5912	14.0		
31	2-Ethyl-1-hexanol	NI	130.2	2.73	-0.7563	15.1		
32	Isobutanol	I	74.1	0.77	-0.5003	15.3		
33	Butanol	I	74.1	0.84	-0.4278	15.2		
34	Hexanol	I	102.2	1.82	-0.5883	15.4		
35	Butyl cellosolve (2-butoxyethanol)	I	118.2	0.57	-1.1650	14.4		
36	Cyclohexanol	I	100.2	1.23	-0.5370	15.3		
37	Methyl trimethyl acetate	NI	116.2	1.74	-1.6992			
38	Ethyl trimethyl acetate	NI	130.2	2.23	-1.7031			
39	Butyl acetate	NI	116.2	1.85	-1.3617			
40	Cellosolve (2-ethoxyethyl) acetate	NI	132.2	0.31	-1.9131			
41	Ethyl acetate	NI	88.1	0.86	-1.1187			
42	Ethyl-2-methyl acetoacetate	NI	144.2	0.21	-2.2473	12.0		
43	Methyl cyanoacetate	I	99.1	-0.47	-1.7946	2.8		
44	Methyl acetate	I	74.1	0.37	-1.0409			
45	Ethyleneglycol diethylether	NI	118.2	0.77	-1.1968			
46	Di-isobutyl ketone (2,6-dimethyl-4-heptanone)	NI	142.2	2.56	-1.2112			
47	Methyl isobutyl ketone (4-methylpentan-2-one)	NI	100.2	1.16	-0.9911			

Table 9.2 continued

No	Chemical	I/NI	MW	LogP	dV1	pKa1	pKa2	pKa3
48	Methyl amyl ketone (heptanone)	NI	114.2	1.73	-0.9020			
49	Methyl ethyl ketone (butanone)	I	72.1	0.26	-0.6864			
50	Acetone	I	58.1	-0.24	-0.6240			
51	2-Bromobutane	NI	137.0	2.58	0.0388			
52	1-Bromo-4-chlorobutane	NI	171.5	2.9	0.3917			
53	Bromooctane	NI	193.1	4.61	0.3484			
54	Bromohexane	NI	165.1	3.63	0.3352			
55	Bromopentane	NI	151.0	3.14	0.3244			
56	1,6-Dibromohexane	NI	244.0	3.97	0.7219			
57	Isopropyl bromide	NI	123.0	2.08	-0.0265			
58	1,4-Dibromobutane	NI	215.9	2.99	0.6889			
59	1,5-Dibromopentane	NI	229.9	3.48	0.7089			
60	1,3-Dibromopropane	NI	201.9	2.5	0.6556			
61	γ-Butyrolactone	I	86.1	-0.31	-0.9722			
62	2,2-Dimethylbutanoic acid	NI	116.2	1.94	-1.5273	5.0		
63	2-Ethoxyethyl methacrylate	NI	158.2	1.49	-2.5281			
64	Ethyl thioethyl methacrylate	NI	174.3	2.34	-1.4107			
65	Nonyl acrylate	NI	198.3	4.66	-2.0517			
66	Nonyl methacrylate	NI	212.3	5.2	-2.1845			
67	Iso-octyl acrylate	NI	184.3	4.09	-2.1516			
68	Heptyl methacrylate	NI	184.3	4.22	-2.1008			
69	Trifluoroethyl methacrylate	NI	168.1	2.18	-3.6654			
70	Allyl methacrylate	NI	126.2	2.12	-2.2401			
71	Glycidyl methacrylate	NI	142.2	0.81	-2.3688			
72	2-Methoxyethyl acrylate	I	128.2	1.05	-2.1169			
73	3-Chloropropionitrile	NI	89.5	0.6	-0.5506			
74	Diethylaminopropionitrile	I	126.2	0.53	-1.1082	9.2		
75	2-Ethylhexylthioglycolate	NI	204.3	3.68	-1.6108	8.2		
76	4-Methylthiobenzaldehyde	NI	152.2	2.31	-1.1904			
77	Iso-octylthioglycolate	NI	204.3	3.68	-1.6868	8.2		
78	Dipropyl disulphide	NI	150.3	3.84	1.5529			
79	Thiodiglycol	NI	122.2	-0.62	-0.4646	14.1	14.8	
80	1,2,3-Trimercaptopropane	NI	140.3	1.57	0.2838	8.7	9.8	10.7
81	Ethylthioglycolate	NI	120.2	0.81	-1.1229	8.4		
82	Methylthioglycolate	I	106.1	0.32	-1.0830	8.2		
83	2,2,3-Trimethyl-3-cyclo-pentene-1-acetaldehyde	NI	152.2	3.31	-1.5834			
84	Silan 108 (trimethoxyoctylsilane)	NI	234.4	2.77	-1.7133			
85	Silan 103 (trimethoxypropylsilane)	NI	164.3	0.31	-1.3920			
86	Methyltriglycol	NI	164.2	-1.46	-2.4692	14.4		
87	Ethylacrolein	I	84.1	1.23	-1.1034			
88	Polyethylene glycol butyl ether	I	162.2	0.29	-1.9350	14.4		
89	1,2-Epoxycyclooctane	NI	126.2	2.64	-0.5902			
90	Methylacetate	I	74.1	0.37	-1.0409			
91	Polysolvan (butyl glycolate)	I	132.2	0.38	-1.8180	13.0		
92	α-Lactid	NI	144.1	1.65	-2.4773			
93	Isopropyl acetoacetate	NI	144.2	0.21	-2.3002	11.7		

Table 9.2 continued

No	Chemical	I/NI	MW	LogP	dV1	pKa1	pKa2	pKa3
94	Butanal	NI	72.1	0.82	-0.6001			
95	3-Methoxybenzaldehyde	NI	136.2	1.79	-2.4763			
96	Butanol	I	74.1	0.84	-0.4278	15.2		
97	Anisole	I	108.1	2.07	-1.6437			
98	4-Anisidine	I	123.2	1.16	-1.9917	5.2		
99	Ethylbutanal	NI	100.2	1.73	-0.8119			
100	Isobutanal	NI	72.1	0.74	-0.7274			
101	1,2-Epoxydodecane	NI	184.3	4.79	-0.7807			
102	Isononylaldehyde	NI	142.2	3.2	-1.0415			
103	2-Methyl-1-propanol	I	74.1	0.77	-0.5720	15.1		
104	2-Hydroxyisobutyric acid ethylester	I	132.2	0.27	-2.0142	13.3		
105	Paraformaldehyde (methanal)	NI	30.0	0.35	-0.1080			
106	3-Cyclohexene-1-methanol	NI	112.2	1.91	-0.9760	15.2		
107	4,4-Dimethyl-3-oxopentanenitrile	NI	125.2	0.28	-1.8293	10.3		
108	2-Hydroxyisobutyric acid	I	104.1	-0.2	-1.6922	4.0	16.6	
109	Hexahydrofarnesyl acetone	NI	268.5	6.91	-1.5106			
110	Trioxane	NI	90.1	-0.56	-1.1713			
111	Butyl carbamate	I	117.1	0.96	-1.4850	-1.4	13.5	
112	Isobornyl acetate	NI	196.3	3.86	-2.1031			
113	Acetoacetic acid glycolester	NI	230.2	-1.73	-4.5517	9.9	10.6	
114	Camphen (2,2-dimethyl-3-methylene bicyclo 2.2.1 heptane)	I	136.2	4.35	-0.6003			
115	Methylpentynol	I	98.1	0.94	-1.2887	13.3		
116	2,5-Dimethylhexanediol	NI	146.2	1.6	-1.5143	14.7	15.3	
117	3,6-Dimethyloctanol	NI	158.3	3.64	-0.9684	15.1		
118	1-Chloroctan-8-ol	NI	164.7	3.06	-0.6731	15.2		
119	Isotridecanal	NI	198.3	5.16	-1.1416			

Footnote to Table 9.2

dV1 = the first-order difference valence connectivity index; I = irritant (EU criteria); logP = logarithm of the octanol-water partition coefficient; MW = molecular weight (g/mol); NI = non-irritant (EU criteria); pKa(n) = the nth value of the negative logarithm of the acid dissociation constant.

Table 9.3 Training set of pH data for 165 chemicals[a]

No	Chemical	pH[b]	EU class
1	Acetic acid	2.50	R41
2	Acrylic anhydride	3.10	R41
3	Amino-11-undecenoic acid	7.10	NI
4	Benzoic acid	2.60	R41
5	Cekanoic C7 acid	4.00	R41
6	Cekanoic C8 acid	4.20	R41
7	Cekanoic C9 acid	4.30	R41
8	Cekanoic C10 acid	4.20	R41
9	Chlorhydric 2,5 N acid	2.00	R41
10	Detartrant BS12	0.90	R41
11	Dithiodipropionic acid	3.30	NI
12	Methacrylic anhydride	3.20	R41
13	N-Heptylamino-11-undecenoic acid	6.20	NI
14	Potassium tetrafluoroborate	3.60	NI
15	Sodium undecylenate	8.80	R36
16	Sulphamic acid	0.90	R41
17	Adamquat MC80	4.50	R36
18	Dicyclopentenyloxyethyl acrylate	4.10	R36
19	Hexyl acrylate	7.30	NI
20	Iso-octyl acrylate	7.20	NI
21	Methoxyethyl acrylate	5.40	R36
22	Nonyl acrylate	6.90	R41
23	Butanol	6.80	R36
24	Glycerol	7.90	NI
25	Isomyristic alcohol	6.30	NI
26	Methanol	5.40	NI
27	2-Methoxyethanol	5.90	NI
28	Octanol	6.40	R41
29	Propanol	5.80	NI
30	Isopropanol	5.50	NI
31	TLH (mixed heavy alcohols)	6.60	R41
32	Acetic hydrazide	7.00	NI
33	Ethyl carbazate	6.90	R36
34	Methanesulphonamide	5.30	NI
35	Methyl carbazate	7.00	NI
36	Pivalic hydrazide	5.10	R36
37	Propionic hydrazide	8.00	R36
38	3(2-Methoxyethoxy)propylamine	11.50	R41
39	Diethylaminopropylamine	11.90	R41
40	Di-isopropylamine	11.70	R41
41	Dimethylaminopropylamine	12.10	R41
42	Dimethyl-N-butylamine	11.50	R41
43	Dimethyldiethylpropylenediamine	11.70	R41
44	Dimethyldipropylenetriamine	12.00	R41

Table 9.3 continued

No	Chemical	pH	EU class
45	Dimethylethylamine	11.10	R41
46	Dimethylisopropylamine	12.20	R41
47	Heptylamine	11.70	R41
48	3-Methoxypropylamine	10.10	R41
49	Tetramethylpropylenediamine	11.20	R41
50	Triethanolamine	10.70	NI
51	Benzylcumene	6.50	NI
52	Benzylpentamethylbenzene	6.40	NI
53	Benzyltoluene	5.60	NI
54	Benzyltoluene/dibenzyltoluene	2.80	NI
55	Bis(p-xylyl)phenylmethane	6.20	NI
56	Dibenzylmesitylene	6.40	NI
57	Dibenzyltoluene	6.60	NI
58	Orthoxylyl-orthoxylene	6.30	NI
59	Toluene	5.90	NI
60	Esterol T	7.80	R41
61	Metiloil A	3.80	NI
62	Metiloil B	3.80	NI
63	Metiloil C	3.70	NI
64	1-Bromo-4-chlorobutane	4.10	NI
65	Chloroform	6.00	NI
66	1,4-Dibromobutane	3.90	NI
67	Dibromodichlorodifluoroethane	3.10	R41
68	1,6-Dibromohexane	6.90	NI
69	1,5-Dibromopentane	4.30	NI
70	1,3-Dibromopropane	3.80	NI
71	Dibromotrichloroethane	2.70	R41
72	Dichlorotoluene	4.70	NI
73	Ethyl bromodifluoroacetate	1.00	R36
74	Phenethyl bromide	4.40	NI
75	Tetrachlorobenzyltoluene	5.90	NI
76	Calcium sulphydrate	10.50	R41
77	Dimethylsulphoxide	5.20	NI
78	Dipropyl disulphide	5.80	NI
79	Mercaptopropanol	6.40	R41
80	4-Methylthiobenzaldehyde	6.20	NI
81	Methylthiodichlorophosphine	0.00	R41
82	Sodium methylmercaptide	12.00	R41
83	Sodium methylmercaptide	12.00	R41
84	TPS 20	5.90	NI
85	TPS 27	4.70	NI
86	TPS 32	5.90	NI
87	TPS 37	4.90	NI
88	Trimercaptopropane	4.60	R36

Table 9.3 continued

No	Chemical	pH	EU class
89	Allyl methacrylate	6.50	NI
90	Dimethylaminoethyl methacrylate	8.80	R41
91	Ethyltriglycol methacrylate	4.50	NI
92	Glycidyl methacrylate	4.00	R36
93	Heptyl methacrylate	7.00	NI
94	Madquat MC	3.90	R36
95	Methoxyethyl methacrylate	6.00	NI
96	Nonyl methacrylate	6.70	NI
97	Trifluoroethanol methacrylate	4.40	NI
98	Chloro-3-propionitrile	4.30	NI
99	Dichloroproprionitrile	2.50	R41
100	Dimethylaminopropionitrile	10.10	R41
101	Forafac 1028	3.50	R41
102	Forafac 1033	1.80	R41
103	Forafac 1057	4.30	NI
104	Forafac 1110 D	4.20	NI
105	Forafac 1159	5.30	R36
106	Forafac 1183 N	6.20	NI
107	Forafac 1185 D	3.60	NI
108	Forafac 1187	5.60	NI
109	Foralkyl AC8N	4.60	R36
110	Foralkyl FH6-11A	4.60	NI
111	Foralkyl MAC-8N	5.40	NI
112	Foraperle 200	4.80	R36
113	Foraperle 215	4.20	NI
114	Foraperle 222	4.90	NI
115	Foraperle 259	6.30	NI
116	Foraperle 263	5.00	NI
117	Foraperle 305	5.90	NI
118	Foraperle 320	4.10	NI
119	Foraperle 390A	5.30	NI
120	Foraperle 390B	6.80	NI
121	Foraperle 440	3.50	R36
122	Foraperle B 208	5.90	NI
123	Foraperle B 244	4.60	R36
124	Perfluoroalkyl iodide	3.40	NI
125	Perfluoroalkyl ethyl iodide	5.00	NI
126	Perfluorohexyl iodide	3.50	NI
127	Perfluorohexyl ethyl iodide	4.40	NI
128	Perfluorooctyl bromide	9.20	NI
129	Perfluorosulphochloride	3.20	NI
130	Perfluoroethanesulphochloride	3.30	NI
131	Orgasol 1002	6.00	NI
132	Orgasol 2002	5.60	NI

Table 9.3 continued

No	Chemical	pH	EU class
133	Resin E561	3.90	NI
134	Resin SR9	7.40	NI
135	Additive A71	7.50	R36
136	Inhibitor RHB	7.50	R36
137	Sodium dodecyl sulphate	6.90	R41
138	Dinoram C	10.70	R41
139	Dinoram S	10.20	R41
140	Dinoram SL	10.50	R41
141	EXP 3830 D	4.20	R41
142	Noram C	10.20	R41
143	Noram S96	10.00	R41
144	Noramac C26	6.40	R41
145	Noramac S	5.80	R41
146	Noramium M2SH 3	4.50	R41
147	Polyram S	10.20	R41
148	Remcopal 011	6.00	R36
149	AC 302 B	8.16	NI
150	AC 551 C	8.44	NI
151	Adine BBH44	3.90	NI
152	B92	7.40	R36
153	Creosote 14130	7.00	NI
154	DA 89	5.40	NI
155	DE 917D	6.20	NI
156	IDPA	7.00	NI
157	Diurethane XPU 6734	4.00	NI
158	Hexane	6.10	NI
159	Resin C11	3.30	NI
160	Sodium hydroxide	12.00	R41
161	Triacetin	5.20	NI
162	Tributyltin chloride	2.50	R41
163	WAC	4.20	R36
164	XT 7660	9.40	R41
165	XT 7661	9.30	R41

Footnote to Table 9.3

[a]The pH data and EU classifications were taken from Régnier & Imbert (1992).
[b]Measurements were made on a 10% solution.

Table 9.4 Test set of pH data for 49 chemicals[a]

No	Chemical	EU class	pH[b]
1	Sodium hydroxide	I	12.66
2	Benzalkonium chloride	I	3.10
3	Cetylpyridinium bromide	I	4.81
4	Captan 90 concentrate	I	7.95
5	Chlorhexidine	I	10.12
6	Cyclohexanol	I	4.54
7	Quinacrine	I	3.77
8	Promethazine HCl	I	4.50
9	4-Fluoroaniline	I	9.02
10	Triton X-100	I	7.18
11	Acetone	I	5.32
12	Hexanol	I	5.48
13	1-Naphthalene acetic acid, Na salt	I	3.34
14	Sodium oxalate	I	9.40
15	Isobutanol	I	5.72
16	Imidazole	I	10.32
17	2-Ethyl hexanol	NI	4.78
18	4-Carboxybenzaldehyde	I	3.12
19	Methyl ethyl ketone	I	5.51
20	Pyridine	I	9.85
21	1-Naphthalene acetic acid	NI	7.98
22	2,2-Dimethylbutanoic acid	NI	3.32
23	γ-Butyrolactone	I	4.51
24	Octanol	I	6.11
25	Methyl acetate	I	4.84
26	L-Aspartic acid	I	2.96
27	Benzoyl-L-tartaric acid	I	2.39
28	Potassium cyanate	I	10.08
29	Isopropanol	I	5.86
30	Sodium perborate	I	9.98
31	Dibenzyl phosphate	I	2.39
32	2,5-Dimethylhexanediol	NI	5.72
33	Methyl cyanoacetate	I	5.76
34	Ethanol	NI	6.24
35	2,6-Dichlorobenzoyl chloride	I	2.50
36	Ammonium nitrate	I	4.78
37	Ethyl-2-methylacetoacetate	NI	7.47
38	Ethyl acetate	NI	4.81
39	Maneb	I	8.39
40	Fomesafen	NI	3.99
41	Tetraaminopyrimidine sulphate	NI	2.49
42	Toluene	NI	5.67
43	Butyl acetate	NI	5.00
44	Methyl isobutyl ketone	NI	4.59
45	Tween 20	NI	3.84
46	Ethyl trimethyl acetate	NI	4.75
47	Methylcyclopentane	NI	6.52
48	Glycerol	NI	5.46
49	Polyethylene glycol 400	NI	3.65

Footnote to Table 9.4

[a]The 49 chemicals are a subset of the 60 chemicals tested in the EC/HO validation study on alternatives to the Draize eye test (Balls *et al*., 1995); the pH data were provided by BIBRA International (Surrey, UK).

[b]Measurements were made on a 10% solution.

Table 9.5 ***In vitro*** **data set of 59 chemicals**[a]

No	Chemical	EU class	SM	RBC Dlow	Dmax	H50	IRE O1	O4	S1	S4	ICE R	O	S	FL	EDE	BCOP O	P
1	Sodium hydroxide (10%)	I	3.3	10000	57.7	3893	2.9	4.0	101.6	138.3	3.0	3.6	111.6	8.8	51.0	216.6	3.7
2	Benzalkonium chloride (10%)	I	2.6	5253	62.5	156	1.7	2.5	36.4	73.1	3.0	2.4	53.6	267.3	50.8	70.8	4.4
3	Trichloroacetic acid (30%)	I	3.2	10000	80.2	19294	3.4	3.7	24.0	77.4	3.0	4.0	92.5	348.8	51.0	212.4	3.4
4	Cetylpyridinium bromide (10%)	I	3.1	10000	57.1	64	0.8	1.9	17.9	43.5	2.3	1.9	28.0	259.3	26.0	23.0	3.4
5	Cetylpyridinium bromide (6%)	I	2.9	10000	62.3	109	0.6	1.8	21.4	32.0	2.4	1.3	28.6	257.0	26.3	29.5	2.8
6	Benzalkonium chloride (5%)	I	3.0	32500	63.5	360	1.3	3.0	32.3	99.2	2.5	2.4	45.2	267.0	48.7	64.9	4.2
7	Captan 90 concentrate	I	nd	370	3.3	370	0.8	1.0	6.5	18.8	0.3	0.9	17.0	175.0	8.3	43.5	0.0
8	Chlorhexidine	I	nd	700	23.7	440	1.3	2.7	26.8	69.2	3.0	3.8	78.4	175.0	51.0	111.5	0.2
9	Cyclohexanol	I	3.5	14000	2.5	9925	1.1	2.5	24.3	82.1	2.8	2.3	52.2	788.0	38.5	11.3	3.3
10	Quinacrine	I	3.0	10000	4.4	13750	0.0	0.2	7.1	8.0	1.1	0.8	8.6	375.0	21.0	1.5	0.0
11	Promethazine HCl	I	3.1	30250	32.9	419	1.5	2.3	44.1	89.8	2.7	2.4	56.9	38.8	51.0	121.7	0.0
12	4-Fluoraniline	I	3.5	7750	23.3	7730	1.2	2.3	27.8	64.4	3.0	2.1	46.6	773.7	50.7	15.7	1.0
13	Triton X-100 (10%)	I	3.3	100000	2.5	967	0.7	2.3	27.1	56.8	1.7	0.8	17.6	298.4	20.5	5.9	4.3
14	Acetone	I	5.2	100000	2.5	100000	0.4	1.1	15.3	31.9	1.8	1.1	20.0	687.3	36.4	78.7	3.0
15	Hexanol	I	nd	2125	2.5	1125	0.7	2.7	18.6	48.3	2.3	2.3	46.8	508.2	39.4	13.3	3.2
16	1-Naphthalene acetic acid, Na salt	I	4.1	10150	2.5	10150	1.0	2.7	57.0	107.6	3.0	2.8	69.2	291.5	37.5	86.4	4.2
17	Sodium oxalate	I	nd	6500	6.8	7000	0.0	0.0	7.3	9.7	0.6	0.3	8.8	337.5	51.0	11.6	0.2
18	Isobutanol	I	4.4	56250	2.5	33800	1.3	2.5	25.1	75.5	3.0	2.4	61.4	597.3	40.2	20.7	2.4
19	Imidazole	I	3.9	100000	54.6	21348	2.5	2.8	44.8	74.8	3.0	3.1	99.8	147.8	51.0	75.0	2.5
20	Sodium lauryl sulphate (15%)	I	2.8	7750	32.5	256	0.1	1.3	16.3	23.4	1.1	0.7	15.4	28.7	39.5	4.5	3.9
21	2-Ethyl-1-hexanol	NI	nd	833	4.3	765	0.3	1.4	10.7	20.3	1.8	2.0	42.7	509.6	24.8	7.1	2.2
22	4-Carboxybenzaldehyde	I	nd	1000	2.5	1000	0.3	0.4	6.2	13.0	1.2	1.3	26.4	175.0	46.7	77.4	0.1
23	Methyl ethyl ketone	I	4.8	100000	2.5	74527	0.9	2.4	21.2	61.3	2.7	2.4	42.4	441.1	38.1	52.6	1.2

Table 9.5 continued

No	Chemical	EU class	SM	RBC Dlow	Dmax	H50	IRE O1	IRE O4	IRE S1	IRE S4	ICER	ICEO	ICES	FL	EDE	BCOP O	BCOP P
24	Pyridine	I	4.1	100000	58.7	48931	1.8	2.8	25.9	54.9	3.0	2.6	60.9	463.0	51.0	76.1	4.8
25	1-Naphthalene acetic acid	NI	nd	3063	2.5	3063	0.3	0.9	12.0	13.7	1.3	1.0	21.3	212.5	49.8	76.6	0.1
26	Benzalkonium chloride (1%)	I	3.7	100000	49.2	1401	0.9	2.4	23.9	52.8	1.9	1.9	21.6	350.3	43.0	38.1	3.4
27	2,2-Dimethylbutanoic acid	NI	nd	1145	3.2	1038	2.8	2.7	33.6	68.1	3.0	2.7	54.1	333.0	47.6	61.4	3.5
28	γ-Butyrolactone	I	4.5	100000	2.5	100000	0.3	1.7	21.4	38.3	1.6	1.7	22.2	231.9	23.9	32.5	1.9
29	Octanol	I	nd	700	2.5	483	0.0	1.5	11.8	21.7	1.6	1.7	45.1	512.7	31.2	9.2	2.1
30	Methyl acetate	I	5.0	100000	2.5	84475	0.5	1.6	15.1	30.6	1.9	2.5	38.3	687.5	31.6	44.1	0.7
31	L-Aspartic acid	I	3.1	3000	39.3	1200	0.3	0.3	5.1	6.1	1.7	1.3	21.0	154.0	50.2	1.3	0.0
32	Benzoyl-L-tartaric acid	I	2.9	2025	45.5	2400	1.0	1.9	18.2	24.5	1.7	2.3	25.2	106.1	51.0	166.9	0.2
33	Triton X-100 (5%)	I	3.5	100000	2.5	2175	0.6	2.0	19.7	33.0	1.3	0.7	22.8	417.7	20.5	6.0	4.8
34	Potassium cyanate	I	4.4	100000	14.0	100000	0.0	0.0	5.2	5.3	1.0	0.9	17.2	625.0	48.0	7.6	0.5
35	Isopropanol	I	5.0	100000	2.5	98250	1.3	1.9	16.0	35.8	2.0	1.8	35.5	718.5	40.0	20.0	2.5
36	Sodium perborate	I	2.8	2800	13.1	3800	0.0	0.0	3.2	5.5	0.8	0.7	12.1	200.1	28.2	10.9	5.8
37	Dibenzyl phosphate	I	2.9	3000	38.2	1563	0.5	1.1	9.5	16.5	1.9	1.4	23.0	136.5	49.7	376.5	0.1
38	2,5-Dimethylhexanediol	NI	4.8	100000	2.5	100000	0.3	0.4	15.3	16.5	2.1	1.7	23.5	625.0	27.3	11.1	0.6
39	Methyl cyanoacetate	I	3.4	30250	19.7	33844	0.1	0.7	5.0	6.9	0.5	0.6	16.0	766.3	48.3	12.3	0.0
40	Sodium hydroxide (1%)	I	4.4	100000	38.0	42001	1.0	2.8	50.2	93.5	1.5	1.9	33.2	129.3	50.1	94.9	3.7
41	Ethanol	NI	5.0	100000	2.5	100000	1.7	2.6	26.8	52.6	2.3	2.6	43.8	701.2	35.6	26.9	2.9
42	2,6-Dichlorobenzoyl chloride	I	nd	700	23.0	650	0.8	1.9	8.5	21.1	2.0	1.1	18.2	1149.5	36.4	9.7	0.0
43	Ammonium nitrate	I	4.7	100000	3.0	100000	0.0	0.0	7.3	10.2	1.6	1.1	16.8	700.3	16.7	6.9	0.2
44	Ethyl-2-methylacetoacetate	NI	2.6	17500	8.0	17500	0.4	1.7	16.3	21.2	0.8	0.5	5.1	786.3	19.0	14.1	0.0
45	Sodium lauryl sulphate (3%)	NI	3.5	7503	29.1	1215	0.0	0.5	9.8	15.4	0.8	0.3	15.4	107.4	36.5	4.4	1.4
46	Ethyl acetate	NI	4.7	21375	2.5	21375	0.0	1.4	14.6	30.6	2.2	2.1	36.5	716.0	28.8	9.5	1.5

Table 9.5 continued

No	Chemical	EU class	SM	RBC			IRE				ICE			FL	EDE	BCOP	
				Dlow	Dmax	H50	O1	O4	S1	S4	R	O	S			O	P
47	Maneb	I	nd	2578	25.0	3892	1.0	1.0	24.0	26.6	0.5	1.0	12.6	337.5	40.4	39.0	0.1
48	Fomesafen	NI	nd	26250	19.0	1915	0.8	1.2	9.2	16.3	0.7	0.7	5.9	337.5	44.9	6.4	3.6
49	Tetraaminopyrimidine sulphate	NI	3.0	1000	21.0	1333	0.8	0.8	4.3	10.3	1.2	1.4	13.7	175.0	48.5	14.8	0.0
50	Toluene	NI	nd	650	2.5	650	0.4	0.5	14.4	22.8	1.4	1.6	26.6	872.5	18.1	5.3	2.0
51	Butyl acetate	NI	nd	1625	2.5	1625	0.0	0.3	6.6	14.7	1.1	2.1	25.7	693.5	24.1	5.9	1.9
52	Trichloroacetic acid (3%)	NI	4.2	100000	77.2	73188	0.7	0.8	8.1	18.5	2.0	1.9	26.4	764.8	35.0	74.5	0.1
53	Methyl isobutyl ketone	NI	2.9	10000	2.5	10000	0.3	1.6	18.2	34.2	2.4	2.3	31.1	840.5	21.6	6.6	0.4
54	Tween 20	NI	3.4	100000	5.1	78400	0.0	0.3	13.5	15.9	1.2	0.6	11.7	876.0	nd	0.0	0.0
55	Ethyl trimethyl acetate	NI	nd	1000	2.5	1000	0.0	0.8	6.6	12.0	1.1	0.7	11.6	873.5	18.5	5.1	0.8
56	Methylcyclopentane	NI	nd	775	2.5	559	0.0	0.0	8.2	9.5	0.6	0.4	7.5	790.3	15.0	1.0	0.1
57	Cetylpyridinium bromide (0.1%)	NI	4.7	75500	2.5	10327	0.0	0.0	14.7	19.8	0.8	0.5	12.0	961.0	15.2	4.2	0.3
58	Glycerol	NI	4.9	100000	2.5	100000	0.0	0.3	7.7	7.6	1.1	0.6	13.4	1113.5	7.6	0.1	0.0
59	Polyethylene glycol 400	NI	5.6	100000	2.5	100000	0.3	0.5	15.0	17.6	1.2	0.6	14.1	852.8	24.9	0.9	0.0

Footnote to Table 9.5

[a]The *in vitro* data were obtained during the EC/HO validation study on alternatives to the Draize eye test (Balls *et al*., 1995).

