

Eye Irritation Testing: The Way Forward

The Report and Recommendations of ECVAM Workshop 34^{1,2}

Michael Balls,³ Ninna Berg,⁴ Leon H. Bruner,⁵ Rodger D. Curren,⁶ Odile de Silva,⁷ Lesley K. Earl,⁸ David J. Esdaile,⁹ Julia H. Fentem,³ Manfred Liebsch,¹⁰ Yasuo Ohno,¹¹ Menk K. Prinsen,¹² Horst Spielmann¹⁰ and Andrew P. Worth³

³ECVAM, JRC Environment Institute, 21020 Ispra (VA), Italy; ⁴Novo Nordisk, Novo Alle, 2880 Bagsvaerd, Denmark; ⁵Procter & Gamble (Health and Beauty Care), Lovett House, Lovett Road, Staines, Middlesex TW18 3AZ, UK; ⁶Institute for In Vitro Sciences, 21 Firstfield Road, Gaithersburg, MD 20878, USA; ⁷L'Oréal Recherche Avancée — Sciences du Vivant, 1 Avenue Eugène Schueller, 93601 Aulnay-sous-Bois Cedex, France; ⁸SEAC Toxicology Unit, Unilever Research, Colworth House, Sharnbrook, Bedford MK44 1LQ, UK; ⁹Rhône-Poulenc, 355 rue Dostoievski, 06903 Sophia Antipolis Cedex, France; ¹⁰ZEBET, BgVV, Diedersdorfer Weg 1, 12277 Berlin, Germany; ¹¹Division of Pharmacology, Biological Safety Research Centre, National Institute of Health Sciences, 1-18-1 Kamiyoga, Setakaya-ku, Tokyo 158, Japan; ¹²TNO Nutrition and Food Research Institute, Division of Toxicology, 3700 AJ Zeist, The Netherlands

Preface

This is the report of the thirty-fourth of a series of workshops organised by the European Centre for the Validation of Alternative Methods (ECVAM). ECVAM's main goal, as defined in 1993 by its Scientific Advisory Committee, is to promote the scientific and regulatory acceptance of alternative methods which are of importance to the biosciences and which reduce, refine or replace the use of laboratory animals. One of the first priorities set by ECVAM was the implementation of procedures which would enable it to become well-informed about the state-of-the-art of non-animal test development and validation, and the potential for the possible incorporation of alternative tests into regulatory procedures. It was decided that this would be best achieved by the organisation of ECVAM workshops on specific topics, at which small groups of invited experts would review the

current status of various types of *in vitro* tests and their potential uses, and make recommendations about the best ways forward (1).

The workshop on Eye Irritation Testing: The Way Forward was held in Egham, UK, on 15–17 June 1998, under the chairmanship of Michael Balls (ECVAM, Italy). The workshop had two aims, the first of which was to review some of the previous multilaboratory validation studies on alternatives to the Draize eye test and assess why many promising alternative methods were not successful in these studies. The second aim was to discuss strategies for making progress toward the short-term reduction, refinement, and eventual replacement, of the Draize test, including: a new approach to the validation of *in vitro* tests for eye irritancy, based on the use of reference standards, which promises to overcome some of the problems encountered in previous studies; the use of

Address for correspondence and reprints: Professor Michael Balls, ECVAM, TP 580, JRC Institute for Health & Consumer Protection, 21020 Ispra (VA), Italy.

¹ECVAM — European Centre for the Validation of Alternative Methods. ²This document represents the agreed report of the participants as individual scientists.

stepwise testing strategies which reduce and refine the use of animals in eye irritation testing; the use of multivariate and other statistical techniques for the further analysis of data generated in previous validation studies; and a programme of research aimed at understanding the underlying mechanisms of eye irritation.