SM = silicon microphysiometer; RBC = red blood cell; IRE = isolated rabbit eye; ICE = isolated chicken eye; FL = fluorescein leakage; EDE = EYTEX Draize equaivalent score; IRE = isolated rabbit eye; BCOP = bovine corneal opacity and permeability.
O1 = opacity at 1h; O4 = opacity at 4h; S1 = swelling at 1h; S4 = swelling at 4h; R = retention; O = opacity; S = swelling; P = permeability.
nd = no data. See chapter 2 for further details on these *in vitro* endpoints.

Table 9.6 ***In vitro*** **data for 143 chemicals obtained in the HETCAM and neutral red uptake tests**[a]

No	Chemical	EU class	TC10	TH10	TL10	TC100	TH100	TL100	logIC50
1	Arcopal	R41	204.0	50.1	119.6	169.1	26.6	80.1	-1.52
2	Kuppler 43	R36	128.0	70.6	155.0	nd	nd	nd	0.05
3	Hoe T 3761	NI	292.6	125.1	210.6	165.5	49.8	124.2	0.54
4	Remoglan	R41	116.5	25.1	126.1	100.1	16.7	85.3	-1.10
5	Genagen	NI	249.2	64.6	171.2	208.7	46.1	180.1	0.11
6	Sept	R41	301.0	54.2	178.8	5.5	19.8	114.9	-0.17
7	RK Blau	NI	301.0	301.0	216.4	nd	nd	nd	0.04
8	Glycediol	NI	301.0	187.7	301.0	301.0	92.0	301.0	0.95
9	Hypo 36	NI	301.0	46.0	301.0	200.3	32.5	104.1	-1.73
10	2,2,3-Trimethyl-3-cyclo-pentene-1-acetaldehyde	NI	301.0	144.8	300.8	200.7	28.6	100.2	-0.87
11	Silan 167	NI	301.0	216.7	301.0	301.0	103.1	301.0	-0.95
12	Silan 108 (trimethoxyoctylsilane)	NI	301.0	141.7	224.0	215.5	54.9	183.7	0.02
13	Hypo 20	NI	301.0	301.0	301.0	301.0	53.1	178.5	-1.50
14	Ede 140	NI	301.0	198.3	301.0	301.0	160.9	301.0	2.99
15	Silan 165	NI	301.0	280.4	301.0	301.0	157.3	148.4	0.18
16	Olak	R41	75.4	15.4	32.1	31.1	11.6	17.7	-0.64
17	Silan 103 (trimethoxypropyl silane)	NI	301.0	71.3	301.0	238.0	26.3	188.9	nd
18	Olesulf	R41	91.9	16.7	48.7	31.3	10.3	29.1	-0.78
19	N-(2-methylphenyl)-imidodicarbonimidic diamide	R41	83.2	18.5	50.5	301.0	48.8	279.0	-0.24
20	Ethiosan	NI	301.0	301.0	301.0	nd	nd	nd	nd
21	Sodium dodecyl ether sulphate	R36	182.8	16.4	50.1	57.0	11.8	30.5	-0.69
22	Tocla	R41	29.8	7.3	157.8	14.5	4.7	154.2	-0.83
23	Methyltriglycol	NI	253.7	52.4	268.4	30.0	12.1	160.7	1.23
24	1-(2,6-Dimethylphenoxy)-2-propanol	NI	193.2	119.6	180.5	156.8	38.3	122.2	-0.12
25	4-([2-Sulphatoethyl]sulphonyl)-aniline	R41	18.1	8.4	11.7	195.0	85.0	146.2	-1.21
26	Hyton	R36	141.3	17.0	55.2	48.4	10.8	43.1	-0.58
27	Caffeine sodium salicylate	NI	293.8	54.6	116.8	84.6	13.0	38.1	0.70
28	Phosphonat A	NI	276.3	58.4	195.0	256.8	121.6	175.1	-0.19
29	Mecre	NI	nd	nd	nd	nd	nd	nd	-0.08
30	Ethylacrolein	R41	301.0	30.0	301.0	44.4	11.9	183.3	-2.27
31	2-Hydroxyethyliminodisodium acetate	NI	244.8	29.5	88.3	97.0	18.7	195.3	0.99
32	Hydo 98	R41	238.8	15.2	75.1	230.2	14.1	78.7	-1.21
33	Hypo 45	NI	301.0	167.9	301.0	217.4	65.3	220.7	-1.69
34	(-)Phenylephrine	NI	169.5	119.3	266.8	301.0	82.2	210.0	-0.12
35	p-Nitrobenzoic acid	NI	165.4	134.9	142.5	301.0	301.0	301.0	0.17
36	Hypo 54	R41	301.0	39.3	301.0	301.0	50.8	301.0	-1.81
37	Piperazine	NI	207.1	187.4	218.2	186.6	61.1	90.9	0.78
38	Xanthinol nicotinate	NI	301.0	37.2	161.7	191.1	40.7	57.0	0.88
39	3-5-Dihydroxyacetophenone	NI	301.0	55.7	301.0	nd	nd	nd	-0.40
40	Caffeine sodium benzoate	NI	275.5	49.4	215.3	152.2	27.1	142.6	0.06

Table 9.6 continued

No	Chemical	EU class	TC10	TH10	TL10	TC100	TH100	TL100	logIC50
41	Genomoll (tris(2-chloroethyl)phosphate)	NI	203.4	59.0	199.2	174.8	43.3	187.1	-0.21
42	Phenylephrine hydrochloride	NI	282.5	25.8	99.7	38.4	8.3	27.9	0.15
43	1,2,6-Hexanetriol	NI	301.0	24.4	159.0	85.7	18.3	49.3	1.22
44	Silan 253	R41	301.0	117.4	301.0	106.8	24.8	219.9	-0.55
45	Polyethylene glycol butyl ether	R36	183.1	30.7	73.5	33.8	9.0	26.3	0.80
46	β-Resorcylic acid (2,4-dihydroxybenzoic acid)	R41	34.7	42.6	163.5	191.6	240.4	203.7	0.28
47	Theophylline sodium acetate	NI	220.0	26.3	198.8	299.3	93.7	155.2	0.42
48	Theophylline sodium	NI	293.0	53.3	227.6	301.0	61.6	251.1	0.30
49	Potato starch	NI	215.4	212.3	301.0	301.0	227.6	301.0	1.78
50	1,2-Epoxycyclooctane	NI	301.0	87.5	301.0	301.0	164.3	248.4	0.05
51	Methylacetate	R36	301.0	38.3	301.0	69.5	6.8	54.1	0.98
52	Polysolvan (butyl glycolate)	R41	80.1	16.3	174.8	38.3	3.9	36.3	0.29
53	Theobromine	NI	301.0	301.0	301.0	301.0	301.0	301.0	-0.04
54	α-Lactid	NI	113.3	148.1	195.6	246.7	194.3	286.0	0.53
55	Isopropyl acetoacetate	NI	301.0	90.2	301.0	188.7	44.4	199.2	-0.06
56	Hnol	NI	301.0	279.3	301.0	301.0	238.3	172.3	-1.80
57	(+)Phenylephrine	NI	67.0	24.2	29.2	266.3	34.4	174.3	-0.02
58	Polyhexamethylene guanidine	R41	283.1	39.5	76.1	98.8	25.3	110.7	-2.02
59	N-Acetyl methionine	R41	198.1	167.0	104.0	219.7	301.0	178.5	0.76
60	Potassium cyanate	R41	67.0	19.1	60.6	40.3	22.3	77.5	-0.58
61	Methyltetraglycol	NI	301.0	127.3	222.7	56.8	7.5	165.1	1.58
62	1,3-Dinitrobenzene	R36	240.3	176.3	301.0	286.5	194.8	301.0	-1.16
63	Butanal	NI	201.0	55.2	190.7	6.7	11.6	62.6	0.20
64	α-Ketoglutaric acid	R41	27.2	11.0	46.2	16.9	9.2	27.4	0.37
65	Sodium disilicate	R41	31.7	5.8	14.7	72.8	23.1	48.1	1.64
66	3-Methoxybenzaldehyde	NI	301.0	111.8	301.0	108.3	31.2	198.1	-0.53
67	Butanol	R41	104.7	11.4	161.8	23.9	4.7	155.5	0.54
68	L-Lysine monohydrate	NI	301.0	39.6	96.9	177.7	22.2	72.4	0.75
69	Anisole	R41	301.0	136.9	230.7	32.3	18.7	38.2	0.15
70	Sodium hydrogensulphate	R41	60.8	30.9	41.7	42.4	29.9	26.8	0.60
71	Triisooctylamine	NI	301.0	269.7	301.0	166.1	52.8	219.4	-2.15
72	4-Anisidine	R36	259.0	69.8	301.0	286.3	64.7	178.3	-0.75
73	Aspartic acid	NI	301.0	220.2	261.1	301.0	301.0	301.0	0.51
74	Ethylbutanal	NI	301.0	190.7	301.0	62.8	14.1	37.0	0.18
75	Potassium hexacyanoferrate II	R41	92.8	29.1	55.7	137.5	100.1	196.9	1.19
76	Isobutanal	NI	301.0	232.8	301.0	18.6	19.2	19.4	0.12
77	1,2-Epoxydodecane	NI	301.0	176.2	301.0	301.0	120.4	216.0	-0.79
78	1,2-Phenylenediamine (1,2-benzenediamine)	R36	125.9	29.9	138.4	197.1	56.4	174.9	-0.22
79	L-Glutamic acid hydrochloride	R41	103.8	30.2	190.6	69.8	37.9	206.6	0.48
80	DTPA pentasodium salt	R41	119.7	47.4	43.0	29.2	7.6	16.3	0.22
81	Potassium hexacyanoferrate III	NI	269.0	44.4	301.0	242.3	104.8	254.9	-0.24
82	Sodium pyrosulphite	R41	29.3	28.7	193.0	58.3	44.6	237.1	-0.20

Table 9.6 continued

No	Chemical	EU class	TC10	TH10	TL10	TC100	TH100	TL100	logIC50
83	Sodium cyanate	NI	220.3	24.8	72.5	233.6	39.8	170.3	-0.73
84	4-Chloro-4'-nitrodiphenylether	NI	301.0	75.8	301.0	301.0	195.5	262.1	1.64
85	4-Amino-5-methoxy-2-methylbenzene sulphonic acid	R41	92.7	143.5	169.4	90.4	216.7	110.8	nd
86	Isononylaldehyde	NI	301.0	301.0	301.0	301.0	89.2	148.1	-0.65
87	Ammonium persulphate	NI	301.0	167.0	223.4	109.8	31.6	174.8	-1.12
88	4-Chloro-methanilic acid	R41	41.9	29.4	137.1	143.9	179.3	273.5	0.60
89	2-Methyl-1-propanol	R41	72.0	24.7	33.5	18.9	6.7	24.8	0.45
90	Diepoxide 126	NI	264.9	101.1	254.7	146.3	48.1	229.9	-0.29
91	2-Hydroxyisobutyric acid ethyl ester	R41	107.8	31.1	60.7	33.4	6.7	20.1	0.87
92	Paraformaldehyde (methanal)	R41	301.0	263.5	265.3	301.0	301.0	301.0	-2.13
93	3-Cyclohexene-1-methanol	R41	102.0	24.0	283.8	25.3	6.1	51.0	-0.24
94	4,4-Dimethyl-3-oxopentanenitrile	NI	301.0	42.3	301.0	301.0	173.8	125.3	0.32
95	Methylphosphonic acid bis(oxianylmethyl) ester	R41	181.7	33.1	165.8	69.7	13.5	116.2	-1.23
96	Hexamethylene tetramine	NI	301.0	172.0	175.3	190.8	20.9	162.8	-0.08
97	3-Mercapto-1,2,4-triazole	R36	301.0	36.7	301.0	301.0	301.0	301.0	-0.85
98	Phenylthiourea	NI	301.0	179.6	301.0	301.0	207.7	301.0	nd
99	1,4-Dibutoxy-benzene	R36	301.0	173.9	301.0	301.0	301.0	301.0	-0.93
100	Ambuphylline	NI	152.3	46.4	127.5	102.6	33.9	125.6	0.16
101	2-Hydroxyisobutyric acid	R41	44.8	182.2	163.7	40.0	169.0	153.3	0.52
102	Chlorhexidine	R41	178.0	95.2	284.8	301.0	151.5	297.8	-2.09
103	Iminodiacetic acid	R41	217.0	174.5	283.0	194.2	170.2	255.4	0.64
104	5-Methyl-1,3-4-thiadiazol-2-amine	NI	301.0	301.0	301.0	301.0	301.0	301.0	0.77
105	Chlorhexidine hydrochloride	NI	66.3	231.7	180.9	301.0	237.0	301.0	-0.44
106	1,2-Dodecanediol	R41	301.0	171.0	301.0	301.0	110.2	262.2	-1.61
107	Sodium monochloroacetate	NI	301.0	172.8	230.0	76.4	26.2	207.3	-0.66
108	Rubinrot Y	NI	301.0	61.5	301.0	287.5	165.0	301.0	0.06
109	Polyethylene glycol dimethylether	NI	301.0	179.3	301.0	152.8	31.9	94.6	1.33
110	Hexahydrofarnesyl acetone	NI	301.0	195.1	301.0	301.0	101.3	182.3	nd
111	Hoe MBF	NI	301.0	301.0	301.0	301.0	290.7	301.0	0.08
112	Sodium bisulphite	NI	82.8	25.0	207.5	47.0	10.9	45.6	0.22
113	TA 01946 alkylsilan	NI	301.0	31.5	121.9	202.7	28.3	61.8	nd
114	Trioxane	NI	230.9	54.4	80.8	63.2	18.0	42.1	1.30
115	Butyl carbamate	R41	208.6	121.6	256.7	80.8	13.7	150.7	0.14
116	Gadopentetic acid dimeglumine salt	NI	289.9	266.5	122.4	258.3	221.7	152.3	1.48
117	Sodium sulphite	NI	110.6	73.4	163.1	130.4	59.7	124.2	-0.09
118	Isobornyl acetate	NI	301.0	124.9	301.0	301.0	44.3	212.6	-0.80
119	Acetoacetic acid glycolester	NI	301.0	205.7	301.0	269.3	177.8	281.0	0.24
120	Camphen (2,2-dimethyl-3-methylene bicyclo 2.2.1 heptane)	R36	301.0	149.1	141.8	301.0	217.5	191.0	-0.63
121	Lial-111-glucoside	R41	212.9	24.1	37.6	93.9	14.2	22.1	-0.65
122	Isodecylglucoside	R41	195.3	9.2	63.2	182.2	10.1	45.1	-0.38
123	2-Pseudojonon	NI	301.0	89.2	180.9	301.0	91.6	205.0	-1.23

Table 9.6 continued

No	Chemical	EU class	TC10	TH10	TL10	TC100	TH100	TL100	logIC50
124	Methylpentynol	R41	84.1	27.8	180.3	15.7	13.6	164.9	0.49
125	PO 2	NI	301.0	39.3	301.0	301.0	292.8	301.0	nd
126	2,5-Dimethylhexanediol	R41	301.0	52.9	191.2	126.5	20.8	167.6	0.81
127	Sacyclo	NI	301.0	98.8	301.0	301.0	69.4	301.0	nd
128	C12/C14-Glucoside	R41	301.0	46.4	76.5	101.4	29.0	39.2	-0.70
129	1,6,7,12-Tetrachloro-3,4,9,10-tetracarbonic acid anhydride	NI	301.0	301.0	301.0	301.0	239.7	301.0	nd
130	Diphocars	R41	54.9	19.8	33.0	54.9	23.7	140.5	-0.39
131	3,6-Dimethyloctanol	R41	301.0	291.4	301.0	281.7	37.2	301.0	-0.32
132	Napt	NI	301.0	171.2	301.0	224.1	192.3	154.5	0.52
133	B 25	NI	301.0	301.0	301.0	301.0	301.0	301.0	nd
134	1-Chloroctan-8-ol	R41	300.3	35.8	252.3	50.9	13.5	120.6	-0.85
135	DC 8	NI	301.0	301.0	301.0	301.0	207.7	258.8	nd
136	Bis-(3-triethoxysilylpropyl)-tetrasulphide	NI	301.0	172.5	301.0	301.0	113.9	301.0	nd
137	Nitro-bis-octylamide	NI	301.0	261.8	292.9	301.0	138.9	221.5	-0.56
138	7-Acetoxyheptanal	R41	301.0	229.3	301.0	113.7	66.9	143.2	-1.18
139	4-Amino-azobenzene-4-sulphonic acid	NI	301.0	301.0	301.0	301.0	301.0	301.0	0.39
140	Oxol 9N	NI	301.0	173.3	301.0	288.3	63.1	301.0	nd
141	Isotridecanal	NI	301.0	301.0	301.0	301.0	203.7	301.0	-0.14
142	Cerium 2-ethylhexanoate	NI	250.6	86.2	201.9	39.8	11.1	110.1	0.53
143	Wessalith slurry	R41	301.0	70.4	153.9	203.1	41.0	137.3	0.41

Footnote to Table 9.6

[a]The *in vitro* data were taken from Spielmann *et al*. (1996).

Effects observed in the hen's egg chorioallantoic membrane: TC10 = detection time for coagulation with a 10% solution; TC100 = detection time for coagulation with a neat substance; TH10 = detection time for hæmorrhage with a 10% solution; TH100 = detection time for hæmorrhage with a neat substance; TL10 =detection time for lysis with a 10% solution; TL100 = detection time for lysis with a neat substance.

logIC50 = logarithm of concentration of test chemical resulting in 50% inhibition of neutral red uptake in 3T3 cells; nd = not determined.

Table 9.7 Performance of classification models for eye irritation potential

Model	Sensitivity (95% CI)	Specificity (95% CI)	Concordance (95% CI)	False positives (95% CI)	False negatives (95% CI)
SAR 1[a]	97	49	61	51	3
	(90-100)	(39-60)	(53-71)	(40-61)	(0-10)
SAR 2[b]	89	79	81	21	11
	(78-100)	(70-87)	(74-88)	(13-30)	(0-22)
PM 1[c]	53	97	76	3	47
	(41-63)	(93-100)	(69-83)	(0-17)	(37-59)
PM 1[d]	44	94	61	6	56
PM 2[e]	46	100	63	0	54
	(31-62)	g	(51-76)	g	(38-69)
PM 3[e]	59	100	72	0	61
	(43-75)	g	(60-84)	g	(16-40)
PM 4[e]	56	94	68	6	44
	(54-83)	(74-100)	(64-86)	(0-26)	(17-46)
PM 5[e]	69	89	75	11	31
	(55-84)	(74-100)	(64-86)	(0-26)	(16-45)
PM 6[f]	67	79	74	21	33
	(55-80)	(69-88)	(66-81)	(12-31)	(20-45)

Footnote to Table 9.7

[a]Statistics based on the application of SAR 1 to its training set of 119 chemicals; [b]Statistics based on the application of SAR 2 to its training set of 117 chemicals; [c]Statistics based on the application of PM 1 to its training set of 165 chemicals; [d]Statistics based on the application of PM 1 to a test set of 49 chemicals; [e]Statistics based on the application of the PM to its training set of 57 chemicals; [f]Statistics based on the application of PM 6 to its training set of 129 chemicals; [g]No variability because all NI chemicals in the data set were correctly identified.

The 95% confidence intervals were derived by bootstrap resampling the appropriate data set 1000 times.

Table 9.8 Variability in the parameters of an elliptic model for eye irritation

parameter	estimate based on full data set	bootstrap mean	bias	bias-corrected estimate	bootstrap lower 95% CL	bootstrap upper 95% CL
centroid along logP	0.655	0.653	-0.002	0.657	0.439	0.867
centroid along dV1	-1.226	-1.233	-0.007	-1.219	-1.438	-1.028
radial axis along logP	1.174	1.137	-0.037	1.211	0.825	1.448
radial axis along dV1	1.090	1.063	-0.027	1.117	0.890	1.236

CHAPTER 10

EVALUATION OF TIERED APPROACHES TO HAZARD CLASSIFICATION

10.1 INTRODUCTION

In chapters 7 and 9, various CMs for predicting acute skin and eye toxicity are reported. These models were developed and evaluated on the basis that they would be applied as stand-alone alternatives to animal experiments. In practice, however, such models are more likely to be used in the context of an integrated testing strategy, such as the OECD testing strategies described in chapter 2. The rationale behind the OECD approach is that alternative methods should be used to identify toxic substances, so that animal experiments are used mainly to confirm predictions of non-toxicity. An alternative rationale, proposed in chapter 9, is that models of high specificity (low false positive rate) could be used to identify toxic substances, whereas models of high sensitivity (low false negative rate) could be used to identify non-toxic substances. Chemicals predicted to be toxic do not undergo further testing, whereas chemicals predicted to be non-toxic are tested directly in animals. In this approach, models that identify toxic chemicals are used to terminate the testing process, whereas models that identify non-toxic substances are used to expedite the process (by skipping intermediate steps based on more time-consuming and expensive alternative methods).

The aim of the work presented in this chapter was to evaluate these two approaches to hazard classification. The OECD approach was evaluated by using models for skin irritation and corrosion reported in chapter 7, whereas the alternative approach was investigated by using models for eye irritation reported in chapter 9. This work represents a development of previous studies (Worth *et al.*, 1998; Worth & Fentem, 1999), which were communicated to the OECD Secretariat before the (then proposed) OECD testing strategies for eye and skin irritation/corrosion were accepted (OECD, 1998).

10.2 METHOD

10.2.1 Evaluation of the OECD approach to hazard classification

The OECD approach was evaluated by simulating possible outcomes obtained when a stepwise sequence of five alternative tests and one animal test (Figure 10.1) is applied to

a heterogeneous set of 51 chemicals (Table 10.1). The decision rules in steps 1 to 5 (Figure 10.1) are based on CMs for skin irritation and corrosion presented in chapter 7.

The 51 chemicals chosen for the evaluation form a subset of 60 test chemicals used in the ECVAM validation study on alternative methods for skin corrosion (Barratt *et al*., 1998; Fentem *et al*., 1998), and were selected on the basis that the *in vivo* effects of each chemical were determined for the neat substance (since the models were parameterised to predict the effects of neat substances, not dilutions). The chemicals were classified for skin irritation and corrosion potential by applying EU classification criteria (see chapter 2) to the animal data.

A number of simulations were performed to assess the effects of applying the different combinations of the six steps outlined in Figure 10.1 (each combination is referred to hereafter as a different sequence). Specifically, assessments were made of: a) six stepwise sequences for predicting whether chemicals are corrosive (and irritant), non-corrosive and irritant, or non-corrosive and non-irritant; b) two sequences for predicting whether chemicals are corrosive or non-corrosive (without regard to irritation potential); and c) one sequence for predicting skin irritation potential (without regard to its corrosion potential). The outcome of each simulation was used to compare the ability of each stepwise sequence to predict EU classifications, and to reduce and refine the use of animals, with the corresponding ability of the EPISKIN test, when used as a stand-alone alternative method.

10.2.2 Evaluation of an alternative approach to hazard classification

The alternative approach was evaluated by simulating possible outcomes obtained when a stepwise sequence of alternative and animal tests (Figure 10.2) is applied to a heterogeneous set of 45 chemicals (Table 10.2). The decision rules in steps 1 to 3 (Figure 10.2) are based on CMs presented in chapter 9.

The 45 chemicals chosen for the evaluation form a subset of 60 test chemicals used in the EC/HO validation study on alternative methods for eye irritation (Balls *et al*., 1995), and were selected on the basis that the *in vivo* effects of each chemical were determined for the neat substance. One chemical, thiourea, was excluded because it was acutely

toxic to the rabbits. The chemicals were classified for eye irritation potential by applying EU classification criteria (see chapter 2) to the animal data.

To assess the alternative approach to hazard classification, a simulation was performed to assess the effects of applying steps 1, 2 and 3 outlined in Figure 10.2. In step 1 of this sequence, a PM based on pH measurements is used as the high-specificity model for identifying irritant chemicals (but not non-irritants), and an elliptic SAR based on logP and dV1 is used as the high-sensitivity model for identifying non-irritants (but not irritants). The PM was applied in the first step, so that chemicals it predicted to be irritant would not proceed to subsequent steps, and the SAR was applied in the second step, so that chemicals it predicted to be non-irritant would undergo animal testing directly, skipping the *in vitro* test in the third step.

The use of the OECD approach for the assessment of eye irritation potential was also simulated. In this case, the PMs based on pH measurements and isolated rabbit eye (IRE) data were applied (i.e. steps 1 and 3). The elliptic SAR (step 2) was omitted because this CM was being used to predict the absence of irritation potential, and was therefore redundant in a strategy based on the sequential elimination of chemicals predicted to be irritant.

The outcome of each simulation was used to compare the ability of the corresponding stepwise sequence to predict EU classifications, and to reduce and refine the use of animals, with the corresponding ability of the IRE test, when used as a stand-alone method.

10.3 RESULTS

10.3.1 The OECD approach to hazard classification

The predicted and known classifications of skin irritation and corrosion potential are given in Table 10.1. Predictions of corrosion potential made by the SAR in step 1 are only made for the 36 single chemicals that are organic liquids, since the domain of the SAR excludes inorganic substances, solids, and mixtures.

When used as a stand-alone test for skin irritation and corrosion, EPISKIN predicts the three classes of toxic potential (corrosive and irritant, non-corrosive and irritant, non-corrosive and non-irritant) with an accuracy of 69% and a kappa statistic (chance-corrected accuracy) of 53% (Table 10.3). If the 22 chemicals predicted to be non-corrosive and non-irritant were subsequently tested on rabbits, three chemicals would be found to be corrosive, five irritant, and 14 to be non-corrosive and non-irritant (Table 10.4).

Although the effects of six stepwise sequences incorporating EPISKIN and other alternative models were investigated, only one sequence, in which the use of EPISKIN is preceded by the use of the pH test for corrosion (step 2), was found to achieve a greater accuracy (73%; Table 10.3). This can be related to the fact that two models, the SAR for corrosion (step 1) and the pH test for irritation (step 4) have relatively high false positive rates (Table 10.4), so the exclusion of these CMs from a stepwise sequence improves its predictive capacity. The effect on animal testing of the best sequence (steps 2-3-5) is essentially the same as the use of EPISKIN alone (steps 3-5), with two corrosive chemicals being tested on rabbits, rather than three (Table 10.4).

When used as a stand-alone test for skin corrosion, EPISKIN has an accuracy of 86% and a kappa statistic of 72% (Table 10.3), predicting 21 chemicals to be corrosive (Table 10.4). Thus, if EPISKIN were used in sequence with the Draize eye test, the 30 remaining chemicals predicted to be non-corrosive would be tested on rabbits. Four of these chemicals would be found to be corrosive, and 26 to be non-corrosive (Table 10.4). If the pH test for corrosion is used in combination with EPISKIN (steps 2-3), a greater accuracy of 90% is obtained (Table 10.3), with just two corrosive chemicals being tested on rabbits, instead of four (Table 10.4).

For the prediction of skin irritation potential, the use of EPISKIN alone is more predictive (accuracy of 73%) than its use in combination with the pH test (accuracy of 61%). Again, this is because the pH test for irritation has a tendency to over-predict.

10.3.2 An alternative approach to hazard classification

The predicted and known classifications of eye irritation potential are given in Table 10.2. As a stand-alone method for predicting eye irritation potential, the IRE test discriminates between the 28 irritants and 17 non-irritants with an accuracy of 71% and a kappa statistic of 42% (Table 10.5), and predicts 29 of the 45 chemicals to be non-irritant (Table 10.6). If the remaining 16 chemicals were tested on rabbits, six would be found irritant, and ten non-irritant.

Adoption of the OECD approach to the tiered testing leads to the slightly increased accuracy of 73% (kappa statistic = 46%; Table 10.5), with five irritant and 10 non-irritant chemicals being tested on rabbits (Table 10.6). Conversely, adoption of the alternative approach increases the accuracy to 76%, with six irritant and 12 non-irritant chemicals being tested on animals.

10.4 DISCUSSION

The results of this study show that stepwise approaches to hazard classification, in which alternative methods are applied before animal tests, provide a promising means of reducing and refining the use of animals, since fewer animal experiments need to be conducted, and of those chemicals tested *in vivo*, the majority are found to be non-toxic. The results also show that the sequential use of several alternative methods can be at least as predictive as the stand-alone use of a single *in vitro* test.

The validity of these conclusions depends on the adequate performance of each alternative method included in the stepwise sequence. In particular, methods that over-predict toxic potential may compromise the performance of strategies in which they are incorporated, since (according to the approaches evaluated) chemicals found to be toxic do not undergo further testing. Thus, when designing a tiered testing strategy, it is important that the models included should have low false positive rates (i.e. high specificities). For example, it might be decided that false positive rates should not exceed 10%. In general, models with lower false positive rates tend to have lower sensitivities. However, as discussed in chapter 9, even models with sensitivities less than or equal to 50% may be useful in the context of a tiered testing strategy, since it is

not important that any single model is capable of identifying a majority of the toxic chemicals in a test set, as long as there is a high degree of certainty associated with the positive predictions.

In this study, the OECD approach to tiered testing was evaluated by using models for skin corrosion and eye irritation, whereas the alternative approach was examined by using models for eye irritation. The rationale behind the alternative approach is that some models are better suited for identifying toxic chemicals, whereas others are better suited for identifying non-toxic chemicals, due to the inescapable overlap between toxic and non-toxic chemicals along certain variables. Although such models may be unacceptable as stand-alone alternatives to animal experiments, their combined use should provide a means of exploiting their strengths while compensating for their weaknesses. In particular, it is foreseen that highly specific methods could be successfully combined with highly sensitive ones. The main difference between the alternative approach and the OECD approach concerns the consequence of negative predictions. In the OECD approach, further tests are conducted to confirm predictions of non-toxicity, which means that there is no useful role to be played by a model that only identifies non-toxic chemicals. In contrast, the alternative approach allows chemicals predicted to be non-toxic by an SAR in a screening step to skip the *in vitro* test in a subsequent step, thereby expediting the assessment process. In the case of the 45 chemicals in Table 10.2, adoption of the OECD approach would lead to *in vitro* testing on 32 chemicals (Table 10.6), whereas adoption of the alternative approach would lead to *in vitro* testing on 22 chemicals. Thus, in the case of the chemicals and CMs studied here, adoption of the alternative approach reduces the burden of *in vitro* testing, without a concomitant reduction in predictive capacity or significant increase in animal testing.

In practice, the desirability of adopting the alternative approach would depend on the time and expense required to perform the *in vitro* test, and the additional information this is expected to generate. In general, *in vitro* tests are included in testing strategies because they are expected to identify mechanisms of toxic action that are not detected by SARs and PMs based on physicochemical data. In such cases, the effect of skipping the *in vitro* test would be to place the burden of identifying toxic chemicals on the animal test, which is unsatisfactory from an animal welfare point-of-view.

Finally, it is important to note that the approaches to hazard classification evaluated in this study represent just two possible ways of integrating the use of different CMs; other designs are conceivable. For example, if each prediction of toxic and non-toxic potential were associated with a probability (e.g. a 70% probability of being corrosive), thresholds other than 50% could be chosen for the identification of toxic and non-toxic chemicals. In fact, models derived by logistic regression or linear discriminant analysis can be used to assign probabilities, but these are likely to be misleading if the assumptions of the statistical methods are not obeyed. Alternatively, the identification of toxic potential could proceed according to a 'majority voting system', in which predictions were made by several models, with classifications being assigned when a majority of models made the same prediction. The OECD approach has now been accepted in OECD Member States for the classification of industrial chemicals on the basis of their acute local toxicity (OECD, 1998). However, variations of this approach could be useful in the industrial setting for non-regulatory testing, for example as a means of priority setting during chemical development.

10.5 CONCLUSIONS

It is concluded that:

a) testing strategies based on the sequential use of alternative methods prior to the use of animal methods provide an effective means of reducing and refining the use of animals, without compromising the ability to classify chemicals on the basis of toxic hazard.
b) CMs incorporated into OECD-style testing strategies should have high specificities (i.e. low false positive rates), but do not necessarily need high sensitivities, if several complementary models are capable of identifying different groups of toxic chemicals.
c) a CM of high sensitivity (but low specificity) can be combined with a CM of high specificity (but low sensitivity) to exploit the strengths, and compensate for the weaknesses, of the two models.

Figure 10.1 A tiered testing strategy for skin irritation and corrosion based on the OECD approach to hazard classification

Step 1:	Apply SAR	⇒	predict C and stop testing
	⇓		
	NC or no SAR		
	⇓		
Step 2:	Apply PM based on pH data	⇒	predict C and stop testing
	⇓		
	NC or no pH data		
	⇓		
Step 3:	Apply PM based on EPISKIN data	⇒	predict C and stop testing
	⇓		
	NC		
	⇓		
Step 4:	Apply PM based on pH data	⇒	predict I and stop testing
	⇓		
	NI		
	⇓		
Step 5:	Apply PM based on EPISKIN data	⇒	predict I and stop testing
	⇓		
	NI		
	⇓		
Step 6:	Perform Draize skin test	⇒	classify as C, I or NC & NI

Footnote to Figure 10.1

Step 1: If MP ≤ 37ºC and MW ≤ 123 g/mol, predict C; otherwise predict NC.

Step 2: If pH < 2.4 or pH > 10.9, predict C; otherwise predict NC.

Step 3: If EPISKIN viability at 4h < 36%, predict C; otherwise predict NC.

Step 4: If pH < 4.4 or pH > 9.2, predict I; otherwise predict NI.

Step 5: If EPISKIN viability at 4h < 67%, predict I; otherwise predict NI.

Figure 10.2 A tiered testing strategy for eye irritation based on an alternative approach to hazard classification

Step 1:	Apply CM for I chemicals (PM based on pH data)	⇒	predict I and stop testing
	⇓		
	no prediction		
	⇓		
Step 2:	Apply CM for NI chemicals (SAR based on logP and dV1)	⇒	predict NI and go to step 4
	⇓		
	no prediction		
	⇓		
Step 3:	Apply *in vitro* test (PM based on IRE data)	⇒	predict I and stop testing
	⇓		
	NI		
	⇓		
Step 4:	Perform Draize eye test	⇒	classify as I or NI

Footnote to Figure 10.2

CM = classification model; IRE = isolated rabbit eye.

Step 1: If pH < 3.2 or pH > 8.6, predict I; otherwise make no prediction.

Step 2: If MP ≤ 37ºC and $(\text{logP-0.66})^2$ / 1.212 + $(\text{dV1} + 1.22)^2$/ 1.122 >1, predict NI; otherwise make no prediction.

Step 3: If isolated rabbit eye swelling at 1h > 18%, predict I; otherwise predict NI.

Table 10.1 Data set of 51 chemicals used to evaluate a tiered testing strategy for skin irritation and corrosion

No	Chemical	Step 1	Step 2	Step 3	Step 4	Step 5	Draize test
1	Hexanoic acid	C	NC	C	I	I	C & I
2	1,2-Diaminopropane	C	C	C	I	I	C & I
3	Carvacrol	NC	NC	C	I	I	C & I
4	Boron trifluoride dihydrate	np	C	C	I	I	C & I
5	Methacrolein	C	NC	C	I	I	C & I
6	Phenethyl bromide	NC	NC	NC	I	NI	NC & NI
7	3,3'-Dithiodipropionic acid	NC	NC	NC	I	NI	NC & NI
8	Isopropanol	C	NC	NC	I	NI	NC & NI
9	2-Methoxyphenol (Guaiacol)	NC	NC	C	I	I	NC & NI
10	2,4-Xylidine (2,4-Dimethylaniline)	NC	NC	NC	I	I	NC & NI
11	2-Phenylethanol (phenylethylalcohol)	C	NC	NC	I	NI	NC & NI
12	Dodecanoic (lauric) acid	NC	NC	NC	I	NI	NC & NI
13	3-Methoxypropylamine	C	C	C	I	I	C & I
14	Allyl bromide	C	NC	C	I	I	C & I
15	Dimethyldipropylenetriamine	NC	C	C	I	I	C & I
16	Methyl trimethylacetate	C	NC	NC	I	NI	NC & NI
17	Dimethylisopropylamine	C	C	C	I	I	C & I
18	Tetrachloroethylene	NC	NC	NC	I	I	NC & I
19	Ferric [iron (III)] chloride	np	C	C	I	I	C & I
20	Butyl propanoate	NC	NC	NC	I	NI	NC & NI
21	2-*tert*-Butylphenol	NC	NC	C	NI	I	C & I
22	Isostearic acid	np	C	C	I	I	NC & I
23	Methyl palmitate	NC	NC	NC	I	NI	NC & I
24	Phosphorus tribromide	NC	NC	NC	I	NI	C & I
25	65/35 Octanoic/decanoic acids	np	C	C	I	I	C & I
26	4,4'-Methylene-bis-(2,6-ditert-butylphenol)	NC	NC	NC	I	NI	NC & NI
27	2-Bromobutane	NC	NC	NC	I	I	NC & NI
28	Phosphorus pentachloride	np	C	C	I	I	C & I
29	4-(Methylthio)-benzaldehyde	NC	NC	NC	I	NI	NC & NI
30	70/30 Oleine/octanoic acid	np	NC	NC	I	I	NC & I
31	Hydrogenated tallow amine	np	NC	NC	I	NI	NC & I
32	2-Methylbutyric acid	C	NC	C	I	I	C & I
33	Tallow amine	np	NC	NC	I	NI	C & I
34	2-Ethoxyethyl methacrylate	NC	NC	NC	I	NI	NC & NI
35	Octanoic acid (caprylic acid)	NC	NC	C	I	I	C & I
36	20/80 Coconut/palm soap	np	NC	NC	I	NI	NC & I
37	Benzyl acetone	NC	NC	NC	I	NI	NC & NI
38	Heptylamine	C	C	NC	I	NI	C & I
39	Cinnamaldehyde	NC	NC	NC	I	I	NC & I
40	60/40 Octanoic/decanoic acids	np	NC	C	I	I	C & I

Table 10.1 continued

No	Chemical	Step 1	Step 2	Step 3	Step 4	Step 5	Draize test
41	Eugenol	NC	NC	NC	I	I	NC & NI
42	55/45 Octanoic/decanoic acids	np	NC	C	I	I	C & I
43	Methyl laurate	NC	NC	NC	I	NI	NC & I
44	Sodium bicarbonate	np	NC	NC	I	NI	NC & NI
45	Sulphamic acid	np	C	C	I	I	NC & NI
46	Sodium bisulphite	np	NC	NC	I	NI	NC & NI
47	1-(2-Aminoethyl)piperazine	NC	C	NC	I	I	C & I
48	1,9-Decadiene	NC	NC	NC	I	NI	NC & I
49	Phosphoric acid	np	C	C	I	I	C & I
50	10-Undecenoic acid	NC	NC	NC	I	I	NC & NI
51	4-Amino-1,2,4-triazole	C	NC	NC	I	NI	NC & NI

Footnote to Table 10.1

Step 1: If MP ≤ 37ºC and MW ≤ 123 g/mol, predict C; otherwise predict NC.

Step 2: If pH < 2.4 or pH > 10.9, predict C; otherwise predict NC.

Step 3: If EPISKIN viability at 4h < 36%, predict C; otherwise predict NC.

Step 4: If pH < 4.4 or pH > 9.2, predict I; otherwise predict NI.

Step 5: If EPISKIN viability at 4h < 67%, predict I; otherwise predict NI.

C = corrosive (R34 or R35); I = irritant (R38); NC = non-corrosive; NI = non-irritant; np = no prediction (chemical outside domain of SAR).

Shading indicates the step at which the assessment would stop and a classification would be assigned to the chemical in the OECD-style tiered testing strategy (Figure 10.1).

Table 10.2 Data set of 45 chemicals used to evaluate a tiered testing strategy for eye irritation

No	Chemical	Step 1	Step 2	Step 3	Draize test
1	Cyclohexanol	np	np	I	I
2	Parafluoraniline	I	np	I	I
3	Acetone	np	np	NI	I
4	Hexanol	np	NI	I	I
5	Isobutanol (butan-2-ol)	np	np	I	I
6	2-Ethyl-1-hexanol	np	NI	NI	NI
7	Methyl ethyl ketone	np	np	I	I
8	Pyridine	I	np	I	I
9	2,2-Dimethylbutanoic acid	np	NI	I	NI
10	γ-Butyrolactone	np	np	I	I
11	Octanol	np	NI	NI	I
12	Methyl acetate	np	np	NI	I
13	Isopropanol (propan-2-ol)	np	np	NI	I
14	Methyl cyanoacetate	np	NI	NI	I
15	Ethanol	np	NI	I	NI
16	2,6-Dichlorobenzoyl chloride	I	NI	NI	I
17	Ethyl-2-methyl acetoacetate	np	np	NI	NI
18	Ethyl acetate	np	np	NI	NI
19	Toluene	np	NI	NI	NI
20	Butyl acetate	np	np	NI	NI
21	Methyl isobutyl ketone	np	np	I	NI
22	Tween 20	np	np	NI	NI
23	Ethyl trimethyl acetate	np	NI	NI	NI
24	Methylcyclopentane	np	NI	NI	NI
25	Glycerol	np	NI	NI	NI
26	Polyethylene glycol 400	np	np	NI	NI
27	Captan 90 concentrate	np	np	I	I
28	Chlorhexidine	I	np	I	I
29	Quinacrine	np	np	I	I
30	Promethazine HCl	np	np	I	I
31	1-Naphthalene acetic acid, Na salt	np	np	I	I
32	Sodium oxalate	I	np	I	I
33	Imidazole	I	np	I	I
34	4-Carboxybenzaldehyde	I	np	I	I
35	1-Naphthalene acetic acid	np	np	I	NI
36	L-Aspartic acid	I	np	I	I
37	Benzoyl-L-tartaric acid	I	np	I	I
38	Potassium cyanate	I	np	I	I
39	Sodium perborate	I	np	I	I
40	Dibenzyl phosphate	I	np	I	I
41	2,5-Dimethylhexanediol	np	np	I	NI
42	Ammonium nitrate	np	np	I	I
43	Maneb	np	np	I	I
44	Fomesafen	np	np	I	NI
45	Tetraaminopyrimidine sulphate	I	np	I	NI

Footnote to Table 10.2

Step 1: If pH < 3.2 or pH > 8.6, predict I; otherwise make no prediction.

Step 2: If MP ≤ 37ºC and $(\text{logP-0.66})^2 / 1.21^2 + (\text{dV1} + 1.22)^2 / 1.12^2 > 1$, predict NI; otherwise make no prediction.

Step 3: If isolated rabbit eye swelling at 1h > 18%, predict I; otherwise predict NI.

I = irritant (R36 or R41); NI = non-irritant; np = no prediction (model used to predict either the presence or the absence of irritation potential).

Table 10.3 Predictive abilities of stepwise sequences for skin irritation and corrosion compared with the stand-alone use of the EPISKIN test

Endpoint(s)	Test or stepwise sequence	Sensitivity	Specificity	Concordance	Kappa
Skin corrosion and irritation	EPISKIN test (3→ 5)	NA	NA	69	53
	Steps: 1→ 2 → 3 → 4 → 5	NA	NA	55	33
	Steps: 1→ 2 → 3 → 5	NA	NA	65	47
	Steps: 1→ 2 → 4 → 5	NA	NA	45	18
	Steps: 1 → 3 → 5	NA	NA	63	44
	Steps: 1 → 4 → 5	NA	NA	55	33
	Steps: 2 → 3 → 5	NA	NA	73	59
Skin corrosion	EPISKIN test (step 3)	82	90	86	72
	Steps: 1 → 2 → 3	91	76	82	64
	Steps: 2 → 3	91	90	90	80
Skin irritation	EPISKIN test (step 5)	74	70	73	46
	Steps: 4 → 5	100	0	61	22

Footnote to Table 10.3

Step 1: If MP ≤ 37ºC and MW ≤ 123 g/mol, predict C; otherwise predict NC.

Step 2: If pH < 2.4 or pH > 10.9, predict C; otherwise predict NC.

Step 3: If EPISKIN viability at 4h < 36%, predict C; otherwise predict NC.

Step 4: If pH < 4.4 or pH > 9.2, predict I; otherwise predict NI.

Step 5: If EPISKIN viability at 4h < 67%, predict I; otherwise predict NI.

NA = not applicable ('sensitivity' and 'specificity' are defined with respect to 2x2 contingency tables, not 3x3 tables).
The most predictive test or sequence for each endpoint is shaded. All performance measures are expressed as percentages.

Table 10.4 Possible outcomes of tiered testing strategies for skin irritation and corrosion

Sequence of steps	No of chemicals entering step	Known toxic potential			No of positive predictions	No of true positives	No of false positives
Assessment of skin corrosion and irritation							
Step 1	51	22 C	9 I	20 NC & NI	12	8	4
Step 2	39	14 C	9 I	16 NC & NI	9	7	2
Step 3	30	7 C	8 I	15 NC & NI	6	5	1
Step 4	24	2 C	8 I	14 NC & NI	24	8	16
Step 5	0	-	-	-	0	0	0
Draize test	0	-	-	-	-	-	-
Step 1	51	22 C	9 I	20 NC & NI	12	8	4
Step 2	39	14 C	9 I	16 NC & NI	9	7	2
Step 3	30	7 C	8 I	15 NC & NI	6	5	1
Step 5	24	2 C	8 I	14 NC & NI	7	3	4
Draize test	17	2 C	5 I	10 NC & NI	-	-	-
Step 2	51	22 C	9 I	20 NC & NI	13	11	2
Step 3	38	11 C	8 I	19 NC & NI	10	9	1
Step 5	28	2 C	8 I	18 NC & NI	7	3	4
Draize test	21	2 C	5 I	14 NC & NI	-	-	-
Step 3	51	22 C	9 I	20 NC & NI	21	18	3
Step 5	30	4 C	8 I	18 NC & NI	8	3	5
Draize test	22	3 C	5 I	14 NC & NI	-	-	-
Assessment of skin corrosion							
Step 1	51	22 C		29 NC	12	8	4
Step 2	39	14 C		25 NC	9	7	2
Step 3	30	7 C		23 NC	6	5	1
Draize test	24	2 C		22 NC	-	-	-
Step 2	51	22 C		29 NC	13	11	2
Step 3	38	11 C		27 NC	10	9	1
Draize test	28	2 C		26 NC	-	-	-
Step 2	51	22 C		29 NC	21	18	3
Draize test	30	4 C		26 NC	-	-	-
Assessment of skin irritation							
Step 4	51	31 I	20 NI		50	30	20
Step 5	1	1 I	0 NI		1	1	0
Draize test	0	-	-		-	-	-
Step 5	51	31 I	20 NI		29	23	6
Draize test	22	8 I	14 NI		-	-	-

Table 10.5 Predictive abilities of stepwise sequences for eye irritation compared with the stand-alone use of the isolated rabbit eye test

Test or stepwise sequence	Sensitivity	Specificity	Concordance	Kappa
Isolated rabbit eye test	79	59	71	42
Steps: $1 \rightarrow 3$ (OECD approach)	82	59	73	46
Steps: $1 \rightarrow 2 \rightarrow 3$ (alternative approach)	79	71	76	52

Footnote to Table 10.5

Step 1: If MP $\leq$ 37ºC and $(\text{logP-0.66})^2 / 1.21^2 + (\text{dV1} + 1.22)^2 / 1.12^2 > 1$, predict NI; otherwise make no prediction.

Step 2: If pH < 3.2 or pH > 8.6, predict I; otherwise make no prediction.

Step 3: If isolated rabbit eye swelling at 1h > 18%, predict I; otherwise predict NI.

All performance measures are expressed as percentages.

Table 10.6 Possible outcomes of tiered testing strategies for eye irritation

Sequence of steps	No of chemicals entering step	Known toxic potential	No of I predictions	No of true positives	No of NI predictions	No of true negatives
Isolated rabbit eye test						
Step 3	45	28 I 17 NI	29	22	16	10
Draize test	16	6 I 10 NI	-	-	-	-
OECD approach						
Step 1	45	28 I 17 NI	13	12	-	-
Step 3	32	16 I 16 NI	17	11	15	10
Draize test	15	5 I 10 NI	-	-	-	-
Alternative approach						
Step 1	45	28 I 17 NI	13	12	-	-
Step 2	32	16 I 16 NI	-	-	10	7
Step 3	22	13 I 9 NI	14	10	8	5
Draize test	18	6 I 12 NI	-	-	-	-

Footnote to Table 10.6

Step 1: If pH < 3.2 or pH > 8.6, predict I; otherwise make no prediction.