Introduction

The Draize rabbit test (2) continues to be the method of choice for the regulatory assessment of eye irritation hazard (3, 4), despite criticism on both scientific and animal welfare grounds. Continued use of the Draize test is not due to a shortage of potentially useful alternative methods, since more effort has probably been put into the development of alternatives to the Draize eye irritation test than into seeking replacements for all the other acute *in vivo* toxicity tests put together. However, no test, combination of tests, or testing strategy has yet been developed which meets all the requirements of the regulatory authorities. There are several possible reasons for this, one of which is that the *in vivo* test, being based on the subjective scoring of tissue lesions in the eye, provides variable estimates of eye irritancy (5). Other reasons for the outcomes of recently completed validation studies could be related to: a) the adequacy of the non-animal method protocols; b) the choice of test substances; and c) the choice of statistical approaches for analysing the data. Thus, although there is much confidence that a number of the alternative tests and testing strategies do work in-house, it has proved impossible to establish this satisfactorily by conducting validation studies in which *in vitro* test results are compared with historical *in vivo* data (6, 7).

To find possible solutions to this impasse, ECVAM organised a workshop on eye irritation testing, which brought together experts in the field from industrial and governmental organisations. The workshop participants reviewed the eye irritation validation studies carried out to date, and discussed a number of initiatives which could lead to the short-term reduction and refinement, and to the long-term replacement, of animal use in eye irritation testing. These initiatives include: a) an evaluation of the use of reference standards (benchmark chemicals) in

the validation process, since this could overcome some of the problems which were encountered in previous studies; b) an evaluation of tiered testing strategies; c) further analyses of the data generated during previous validation studies; and d) research into the mechanistic basis of eye irritation. This report summarises the workshop discussions, and presents the conclusions and recommendations of the workshop participants.

Review of Validation Studies

The EC/HO study

The validation study which has become known as the European Commission/British Home Office (EC/HO) study (6) was set up in the light of an EC-funded pilot study (8, 9) to establish whether one or more of nine tests could be used to replace the Draize test for: a) all severely irritating substances or for severely irritating substances belonging to specific chemical classes; and b) for all levels of eye irritancy with or without regard to the chemical class. The nine tests included four cell culture methods (based on red blood cell [RBC] haemolysis, neutral red uptake [NRU], fluorescein leakage [FL], and the use of the silicon microphysiometer [SM]), three *ex vivo* tests (the isolated rabbit eye [IRE] test, the isolated chicken eye [ICE] test, and the bovine corneal opacity/permeability [BCOP] test), the hen's egg chorio-allantoic membrane (HET-CAM) test, and a physico-chemical method based on protein precipitation (EYTEX™).

The relevance and reliability of the nine test methods were assessed under blind conditions by using a test set of 60 single chemicals, which were independently selected, coded and supplied to 37 laboratories. The data generated by the laboratories were analysed independently. The reliability of each test was assessed by determining the interlaboratory Pearson's correlation coefficients of the *in vitro* scores for each endpoint. These analyses indicated that there was good reproducibility between the laboratories conducting the same test. The predictive capacity of each test was assessed by: a) preparing scatter plots showing the relationship between the *in vitro* test scores and *in vivo* irritancy; b) calculating the Pearson's and Spearman's correlation coefficients for the relationship between each alternative

test endpoint and the Modified Maximum Average Draize Test Score (MMAS); and c) deriving a linear regression equation to predict the MMAS from each alternative test endpoint and to determine a 95% confidence interval (CI) for this prediction. These analyses showed that, for the full set of test chemicals, the *in vitro*–*in vivo* correlations were generally low (typically less than 0.6) and the 95% CIs were generally wide (often greater than ± 40 MMAS units). Analyses were also carried out for six (overlapping) subsets of chemicals (30 water-soluble chemicals, 18 water-insoluble chemicals, 12 surfactants, 20 solids, 14 solutions made from solids, and 26 liquids). The results for the surfactant subset were considered to be more encouraging as the correlation coefficients were generally higher (greater than 0.8 for some endpoints) and the 95% CIs tended to be narrower. However, the relatively high correlation coefficients for the surfactants could be partly as a result of dose–response effects, since the “12” surfactants were in fact six different surfactants tested at various concentrations.