Step 2: If MP ≤ 37ºC and $(\text{logP}-0.66)^2 / 1.21^2 + (\text{dV1} + 1.22)^2 / 1.12^2 > 1$, predict NI; otherwise make no prediction.

Step 3: If isolated rabbit eye swelling at 1h > 18%, predict I; otherwise predict NI.

CHAPTER 11

SUMMARY AND GENERAL DISCUSSION

11.1 INTRODUCTION

The aim of this final chapter is to summarise the outcome of this research project, highlighting novel aspects and contributions to existing knowledge. In addition, some perspectives are offered regarding the 'integrated approach to toxicity testing', and an assessment is made of current research needs and future prospects in this field.

11.2 SUMMARY OF THE THESIS

The background to the research project is described in chapters 2-4: chapter 2 covers the underlying biology and toxicology, whereas chapters 3 and 4 describe relevant methods of computational chemistry and statistics, respectively.

The main research findings are reported in chapters 5-10. Chapter 5 describes several applications of the bootstrap resampling method that have not been reported previously in the field of alternative toxicology. In particular, algorithms have been devised to provide a means of assessing the uncertainty in Cooper statistics that arises from chemical variation in the test set. The usefulness of these algorithms in validation studies has been illustrated by applying the algorithms to the high-quality data generated in a validation study. In addition, algorithms have been formulated to bootstrap the raw data obtained in the Draize eye and skin irritation tests, and the bootstrap distributions generated by these algorithms have been used to estimate the variability in the animal endpoints resulting from biological and temporal variation. These estimates are of value in placing an upper limit on the predictive capacity that can be expected of any model that aims to predict the results of the Draize skin or eye test. Furthermore, bootstrap distributions of specific tissue scores (such as conjunctival erythema) have been shown to provide a means of assessing whether or not chemicals are borderline (with respect to the boundaries imposed by regulatory classification systems), and the usefulness of this information has been discussed.

Chapter 6 describes a novel statistical method that the author has termed 'embedded cluster modelling' (ECM). This method has been devised to generate elliptic models of biological activity from embedded data sets, and its usefulness has been illustrated by

deriving a model for eye irritation potential. The algorithm for ECM incorporates an option for bootstrapping the parameters of the elliptic model, so that the uncertainty associated with these parameters can be estimated.

New models for acute dermal and ocular toxicity are presented in Chapters 7 and 9. In the development of these models, it was considered important that: a) the models should have a clear mechanistic basis; and b) should provide an objective means of predicting the toxic potential of chemicals. With respect to the second point, the models are considered to constitute an advance over previous work (see chapter 2), in which appropriate predictor variables were identified, but no explicit prediction models reported. In chapter 9, the method of ECM was used again to develop an SAR for eye irritation potential.

Chapter 8 reports the development of two structure-permeability relationships (SPRs) for predicting the corneal penetration of chemicals. These represent an advance over previous attempts to model corneal penetration (considered to be the rate-limiting step for ocular penetration) in terms of their relative simplicity and broader domains of application. However, due to the variable quality of the data used, the predictive capacity of the models is necessarily limited. Nevertheless, the models themselves could provide the starting point for the further development of ocular SPRs. It was anticipated that the corneal permeabilities of chemicals could be used to predict their potencies in the Draize eye irritation test. However, the data examined did not provide strong evidence in support of this hypothesis. Similarly, in chapter 7, no apparent relationship was observed between the dermal penetration of chemicals and their potency in the Draize skin test.

Whereas chapters 7 and 9 describe the development and assessment of individual classification models, chapter 10 focuses on the integrated use of some of these models in the context of tiered testing strategies for acute dermal and ocular toxicity. The simulations in chapter 7 represent a development of work already published by the author (Worth *et al.*, 1998; Worth & Fentem, 1999). This approach to the assessment of tiered testing strategies had not been published previously. The main conclusion from the simulations is that the OECD approach to hazard identification provides a reliable means of reducing and refining animal testing, without compromising our ability to

identify toxic chemicals. The original work on which chapter 10 is based was communicated to the OECD Secretariat at a time when the (then proposed) OECD testing strategies for skin and eye irritation/corrosion were being discussed.

11.3 PERSPECTIVES ON THE INTEGRATED APPROACH TO TOXICITY TESTING

11.3.1 A unifying view of alternative methods, models and testing strategies

When reviewing the literature in the field of this thesis, the author noted that different terms for (mathematical and statistical) models tend to be used, depending on the subject area (structure-activity relationship chemistry or *in vitro* toxicology). The emerging field of 'integrated toxicity testing' draws on both of these traditional subject areas, having imported the concepts of structure-activity relationship (SAR) and prediction model (PM), and introduced the concept of integrated testing strategy. Distinctions should be drawn between three types of model for predicting chemical toxicity, according to the nature of the predictor variable(s) in the model: a) SARs, based on physicochemical properties that can be calculated; b) PMs, based on physicochemical properties or *in vitro* endpoints that need to be determined experimentally; and c) integrated models (IMs), based on both physicochemical properties and *in vitro* endpoints. The distinction between SARs and PMs / IMs reflects the relative ease with which the predictor variables can be obtained. These models can be classification models or regression models (see chapter 4), depending on the nature of the response variable. Biokinetic (BK) and biodynamic (BD) models are treated as types of IM. BK models predict the time-dependent and tissue-dependent concentrations of chemicals resulting from the physiological processes of absorption, distribution, metabolism and excretion. In effect, BK models generate the internal dose levels to which target tissues and organs are exposed. BD models go one step further, by applying dose-response models to the target tissue/organ concentration to predict the level of response.

The domain (range of application) of a model for predicting chemical toxicity is often restricted to one or more classes of chemicals, defined according to their structural and/or physicochemical properties. In contrast, an integrated testing strategy can be viewed as a more comprehensive decision-making tool, constructed from a range of

SARs, PMs, IMs, in such a way that predictions of an *in vivo* toxicological effect can be made for a broad range of chemical types with the greatest accuracy, efficiency, and animal welfare benefits. This description of the integrated testing strategy emphasises the fact that predictions are made by combining mathematical models (rather than computer-based and experimental methods) that extrapolate from the molecular and cellular levels to the level of the whole organism. The toxicological effect being predicted (e.g. skin irritation) could be represented by one or more of the endpoints (e.g. erythema and œdema) observed in an animal experiment, or by one or more clinical symptoms observed in humans.

At present, integrated testing strategies are becoming accepted at the regulatory level as a means of reducing our reliance on existing animal procedures for hazard assessment, but not as a means of replacing them. Ultimately, there may be sufficient confidence in the predictions made by such strategies that the animal testing steps (used for confirmatory testing) could be omitted. Relevant to such a decision is the ongoing debate on the extent to which traditional animal models are relevant to the assessment of human health effects, and therefore whether the alternative models currently being designed to predict animal data will ultimately be of any use. However, if alternative methods are developed with a sound mechanistic basis, their usefulness may well outlive the use of current animal methods. This will almost certainly require the mathematical models associated with the alternative methods to be reparameterised, or even completely reformulated (in terms of their mathematical structure), but the predictor variables themselves should continue to be relevant.

It is important to note that the integrated testing strategies discussed in this thesis are designed for hazard assessment, not risk assessment. Toxicological hazard refers to the inherent ability of a substance to cause an adverse effect in the body, whereas toxicological risk refers to the extent or incidence of an adverse effect under defined conditions of exposure (typically the exposure conditions of a societal group, such as workers or consumers). Hazard is often quantified by generating a dose-response curve, which is then used to determine the concentration (or dose) of test chemical required to produce a fixed response; for example, the LD_{50} value is the dose required to cause 50% mortality in an experimental group of animals subjected to a specified test protocol. In the EU and OECD hazard classification systems, chemicals are ranked according to

their LD_{50} values for certain toxicological endpoints (e.g. acute toxicity), whereas for other endpoints (e.g. skin and eye irritation/corrosion), a fixed dose procedure is adopted, and ranking is based on the severity or onset time of the response (recovery may also be taken into account). Thus, hazard assessment may or may not involve the generation of a dose-response curve. In contrast, risk assessment necessarily involves the generation (or mathematical simulation) of a dose-response (or exposure-incidence) curve, so that the extent or incidence of an adverse effect can be related to exposure. As a final point on this matter, it should be remembered that hazard classification schemes reflect the consensus of scientific expert groups, and may be altered in the light of scientific and technological progress, or in the interests of international harmonisation. Thus, any model that is developed for hazard classification will need to be reparameterised in the event of changes to the relevant classification system.

11.3.2 How should the performance of integrated testing strategies be assessed ?

To date, physicochemical and *in vitro* methods have been validated as stand-alone methods for predicting the results of specific animal experiments, but not as component parts in an integrated testing strategy. In the few regulations and guidelines for chemicals testing where integrated approaches are allowed for (e.g. OECD, 1998), it is clearly stated that any alternative method incorporated should have been validated. However, if the validation status of a method continues to depend on its stand-alone merits, some methods may never be accepted as components in a testing strategy. For example, an SAR based on pKa values, or a PM based on pH measurements, may not be capable of identifying a majority of corrosive chemicals, but if a high degree of confidence is associated with the corrosive predictions that it makes, then such a model can still play a useful role in a tiered testing strategy. It follows that different acceptance criteria should be applied during the validation of alternative methods, depending on whether they are intended as stand-alone replacements for an animal test, or as components of an integrated strategy. It is impossible to generalise as to what the acceptance criteria should be for any strategy, since this will depend on the nature of the toxicity testing (which varies from the labelling of industrial chemicals to the safety assessment of pharmaceuticals). However, a minimal requirement is that all the experimental data used should be reproducible (i.e. generated by reliable protocols). In

addition, it is recommended that the criteria for predictive performance should be set at realistic levels, taking into account the best possible predictivity that can be expected on theoretical grounds.

Two possible approaches can be foreseen for the formal assessment of integrated testing strategies. Firstly, they could be assessed by the usual validation process. In this case, a successfully validated strategy would be one that had been fully stipulated (in terms of its basic design and component models) in advance of a validation study, and then subjected to an independent assessment in accordance with pre-defined acceptance criteria. Secondly, integrated strategies could be designed and assessed by using existing data, including data generated in previous validation studies of individual test methods. In practice, the second approach is more likely to be adopted, since testing strategies are, by their very nature, designed to accommodate the strengths and weaknesses of individual methods, so it is logical that the component parts of testing strategies should be assessed before the strategies themselves. If the second approach is adopted, a distinction should be drawn between strategies that have been *evaluated* (using data that were also used in their design), and those strategies that have been *validated* (using independent data).

11.3.3 The reductionist nature of alternative methods and models

The models and strategies discussed in this thesis are all based on the reductionist approach; that is, they aim to extrapolate from lower levels of biological organisation to the level of the whole organism. The philosophical doctrine of reductionism is so deeply engrained in scientific culture, and has formed the basis of so many successful theories, that it would be perverse to question its value. However, one could reasonably question whether reductionist models are sufficient to reproduce the toxicological effects observed in whole organisms. In recent years, it has become apparent that the properties of biological systems cannot always be explained in terms of the properties of their component parts alone. Biological systems are therefore said to be ‘complex systems’, because they exhibit effects that only emerge when many subprocesses interact. For example, the glycolytic cycle (in which the conversion of glucose to pyruvate is coupled with the production of the adenosine triphosphate and the reduced form of nicotinamide adenine dinucleotide) involves 10 enzymes, whose interactions (by feedback inhibition)

can cause oscillations in the concentrations of the glycolytic intermediates (Whitesides & Ismagilov, 1999). While the complexity of the glycolytic cycle has been modelled with a certain amount of success, it is not yet possible to simulate the behaviour of other biological systems, such as intracellular signalling networks that mediate the effects of extracellular chemicals (Weng *et al.*, 1999). The inability to model such systems is partly due to an incomplete knowledge of their component parts, their functions and interconnections, and partly to an incomplete knowledge of the laws of complexity. Indeed, it is not even clear whether there *are* any general laws of complexity (Goldenfeld & Kadanoff, 1999). An important feature of complex systems is that in addition to producing regular patterns of behaviour, they can also produce chaotic patterns, under the appropriate conditions. Chaos is the sensitive dependence of a final result upon the initial conditions that bring it about. Due to the inescapable uncertainty in these initial conditions, the onset of chaos effectively means that the behaviour of a system rapidly becomes unpredictable, since the uncertainties grow exponentially with time.

Given the complexity of biological systems, and our limited ability to model them, the question arises as to what extent the reductionist approach should be applied. In other words, how far up the hierarchy of biological organisation is it reasonable to extrapolate? Goldenfeld & Kadanoff (1999) advise the modeller to choose the 'correct' level of detail to catch the phenomenon of interest, and warn against the modelling of 'bulldozers with quarks'. So, is it reasonable to model the toxicological responses observed in living organisms from the properties of molecules? The responses modelled in this project were all acute local responses in the tissues of the skin and eye, and the ability to model these by using SARs can be related to the fact that the rate-limiting step in the production of the toxicological responses is the partitioning of the chemical into the target tissue. Thus, the author's view is that SARs can be used to extrapolate from physicochemical properties to biological responses to the extent that these responses are 'driven' by physical effects, such as partitioning. It seems less likely that SARs alone will successfully reproduce the various manifestations of systemic toxicity, especially when these are chronic effects, possibly depending on multiple exposures. For such effects, the additional use of *in vitro* models and the integration of *in vitro* data into biokinetic models should be of great value. This is not to say that SARs will have no role to play in such models, but only that their extrapolations will be less extensive.

They could be used, for example, to estimate the various biokinetic parameters that need to be incorporated into the larger mathematical frameworks of the biokinetic models. In addition, more emphasis could be placed on the use of SARs to model *in vitro* endpoints for defined subsets of chemicals.

11.4 CURRENT RESEARCH NEEDS

Strictly speaking, 'integrated toxicity testing' is not so much a new field of toxicology as a new approach to hazard and risk assessment. The need for this approach is widely acknowledged in the toxicological literature, but often little more than lip service is paid to it. A great deal of research is still required to form the scientific and technological basis for the design of detailed integrated strategies capable of predicting well-defined *in vivo* endpoints. In this section, a number of examples are given.

11.4.1 Structure-activity relationship modelling

In the field of SAR modelling, there is scope for the identification of descriptor variables that capture the essential features of specific biochemical processes. For example, one can envisage the development of new, and identification of existing, molecular connectivity indices suitable for identifying the substrates of specific proteins, such as membrane receptors, ion channels and enzymes. Topological descriptors (see chapter 3) have the advantage that they are non-empirical and easy to calculate. At present, a wide variety of connectivity descriptors have already been developed, so there is considerable scope for their further interpretation in the context of specific biochemical processes. Another interesting and potentially fruitful area of descriptor development concerns the use of information from infra-red spectra, since these serve as highly specific molecular 'fingerprints' (Benigni *et al*., 1999).

11.4.2 *In vitro* toxicology

In the field of *in vitro* toxicology, there is an ongoing need to develop and optimise protocols based on endpoints relevant to toxicological processes. For the prediction of skin and eye irritation, a wide range of systems have already been developed, so perhaps in these areas, the challenge is not so much new development, but the retrospective analysis of existing data and the refinement of existing systems. Similarly, for the

prediction of acute systemic toxicity, it is noteworthy that a recently completed multi-centre study reported that the lethal concentrations of chemicals in (human) blood can be predicted from basal cytotoxicity measurements, regardless of the endpoint or cell type used (Clemedson *et al*., 2000). Given that a wealth of *in vitro* cytotoxicity data exists, the successful development of an integrated testing strategy for acute systemic toxicity appears to be a realistic, short-term goal. If such a strategy could successfully predict LC_{50} values (for the inhalation exposure route) and LD_{50} values (for the oral and dermal exposure routes), there would be a chance to significantly reduce animal testing in this area.

Another field of toxicology that is ripe for the design and evaluation of integrated testing strategies is neurotoxicology. A wide variety of neurospecific endpoints have been developed (reviewed in Pentreath, 1999), so the details of an integrated approach for neurotoxicity testing could now be formulated. An encouraging development in this field is the demonstration that a biokinetic model could successfully extrapolate from critical neurotoxic concentrations *in vitro* to lowest effective doses *in vivo* (de Jongh *et al*., 1999). Thus, there is evidence that the integrated approach to hazard assessment, already accepted as a means of local toxicity testing, can also be applied to systemic toxicity testing.

The prediction of other types of toxic effect represents a greater challenge. An example is embryotoxicity, which covers a diverse range of physiological and anatomical aberrations. In this case, it is important to distinguish between maternal toxicity and embryonic/fœtal toxicity, and to consider factors such as metabolism and placental transfer. An area of current research, in which the author is involved, is aimed at the development of novel *in vitro* endpoints for embryotoxic potential using embryonic stem (ES) cells (Bremer *et al*., 1999). This work has shown that it is possible to identify strongly embryotoxic substances (i.e. chemicals that adversely affect most, if not all, of the developing tissues) by using simple endpoints, such as cytotoxicity, or gene-expression endpoints. However, further work is required to identify substances that exert their embryotoxic effects by tissue-specific mechanisms.

11.4.3 Modelling methods

There is also a need for the further development and investigation of modelling methodologies. For example, most SARs have been developed for predicting the efficacies and toxicities of single chemical entities, and yet many substances of interest in toxicology, such as cosmetics and pharmaceuticals, are based on mixtures (formulations) of chemicals. Recently, Patel *et al.* (1999) reported the development of a method called 'quantitative component analysis of mixtures' (QCAM), which enables the toxicity profile of each component in a mixture to be extracted from knowledge of the relative concentrations of all components and the net toxicity of the entire mixture. Thus, QCAM models could provide a means of predicting the toxicities of new formulations (based on the same components). A slightly different challenge is posed by complex mixtures with undefined compositions, such as petroleum substances. To model this kind of mixture, Verburgh *et al.* (1996) developed a 'blocking method' in which components with similar physicochemical properties are treated together (as a block), and the toxicity of the mixture is calculated as a weighted average of the toxicities of the blocks.

Another approach that merits further investigation is the combined use of traditional statistical methods (see chapter 4) and artificial neural networks (ANNs; Bishop, 1995). A common type of ANN is composed of a layered structure of 'neurons': an input layer, in which the descriptor data are represented; one or more hidden layers with a variable number of neurons; and an output layer, which is trained to match a target set of data representing the biological response. Each neuron in the hidden and output layers is connected to the neurons in the input and hidden layers, respectively. The direction and magnitude of the signal sent from one neuron to another (the connection weight) is a variable, and the sum total of all inputs to a neuron determines whether or not it will 'fire' and therefore send outputs to neurons in the next layer. Training of the network involves the iterative setting of the connection weights, so that a given input (set of descriptor data) generates a desired output (response value). This type of ANN is called a feed-forward back-propagation network, because all connections are made in the forward direction (from the input layer to the output layer) and because the connection weights are set by a back-propagation algorithm. In terms of predictive performance, ANNs can outperform traditional statistical methods, without overfitting the training set

of data (Manallack & Livingstone, 1999). However, the model generated by an ANN is characterised in terms of its network architecture and its entire set of connection weights. Thus, ANN models are much less interpretable than traditional statistical models that take the form of explicit algorithms for converting descriptor data into a predicted response. One way of combining the strengths of ANNs and traditional statistical methods would be use the ANN as a variable selection method, to identify the best variables without having to specify the mathematical form of the model, and then to use the traditional methods to develop a simple and explicit statistical model.

11.5 FUTURE PROSPECTS

In the short term, one could expect the development of integrated toxicity testing to proceed mainly in terms of new and refined strategies for hazard assessment, with strategic approaches for risk assessment being developed later. Furthermore, as the scientific knowledge base expands, one could expect that current strategies based largely on empirical models will gradually give way to strategies based on mathematical models with a stronger theoretical basis. At that stage, the SARs and PMs of today will seem relatively crude, but they will have served a valuable temporary purpose, forming the scaffolds on which a more sophisticated predictive framework will ultimately be based.

The greatest impact on the field of predictive toxicology is likely to come from the emerging fields of genomics and proteomics. Genomics is the study of the entire genetic make-up of an organism, whereas proteomics is the study of the entire protein complement of a cell or tissue at a particular stage in its development. Recent technological advances in these fields have made it possible to simultaneously measure the expression levels of thousands of different genes. It is therefore possible to predict the effects of chemicals on multiple gene expression, thereby integrating information at the genetic level. At the time of writing (June 2000), scientists working for the Human Genome Project reported that they had completed the sequencing of the entire human genome (Macilwain, 2000). Following this milestone achievement in biology, attention will focus on the interpretation of DNA sequences in terms of functional proteins and RNA molecules. Ultimately, this will create the entire knowledge base required to model biochemical systems, such as the intracellular pathways that mediate the effects

of extracellular signals, and the extracellular pathways that maintain homeostasis in the entire organism. Hopefully, advances in the study of complexity will provide the mathematical basis required to integrate this knowledge base in the form of complex models of living organisms. These models should not only reproduce the homeostatic state of an individual, but also predict the perturbations of that state, caused by the combined effects of environmental chemicals, pathogens and the genetic constitution, that mark the transition from normal physiology to pathophysiology. As argued by Spence & Aurora (1999), the full value of the human genome will only be exploited if we integrate. And if we ever reach that point in biology, there will be no place for animal testing in toxicology.

CHAPTER 12
REFERENCES

Abraham, M.H. (1993). Scales of solute hydrogen-bonding - their construction and application to physicochemical and biochemical processes. *Chemical Society Reviews* **22**, 73-83.

Abraham, M.H., Chadha, H.S. & Mitchell, R.C. (1995). The factors that influence skin penetration of solutes. *Journal of Pharmacy and Pharmacology* **47**, 8-16.

Abraham, M.H., Kumarsingh, R., Cometto-Muniz, J.E. & Cain, W.S (1998a). A quantitative structure-activity relationship (QSAR) for a Draize eye irritation database. *Toxicology in Vitro* **12**, 201-207.

Abraham, M.H., Kumarsingh, R., Cometto-Muniz, J.E. & Cain, W.S (1998b). Draize eye scores and eye irritation thresholds in man combined into one quantitative structure-activity relationship. *Toxicology in Vitro* **12**, 403-408.

Alberts, A.L. & Handy, N.C. (1988). Møller-Plesset third order calculations with large basis sets. *Journal of Chemical Physics* **89**, 2107-2115.

Allinger, N.L. (1977). Conformational analysis. 130. MM2. A hydrocarbon force field utilising V1 and V2 torsional terms. *Journal of the American Chemical Society* **99**, 8127-8134.

Balls M., Berg N., Bruner L.H., Curren R.D., de Silva O., Earl L.K., Esdaile D.J., Fentem J.H., Liebsch M., Ohno Y., Prinsen M.K., Spielmann H. & Worth A.P. (1999). Eye irritation testing: the way forward. The report and recommendations of ECVAM workshop 34. *Alternatives to Laboratory Animals* **27**, 53-77.

Balls, M., Botham, P.A., Bruner, L.H. & Spielmann, H. (1995). The EC/HO international validation study on alternatives to the Draize eye irritation test. *Toxicology in Vitro* **9**, 871-929.

Balls, M. & Karcher, W. (1995) The validation of alternative test methods. *Alternatives to Laboratory Animals* **23**, 884-886.

Barcikowski, R. & Stevens, J. P. (1975). A Monte Carlo study of the stability of canonical correlations, canonical weights, and canonical variate-variable correlations. *Multivariate Behavioral Research* **10**, 353-364.

Barlow, A., Hirst, R.A., Pemberton, M.A., Rigden, A., Hall, T.J., Botham, P.A. & Oliver, G.J.A. (1991). Refinement of an *in vitro* test for the identification of skin corrosive chemicals. *Toxicology Methods* **1**, 106-115.

Barratt, M.D. (1995a). Quantitative structure-activity relationships for skin permeability. *Toxicology in Vitro* **9**, 27-37.

Barratt, M.D. (1995b). Quantitative structure-activity relationships for skin corrosivity. Appendix A to the report of ECVAM Workshop 6. *Alternatives to Laboratory Animals* **23**, 219-255.

Barratt, M.D. (1996a). Quantitative structure-activity relationships for skin irritation and corrosivity of neutral and electrophilic organic chemicals. *Toxicology in Vitro* **10**, 247-256.

Barratt, M.D. (1996b). Quantitative structure-activity relationships (QSARs) for skin corrosivity of organic acids, bases and phenols: principal components and neural network analysis of extended datasets. *Toxicology in Vitro* **10**, 85-94.

Barratt, M.D. (1997). QSARs for the eye irritation potential of organic chemicals. *Toxicology in Vitro* **11**, 1-8.

Barratt, M.D., Brantom, P.G., Fentem, J.H., Gerner, I., Walker, A.P. & Worth, A.P. (1998). The ECVAM international validation study on *in vitro* tests for skin corrosivity. 1. Selection and distribution of the test chemicals. *Toxicology in Vitro* **12**, 471-482.

Barratt, M.D., Castell, J.V., Chamberlain, M., Combes, R.D., Dearden, J.C., Fentem, J.H., Gerner, I., Giuliani, A., Gray, T.J.B., Livingstone, D.J., Provan, W.M., Rutten, F.A.J.J.L., Verhaar, H.J.M. & Zbinden, P. (1995). The integrated use of alternative approaches for predicting toxic hazard. The report and recommendations of ECVAM workshop 8. *Alternatives to Laboratory Animals* **23**, 410-429.

Barratt, M.D. & Chamberlain, M. (1997). Integration of QSAR and *in vitro* toxicology. In *In Vitro Methods in Pharmaceutical Research* (ed. J.V. Castell & M.J. Gómez-Lechón). pp. 16-31. Academic Press, New York.

Barratt, M.D., Dixit, M.B., Jones, P.A. (1996). The use of *in vitro* cytotoxicity measurements in QSAR methods for the prediction of the skin corrosivity potential of acids. *Toxicology in Vitro* **10**, 283-290.

Basketter, D.A., Scholes, E.W., Chamberlain, M. & Barratt, M.D. (1995). An alternative strategy to the use of guinea pigs for the identification of skin sensitization hazard. *Food & Chemical Toxicology* **33**, 1051-1057.

Basketter, D.A., Whittle, E. & Chamberlain, M. (1994). Identification of irritation and corrosion hazards to skin: an alternative strategy to animal testing. *Food and Chemical Toxicology* **32**, 539-542.

Bell, E., Parenteau, N., Gay, R., Nolte, C., Kemp, P., Ekstein B. & Johnson, E. (1991). The living skin equivalent: its manufacture, its organotypic properties and its response to irritants. *Toxicology In Vitro* **5**, 591-596.

Benigni, R., Passerini, L., Livingstone, D.J., Johnson, M.A. & Giuliani, A. (1999). Infrared spectra information and their correlation with QSAR descriptors. *Journal of Chemical Information and Computer Sciences* **39**, 558-562.

Berner, B., Wilson, D.R., Guy, R.H., Mazzenga, C., Clarke, F.H. & Maibach, H.I. (1988). The relationship of pKa and acute skin irritation in man. *Pharmaceutical Research* **5**, 660-663.

Bingham, R.C., Dewar, M.J.S. & Lo, D.H. (1975). Ground states of molecules. XXV. MINDO/3. An improved version of the MINDO semiempirical SCF-MO method. *Journal of the American Chemical Society* **97**, 1285-1293.

Binkley, J.S., Pople, J.A. & Hehre, W.J. (1980). Self-consistent molecular orbital methods. 21. Small split-valence basis sets for first row elements. *Journal of the American Chemical Society* **102**, 939-947.

Bishop, C. (1995). *Neural Networks for Pattern Recognition*. Oxford University Press, Oxford.

Blaauboer, B.J., Barratt, M.D. & Houston, J.B. (1999). The Integrated Use of Alternative Methods in Toxicological Risk Evaluation. ECVAM Integrated Testing Strategies Task Force Report 1. *Alternatives to Laboratory Animals* **27**, 229-237.

Blein, O., Adolphe, M., Lakhdar, B., Cambar, J., Gubanski, G., Castelli, D., Contie, C., Hubert, F., Latrille, F., Masson, P., Clouzeau, J., Le Bigot, J.F., De Silva, O. & Dossou, K.G. (1991). Correlation and validation of alternative methods to the Draize eye irritation test (OPAL project). *Toxicology in Vitro* **5**, 555-557.

Bodor, N., Gabanyi, Z. & Wong, C.K. (1989). A new method for the estimation of partition coefficients. *Journal of the American Chemical Society* **111**, 3783-3786.

Botham, P.A., Chamberlain, M., Barratt, M.D., Curren, R.D., Esdaile, D.J., Gardner, J.R., Gordon, V.C., Hildebrand, B., Lewis, R.W., Liebsch, M., Logemann, P., Osborne, R., Ponec, M., Régnier, J-F., Steiling, W., Walker, A.P. & Balls, M. (1995). A prevalidation study on *in vitro* skin corrosivity testing. The report and recommendations of ECVAM Workshop 6. *Alternatives to Laboratory Animals* **23**, 219-255.

Botham, P.A., Earl, L.K., Fentem, J.H., Roguet, R. & van de Sandt, J.J.M. (1998). Alternative methods for skin irritation: the current status. ECVAM skin irritation task force report 1. *Alternatives to Laboratory Animals* **26**, 195-211.

Boyce, S., Michel., S., Reichert, U., Shroot, B. & Schmidt, R. (1990). Reconstructed skin from cultured human keratinocytes and fibroblasts on a collagen-glycosaminoglycan biopolymer substrate. *Skin Pharmacology* **2**, 136-143.

Brantom, P.G., Bruner, L.H., Chamberlain, M., De Silva, O., Dupuis, J., Earl, L.K., Lovell, D.P., Pape, W.J.W., Uttley, M., Bagley, D.M., Baker, F., Bracher, M., Courtellemont, P., Declercq, L., Freeman, S., Steiling, W., Walker, A.P., Carr, G.J., Dami, N. & Thomas, G. (1997). A summary report of the COLIPA international validation study on alternatives to the Draize rabbit eye irritation test. *Toxicology in Vitro* **11**, 141-179.

Breiman, L., Friedman, J. H., Olshen, R. A. & Stone, C. J. (1984). *Classification and regression trees.* Wadsworth, Monterey, CA.

Bremer, S., Worth, A.P., Paparella, M., Hescheler, J., Fleischmann, B.K. & Kolossov, E. (1999). Establishment of an embryotoxic screening system using genetically engineered embryonic stem cells. *Alternatives to Laboratory Animals* **27** (Special Issue, July 1999), 292.

Bronaugh, R.L & Barton, C.N. (1993). Prediction of human percutaneous absorption with physicochemical data. In *Health Risk Assessment. Dermal and Inhalation Exposure and Absorption of Toxicants* (ed. R.G.M Wang, J.B. Knaak & H.I. Maibach). pp 117-126. CRC Press, Boca Raton, FL.

Bronaugh, R.L. & Maibach, H.I. (1985). Percutaneous absorption of nitroaromatic compounds: *in vivo* and *in vitro* studies in human and monkey skin. *Journal of Investigative Dermatology* **84**, 180-183.

Bronaugh, R.L. & Stewart, R.F. (1984). Methods for *in vitro* percutaneous absorption studies. III. Hydrophobic compounds. *Journal of Pharmaceutical Sciences* **73**, 1255-1259.

Bronaugh, R.L., Stewart, R.F. & Congdon, E.R. (1982). Methods for *in vitro* absorption studies. II. Animal models for human skin. *Toxicology and Applied Pharmacology* **62**, 481-488.

Bruner, L.H., Carr, G.J., Chamberlain, M. & Curren, R.D. (1996). Validation of alternative methods for toxicity testing. *Toxicology in Vitro* **10**, 479-501.

Buchwald, P. & Bodor, N. (1998). Octanol-water partition: searching for predictive models. *Current Medicinal Chemistry* **5**, 353-380.

Cannon, C.L., Neal, P.J., Southee, J.A., Kubilus, J. & Klausner, M. (1994). New epidermal model for dermal irritancy testing. *Toxicology in Vitro* **8**, 889-891.

Cantor, R.S. (1997). The lateral pressure profile in membranes: a physical mechanism of general anesthesia. *Biochemistry* **36**, 2339-2344.

Caulcutt, R. (1991). *Statistics in Research and Development*. Chapman and Hall, London.

Charton, M. & Charton, B.I. (1982). The structural dependence of amino acid hydrophobicity parameters. *Journal of Theoretical Biology* **99**, 629-644.

Clemedson, C., Barile, F.A., Chesné, C. *et al*. (2000). MEIC evaluation of acute systemic toxicity. Part VII. Prediction of human toxicity by results from testing of the first 30 reference chemicals with 27 further *in vitro* assays. *Alternatives to Laboratory Animals* **28**, Supplement 1, 159-200.

Cooper, J.A., Saracci., R. & Cole, P. (1979). Describing the validity of carcinogen screening tests. *British Journal of Cancer* **39**, 87-89.

Coquette, A., Berna, N., Vandenbosch, A., Rosdy, M. & Poumay, Y. (1999). Differential expression and release of cytokines by an *in vitro* reconstructed human epidermis following exposure to skin irritant and sensitizing chemicals. *Toxicology in Vitro* **13**, 867-877.

Corsini, E. & Galli, C.L. (2000). Epidermal cytokines in experimental contact dermatitis. *Toxicology* **142**, 203-211.

Cramer, C.J., Famini, G.R & Lowrey, A.H. (1993). Use of calculated quantum chemical properties as surrogates for solvatochromic parameters in structure-activity relationships. *Accounts of Chemical Research* **26**, 599-605.

Cronin, M.T.D. (1996). The use of cluster significance analysis to identify asymmetric QSAR data sets in toxicology. An example with eye irritation data. *SAR and QSAR in Environmental Research* **5**, 167-175.

Cronin, M.T.D., Basketter, D.A. & York, M. (1994). A quantitative structure-activity relationship (QSAR) investigation of a Draize eye irritation database. *Toxicology in Vitro* **8**, 21-28.

Cronin, M.T.D., Dearden, J.C., Gupta, R. & Moss, G.P. (1998). An investigation of the mechanism of flux across polydimethylsiloxane membranes by use of quantitative structure-permeability relationships. *Journal of Pharmacy and Pharmacology* **50**, 143-152.

Cronin, M.T.D., Dearden, J.C., Moss, G.P. & Murray-Dickson, G. (1999). Investigation of the mechanism of flux across human skin *in vitro* by quantitative structure-permeability relationships. *European Journal of Pharmaceutical Sciences* **7**, 325-330.

Cullander, C. & Guy, R.H. (1991). Visualising the pathways of iontophoretic current flow in real time with laser scanning confocal microscopy and the vibrating probe electrode. Percutaneous penetration enhancers: mode of action. In *Prediction of Percutaneous Penetration*, Volume 2 (ed. R.C. Scott, R.H. Guy, J. Hadgraft & H.E. Boddé). pp. 229-237. IBC Technical Services, London.

Czismadia, F., Tsantili-Kakoulidou, A., Panderi, I. & Darvas, F. (1997). Prediction of distribution coefficient from structure. 1. Estimation method. *Journal of Pharmaceutical Sciences* **86**, 865-869.

Dearden, J.C. (1990). Physicochemical descriptors. In *Practical Applications of QSAR in Environmental Chemistry and Toxicology* (eds. W. Karcher & J. Devillers). pp. 25-29. Kluwer, Dordrecht.

Devillers, J. & Karcher, W. (1990). Correspondence factor analysis as a tool in environmental SAR and QSAR studies. In *Practical Applications of QSAR in Environmental Chemistry and Toxicology* (eds. W. Karcher & J. Devillers). pp. 181-195. Kluwer, Dordrecht.

Dewar, M.J.S. & Thiel, W. (1977). Ground states of molecules. 38. The MNDO method. Approximation and parameters. *Journal of the American Chemical Society* **99**, 4899-4907.

Dewar, M.J.S., Zoebisch, E.G., Healy, E.F. & Stewart, J.J.P. (1985). The development and use of quantum-mechanical models. 76. AM1 - A new general-purpose quantum-mechanical molecular model. *Journal of the American Chemical Society* **107**, 3902-3909.

De Wever, B. & Rheins, L.A. (1994). Skin^{2TM}: an *in vitro* human skin analogue. In *In Vitro Skin Toxicology - Irritation, Phototoxicity, Sensitization.* Alternative Methods in Toxicology, Volume 10 (ed. A. Rougier, A.M. Goldberg & H.I. Maibach). pp. 121-131. Mary Ann Liebert, New York.

Draize, J.H., Woodard, G. & Calvery, H.O. (1944). Methods for study of irritation and toxicity of substances applied topically to the skin and mucous membranes. *Journal of Pharmacology and Experimental Therapeutics* **82**, 377-390.

Earl, L.K., Dickens, A.D. & Rowson, M.J. (1997). A critical analysis of the rabbit eye irritation test variability and its impact on the validation of alternative methods. *Toxicology in Vitro* **11**, 295-304.

EC (1983). Guide to the classification and labelling of dangerous substances and preparations. Criteria for the choice of chemicals indicating special risks (R phrases)

and safety advice (S phrases). *Official Journal of the European Communities* **L257**, 13-20.

EC (1986). Council Directive *86/609/EEC* of 24 November 1986 on the Approximation of Laws, Regulations and Administrative Provisions of the Member States Regarding the Protection of Animals used for Experimental and other Scientific Purposes. *Official Journal of the European Communities* **L358**, 1-29.

EC (1992). Council Directive *93/69/EEC* of 31 July 1992 adapting to technical progress for the 17th time Council Directive *67/548/EEC* on the approximation of laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances. *Official Journal of the European Communities* **L383**, 113-114.

EC (1993) Commission Directive *93/21/EEC* of 27 April 1993 adapting to technical progress for the 18th time Council Directive *67/548/EEC* on the approximation of laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances. *Official Journal of the European Communities* **L110A**, 1-86.

EC (1998). Non-clinical local tolerance testing of medicinal products. In *The Rules Governing Medicinal Products in the European Union, Volume 3b.* pp. 109-116. European Community, Luxembourg.

EC (2000). Commission Directive *2000/33/EC* of 25 April 2000 adapting to technical progress for the 27th time Council Directive *67/548/EEC* on the approximation of laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances. *Official Journal of the European Communities* **L136A**, 90-97.

ECETOC (1993). *Percutaneous Absorption*. ECETOC Monograph No. 20. European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels.

ECETOC (1995). *Skin Irritation and Corrosion: Reference Chemicals Data Bank.* ECETOC Technical Report No. 66. European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels.

ECETOC (1998a). *Eye Irritation: Reference Chemicals Data Bank*. Second Edition. ECETOC Technical Report No. 48. European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels.

ECETOC (1998b). *QSARs in the Assessment of the Environmental Fate and Effects of Chemicals.* ECETOC Technical Report No. 74. European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels.

Efron, B. & Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall, London.

Emmett, E.A. (1986). Toxic responses of the skin. In *Casarett and Doull's Toxicology* (eds. C.D. Klaassen, M.O. Amdur & J. Doull). pp. 412-431. Macmillan, New York.

Fentem, J.H., Archer, G.E.B., Balls, M., Botham, P.A., Curren, R.D., Earl, L.K., Esdaile, D.J., Holzhütter, H.G. & Liebsch, M. (1998) The ECVAM international validation study on *in vitro* tests for skin corrosivity. 2. Results and evaluation by the management team. *Toxicology in Vitro* **12**, 483-524.

Fentem, J.H. & Balls, M. (1997). The ECVAM approach to validation. In *Animal Alternatives, Welfare and Ethics* (ed. L.F.M. van Zutphen & M. Balls). pp. 1083-1089. Elsevier, Amsterdam.

Flynn, G.L. (1990). Physicochemical determinants of skin absorption. In *Principles of Route-to-Route Extrapolation for Risk Assessment* (eds. T.R. Gerrity & C.J. Henry), pp. 93-127. Elsevier, New York.

Fujita, T., Iwasa, J. & Hansch, C. (1964). A new substituent constant, derived from partition coefficients. *Journal of the American Chemical Society* **86**, 5175-5180.

Fujita, T., Nishioka, T. & Nakajima, M. (1977). Hydrogen-bonding parameter and its significance in quantitative structure-activity studies. *Journal of Medicinal Chemistry* **20**, 1071-1081.

Gamache, D.A., Dimitrijevich, S.D., Weimer, L.K., Lang, L.S., Spellman, J.M., Graff, G. & Yanni, J.M. (1997). Secretion of proinflammatory cytokines by human conjunctival epithelial cells. *Ocular Immunology & Inflammation* **5**, 117-128.

Geladi, P. & Kowalski, B.R. (1986). Partial least-squares regression: a tutorial. *Analytica Chimica Acta* **185**, 1-17.

Gerner, I., Graetschel, G., Kahl, J & Schlede, E. (2000). Development of a decision support system for the introduction of alternative methods into local irritancy/corrosivity testing strategies. Development of a relational database. *Alternatives to Laboratory Animals* **28**, 11-28.

Goldenfeld, N. & Kadanoff, L.P. (1999). Simple lessons from complexity. *Science* **284**, 87-89.

Gordon, V.C., Harvell, J.D. & Maibach, H.I. (1994). Dermal corrosion, the CORROSITEX^{TM} system: a DOT accepted method to predict corrosivity potential of test materials. In *In Vitro Skin Toxicology - Irritation, Phototoxicity, Sensitization.* Alternative Methods in Toxicology, Volume 10 (eds. A. Rougier, A.M. Goldberg & H.I. Maibach). pp. 37-45. Mary Ann Liebert, New York.

Goskonda, V.R., Khan, M.A., Hutak, C.M. & Reddy, I.K. (1999). Permeability characteristics of novel mydriatic agents using an *in vitro* culture model that utilizes SIRC rabbit corneal cells. *Journal of Pharmaceutical Sciences* **88**, 180-184.

Grass, G.M., Cooper, E.R. & Robinson, J.R. (1988). Mechanisms of corneal penetration III: modeling of molecular transport. *Journal of Pharmaceutical Sciences* **77**, 24-26.

Greenacre, M. J. (1984). *Theory and Applications of Correspondence Analysis.* Academic Press, New York.

Greenacre, M. J. (1993). *Correspondence Analysis in Practice*. Academic Press, New York.

Griffith, M., Osborne, R., Munger, R., Xiong, X., Doillon, C.J., Laycock, N.L.C., Hakim, M., Song, Y. & Watsky, M.A. (1999). Functional human corneal equivalents constructed from cell lines. *Science* **286**, 2169-2172.

Guy, R.H., Mak, V.H.W, Kai, T., Bommannan, D. & Potts, R.O. (1990). Percutaneous penetration enhancers: mode of action. In *Prediction of Percutaneous Penetration*, Volume 1 (eds. R.C. Scott, R.H. Guy & J. Hadgraft). pp. 213-223. IBC Technical Services, London.

Hall, L.H. & Kier, L.B. (1991). The molecular connectivity chi indices and kappa shape indexes in structure-property modeling. In *Reviews of Computational Chemistry, Volume 2* (eds. D.B. Boyd & K. Lipkowitz.). pp. 367-422. VCH Publishers, New York.

Hamalainen, K.M., Kontturi, K., Auriola, S., Murtomaki, L. & Urtti, A. (1997a). Estimation of pore size and pore density of biomembranes from permeability measurements of polyethylene glycols using an effusion-like approach. *Journal of Controlled Release* **49**, 97-104.

Hamalainen, K.M., Kananen, K., Auriola, S., Kontturi, K. & Urtti, A. (1997b). Characterization of paracellular and aqueous penetration routes in cornea, conjunctiva and sclera. *Investigative Ophthalmology & Visual Science* **38**, 627-624.

Hammett, L.P. (1937). The effect of structure upon the reactions of organic compounds. Benzene derivatives. *Journal of the American Chemical Society* **59**, 96-103.

Hansch, C. & Fujita, J. (1964). ρ-σ-π analysis. A method for the correlation of biological activity and chemical structure. *Journal of the American Chemical Society* **86,** 1616-1626.

Hansch, C. & Leo, A. (1995). *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*. American Chemical Society, Washington, DC.

Hansch, C., Leo, A. & Hoekman, D. (1995). *Exploring QSAR: Hydrophobic, Electronic, and Steric Constants*. American Chemical Society, Washington, DC.

Hansch, C., Maloney, P.P., Fujita, T. & Muir, R.M. (1962). Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients. *Nature* **194**, 178-180.

Hartigan, J.A. (1975). *Clustering algorithms.* Wiley, New York.

Hartigan, J.A. & Wong, M.A. (1978). Algorithm 136. A *k*-means clustering algorithm. *Applied Statistics* **28**, 100.

Hayashi, H., Nakamura, Y., Higashi, K., Kato, H., Kishida, F. & Kaneko, H. (1999). A quantitative structure-Activity relationship study of the skin irritation potential of phenols. *Toxicology in Vitro* **13**, 915-922.

Hoffman, R. (1963). An extended Hückel theory. I. Hydrocarbons. *Journal of Chemical Physics* **39**, 1397-1412.

Home Office (1986). *Animals (Scientific Procedures) Act.* Her Majesty's Stationery Office, London.

Howes, D., Guy, R., Hadgraft, J., Heylings, J., Hoeck, U., Kemper, F., Maibach, H., Marty, J-P., Merk, H., Parra, J., Rekkas, D., Rondelli, I., Schaefer, H., Täuber, U. & Verbiese, N. (1996). Methods for assessing percutaneous absorption. The report and recommendations of ECVAM Workshop 13. *Alternatives to Laboratory Animals* **24**, 81-106.

Hueber, F., Schaefer, H. & Wepierre, J. (1994). Role of transepidermal and transfollicular routes in percutaneous absorption of steroids: *in vitro* studies on human skin. *Skin Pharmacology* **7**, 237-244.

Hückel, E. (1931). *Zeitschrift für Physik* **70**, 204.

Igarashi, H., Sato, Y., Hamada, S. & Kawasaki, I. (1984). Studies on rabbit corneal permeability of local anesthetics (I). *Japanese Journal of Pharmacology* **34**, 429-434.

de Jongh, J., Forsby, A., Houston, J.B., Beckman, M., Combes, R. & Blauuboer, B.J. (1999). An integrated approach to the prediction of systemic toxicity using computer-based biokinetic models and biological *in vitro* test methods: overview of a prevalidation study based on the ECITTS project. *Toxicology in Vitro* **13**, 549-554.

Kamlet, M.J., Abboud, J.L., Abraham, M.H. & Taft, R.W. (1983). Linear solvation energy relationships. 23. A comprehensive collection of solvatochromic parameters, π^*, α, and β, and some methods for simplifying the generalized solvation equation. *Journal of Organic Chemistry* **48**, 2877-2887.

Kang, F., Kuang, K., Li, J. & Fischbarg, J. (1999). Cultured bovine corneal epithelial cells express a functional aquaporin water channel. *Investigative Ophthalmology & Visual Science* **40**, 253-257.

Kasting, G.B., Smith, R.L. & Cooper, E.R. (1987). Effect of lipid solubility and molecular size on percutaneous absorption. In *Pharmacology and the Skin, Volume 1. Skin Pharmacokinetics* (eds. B. Shroot & H. Schaefer). pp. 138-153. Karger, Basel.

Kawazu, K., Yamada, K., Nakamura, M. & Ota, A. (1999). Characterization of cyclosporin A transport in cultured rabbit corneal epithelial cells: P-glycoprotein transport activity and binding to cyclophilin. *Investigative Ophthalmology & Visual Science* **40**, 1738-1744.

Kay, J.H. & Calandra, I.C. (1962). Interpretation of eye irritation tests. *Journal of the Society of Cosmetic Chemists* **3**, 281-289.

Kier, L.B. (1985). A shape index from molecular graphs. *Quantitative Structure-Activity Relationships* **4**, 109-116.

Kier, L.B. (1990). Indexes of molecular shape from chemical graphs. *In Computational Chemical Graph Theory* (ed. D.H. Rouvray). pp. 151-174. Nova Press, Los Angeles, CA.

Kier, L.B. & Hall, L.H. (1986). *Molecular Connectivity in Structure-Activity Analysis*. John Wiley & Sons, New York.

Kier, L.B. & Hall, L.H. (1991). A differential molecular connectivity index. *Quantitative Structure-Activity Relationships* **10**, 134.

Kirchner, L.A., Moody, R.P., Doyle, E., Bose, R., Jeffery, J. & Chu, I. (1997). The prediction of skin permeability by using physicochemical data. *Alternatives to Laboratory Animals* **25**, 359-370.

Kompella, U.B., Sunkara, G., Thomas, E., Clark, C.R. & Deruiter, J. (1999). Rabbit corneal and conjuntival permeability of the novel aldose reductase inhibitors: *N*-{[4-(benzoylamino)phenyl]sulphonyl}glycines and *N*-benzoyl-*N*-phenylglycines. *Journal of Pharmacy and Pharmacology* **51**, 921-927.

Koopmans, T. (1933). On the relation of wavefunctions and eigenvalues with the individual electrons of atoms. *Physica* **1**, 104.

Kulkarni, A.S. & Hopfinger, A.J. (1999). Membrane-interaction QSAR analysis: application to the estimation of eye irritation by organic compounds. *Pharmaceutical Research* **16**, 1245-1253.

Labanowski, J.K. & Andzelm, J.W. (1991). *Density Functional Methods in Chemistry*. Springer-Verlag, New York.

Lawrence, J.N., Dickson, F.M. & Benford, D.J. (1997). Skin irritant-induced cytotoxicity and prostaglandin E_2 release in human skin keratinocyte cultures. *Toxicology in Vitro* **11**, 627-631.

Lewis, R.W. & Botham, P.A. (1994). Measurement of transcutaneous electrical resistance to assess the skin corrosivity of chemicals. In *In Vitro Skin Toxicology - Irritation, Phototoxicity, Sensitization.* Alternative Methods in Toxicology, Volume 10 (eds. A. Rougier, A.M. Goldberg & H.I. Maibach). pp. 97-105. Mary Ann Liebert, New York.

Lieb, W.R. & Stein, W.D. (1971). Implications of two different types of diffusion for biological membranes. *Nature* **234**, 219-222.

Lindman, H.R. (1974). *Analysis of Variance in Complex Experimental Designs*. W. H. Freeman & Co., San Francisco.

Livingstone, D. (1995). *Data Analysis for Chemists. Applications to QSAR and Chemical Product Design*. Oxford University Press, Oxford.

Loh, W-Y.& Shih, Y-S. (1997). Split selection methods for classification trees. *Statistica Sinica* **7**, 815-840.

Macilwain, C. (2000). World leaders heap praise on human genome landmark. *Nature* **405**, 983-985.

Manallack, D.T. & Livingstone, D.J. (1999). Neural networks in drug discovery: have they lived up to their promise ? *European Journal of Medicinal Chemistry* **34**, 195-208.

McFarland, J.W. & Gans, D.J. (1986a). On the significance of clusters in the graphical display of structure-activity data. *Journal of Medicinal Chemistry* **29**, 505-514.

McFarland, J.W. & Gans, D.J. (1986b). Cluster significance analysis contrasted with three other quantitative structure-activity relationship methods. *Journal of Medicinal Chemistry* **30**, 46-49.

McFarland, J.W. & Gans, D.J. (1990). Cluster significance analysis: a new QSAR tool for asymmetric data sets. *Drug Information Journal* **24**, 705-711.

McFarland, J.W. & Gans, D.J. (1994). On identifying likely determinants of biological activity in high dimensional QSAR problems. *Quantitative Structure-Activity Relationships* **13**, 11-17.

McKinney, J.D., Richard, A., Waller, C., Newman, M.C. & Gerberick, F. (2000). The practice of structure activity relationships (SAR) in Toxicology. *Toxicological Sciences* **56**, 8-17.

Meylan, W.M. & Howard, P.H. (1995). Atom/fragment contribution method for estimating octanol-water partition coefficients. *Journal of Pharmaceutical Sciences* **84**, 83-92.

Meylan, W.M., Howard, P.H. & Boethling, R.S. (1996). Improved method for estimating water solubility from octanol/water partition coefficient. *Environmental Toxicology and Chemistry* **15**, 100-106.