In addition to the comparison of single *in vitro* endpoints to the MMAS, a multivariate analysis of the EC/HO study has been undertaken. The purpose of this analysis was to determine whether the combined use of more than one non-animal method could be useful for predicting eye irritation potential, to determine which combinations of tests may provide improved predictions, and to assess the utility of the rabbit eye irritation test for the evaluation of non-animal methods. The multivariate analysis involved an examination of 20 non-animal test measures and *in vivo* scores for 59 test substances. The chemicals were also split into training sets and test sets to avoid overfitting the data. Principal components analysis (PCA) was used to identify the endpoints which explained the greatest variation in the data, and partial least squares (PLS) was used to develop models for predicting eye irritation potential from a combination of *in vitro* test results. An initial analysis was conducted to determine whether a combination of ten endpoints would improve the predictive capacity of the *in vitro* methods. The outcome of this study was encouraging, so additional analyses were conducted to determine whether smaller sets of *in vitro* endpoints would provide improved predictions.

The additional analyses showed that combinations of data from assays of epithelial integrity (FL test), *ex vivo* models (IRE, ICE) and a cytotoxicity test (NRU) explained more of the variability in the data than any single test used alone. The prediction models (PMs) derived could be evaluated in future validation studies. The multivariate analysis was also useful for the identification of outliers, for the identification of the most mechanistically sound combination of *in vitro* tests to include in a battery, and to generate ideas for future research.

Several factors could account for the low precision of the predictions observed in this study: a) the choice of test chemicals; b) the test methods which were evaluated; c) the variability in the *in vivo* data; d) the use of the MMAS as the *in vivo* endpoint; and e) the choice of statistical methods (correlation analysis and linear regression). It is now recognised that the variability of the *in vivo* data should be considered during the assessment of method performance. Computer simulations carried out by Bruner *et al.* (10) have shown that, even if the alternative methods were perfectly reproducible (if their coefficients of variation were 0), the variability in the Draize scores alone would restrict the Pearson's correlation coefficients to the range 0.89–0.95 when the Draize scores are between 0 and 100, and to the range 0.65–0.80 when the Draize scores are between 0 and 40 (typical of cosmetics ingredients).

In summary, none of the nine tests used alone was sufficiently predictive of *in vivo* eye irritancy for the full set of test chemicals, even though some of the tests were sufficiently reproducible. In spite of the disappointing results, the EC/HO study made a valuable contribution to the validation process by highlighting the importance of optimising the protocols and refining the PMs of alternative methods before entering them into a large-scale validation study (10). The optimisation of protocols and refinement of PMs is now carried out routinely as part of the prevalidation process (11).

The COLIPA study

The European Cosmetic, Toiletry and Perfumery Association (COLIPA) established a validation study to determine whether currently available *in vitro* methods are valid for predicting the eye irritation potential of cos-

metic ingredients and formulations (7). Specifically, the study was designed to determine whether the data obtained by alternative methods could provide acceptable agreement with the MMAS, provide acceptable agreement with individual tissue scores and recovery time in the Draize test, or correctly predict eye irritation potential in the rabbit eye. The COLIPA validation study was designed to build on the lessons learned in the EC/HO study, for example, by ensuring that PMs were defined before the validation study began (10).

Ten alternative methods were assessed in the COLIPA study: the chorio-allantoic membrane vascular assay (CAMVA), EYTEX, the FL test, the HET-CAM test, the NRU assay, the pollen tube growth (PTG) assay, the neutral red release test (Predi-Safe™), the RBC assay, the SM assay, and the tissue equivalent assay (TEA). Five of these tests had protocols in common with the EC/HO study (the EYTEX, HET-CAM, NRU, RBC and SM tests).

The alternative methods were evaluated under blind conditions in 32 participating laboratories by using 55 test substances, of which 23 were cosmetic ingredients and 32 were formulations. Twenty of the cosmetic ingredients have also been assessed in the EC/HO study, so that the data from both studies could be pooled and analysed in greater detail in the future. The formulations included make-up products, skin cleansers, sunscreens, hair dyes, shampoos, deodorants and toothpastes.

The COLIPA study was carried out in two stages: a dry run on ten test substances to ensure compliance with the Standard Operating Procedures, and a main run on all test substances. Good Laboratory Practice (GLP) was used at all stages of sample coding, randomisation and supply to the participating laboratories. The raw data were collected centrally and received a quality assurance check before they were independently analysed with statistical methods which had been agreed before the start of the study.

By using predefined criteria of reliability, the results indicated that none of the methods entered into the study could be confirmed as a valid replacement for the Draize eye test across the full range of irritancy. However, three methods (the FL test, the RBC assay and the TEA) each satisfied one criterion of reliability. The FL test and the

TEA were conducted in only two laboratories, so it was concluded that their reproducibility should be checked in a further study. The predictivity of each method was assessed against the PM derived by the lead laboratory for that method. The PM for the TEA was a mathematical equation for the prediction of MMAS associated with a 95% CI. The PMs for the RBC assay and the FL test were classification models (predicting three levels of irritancy), associated with a 95% CI of the kappa (κ) statistic (the κ statistic is a chance-corrected measure of agreement ranging from zero [no agreement] to one [perfect agreement]). For the TEA, all of the data points fitted entirely within the 95% prediction intervals of the PM in one laboratory. This was the only case in the entire study in which such a close fit was observed. The results from the other laboratory tended to over-predict eye irritancy in that seven of the data points did not fit within the confines of the 95% prediction intervals. The TEA PM tested covered a range of response between MMAS 0 and MMAS 86. Practical constraints of taking samples at very short time intervals prevent assessment of more severely irritating materials. The FL test PM was adequate for distinguishing non-irritating substances from strongly irritating substances. However, since substances of moderate irritancy were under-represented, it was concluded that further work should be conducted to determine whether the FL test is capable of distinguishing between substances of moderate irritancy and substances at the extremes of the irritancy scale.

A multivariate analysis of the data from the COLIPA validation study has also been conducted. The approach used both PCA and PLS techniques similar to those used for the EC/HO study. The results of this analysis were similar to those observed for the EC/HO study, in that improved PMs could be developed based on combinations of *in vitro* endpoints. However, these models were heavily weighted on the TEA because of the good fit of the data to the PM. Excluding TEA data from the analysis considerably reduced the predictive capacity of the models.

In summary, the COLIPA validation study was designed to build on the results obtained in the EC/HO study, taking into account the lessons already learned. The outcome was promising for three methods (the FL test,

the TEA and the RBC assay), but firm conclusions regarding their validity could not be made, indicating a need for follow-up studies.

The BGA/BMBF study

During 1988–1994, a validation study was carried out in Germany (12, 13) to evaluate the suitability of two *in vitro* tests to replace the Draize eye test for severe eye irritants, i.e. the HET-CAM test and the NRU test using 3T3 (mouse fibroblast) cells (3T3 NRU). These tests were chosen for validation because they had been identified as the most promising tests for identifying severe eye irritants in an earlier project (14, 15). The validation study was coordinated by the Centre for Documentation and Evaluation of Alternative Methods to Animal Experiments (ZEBET) at the Bundesgesundheitsamt (BGA), and was supported financially by the German Department of Research and Technology (BMBF). The study was conducted in two phases: Phase I (1988–1990) consisted of a prevalidation study and a blind trial, and Phase II (1990–1994) consisted of a database development phase and biometrical analysis.

During Phase I, standardised protocols for the 3T3 NRU and HET-CAM tests were developed, and the two tests were established in 13 laboratories. Following an independent assessment of the intralaboratory and interlaboratory reproducibilities of the two tests, 34 test chemicals were selected for the blind trial. These chemicals were supported by high quality *in vivo* data, and included chemicals outside the limited group of surfactants for which the tests had been developed. The 34 chemicals were coded, and the two tests were assessed under blind conditions in 13 laboratories. Both tests had satisfactory intralaboratory and interlaboratory reproducibilities, although the 3T3 NRU test was more reproducible than the HET-CAM test. In contrast, the HET-CAM test was better at identifying severe eye irritants (chemicals classified as R41 according to EU guidelines). Both tests were capable of ranking the test chemicals in a similar order to that derived from *in vivo* data.