Nangia, A., Anderson, P.H., Berner, B. & Maibach, H.I. (1996). High dissociation constants (pKa) of basic permeants are associated with in vivo skin irritation in man. *Contact Dermatitis* **34**, 237-242.

NIH (1999). Corrositex: an *in vitro* test method for assessing dermal corrosivity potential of chemicals. NIH Publication No. 99-4495. National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM), NC, USA.

Nys, G.G. & Rekker, R.F. (1973). Statistical analysis of a series of partition coefficients with special reference to the predictability of folding of drug molecules. The

introduction of hydrophobic fragmental constants (f values). *European Journal of Medicinal Chemistry (Chimie Thérapeutique)* **8**, 521-573.

OECD (1987). *OECD Guideline for the Testing of Chemicals. No 405. Acute Eye Irritation/Corrosion.* Organisation for Economic Cooperation and Development, Paris.

OECD (1992). *OECD Guideline for the Testing of Chemicals. No 404. Acute Dermal Irritation/Corrosion.* Organisation for Economic Cooperation and Development, Paris.

OECD (1994). *Proposal for Two New Guidelines on Percutaneous Absorption.* OECD Environmental Health and Safety Division Discussion Document, ref. ENV/EHS/HK/mc/94.103. Organisation for Economic Cooperation and Development, Paris.

OECD (1996). *Final Report of the OECD Workshop on Harmonization of Validation and Acceptance Criteria for Alternative Toxicological Test Methods*. Organisation for Economic Cooperation and Development, Paris.

OECD (1998). *Harmonized Integrated Hazard Classification System for Human Health and Environmental Effects of Chemical Substances.* Organisation for Economic Cooperation and Development, Paris.

Oliver, G.J.A., Pemberton, M.A. & Rhodes, C. (1986). An *in vitro* skin corrosivity test - modifications and validation. *Food and Chemical Toxicology* **24**, 507-512.

Oliver, G.J.A., Pemberton, M.A. & Rhodes, C. (1988). An *in vitro* model for identifying skin corrosive chemicals. I. Initial validation. *Toxicology in Vitro* **2**, 7-17.

Osborne, R. & Perkins, M.A. (1991). *In vitro* skin irritation testing with human skin cell cultures. *Toxicology in Vitro* **5**, 563-567.

Patel, H.C., Duca, J.S. & Hopfinger, A.J. (1999). Quantitative component analysis of mixtures for risk assessment: application to eye irritation. *Chemical Research in Toxicology* **12**, 1050-1056.

Patlewicz, G.Y., Rodford, R.A., Ellis, G. & Barratt, M.D. (2000). A QSAR model for the eye irritation of cationic surfactants. *Toxicology in Vitro* **14**, 79-84.

Pearson, R.G. (1993). The principle of maximum hardness. *Accounts of Chemical Research* **26**, 250-255.

Pentreath, V.W. (1999). *Neurotoxicology In Vitro*. Taylor & Francis, London.

Perkins, M.A. & Osborne, R. (1993), Development of an *in vitro* method for skin corrosion testing. *Journal of Investigative Dermatology* **100**, 535.

Perkins M.A., Osborne R. & Johnson G.R. (1996). Development of an *in vitro* method for skin corrosion testing. *Fundamental and Applied Toxicology* **31**, 9-18.

Perrin, D.D., Dempsey, B. & Serjeant, E.P. (1981). *pKa Prediction for Organic Acids and Bases*. Chapman and Hall, London.

Planck, S.R., Rich, L.F., Ansel, J.C., Huang, X.N. & Rosenbaum, J.T. (1997). Trauma and alkali burns induce distinct patterns of cytokine gene expression in the rat cornea. *Ocular Immunology & Inflammation* **5**, 95-100.

Pohorille, A. & Wilson, M.A. (1996). Excess chemical potential of small solutes across water-membrane and water-hexane interfaces. *Journal of Chemical Physics* **104**, 3760-3773.

Pople, J.A., Beveridge, D.L. & Dobosh, P.A. (1967). Approximate self-consistent molecular-orbital theory. V. Intermediate neglect of differential overlap. *Journal of Chemical Physics* **47**, 2026-2033.

Pople, J.A. & Beveridge, D.L. (1970). *Approximate Molecular Orbital Theory*. McGraw-Hill, New York.

Potts, R.O. & Guy, R.H. (1992). Predicting skin permeability. *Pharmaceutical Research* **9**, 663-669.

Prausnitz, M.R. & Noonan, J.S. (1998). Permeability of cornea, sclera and conjunctiva: a literature analysis for drug delivery to the eye. *Journal of Pharmaceutical Sciences* **87**, 1479-1488.

Prinsen, M.K. (1999). An evaluation of the OECD proposal for the harmonised classification of eye irritants and corrosives. Appendix 1 to the report of ECVAM Workshop 34. *Alternatives to Laboratory Animals* **27**, 72-77.

Protic, M. & Sabljic, A. (1989). Quantitative structure-activity relationships of acute toxicity of commercial chemicals on fathead minnows: effect of molecular size. *Aquatic Toxicology* **14**, 47-64.

Pugh, W.J. & Hadgraft, J. (1994). *Ab initio* prediction of human skin permeability coefficients. *International Journal of Pharmaceutics* **103**, 163-178.

Pugh, W.J., Roberts, M.S. & Hadgraft, J. (1996). Epidermal permeability - penetrant structure relationships: 3. The effect of hydrogen bonding interactions and molecular size on diffusion across the stratum corneum. *International Journal of Pharmaceutics* **138**, 149-165.

Régnier, J-F. & Imbert, C. (1992). Contributions of physicochemical properties to the evaluation of ocular irritation. *Alternatives to Laboratory Animals* **20**, 457-465.

Roberts, M.S., Anderson, R.A. & Swarbrick, J. (1977). Permeability of human epidermis to phenolic compounds. *Journal of Pharmacy and Pharmacology* **29**, 677.

Robinson, M.K., Osborne, R. & Perkins, M.A. (1999). Strategies for the assessment of acute skin irritation potential. *Journal of Pharmacological and Toxicological Methods* **42**, 1-9.

Rosdy, M. & Clauss, L.C. (1990). Complete human epidermal cell differentiation in chemical defined medium at the air-liquid interface on inert filter substrates. *Journal of Investigative Dermatology* **95**, 409-414.

Rose, V.S. & Wood, J. (1998). Generalized cluster significance analysis and stepwise cluster significance analysis with conditional probabilities. *Quantitative Structure-Activity Relationships* **17**, 348-356.

Rose, V.S., Wood, J. & MacFie, H.J.H. (1991). Single class discrimination using principal components analysis (SCD-PCA). A new method for the analysis of embedded structure-activity relationships. *Quantitative Structure-Activity Relationships* **10**, 359-368.

Rose, V.S., Wood, J. & MacFie, H.J.H. (1992). Generalized single class discrimination (GSCD). A new method for the analysis of embedded structure-activity relationships. *Quantitative Structure-Activity Relationships* **11**, 492-504.

Rosenkranz, H.S., Zhang, Y.P. & Klopman, G. (1998). The development and characterisation of a structure-activity relationship model of the Draize eye irritation test. *Alternatives to Laboratory Animals* **26**, 779-809.

Russell, W.M.S. & Burch, R.L. (1959). *The Principles of Humane Experimental Technique*. Methuen, London.

Sabljic, A. (1987). On the prediction of soil sorption coefficients of organic pollutants from molecular structure: application of molecular topology model. *Environmental Science and Technology* **21**, 238-366.

Sabljic, A. (1990). Topological indices and environmental chemistry. In *Practical Applications of QSAR in Environmental Chemistry and Toxicology* (eds. W. Karcher & J. Devillers). pp 61-82. Kluwer, Dordrecht.

Sasaki, H., Ichikawa, M., Yamamura, K., Nishida, K. & Nakamura, J. (1997). Ocular membrane permeability of hydrophilic drugs for ocular peptide delivery. *Journal of Pharmacy and Pharmacology* **49**, 135-139.

Saunders, M., Houk, K.N., Wu, Y-D., Still, W.C., Lipton, M., Chang, G. & Guida, W.C. (1990). Conformations of cycloheptadecane. A comparison of methods for conformational searching. *Journal of the American Chemical Society* **112**, 1419-1427.

Scherrer, R.A. & Howard, S.M. (1977). Use of distribution coefficients in quantitative structure-activity relationships. *Journal of Medicinal Chemistry* **20**, 53-58.

Scheuplein, R.J. (1965). Mechanism of percutaneous absorption. I. Routes of penetration and influence of solubility. *Journal of Investigative Dermatology* **45**, 334.

Scheuplein, R.J., Blank, I.H., Brauner, G.J. & MacFarlane, D.J. (1969). Percutaneous absorption of steroids. *Journal of Investigative Dermatology* **52**, 63.

Schoenwald, R.D. & Ward, R.L. (1978). Relationship between steroid permeability across excised rabbit cornea and octanol-water partition coefficients. *Journal of Pharmaceutical Sciences* **67**, 786-788.

Schoenwald, R.D. & Huang, H-S. (1983). Corneal penetration behaviour of beta-blocking agents I: physicochemical factors. *Journal of Pharmaceutical Sciences* **72**, 1266-1272.

Schrödinger, E. (1926). Quantisierung als Eigenwertproblem. 1. *Annalen der Physik* **79**, 361.

Scott, R.C., Batten, P.L., Clowes, H.M., Jones, B.K. & Ramsey, J.D. (1992). Further validation of an *in vitro* method to reduce the need for *in vivo* studies for measuring the absorption of chemicals through the skin. *Fundamental and Applied Toxicology* **19**, 484-492.

Smith, J.S., Macina, O.T., Sussman, N.B., Luster, M.I. & Karol, M.H. (2000). A robust structure-activity relationship (SAR) model for esters that cause skin irritation in humans. *Toxicological Sciences* **55**, 215-222.

Smyth, H.F. & Carpenter, C.P. (1946). Chemical burns of the rabbit cornea. *American Journal of Opthalmology* **29**, 1363-1372.

Spence, P. & Aurora, R. (1999). From reductionist to constructionist, but only if we integrate. In *Pharmainformatics: a Trends Guide. Trends Supplement 1999*. pp. 37-39. Elsevier, Amsterdam.

Spielmann, H., Liebsch, M., Kalweit, S., Moldenhauer, F., Wirnsberger, T., Holzhütter, H-G., Schneider, B., Glaser, S., Gerner, I., Pape, W.J.W., Kreiling, R., Krauser, K. & Miltenburger, H.G., Steiling, W., Luepke, N.P., Muller, N., Kreuzer, H., Murmann, P., Spengler, J., Bertram-Neis, E., Siegemund, B. & Wiebel, F.J. (1996). Results of a validation study in Germany on two *in vitro* alternatives to the Draize eye irritation test, the HET-CAM test and the 3T3-NRU cytotoxicity test. *Alternatives to Laboratory Animals* **24**, 741-858.

Stern, M., Klausner, M., Alvarado, R., Renskers, K. & Dickens, M. (1998). Evaluation of the EpiOcular tissue model as an alternative to the Draize eye irritation test. *Toxicology in Vitro* **12**, 455-461.

Stevens, J. (1986). *Applied Multivariate Statistics for the Social Sciences.* Erlbaum, Hillsdale, NJ.

Stewart, J.J.P. (1989a). Optimization of parameters for semiempirical methods. I. Method. *Journal of Computational Chemistry* **10**, 209-220.

Stewart, J.J.P. (1989b). Optimization of parameters for semiempirical methods. II. Applications. *Journal of Computational Chemistry* **10**, 221.

Sugai, S., Murata, K., Kitagaki, T. & Tomita, I. (1991). Studies on eye irritation caused by chemicals in rabbits – II. Structure-activity relationships and *in vitro* approach to

primary eye irritation of salicylates in rabbits. *Journal of Toxicological Sciences* **16**, 111-130.

Tabachnick, B.G. & Fidell, L.S. (1996). *Using Multivariate Statistics*. Harper Collins, New York.

Taft, R.W. (1956). Separation of polar, steric and resonance effects in reactivity. In *Steric Effects in Organic Chemistry* (ed. M.S. Newman). John Wiley & Sons, New York.

Takahashi, H., Kaminski, A.E. & Zieske, J.D. (1996). Glucose transporter 1 expression is enhanced during corneal epithelial wound repair. *Experimental Eye Research* **63**, 649-659.

Tinois, E., Tillier, J., Gaucherand, M., Dumas, H., Tardy, M. & Thivolet, J. (1991). *In vitro* and post-transplantation differentiation of human keratinocytes grown on the human type IV collagen film of a bilayered dermal substitute. *Experimental Cell Research* **193**, 310-319.

Topliss, J.G. & Costello, R.J. (1972). Chance correlations in structure-activity studies using multiple regression analysis. *Journal of Medicinal Chemistry* **15**, 1066-1068.

Torres, P., De Vos, A.F., Van der Gaag, R. & Kijlstra, A. (1994). Expression of the interleukin 1 receptor antagonist in the normal human cornea. *Ocular Immunology & Inflammation* **2**, 217-222.

UN (1977). United Nations Economic and Social Council. Joint meeting of the RID safety committee and the group of experts on the transportation of dangerous goods. *Trans/GE 15/R* **274**, 2.

van de Sandt, J., Roguet, R., Cohen, C., Esdaile, D., Ponec, M., Corsini, E., Barker, C., Fusenig, N., Liebsch, M., Benford, D., de Brugerolle de Fraissinette, A. & Fartasch, M. (1999). The use of human keratinocytes and human skin models for predicting skin

irritation. The report and recommendations of ECVAM workshop 3. *Alternatives to Laboratory Animals* **27**, 723-743.

Verburgh, J., Hansen, B., Worth, A. & Karcher, W. (1996). Risk assessment of PAH-containing petroleum substances in the EU. *Polycyclic Aromatic Compounds* **11**, 169-176.

Verloop, A., Hoogeenstraaten, W. & Tipker, J. (1976). Development and application of new steric substiutent parameters in drug design. In *Drug Design, Volume 7* (ed. E.J. Ariens). pp. 165-207. Academic Press, New York.

Viswanadhan, V.N., Ghose, A.K., Revankar, G.R. & Robins, R.K. (1989). Atomic physicochemical parameters for three dimensional structure directed quantitative structure-activity relationships. 4. Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain naturally occurring nucleoside antibiotics. *Journal of Chemical Information and Computer Sciences* **29**, 163-172.

Weil, C.S. & Scala, A. (1971). Study of intra- and inter-laboratory variability in the results of rabbit eye and skin irritation tests. *Toxicology and Applied Pharmacology* **19**, 276-360.

Weiner, S.J., Kollman, P.A., Case, D.A., Songh, U.C., Ghio, C., Alagona, G., Profeta, S. & Weiner, P. (1984). A new force field for molecular mechanical simulation of nucleic acids and proteins. *Journal of the American Chemical Society* **106**, 765-784.

Weininger, D. (1988). SMILES, a chemical language and information-system 1. Introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences* **28**, 31-36.

Weng, G., Bhalla, U.S. & Iyengar, R. (1999). Complexity in biological signaling systems. *Science* **284**, 92-95.

Whitesides, G.M. & Ismagilov, R.F. (1999). Complexity in chemistry. *Science* **284**, 89-92.

Whittle, E.G., Barratt, M.D., Carter, J.A., Basketter, D.A. & Chamberlain, M. (1996). The skin corrosivity potential of fatty acids: *in vitro* rat and human skin testing and QSAR studies. *Toxicology in Vitro* **10**, 95-100.

Wilhelm, K.P., Samblebe, M. & Siegers, C.P. (1994). Quantitative *in vitro* assessment of N-alkyl sulphate-induced cytotoxicity in human keratinocytes (HaCaT). Comparison with *in vivo* human irritation tests. *British Journal of Dermatology* **130**, 18-23.

Williams, I.R. & Kupper, T.S. (1996). Immunity at the surface: homeostatic mechanisms of the skin immune system. *Life Sciences* **58**, 1485-1507.

Worth, A.P., Barratt, M.D. & Houston, J.B. (1998). The validation of computational prediction techniques. *Alternatives to Laboratory Animals* **26**, 241-247.

Worth, A.P. & Cronin, M.T.D. (1999). Embedded cluster modelling – a novel method for analysing embedded data sets. *Quantitative Structure-Activity Relationships* **18**, 229-235.

Worth A.P. & Cronin, M.T.D. (2000a). Embedded cluster modelling: a novel QSAR method for generating elliptic models of biological activity. In *Progress in the Reduction, Refinement and Replacement of Animal Experimentation* (eds. M. Balls, A-M. van Zeller & M.E. Halder). pp. 479-491. Elsevier Science, Amsterdam.

Worth, A.P. & Cronin, M.T.D. (2000b). Structure-permeability relationships for transcorneal penetration. *Alternatives to Laboratory Animals* **28**, 403-413.

Worth A.P. & Fentem J.H. (1999). A general approach for evaluating stepwise testing strategies. *Alternatives to Laboratory Animals* **27**, 161-177.

Worth A.P., Fentem J.H., Balls M., Botham P.A., Curren R.D., Earl L.K., Esdaile D.J. & Liebsch M. (1998). An evaluation of the proposed OECD testing strategy for skin corrosion. *Alternatives to Laboratory Animals* **26**, 709-720.

Yalkowsky, S.H. & Valvani, S.C. (1980). Solubility and Partitioning I: Solubility of Nonelectrolytes in Water. *Journal of Pharmaceutical Sciences* **69**, 912-922.

Yoshida, F. & Topliss, J.G. (1996). Unified model for the corneal permeability of related and diverse compounds with respect to their physicochemical properties. *Journal of Pharmaceutical Sciences* **85**, 819-823.

Young, J.R. & How, M.J. (1994). Product classification as corrosive or irritant by measuring pH and acid/alkali reserve. In *In Vitro Methods in Toxicology - Irritation, Phototoxicity, Sensitization.* Alternative Methods in Toxicology, Volume 10 (eds. A. Rougier, A.M.Goldberg & H.I. Maibach). pp. 23-35. Mary Ann Liebert, New York.

Young, J.R., How, M.J., Walker, A.P. & Worth, W.M.H. (1988). Classification as corrosive or irritant to skin of preparations containing acidic or alkaline substances without testing on animals. *Toxicology in Vitro* **2**, 19-26.

Zinke, S., Gerner, I., Graetschel, G. & Schlede, E. (2000). Local irritation/corrosion testing strategies: development of a decision support system for the introduction of alternative methods. *Alternatives to Laboratory Animals* **28**, 29-40.

APPENDIX A
MINITAB MACROS

A1 CONSTRUCTION OF 2 x 2 CONTINGENCY TABLE AND CALCULATION OF COOPER STATISTICS

```
# GLOBAL MACRO CTABLE.MAC
#
# Constructs 2x2 contingency table, performs Chi-square significance test and calculates
# Cooper statistics
#
# MINITAB for Windows version 11 and above
#
# Place the chemical names in C1, the known classifications in C2 and the
# predicted classifications in C4.
# Type the symbol for toxic chemicals in C5(1) and the symbol for non-toxic chemicals in C6(1)
# Type %ctable at the command line prompt.

GMACRO
CTABLE

NAME C1='CHEMICAL' C2='OBSERVED' C3='PREDICTED' C5='ACTIVE' C6='INACTIVE'

NOTE
NOTE Have you entered the appropriate symbols in C5(1) and C6(1) ?
NOTE
YESNO K1
IF K1=1
 CALL TABLE
ELSE
 NOTE
 NOTE Then please do so, and rerun the macro
 EXIT
ENDIF
ENDMACRO

GMACRO
TABLE

LET C4(2)=C5(1)
LET C4(3)=C6(1)
TEXT C4 C4
LET K1=COUNT(C1)
LET K2=0
LET K3=0
LET K4=0
LET K5=0
LET K6=0
LET K7=0
LET K8=0
LET K9=0
LET K10=0

DO K10=1:K1
 IF C2(K10)=C3(K10)
  LET K2=K2+1                    # total no of correct predictions
 ENDIF
 IF C2(K10)=C3(K10) AND C3(K10)=C5(1)
  LET K3=K3+1                    # no of true positives
 ENDIF
 IF C2(K10)=C3(K10) AND C3(K10)=C6(1)
  LET K4=K4+1                    # no of true negatives
```

```
 ENDIF
 IF C2(K10)~=C3(K10) AND C3(K10)=C5(1)
  LET K5=K5+1                    # no of false positives
 ENDIF
 IF C2(K10)~=C3(K10) AND C3(K10)=C6(1)
  LET K6=K6+1                    # no of false negatives
 ENDIF
ENDDO

LET K7=K3+K6                     # total no of known positives
LET K8=K4+K5                     # total no of known negatives
LET K9=K3+K5                     # total no of positive predictions
LET K10=K4+K6                    # total no of negative predictions
DELETE 1 C5 C6
LET C5(2)=K3
LET C5(3)=K5
LET C6(2)=K6
LET C6(3)=K4
LET C7(2)=K7
LET C7(3)=K8
LET C5(4)=K9
LET C6(4)=K10
LET K20=C5(4)
LET K21=C6(4)
LET C7(4)=C5(4)+C6(4)
LET K11=C7(4)

TEXT C5 C5
TEXT C6 C6

LET C5(1)=C4(2)
LET C6(1)=C4(3)
LET C5(5)="PREDICTED"
LET C6(5)="PREDICTED"
LET C8(2)="OBSERVED"
LET C8(3)="OBSERVED"
LET C6(7)="Sensitivity"
LET C6(8)="Specificity"
LET C6(9)="Concordance"
LET C6(10)="Pos predict"
LET C6(11)="Neg predict"
LET C6(12)="False pos rate"
LET C6(13)="False neg rate"
LET C7(7)=(K3/K7)*100
LET C7(8)=(K4/K8)*100
LET C7(9)=(K2/K1)*100
LET C7(10)=(K3/K9)*100
LET C7(11)=(K4/K10)*100
LET C7(12)=100-C7(8)
LET C7(13)=100-C7(7)
LET C10(1)=C5(2)
LET C10(2)=C5(3)
LET C11(1)=C6(2)
LET C11(2)=C6(3)

NUMERIC C10 C10
NUMERIC C11 C11
CHISQUARE C10 C11

ENDMACRO
```

A2 LEAVE-ONE-OUT CROSS-VALIDATION DURING MULTIPLE REGRESSION

```
# GLOBAL MACRO CVAL2.MAC
#
# Performs leave-one out cross-validation in two-variable multiple regression
#
# MINITAB for Windows version 11 and above
#
# Place the response in C1, and the two predictor variables in C2 and C3.
# Type %cval2 at the command line prompt.

GMACRO
CVAL2

NAME C1 = 'y' C2 = 'a' C3 = 'b' C5 = 'pred y' C6 = '(y-pred y)^2' C7 = '(y-mean y)^2' C8 = 'PRESS' C9 =
'TSS' C10 ='r(cv)^2'
NAME C20 = 'lower CL' C21 = 'upper CL' C22 = 'lower PL' C23 = 'upper PL' C24 ='RMS' C25 ='s'

MEAN C1 K100
LET K101 = COUNT (C1)
REGRESS C1 2 C2 C3;
  CONSTANT.

DO K102 = 1 : K101
 LET K103 = C2(K102)
 LET K104 = C3(K102)
 LET K105 = C1(K102)-K100
 LET C7(K102) = K105**2
 COPY C1 C2 C3 C12 C13 C14;
  OMIT K102.
 REGRESS C12 2 C13 C14;
  CONSTANT;
  MSE K120;
  PREDICT K103 K104;
  PFITS C99;
  CLIMITS C100 C101;
  PLIMITS C102 C103.
 LET C5(K102)=C99
 LET C20(K102)=C100
 LET C21(K102)=C101
 LET C22(K102)=C102
 LET C23(K102)=C103
 LET C24(K102)=K120
 LET C25(K102)=SQRT(K120)
 LET K106 = C1(K102)-C5(K102)
 LET C6(K102) = K106**2
ENDDO

LET K107=SUM(C6)
LET K108=SUM(C7)
LET K109 = 1-(K107/K108)
LET C8(1)=K107
LET C9(1)=K108
LET C10(1)=K109

ENDMACRO
```

A3 CLUSTER SIGNIFICANCE ANALYSIS

```
# GLOBAL MACRO CSA2.MAC
#
# Carries out cluster significance analysis on two variables
#
# MINITAB for Windows version 11 and above
#
# Place the chemical names (or numbers) in C1, chemical activities (0/1)
# in C2, and the variables in C3 and C4.
# Type %csa2 at the command line prompt.

GMACRO
Csa2

NAME C1='CHEMICAL' C2='ACTIVITY' C3='DESCRIPTOR 1' C4='DESCRIPTOR 2'
NAME C7='VARIANCE' C8='PROBABILITY' C9='LOWER 95% CL' C10='UPPER 95% CL'
C20='DISTRIBUTION'
NAME C21='RAND SUBSET1' C22='RAND SUBSET2' C23='SUBSETS'

NOTE
NOTE Do you wish to standardise the data in C3 and C4 ?
NOTE
YESNO K50
IF K50=1
 CENTRE C3-C4 C3-C4.
ENDIF

UNSTACK C3 C5-C6;
 SUBSCRIPTS C2.

LET K1=(STDEV (C6))**2
LET C7(1)=K1

ERASE C5 C6
UNSTACK C4 C5-C6;
 SUBSCRIPTS C2.

LET K2=(STDEV (C6))**2
LET C7(2)=K2
LET K3=K1+K2
LET C7(3)=K3
LET K4=COUNT (C6)
ERASE C5 C6

LET C6(1)="1st variable"
LET C6(2)="2nd variable"
LET C6(3)="1st and 2nd"
NOTE
NOTE How many subsets do you want to sample?
SET C23;
 FILE "TERMINAL";
 NOBS=1.
COPY C23 K5

LET K6=0
LET K7=0
LET K8=0

DO K9=1:K5
```

```
 SAMPLE K4 C3 C4 C21 C22
 LET K10=(STDEV (C21))**2
 LET K11=(STDEV (C22))**2
 LET K12=K10+K11
 LET C20(K9)=K12
 IF K10<=K1
  LET K6=K6+1
 ENDIF
 IF K11<=K2
  LET K7=K7+1
 ENDIF
 IF K12<=K3
  LET K8=K8+1
 ENDIF
ENDDO

LET K13=K6/K5
LET K14=K7/K5
LET K15=K8/K5
LET K16=(K13+K15)/2
IF K13>K15
 LET K17=K13-K15
ELSE
 LET K17=K15-K13
ENDIF
LET K18=2/K5
LET K19=SQRT(K16*(1-K16)*K18)
LET K20=K17/K19
LET K21=SQRT((K13*(1-K13))/K5)
LET K22=0.5/K5
LET K23=(1.96*K21)+K22
LET K24=SQRT((K14*(1-K14))/K5)
LET K25=0.5/K5
LET K26=(1.96*K24)+K25
LET K27=SQRT((K15*(1-K15))/K5)
LET K28=0.5/K5
LET K29=(1.96*K27)+K28

LET C8(1)=K13
LET C8(2)=K14
LET C8(3)=K15
LET C9(1)=K13-K23
LET C9(2)=K14-K26
LET C9(3)=K15-K29
LET C10(1)=K13+K23
LET C10(2)=K14+K26
LET C10(3)=K15+K29

IF K20>=1.96
 NOTE
 NOTE The difference in the P-value caused by adding the second variable is
 NOTE SIGNIFICANT at the 95% level
ELSE
 NOTE
 NOTE The difference in the P-value caused by adding the second variable is
 NOTE NOT SIGNIFICANT at the 95% level
ENDIF

ENDMACRO
```

A4 GENERALISED CLUSTER SIGNIFICANCE ANALYSIS

```
# GLOBAL MACRO GCSA2.MAC
#
# Carries out generalised cluster significance analysis on two variables
#
# MINITAB for Windows version 11 and above
#
# Place the chemical names (or numbers) in C1, the activity in C2
# and the variables in C3 and C4.
# Type %gcsa2 at the command line prompt.

GMACRO
Gcsa2

NAME C1='CHEMICAL' C2='ACTIVITY' C3='DESCRIPTOR 1' C4='DESCRIPTOR 2'
NAME C5='WEIGHT' C6='WTD DESCRIP1' C7='WTD DESCRIP2'
NAME C8='VARIANCE' C9='PROBABILITY' C10='95% C.I.'  C11='DISTRIBUTION'
NAME C12='RESAMP WT' C13='RESAMP WT D1' C14='RESAMP WT D2' C15='SUBSETS'

NOTE
NOTE Do you wish to standardise the data in C3 and C4 ?
NOTE
YESNO K50
IF K50=1
 CENTRE C3-C4 C3-C4.
ENDIF

CENTRE C2 C5;
 MINMAX 0 1.0.
LET C6=C3*C5
LET C7=C4*C5

LET K1=(STDEV (C6))**2
LET K2=(STDEV (C7))**2
LET K3=K1+K2
LET C8(1)=K1
LET C8(2)=K2
LET C8(3)=K3
LET K4=COUNT(C1)

ERASE C6-C7

LET C7(1)="1st variable"
LET C7(2)="2nd variable"
LET C7(3)="1st and 2nd"

NOTE
NOTE How many times do you want to resample?
SET C15;
 FILE "TERMINAL";
 NOBS=1.
COPY C15 K5

LET K6=0
LET K7=0
LET K8=0

DO K9=1:K5
 SAMPLE K4 C5 C12
```

```
 LET C13=C3*C12
 LET C14=C4*C12
 LET K10=(STDEV (C13))**2
 LET K11=(STDEV (C14))**2
 LET K12=K10+K11
 LET C11(K9)=K12
 IF K10<=K1
  LET K6=K6+1
 ENDIF
 IF K11<=K2
  LET K7=K7+1
 ENDIF
 IF K12<=K3
  LET K8=K8+1
 ENDIF
ENDDO

LET K13=K6/K5
LET K14=K7/K5
LET K15=K8/K5
LET K16=(K13+K15)/2
IF K13>K15
 LET K17=K13-K15
ELSE
 LET K17=K15-K13
ENDIF

LET K18=2/K5
LET K19=SQRT(K16*(1-K16)*K18)
LET K20=K17/K19
LET K21=SQRT((K13*(1-K13))/K5)
LET K22=0.5/K5
LET K23=(1.96*K21)+K22
LET K24=SQRT((K14*(1-K14))/K5)
LET K25=0.5/K5
LET K26=(1.96*K24)+K25
LET K27=SQRT((K15*(1-K15))/K5)
LET K28=0.5/K5
LET K29=(1.96*K27)+K28

LET C9(1)=K13
LET C9(2)=K14
LET C9(3)=K15
LET C10(1)=K23
LET C10(2)=K26
LET C10(3)=K29

IF K20>=1.96
 NOTE
 NOTE The difference in the P-value caused by adding the second variable is
 NOTE SIGNIFICANT at the 95% level
ELSE
 NOTE
 NOTE The difference in the P-value caused by adding the second variable is
 NOTE NOT SIGNIFICANT at the 95% level
ENDIF

ENDMACRO
```

A5 BOOTSTRAP ANALYSIS OF COOPER STATISTICS I: APPLICATION OF THE STANDARD BOOTSTRAP METHOD

```
# GLOBAL MACRO BOOT.MAC
#
# Estimates confidence intervals for two-group classification models
#
# MINITAB for Windows version 11 and above
#
# Place the chemical names, the known classifications in C2, and the predicted classifications
# made by different labs/runs in column C3-C11.
# Enter the symbols for toxic, non-toxic and non-qualifying chemicals in C13(1), C13(2) and C13(3)
# Type %cval2 at the command line prompt.

GMACRO
boot

NAME C1='CHEMICAL' C2='OBS CLASS' C3='PRED1' C4='PRED2' C5='PRED3' C6='PRED4'
NAME C7='PRED5' C8='PRED6' C9='PRED7' C10='PRED8' C11='PRED9'
NAME C13='TOXIC' C14='NON-TOXIC' C15='NON-QUALIFIER'

NOTE
NOTE Have you entered the appropriate symbols in C13(1), C14(1) and C15(1)?
NOTE
YESNO K1
IF K1=1
 CALL RESAMPLE
ELSE
 NOTE
 NOTE Then please do so, and rerun the macro
 EXIT
ENDIF
ENDMACRO

GMACRO
RESAMPLE

NOTE
NOTE How many labs/runs are there ?
SET C100;
 FILE "TERMINAL";
 NOBS=1.
COPY C100 K7                    # no of labs/runs K7
ERASE C100


NAME C32='RAND OBS' C33='RAND PRED1' C34='RAND PRED2' C35='RAND PRED3'
C36='RAND PRED4'
NAME C37='RAND PRED5' C38='RAND PRED6' C39='RAND PRED7' C40='RAND PRED8'
C41='RAND PRED9'
NAME C45='CONCORD 1' C46='CONCORD 2' C47='CONCORD 3' C48='CONCORD 4'
C49='CONCORD 5'
NAME C50='CONCORD 6' C51='CONCORD 7' C52='CONCORD 8' C53='CONCORD 9'
NAME C55='SENSITIV 1' C56='SENSITIV 2' C57='SENSITIV 3' C58='SENSITIV 4' C59='SENSITIV
5'
NAME C60='SENSITIV 6' C61='SENSITIV 7' C62='SENSITIV 8' C63='SENSITIV 9'
NAME C65='SPECIFIC 1' C66='SPECIFIC 2' C67='SPECIFIC 3' C68='SPECIFIC 4' C69='SPECIFIC 5'
NAME C70='SPECIFIC 6' C71='SPECIFIC 7' C72='SPECIFIC 8' C73='SPECIFIC 9'
```

```
LET K1=COUNT(C1)

NOTE
NOTE How many times do you want to resample?
SET C100;
 FILE "TERMINAL";
 NOBS=1.
COPY C100 K3                    # K3 = no of bootstrap resamples
ERASE C100

DO K2=1:K3                      # start bootstrap routine
 LET K8=0
 LET K9=0
 LET K12=0
 LET K13=0
 LET K14=0
 LET K15=0
 LET K16=0
 LET K22=0
 LET K23=0
 LET K24=0
 LET K25=0
 LET K26=0
 LET K32=0
 LET K33=0
 LET K34=0
 LET K35=0
 LET K36=0
 LET K42=0
 LET K43=0
 LET K44=0
 LET K45=0
 LET K46=0
 LET K52=0
 LET K53=0
 LET K54=0
 LET K55=0
 LET K56=0
 LET K62=0
 LET K63=0
 LET K64=0
 LET K65=0
 LET K66=0
 LET K72=0
 LET K73=0
 LET K74=0
 LET K75=0
 LET K76=0
 LET K82=0
 LET K83=0
 LET K84=0
 LET K85=0
 LET K86=0
 LET K92=0
 LET K93=0
 LET K94=0
 LET K95=0
 LET K96=0

 LET K131=0         # correction for non-qualifying chemicals
```

```
LET K132=0        # observed='toxic'
LET K133=0
LET K134=0
LET K135=0
LET K136=0
LET K137=0
LET K138=0
LET K139=0

LET K141=0        # correction for non-qualifying chemicals
LET K142=0        # observed='non-toxic'
LET K143=0
LET K144=0
LET K145=0
LET K146=0
LET K147=0
LET K148=0
LET K149=0

IF K7=1
 SAMPLE K1 C2-C3 C32-C33;
  REPLACE.
ENDIF
IF K7=2
 SAMPLE K1 C2-C4 C32-C34;
  REPLACE.
ENDIF
IF K7=3
 SAMPLE K1 C2-C5 C32-C35;
  REPLACE.
ENDIF
IF K7=4
 SAMPLE K1 C2-C6 C32-C36;
  REPLACE.
ENDIF
IF K7=5
 SAMPLE K1 C2-C7 C32-C37;
  REPLACE.
ENDIF
IF K7=6
 SAMPLE K1 C2-C8 C32-C38;
  REPLACE.
ENDIF
IF K7=7
 SAMPLE K1 C2-C9 C32-C39;
  REPLACE.
ENDIF
IF K7=8
 SAMPLE K1 C2-C10 C32-C40;
  REPLACE.
ENDIF
IF K7=9
 SAMPLE K1 C2-C11 C32-C41;
  REPLACE.
ENDIF

DO K4=1:K1                    # start chemicals routine
IF C32(K4)=C13(1)
 LET K8=K8+1
ENDIF
```

```
 IF C32(K4)=C14(1)
  LET K9=K9+1
 ENDIF

 # LABORATORY 1

  IF C32(K4)=C33(K4)
   LET K12=K12+1                                  # total no of correct predictions
  ENDIF
  IF C32(K4)=C33(K4) AND C33(K4)=C13(1)
   LET K13=K13+1                                  # no of true positives
  ENDIF
  IF C32(K4)=C33(K4) AND C33(K4)=C14(1)
   LET K14=K14+1                                  # no of true negatives
  ENDIF
  IF C32(K4)=C13(1) AND C33(K4)=C15(1)       # correction for non-qualifiers
   LET K131=K131+1
  ENDIF
  IF C32(K4)=C14(1) AND C33(K4)=C15(1)       # correction for non-qualifiers
   LET K141=K141+1
  ENDIF

IF K7=1
 GOTO 1
ENDIF

# LABORATORY 2

  IF C32(K4)=C34(K4)
   LET K22=K22+1                                  # total no of correct predictions
  ENDIF
  IF C32(K4)=C34(K4) AND C34(K4)=C13(1)
   LET K23=K23+1                                 # no of true positives
  ENDIF
  IF C32(K4)=C34(K4) AND C34(K4)=C14(1)
   LET K24=K24+1                                 # no of true negatives
  ENDIF
  IF C32(K4)=C13(1) AND C34(K4)=C15(1)     # correction for non-qualifiers
   LET K132=K132+1
  ENDIF
  IF C32(K4)=C14(1) AND C34(K4)=C15(1)     # correction for non-qualifiers
   LET K142=K142+1
  ENDIF

IF K7=2
 GOTO 1
ENDIF

  # LABORATORY 3

  IF C32(K4)=C35(K4)
   LET K32=K32+1                                  # total no of correct predictions
  ENDIF
  IF C32(K4)=C35(K4) AND C35(K4)=C13(1)
   LET K33=K33+1                                  # no of true positives
  ENDIF
  IF C32(K4)=C35(K4) AND C35(K4)=C14(1)
   LET K34=K34+1                                  # no of true negatives
  ENDIF
  IF C32(K4)=C13(1) AND C35(K4)=C15(1)     # correction for non-qualifiers
```

```
  LET K133=K133+1
 ENDIF
 IF C32(K4)=C14(1) AND C35(K4)=C15(1)     # correction for non-qualifiers
  LET K143=K143+1
 ENDIF

IF K7=3
 GOTO 1
ENDIF

# LABORATORY 4

 IF C32(K4)=C36(K4)
  LET K42=K42+1                           # total no of correct predictions
 ENDIF
 IF C32(K4)=C36(K4) AND C36(K4)=C13(1)
  LET K43=K43+1                           # no of true positives
 ENDIF
 IF C32(K4)=C36(K4) AND C36(K4)=C14(1)
  LET K44=K44+1                           # no of true negatives
 ENDIF
 IF C32(K4)=C13(1) AND C36(K4)=C15(1)    # correction for non-qualifiers
  LET K134=K134+1
 ENDIF
 IF C32(K4)=C14(1) AND C36(K4)=C15(1)    # correction for non-qualifiers
  LET K144=K144+1
 ENDIF

IF K7=4
 GOTO 1
ENDIF

# LABORATORY 5

 IF C32(K4)=C37(K4)
  LET K52=K52+1                           # total no of correct predictions
 ENDIF
 IF C32(K4)=C37(K4) AND C37(K4)=C13(1)
  LET K53=K53+1                           # no of true positives
 ENDIF
 IF C32(K4)=C37(K4) AND C37(K4)=C14(1)
  LET K54=K54+1                           # no of true negatives
 ENDIF
 IF C32(K4)=C13(1) AND C37(K4)=C15(1)    # correction for non-qualifiers
  LET K135=K135+1
 ENDIF
 IF C32(K4)=C14(1) AND C37(K4)=C15(1)    # correction for non-qualifiers
  LET K145=K145+1
 ENDIF

IF K7=5
 GOTO 1
ENDIF

# LABORATORY 6

 IF C32(K4)=C38(K4)
  LET K62=K62+1                           # total no of correct predictions
 ENDIF
 IF C32(K4)=C38(K4) AND C38(K4)=C13(1)
```

```
  LET K63=K63+1                              # no of true positives
 ENDIF
 IF C32(K4)=C38(K4) AND C38(K4)=C14(1)
  LET K64=K64+1                              # no of true negatives
 ENDIF
 IF C32(K4)=C13(1) AND C38(K4)=C15(1)     # correction for non-qualifiers
  LET K136=K136+1
 ENDIF
 IF C32(K4)=C14(1) AND C38(K4)=C15(1)     # correction for non-qualifiers
  LET K146=K146+1
 ENDIF

IF K7=6
 GOTO 1
ENDIF

# LABORATORY 7

 IF C32(K4)=C39(K4)
  LET K72=K72+1                             # total no of correct predictions
 ENDIF
 IF C32(K4)=C39(K4) AND C39(K4)=C13(1)
  LET K73=K73+1                             # no of true positives
 ENDIF
 IF C32(K4)=C39(K4) AND C39(K4)=C14(1)
  LET K74=K74+1                             # no of true negatives
 ENDIF
 IF C32(K4)=C13(1) AND C39(K4)=C15(1)    # correction for non-qualifiers
  LET K137=K137+1
 ENDIF
 IF C32(K4)=C14(1) AND C39(K4)=C15(1)    # correction for non-qualifiers
  LET K147=K147+1
 ENDIF

IF K7=7
 GOTO 1
ENDIF

# LABORATORY 8

 IF C32(K4)=C40(K4)
  LET K82=K82+1                             # total no of correct predictions
 ENDIF
 IF C32(K4)=C40(K4) AND C40(K4)=C13(1)
  LET K83=K83+1                             # no of true positives
 ENDIF
 IF C32(K4)=C40(K4) AND C40(K4)=C14(1)
  LET K84=K84+1                             # no of true negatives
 ENDIF
 IF C32(K4)=C13(1) AND C40(K4)=C15(1)    # correction for non-qualifiers
  LET K138=K138+1
 ENDIF
 IF C32(K4)=C14(1) AND C40(K4)=C15(1)    # correction for non-qualifiers
  LET K138=K138+1
 ENDIF

IF K7=8
 GOTO 1
ENDIF
```

```
# LABORATORY 9

  IF C32(K4)=C41(K4)
   LET K92=K92+1                              # total no of correct predictions
  ENDIF
  IF C32(K4)=C41(K4) AND C41(K4)=C13(1)
   LET K93=K93+1                              # no of true positives
  ENDIF
  IF C32(K4)=C41(K4) AND C41(K4)=C14(1)
   LET K94=K94+1                               # no of true negatives
  ENDIF
  IF C32(K4)=C13(1) AND C41(K4)=C15(1)      # correction for non-qualifiers
   LET K139=K139+1
  ENDIF
  IF C32(K4)=C14(1) AND C41(K4)=C15(1)      # correction for non-qualifiers
   LET K149=K149+1
  ENDIF

MLABEL 1
ENDDO                                        # stop chemicals routine

ERASE C32-C41

# Laboratory 1

 LET C45(K2)=(K12/(K1-K131-K141))*100     # concordance
 LET C55(K2)=(K13/(K8-K131))*100          # sensitivity
 LET C65(K2)=(K14/(K9-K141))*100          # specificity

# Laboratory 2

 LET C46(K2)=(K22/(K1-K132-K142))*100     # concordance
 LET C56(K2)=(K23/(K8-K132))*100          # sensitivity
 LET C66(K2)=(K24/(K9-K142))*100          # specificity

# Laboratory 3

 LET C47(K2)=(K32/(K1-K133-K143))*100     # concordance
 LET C57(K2)=(K33/(K8-K133))*100          # sensitivity
 LET C67(K2)=(K34/(K9-K143))*100          # specificity

# Laboratory 4

 LET C48(K2)=(K42/(K1-K134-K144))*100     # concordance
 LET C58(K2)=(K43/(K8-K134))*100          # sensitivity
 LET C68(K2)=(K44/(K9-K144))*100          # specificity

# Laboratory 5

 LET C49(K2)=(K52/(K1-K135-K145))*100     # concordance
 LET C59(K2)=(K53/(K8-K135))*100          # sensitivity
 LET C69(K2)=(K54/(K9-K145))*100          # specificity

# Laboratory 6

 LET C50(K2)=(K62/(K1-K136-K146))*100     # concordance
 LET C60(K2)=(K63/(K8-K136))*100          # sensitivity
 LET C70(K2)=(K64/(K9-K146))*100          # specificity

IF K7=6
```

```
 GOTO 2
ENDIF

# Laboratory 7

 LET C51(K2)=(K72/(K1-K137-K147))*100     # concordance
 LET C61(K2)=(K73/(K8-K137))*100          # sensitivity
 LET C71(K2)=(K74/(K9-K147))*100          # specificity

IF K7=7
 GOTO 2
ENDIF

# Laboratory 8

 LET C52(K2)=(K82/(K1-K138-K148))*100     # concordance
 LET C62(K2)=(K83/(K8-K138))*100          # sensitivity
 LET C72(K2)=(K84/(K9-K148))*100          # specificity

IF K7=8
 GOTO 2
ENDIF

# Laboratory 9

 LET C53(K2)=(K92/(K1-K139-K149))*100     # concordance
 LET C63(K2)=(K93/(K8-K139))*100          # sensitivity
 LET C73(K2)=(K94/(K9-K149))*100          # specificity

MLABEL 2
ENDDO                                     # stop bootstrap routine

# Average statistics over laboratories

NAME C80='MEAN CONC' C81='MEAN SENS' C82='MEAN SPEC'

IF K7=1
 LET C80=C45
 LET C81=C55
 LET C82=C65
ENDIF

IF K7=2
 LET C80=(C45+C46)/K7
 LET C81=(C55+C56)/K7
 LET C82=(C65+C66)/K7
ENDIF

IF K7=3
 LET C80=(C45+C46+C47)/K7
 LET C81=(C55+C56+C57)/K7
 LET C82=(C65+C66+C67)/K7
ENDIF

IF K7=4
 LET C80=(C45+C46+C47+C48)/K7
 LET C81=(C55+C56+C57+C58)/K7
 LET C82=(C65+C66+C67+C68)/K7
ENDIF
```

```
IF K7=5
 LET C80=(C45+C46+C47+C48+C49)/K7
 LET C81=(C55+C56+C57+C58+C59)/K7
 LET C82=(C65+C66+C67+C68+C69)/K7
ENDIF

IF K7=6
 LET C80=(C45+C46+C47+C48+C49+C50)/K7
 LET C81=(C55+C56+C57+C58+C59+C60)/K7
 LET C82=(C65+C66+C67+C68+C69+C70)/K7
ENDIF

IF K7=7
 LET C80=(C45+C46+C47+C48+C49+C50+C51)/K7
 LET C81=(C55+C56+C57+C58+C59+C60+C61)/K7
 LET C82=(C65+C66+C67+C68+C69+C70+C71)/K7
ENDIF

IF K7=8
 LET C80=(C45+C46+C47+C48+C49+C50+C51+C52)/K7
 LET C81=(C55+C56+C57+C58+C59+C60+C61+C62)/K7
 LET C82=(C65+C66+C67+C68+C69+C70+C71+C72)/K7
ENDIF

IF K7=9
 LET C80=(C45+C46+C47+C48+C49+C50+C51+C52+C53)/K7
 LET C81=(C55+C56+C57+C58+C59+C60+C61+C62+C63)/K7
 LET C82=(C65+C66+C67+C68+C69+C70+C71+C72+C73)/K7
ENDIF

NAME C20='BOOTSTRAP MEAN' C21='LOWER CL' C22='UPPER CL'

LET C19(1)="CONCORDANCE"
LET C19(2)="SENSITIVITY"
LET C19(3)="SPECIFICITY"

LET K101=MEAN(C80)
LET K102=MEAN(C81)
LET K103=MEAN(C82)

LET C20(1)=K101
LET C20(2)=K102
LET C20(3)=K103

LET K110=STDEV(C80)
LET K111=STDEV(C81)
LET K112=STDEV(C82)
LET K120=1.96*K110
LET K121=1.96*K111
LET K122=1.96*K112

LET C21(1)=K101-K120
LET C21(2)=K102-K121
LET C21(3)=K103-K122
LET C22(1)=K101+K120
LET C22(2)=K102+K121
LET C22(3)=K103+K122

ENDMACRO
```

A6 BOOTSTRAP ANALYSIS OF COOPER STATISTICS II: APPLICATION OF THE BIAS-CORRECTED PERCENTILE METHOD

```
# GLOBAL MACRO BIAS.MAC
#
# Applies the bias-corrected percentile bootstrap method to bootstrap distributions
# of Cooper statistics generated by BOOT.MAC
#
# MINITAB for Windows version 11 and above
#
# Place the bootstrap distribution of Cooper statistics in C1
# Type %bias at the command line prompt.


GMACRO
bias

NAME C1='Bootstrap distribution' C3='PERCENTILES'

LET K1=COUNT(C1)

NOTE
NOTE What is the value of the Cooper statistic based on the original sample ?
NOTE
SET C10;
 FILE "TERMINAL";
 NOBS=1.
COPY C10 K100

LET K2=0

DO K3=1:K1
 IF C1(K3)>K100
  LET K2=K2+1
 ENDIF
ENDDO

# calculate p, proportion of times that bootstrap estimate > original sample estimate

LET K4=K2/K1

# calculate Zp, value of normal distribution exceeded with probability p

LET K4=1-K4

INVCDF K4 K5;
 NORMAL 0.0 1.0.

# calculate proportion of standard normal distribn less than (2Zp-1.96)

LET K6=(2*K5)-1.96
CDF K6 K8;
 NORMAL 0.0 1.0.


# calculate proportion of standard normal distribn less than (2Zp+1.96)

LET K6=(2*K5)+1.96
CDF K6 K9;
```

```
 NORMAL 0.0 1.0.

LET C3(1)=K8    # lower cumulative probability
LET C3(2)=K9    # upper cumulative probability
LET C3(3)=0.5    # median

LET K8=K8*100
LET K9=K9*100

LET K10=MINIMUM(C1)
LET K11=MAXIMUM(C1)
LET K12=K11/10

SET C10
 1( K10 : K11 / 5 )1
 END.

LET C10=ROUN(C10)

HISTOGRAM C1;
 CUMULATIVE;
 PERCENT;
 MIDPOINT;
 CONNECT;
 AXIS 1;
 AXIS 2;
 TICK 1 C10;
 TICK 2 K8 K9 50;
 GRID 2;
 GRID 1.

ENDMACRO
```

A7 BOOTSTRAP ANALYSIS OF DRAIZE TEST SKIN SCORES

```
# GLOBAL MACRO PII.MAC
#
# Bootstrap resampling of Draize skin scores
#
# MINITAB for Windows version 11 and above
#
# Place the chemical names in C1, and the tissue scores in columns C2-C37, as indicated in column headings  (e.g. Ery12 = erythema score in rabbit no 1 at the second time-point). Maximum of 6 rabbits.
# Type %pii at the command line prompt.

GMACRO
PII

NAME C1='Chemical'
NAME C2='Ery11' C3='Ery12' C4='Ery13'
NAME C5='Oed11' C6='Oed12' C7='Oed13'
NAME C8='Ery21' C9='Ery22' C10='Ery23'
NAME C11='Oed21' C12='Oed22' C13='Oed23'

NAME C50='Erythema' C51='Oedema'
NAME C60='Resamp Eryth' C61='Resamp Oed' C62='Resamp PII'

NOTE
NOTE How many rabbits (1-6) ?
SET C100;
 FILE "TERMINAL";
 NOBS=1.
COPY C100 K1                                    # no of rabbits K1
ERASE C100

IF K1=1
 GOTO 1
ELSEIF K1=2
 GOTO 2
ELSEIF K1=3
 GOTO 3
ELSEIF K1=4
 GOTO 4
ELSEIF K1=5
 GOTO 5
ELSEIF K1=6
 GOTO 6
ENDIF

MLABEL 1                                        # One rabbit
COPY C2-C4 M1
COPY C5-C7 M2
GOTO 10

MLABEL 2                                         # Two rabbits
COPY C2-C4, C8-C10 M1
COPY C5-C7, C11-C13 M2
GOTO 10

MLABEL 3                                        # Three rabbits
NAME C14='Ery31' C15='Ery32' C16='Ery33'
```

```
NAME C17='Oed31' C18='Oed32' C19='Oed33'

COPY C2-C4, C8-C10, C14-C16 M1
COPY C5-C7, C11-C13, C17-C19 M2
GOTO 10

MLABEL 4                                    # Four rabbits
NAME C20='Ery41' C21='Ery42' C22='Ery43'
NAME C23='Oed41' C24='Oed42' C25='Oed43'
COPY C2-C4, C8-C10, C14-C16, C20-C22 M1
COPY C5-C7, C11-C13, C17-C19, C23-C25 M2
GOTO 10

MLABEL 5                                    # Five rabbits
NAME C20='Ery41' C21='Ery42' C22='Ery43'
NAME C23='Oed41' C24='Oed42' C25='Oed43'
NAME C26='Ery51' C27='Ery52' C28='Ery53'
NAME C29='Oed51' C30='Oed52' C31='Oed53'
COPY C2-C4, C8-C10, C14-C16, C20-C22, C26-C28 M1
COPY C5-C7, C11-C13, C17-C19, C23-C25, C29-C31 M2
GOTO 10

MLABEL 6                                    # Six rabbits
NAME C20='Ery41' C21='Ery42' C22='Ery43'
NAME C23='Oed41' C24='Oed42' C25='Oed43'
NAME C26='Ery51' C27='Ery52' C28='Ery53'
NAME C29='Oed51' C30='Oed52' C31='Oed53'
NAME C32='Ery61' C33='Ery62' C34='Ery63'
NAME C35='Oed61' C36='Oed62' C37='Oed63'
COPY C2-C4, C8-C10, C14-C16, C20-C22, C26-C28, C32-C34 M1
COPY C5-C7, C11-C13, C17-C19, C23-C25, C29-C31, C35-C37 M2
GOTO 10

MLABEL 10
TRANSPOSE M1 M11
TRANSPOSE M2 M12
COPY M11 C50
COPY M12 C51
LET K1=3*K1

NOTE
NOTE How many times do you want to resample?
SET C100;
 FILE "TERMINAL";
 NOBS=1.
COPY C100 K3                        # K3 = no of bootstrap resamples
ERASE C100

DO K2=1:K3                          # start bootstrap routine
 SAMPLE K1 C50-C51 C55-C56;
  REPLACE.
 LET C60(K2)=MEAN(C55)
 LET C61(K2)=MEAN(C56)
 LET C62(K2)=C60(K2)+C61(K2)
ENDDO                               # stop bootstrap routine

LET K10=MEAN(C60)
LET K11=MEAN(C61)
LET K12=MEAN(C62)
LET K15=STDEV(C60)
```

```
LET K16=STDEV(C61)
LET K17=STDEV(C62)

LET K18=MEDIAN(C60)
LET K19=MEDIAN(C61)
LET K20=MEDIAN(C62)
LET C1(4)="Erythema"
LET C1(5)="Oedema"
LET C1(6)="PII"
LET C2(4)=K10
LET C2(5)=K11
LET C2(6)=K12
LET C3(4)=K15
LET C3(5)=K16
LET C3(6)=K17
LET C4(4)=K10-(1.96*K15)
LET C4(5)=K11-(1.96*K16)
LET C4(6)=K12-(1.96*K17)
LET C5(4)=K10+(1.96*K15)
LET C5(5)=K11+(1.96*K16)
LET C5(6)=K12+(1.96*K17)
LET C7(4)=K18
LET C7(5)=K19
LET C7(6)=K20

STATS C60-C62;
 QONE C70-C72.
LET C8(4)=C70
LET C8(5)=C71
LET C8(6)=C72

ERASE C70-C72

STATS C60-C62;
 QTHREE C70-C72.
LET C9(4)=C70
LET C9(5)=C71
LET C9(6)=C72

ERASE C70-C72

TEXT C2 C2
TEXT C3 C3
TEXT C4 C4
TEXT C5 C5
TEXT C7 C7
TEXT C8 C8
TEXT C9 C9

LET C2(3)="mean"
LET C3(3)="std err"
LET C4(3)="lower CI"
LET C5(3)="upper CI"
LET C7(3)="median"
LET C8(3)="Q1"
LET C9(3)="Q3"

ENDMACRO
```

A8 BOOTSTRAP ANALYSIS OF DRAIZE TEST EYE SCORES

```
# GLOBAL MACRO MMAS.MAC
#
# Bootstrap resampling of Draize eye scores
#
# MINITAB for Windows version 11 and above
#
# Place the chemical names in C1, and the tissue scores in columns C2-C109, as indicated.
# (e.g. red12 = redness score in rabbit no 1 at the secondtime-point). Maximum of 6 rabbits.
# Type %mmas at the command line prompt.

GMACRO
MMAS

NAME C1='Chemical'
NAME C2='red11' C3='red12' C4='red13'
NAME C5='red21' C6='red22' C7='red23'
NAME C8='red31' C9='red32' C10='red33'
NAME C11='red41' C12='red42' C13='red43'
NAME C14='red51' C15='red52' C16='red53'
NAME C17='red61' C18='red62' C19='red63'

NAME C20='chem11' C21='chem12' C22='chem13'
NAME C23='chem21' C24='chem22' C25='chem23'
NAME C26='chem31' C27='chem32' C28='chem33'
NAME C29='chem41' C30='chem42' C31='chem43'
NAME C32='chem51' C33='chem52' C34='chem53'
NAME C35='chem61' C36='chem62' C37='chem63'

NAME C38='dis11' C39='dis12' C40='dis13'
NAME C41='dis21' C42='dis22' C43='dis23'
NAME C44='dis31' C45='dis32' C46='dis33'
NAME C47='dis41' C48='dis42' C49='dis43'
NAME C50='dis51' C51='dis52' C52='dis53'
NAME C53='dis61' C54='dis62' C55='dis63'

NAME C56='opac11' C57='opac12' C58='opac13'
NAME C59='opac21' C60='opac22' C61='opac23'
NAME C62='opac31' C63='opac32' C64='opac33'
NAME C65='opac41' C66='opac42' C67='opac43'
NAME C68='opac51' C69='opac52' C70='opac53'
NAME C71='opac61' C72='opac62' C73='opac63'

NAME C74='area11' C75='area12' C76='area13'
NAME C77='area21' C78='area22' C79='area23'
NAME C80='area31' C81='area32' C82='area33'
NAME C83='area41' C84='area42' C85='area43'
NAME C86='area51' C87='area52' C88='area53'
NAME C89='area61' C90='area62' C91='area63'

NAME C92='iris11' C93='iris12' C94='iris13'
NAME C95='iris21' C96='iris22' C97='iris23'
NAME C98='iris31' C99='iris32' C100='iris33'
NAME C101='iris41' C102='iris42' C103='iris43'
NAME C104='iris51' C105='iris52' C106='iris53'
NAME C107='iris61' C108='iris62' C109='iris63'

NAME C115='redness' C116='chemosis' C117='discharge' C118='opacity' C119='area' C120='iris'
```

```
NAME C125='resamp red' C126='resamp chem' C127='resamp dis' C128='resamp opac' C129='resamp area' C130='resamp iris'

NOTE
NOTE How many rabbits (1-6) ?
SET C200;
 FILE "TERMINAL";
 NOBS=1.
COPY C200 K1                    # no of rabbits K1
ERASE C200

IF K1=1
 GOTO 1
ELSEIF K1=2
 GOTO 2
ELSEIF K1=3
 GOTO 3
ELSEIF K1=4
 GOTO 4
ELSEIF K1=5
 GOTO 5
ELSEIF K1=6
 GOTO 6
ENDIF

MLABEL 1                        # One rabbits
COPY C2-C4 M1
COPY C20-C22 M2
COPY C38-C40 M3
COPY C56-C58 M4
COPY C74-C76 M5
COPY C92-C94 M6
GOTO 10

MLABEL 2                        # Two rabbits
COPY C2-C7 M1
COPY C20-C25 M2
COPY C38-C43 M3
COPY C56-C61 M4
COPY C74-C79 M5
COPY C92-C97 M6
GOTO 10

MLABEL 3                        # Three rabbits
COPY C2-C10 M1
COPY C20-C28 M2
COPY C38-C46 M3
COPY C56-C64 M4
COPY C74-C82 M5
COPY C92-C100 M6
GOTO 10

MLABEL 4                        # Four rabbits
COPY C2-C13 M1
COPY C20-C31 M2
COPY C38-C49 M3
COPY C56-C67 M4
COPY C74-C85 M5
COPY C92-C103 M6
GOTO 10
```

```
MLABEL 5                                # Five rabbits
COPY C2-C16 M1
COPY C20-C34 M2
COPY C38-C52 M3
COPY C56-C70 M4
COPY C74-C88 M5
COPY C92-C106 M6
GOTO 10

MLABEL 6                                # Six rabbits
COPY C2-C19 M1
COPY C20-C37 M2
COPY C38-C55 M3
COPY C56-C73 M4
COPY C74-C91 M5
COPY C92-C109 M6
GOTO 10

MLABEL 10
TRANSPOSE M1 M11
TRANSPOSE M2 M12
TRANSPOSE M3 M13
TRANSPOSE M4 M14
TRANSPOSE M5 M15
TRANSPOSE M6 M16
COPY M11 C115
COPY M12 C116
COPY M13 C117
COPY M14 C118
COPY M15 C119
COPY M16 C120

LET K1=3*K1

NOTE
NOTE How many times do you want to resample?
SET C200;
 FILE "TERMINAL";
 NOBS=1.
COPY C200 K3                            # K3 = no of bootstrap resamples
ERASE C200

DO K2=1:K3                              # start bootstrap routine
 SAMPLE K1 C115-C120 C125-C130;
  REPLACE.
 LET C140(K2)=MEAN(C125)                # redness distribution
 LET C141(K2)=MEAN(C126)                # chemosis distribution
 LET C142(K2)=MEAN(C127)                # discharge distribution
 LET C143(K2)=MEAN(C128)                # opacity distribution
 LET C144(K2)=MEAN(C129)                # area distribution
 LET C145(K2)=MEAN(C130)                # iritis distribution
ENDDO                                   # stop bootstrap routine

LET C146=(2*(C140+C141+C142))+(5*(C143*C144))+(5*C145)        # MMAS distribution

NAME C140='red dist' C141='chem dist' C142='disc dist' C143='opac dist'
NAME C144='area dist' C145='iris dist' C146='MAS dist'
```

```
LET K10=MEAN(C140)
LET K11=MEAN(C141)
LET K12=MEAN(C142)
LET K13=MEAN(C143)
LET K14=MEAN(C144)
LET K15=MEAN(C145)
LET K16=MEAN(C146)

LET K17=STDEV(C140)
LET K18=STDEV(C141)
LET K19=STDEV(C142)
LET K20=STDEV(C143)
LET K21=STDEV(C144)
LET K22=STDEV(C145)
LET K23=STDEV(C146)

LET K25=MEDIAN(C140)
LET K26=MEDIAN(C141)
LET K27=MEDIAN(C142)
LET K28=MEDIAN(C143)
LET K29=MEDIAN(C144)
LET K30=MEDIAN(C145)
LET K31=MEDIAN(C146)

LET K32=MAXIMUM(C140)
LET K33=MAXIMUM(C141)
LET K34=MAXIMUM(C142)
LET K35=MAXIMUM(C143)
LET K36=MAXIMUM(C144)
LET K37=MAXIMUM(C145)
LET K38=(2*(K32+K33+K34))+(5*(K35*K36))+(5*K37)

LET C1(4)="redness"
LET C1(5)="chemosis"
LET C1(6)="discharge"
LET C1(7)="opacity"
LET C1(8)="area"
LET C1(9)="iritis"
LET C1(10)="MMAS"

LET C2(4)=K10
LET C2(5)=K11
LET C2(6)=K12
LET C2(7)=K13
LET C2(8)=K14
LET C2(9)=K15
LET C2(10)=K16

LET C3(4)=K17
LET C3(5)=K18
LET C3(6)=K19
LET C3(7)=K20
LET C3(8)=K21
LET C3(9)=K22
LET C3(10)=K23

LET C4(4)=K10-(1.96*K17)
LET C4(5)=K11-(1.96*K18)
LET C4(6)=K12-(1.96*K19)
LET C4(7)=K13-(1.96*K20)
```

```
LET C4(8)=K14-(1.96*K21)
LET C4(9)=K15-(1.96*K22)
LET C4(10)=K16-(1.96*K23)

LET C5(4)=K10+(1.96*K17)
LET C5(5)=K11+(1.96*K18)
LET C5(6)=K12+(1.96*K19)
LET C5(7)=K13+(1.96*K20)
LET C5(8)=K14+(1.96*K21)
LET C5(9)=K15+(1.96*K22)
LET C5(10)=K16+(1.96*K23)

LET C7(4)=K25
LET C7(5)=K26
LET C7(6)=K27
LET C7(7)=K28
LET C7(8)=K29
LET C7(9)=K30
LET C7(10)=K31

STATS C140-C146;
 QONE C150-C156.
LET C8(4)=C150
LET C8(5)=C151
LET C8(6)=C152
LET C8(7)=C153
LET C8(8)=C154
LET C8(9)=C155
LET C8(10)=C156

ERASE C150-C156

STATS C140-C146;
 QTHREE C150-C156.
LET C9(4)=C150
LET C9(5)=C151
LET C9(6)=C152
LET C9(7)=C153
LET C9(8)=C154
LET C9(9)=C155
LET C9(10)=C156

LET C10(4)=K32
LET C10(5)=K33
LET C10(6)=K34
LET C10(7)=K35
LET C10(8)=K36
LET C10(9)=K37
LET C10(10)=K38

ERASE C150-C156

TEXT C2 C2
TEXT C3 C3
TEXT C4 C4
TEXT C5 C5
TEXT C7 C7
TEXT C8 C8
TEXT C9 C9
TEXT C10 C10
```

```
LET C2(3)="mean"
LET C3(3)="std err"
LET C4(3)="lower CI"
LET C5(3)="upper CI"
LET C7(3)="median"
LET C8(3)="Q1"
LET C9(3)="Q3"
LET C10(3)="max"

ENDMACRO
```

A9 EMBEDDED CLUSTER MODELLING

```
# GLOBAL MACRO ECM2.MAC
#
# Carries out embedded cluster modelling on two variables
#
# MINITAB for Windows version 11 and above
#
# Place the chemical names (or numbers) in C1, the activity (0/1) in C2
# and the variables in C3 and C4.
# Type %ecm2 at the command line prompt.

GMACRO
ECM2

NAME C1='CHEMICAL' C2='ACTIVITY (0/1)' C3='DESCRIP 1' C4='DESCRIP 2'
NAME C11='CENTROID' C12='RADIUS'
NAME C21='ACTIVE 1' C23='ACTIVE 2'
LET C10(1)="descrip 1"
LET C10(2)="descrip 2"

LET K1=COUNT (C1)

NOTE
NOTE Do you wish to rotate the axes by PCA (y/n) ?
NOTE
YESNO K50
IF K50=1
 CALL ROTATE
ELSE
 CALL CONTINUE
ENDIF

ENDMACRO

GMACRO
ROTATE

NAME C51='PC1' C52='PC2'

UNSTACK C3 C20 C21;
 SUBSCRIPTS C2.
ERASE C20
UNSTACK C4 C22 C23;
 SUBSCRIPTS C2.
ERASE C22

PCA C21 C23;
 NCOMPONENTS 2;
 COVARIANCE;
 COEFFICIENTS C51-C52.

COPY C3-C4 C90-C91
ERASE C3-C4
NAME C3='ROTATED 1' C4='ROTATED 2'

DO K2=1:K1
 LET K100=C90(K2)*C51(1)
 LET K101=C91(K2)*C51(2)
```

```
 LET K103=K100+K101
 LET C3(K2)=K103
 LET K105=C90(K2)*C52(1)
 LET K106=C91(K2)*C52(2)
 LET K108=K105+K106
 LET C4(K2)=K108
ENDDO

CALL CONTINUE
ENDMACRO

GMACRO
CONTINUE

UNSTACK C3 C20 C21;
 SUBSCRIPTS C2.
ERASE C20
UNSTACK C4 C22 C23;
 SUBSCRIPTS C2.
ERASE C22

LET K3=MEAN(C21)
LET K4=MEAN(C23)
LET K6=STDEV(C21)
LET K7=STDEV(C23)
LET K9=1.96*K6
LET K10=1.96*K7

LET C11(1)=K3
LET C11(2)=K4
LET C12(1)=K9
LET C12(2)=K10

NAME C7='PRED ACTIVITY'
NAME C8='PRED (XVAL)'

DO K2=1:K1
 LET K12=(C3(K2)-K3)**2
 LET K13=(C4(K2)-K4)**2
 LET K15=K12/(K9**2)
 LET K16=K13/(K10**2)
 LET K18=K15+K16
 IF K18<=1
  LET C7(K2)=1
 ELSE
  LET C7(K2)=0
 ENDIF
ENDDO

# Cross-validation routine

NAME C32='XVAL ACT' C33='XVAL 1' C34='XVAL 2'

DO K2=1:K1
 COPY C2-C4 C32-C34;
  OMIT K2.
 UNSTACK C33 C40 C41;
  SUBSCRIPTS C32.
 ERASE C40
 UNSTACK C34 C42 C43;
```

```
 SUBSCRIPTS C32.
 ERASE C42
 LET K3=MEAN(C41)
 LET K4=MEAN(C43)
 LET K6=STDEV(C41)
 LET K7=STDEV(C43)
 LET K9=1.96*K6
 LET K10=1.96*K7
 LET K12=(C3(K2)-K3)**2
 LET K13=(C4(K2)-K4)**2
 LET K15=K12/(K9**2)
 LET K16=K13/(K10**2)
 LET K18=K15+K16
 IF K18<=1
  LET C8(K2)=1
 ELSE
  LET C8(K2)=0
 ENDIF
ENDDO

NOTE
NOTE Do you wish to bootstrap the parameter estimates (y/n) ?
NOTE
YESNO K50
IF K50=1
 CALL RESAMPLE
ELSE
 EXIT
ENDIF

ENDMACRO

GMACRO
RESAMPLE
NOTE
NOTE How many times do you wish to resample ?
NOTE
SET C50;
 FILE "TERMINAL";
 NOBS=1.
COPY C50 K50
LET K60=SUM(C2)

NAME C14='LOWER CENTROID' C15='UPPER CENTROID' C16='LOWER RADIUS' C17='UPPER
RADIUS'
NAME C22='RESAMP ACT1' C24='RESAMP ACT2'
NAME C42='CENTROID1 DISTRIBN' C43='CENTROID2 DISTRIBN'
NAME C45='RADIUS1 DISTRIBN' C46='RADIUS2 DISTRIBN'

DO K2=1:K50
 SAMPLE K60 C21 C23 C22 C24 ;
  REPLACE.
 LET C42(K2)=MEAN(C22)
 LET C43(K2)=MEAN(C24)
 LET C45(K2)=1.96*STDEV(C22)
 LET C46(K2)=1.96*STDEV(C24)
ENDDO

LET K70=MEAN(C42)
LET K71=MEAN(C43)
```

```
LET K73=MEAN(C45)
LET K74=MEAN(C46)

LET C14(1)=K70-(1.96*STDEV(C42))
LET C14(2)=K71-(1.96*STDEV(C43))
LET C15(1)=K70+(1.96*STDEV(C42))
LET C15(2)=K71+(1.96*STDEV(C43))
LET C16(1)=K73-(1.96*STDEV(C45))
LET C16(2)=K74-(1.96*STDEV(C46))
LET C17(1)=K73+(1.96*STDEV(C45))
LET C17(2)=K74+(1.96*STDEV(C46))

ENDMACRO
```

A10 BOOTSTRAP ANALYSIS OF REGRESSION PARAMETERS

```
# GLOBAL MACRO REG2.MAC
#
# Performs bootstrap resampling of the parameters in two-variable
# regression
#
# MINITAB for Windows version 11 and above
#
# Place the chemical names in C1, the response in C2, and the predictor
# variables in C3 and C4.
# Type %reg2 at the command line prompt.

GMACRO
REG2

NAME C1='CHEMICAL' C2='RESPONSE' C3='DESCRIPTOR 1' C4='DESCRIPTOR 2'
NAME C9='CONSTANT' C10='COEFF 1' C11='COEFF 2' C15='CONST' C16='COEFF1'
C17='COEFF2'

NOTE
NOTE How many observations do you want to sample?
SET C20;
 FILE "TERMINAL";
 NOBS=1.
COPY C20 K1

NOTE
NOTE How many times do you want to resample?
SET C20;
 FILE "TERMINAL";
 NOBS=1.
COPY C20 K2

DO K5=1:K2
 SAMPLE K1 C2 C3 C4 C22 C23 C24
 REGRESS C22 2 C23 C24;
  COEFFICIENTS C30;
  CONSTANT.
 LET K6=C30(1)
 LET K7=C30(2)
 LET K8=C30(3)
 LET C15(K5)=K6
 LET C16(K5)=K7
 LET C17(K5)=K8
ENDDO

LET C8(1)="median"
LET C8(2)="std dev"

LET C9(1)=MEDIAN(C15)
LET C10(1)=MEDIAN(C16)
LET C11(1)=MEDIAN(C17)
LET C9(2)=STDEV(C15)
LET C10(2)=STDEV(C16)
LET C11(2)=STDEV(C17)

ENDMACRO
```

APPENDIX B
LIST OF PUBLICATIONS

B1 ARTICLES RELEVANT TO THE THESIS

Balls, M., Berg, N., Bruner, L.H., Curren, R.D., de Silva, O., Earl, L.K., Esdaile, D.J., Fentem, J.H., Liebsch, M., Ohno, Y., Prinsen, M.K., Spielmann, H. & Worth, A.P. (1999). Eye irritation testing: the way forward. The report and recommendations of ECVAM workshop 33. *Alternatives to Laboratory Animals* **27**, 53-77.

Barratt, M.D., Brantom, P.G., Fentem, J.H., Gerner, I., Walker, A.P. & Worth, A.P. (1998). The ECVAM international validation study on *in vitro* tests for skin corrosivity. 1. Selection and distribution of the test chemicals. *Toxicology in Vitro* **12**, 471-482.

Worth, A.P., Barratt, M.D. & Houston, J.B. (1998). The validation of computational prediction techniques. *Alternatives to Laboratory Animals* **26**, 241-247.

Worth, A.P. & Cronin, M.T.D. (1999). Embedded cluster modelling – a novel method for analysing embedded data sets. *Quantitative Structure-Activity Relationships* **18**, 229-235.

Worth, A.P. & Cronin, M.T.D. (2000). Structure-permeability relationships for transcorneal penetration. *Alternatives to Laboratory Animals* **28**, 403-413.

Worth, A.P. & Cronin, M.T.D. (2000). Embedded cluster modelling: a novel QSAR method for generating elliptic models of biological activity. In *Progress in the Reduction, Refinement and Replacement of Animal Experimentation* (eds. M. Balls, A-M. van Zeller & M.E. Halder). pp. 479-491. Elsevier Science, Amsterdam.

Worth, A.P. & Fentem, J.H. (1999). A general approach for evaluating stepwise testing strategies. *Alternatives to Laboratory Animals* **27**, 161-177.

Worth, A.P., Fentem, J.H., Balls, M., Botham, P.A., Curren, R.D., Earl, L.K., Esdaile, D.J. & Liebsch, M. (1998). An evaluation of the proposed OECD testing strategy for skin corrosion. *Alternatives to Laboratory Animals* **26**, 709-720.

B2 OTHER PUBLICATIONS (INCLUDING POSTER PRESENTATIONS)

Bremer, S., Worth, A.P., Paparella, M., Hescheler, J., Fleischmann, B.K. & Kolossov, E. (1999). Establishment of an embryotoxic screening system using genetically engineered embryonic stem cells. Poster presentation at the Third World Congress on Alternatives and Animal Use in the Life Sciences. Bologna, Italy; 29 August – 2 September 1999. *Alternatives to Laboratory Animals* **27** (Special Issue, July 1999), 292.

Coecke S., Bogni A., Langezaal I., **Worth A.**, Hartung, T., Monshouwer M., & Balls M. (2000). Integrated testing strategies for newly developed technologies in systemic toxicity testing. INVITOX 2000. Invited presentation at the Meeting of the European Society of Toxicology in Vitro. Alicante, Spain; 25-28 October, 2000.

Cronin M.T.D., Duffy J.C. & **Worth A.P.** (2000). QSARs for the prediction of toxic effects to the liver. Poster presentation at the Thirteenth European Symposium on Quantitative Structure-Activity Relationships. Düsseldorf, Germany; 27 August – 1 September, 2000.

Cronin M.T.D., Duffy J.C. & **Worth A.P.** (2000). QSARs for the hepatotoxicity of aliphatic alcohols. Poster presentation at the Ninth International Workshop on Quantitative Structure Activity Relationships in Environmental Sciences. Bourgas, Bulgaria; 16-20 September, 2000.

Dearden, J.C., Barratt, M.D., Benigni, R., Bristol, D.W., Combes, R.D., Cronin, M.T.D., Judson, P.N., Payne, M.P., Richard, A.M., Tichy, M., Worth A.P. & Yourick, J.J. (1997). The development and validation of expert systems for predicting toxicity. The report and recommendations of an ECVAM/ECB workshop (ECVAM workshop 24). *Alternatives to Laboratory Animals* **25**, 223-252.

Lindgren, F., Nouwen, J., Loonen, H., Worth, A., Hansen, B. & Karcher, W. (1996). Environmental modelling based on a structural fragments approach. *Indoor & Built Environment* **5**, 334-340.

Verburgh, J., Hansen, B., Worth, A. & Karcher, W. (1996). Risk assessment of PAH-containing petroleum substances in the EU. *Polycyclic Aromatic Compounds* **11**, 169-176.