During Phase II, further evaluations of the HET-CAM and 3T3 NRU tests were conducted by testing each method with 166 industrial chemicals under blind conditions in two laboratories. These chemicals, chosen to be representative of the chemicals pro-

duced by the pharmaceutical and chemical industries, were different to the 34 chemicals tested in Phase I. Thus, the HET-CAM and 3T3 NRU tests were evaluated with a total of 200 chemicals (147 new chemicals and 53 existing chemicals). During an independent quality control of the database, 57 chemicals were excluded from further analysis because of the unacceptable quality of their *in vitro* or *in vivo* data, leaving 143 chemicals for analysis. The PMs which had been developed in Phase I for the two *in vitro* tests were found to be insufficiently predictive of severe eye irritancy because a new criterion had been introduced in the EU classification system (16), i.e. the presence of irreversible damage within a 21-day observation period was now sufficient for an R41 classification to be assigned. Therefore, some *post hoc* data analyses were performed by using linear discriminant analysis (LDA) to obtain the optimal combination of *in vitro* endpoints for discriminating between severe (R41) and non-severe irritants. During the database development phase, a total of ten *in vitro* endpoints had been determined (nine for the HET-CAM test and one for the 3T3 NRU test), but since the full set of values for the ten endpoints was not available for 27 of the 143 chemicals, the LDA was based on 116 chemicals. This revealed that the best endpoint for identifying severe irritants was the detection time of coagulation. Not only was this more predictive than the other nine endpoints, it was also more predictive than the traditional HET-CAM endpoint based on the combined use of haemorrhage, lysis and coagulation. For water-soluble chemicals, it was found that the detection time of coagulation using a 10% solution had the highest discriminating power, whereas for less water-soluble chemicals, the detection time of coagulation using the undiluted chemical was more appropriate. The classification of water-soluble chemicals was improved further by combining the time-to-coagulation endpoint with the 3T3 NRU endpoint (IC50; the concentration of test chemical resulting in a 50% inhibition of neutral red uptake). The classification models derived by LDA were confirmed by cross-validation.

The authors of the BGA/BMBF study concluded that chemicals can be classified as severe irritants (R41) with sufficient reliability by the combined use of the HET-CAM test and the 3T3 NRU test, both of which are

well-validated tests, as required by OECD Guideline 405 (4). Since 1992, the German authorities have accepted the use of HET-CAM data for the classification of R41 chemicals in the notification of new industrial chemicals. Finally, the report of a validation study (13) provides a good illustration of the use of multivariate statistics in the development of PMs and in the design of tiered testing strategies.

The CTFA study

The Cosmetics, Toiletries and Fragrance Association (CTFA) conducted a six-year programme (1990–1996) to evaluate promising *in vitro* alternatives to the Draize eye irritation test (17–19). The programme was carried out in three phases, each phase serving to investigate the performance of approximately 24 *in vitro* tests (not counting variations of each test) with respect to a specific group of products: Phase I tested hydro-alcoholic formulations (10 materials); Phase II tested oil–water emulsions (18 materials); and Phase III tested surfactant-based formulations (25 materials). All test materials were coded by an independent laboratory so that both the animal tests and the *in vitro* tests could be conducted in a blind fashion. The animal experiments, which generally used six rabbits per test material, were carried out either in parallel (Phase I), or according to a randomised block design (Phases II and III). The latter method enabled the *in vivo* variability of the MMAS to be assessed more realistically. In all of the animal experiments, anaesthesia was applied to the eyes prior to dosing.

When the experimental stage of each phase of the study had been completed, the chemical identities were revealed, and the relationship between the *in vivo* and *in vitro* data was analysed by statistical methods. First, a “concordance analysis” was carried out in which a comparison was made between the *in vitro* and *in vivo* rankings of the materials. The *in vitro* tests which performed to a certain level in the concordance analysis were subsequently analysed by non-linear regression to approximate the relationship between the *in vitro* and *in vivo* scores. A novel feature of this analysis was the inclusion of a 95% prediction interval. This reflects the variability of both the *in vitro* test and the *in vivo* test, and enables the observer to visualise the range of *in vivo*

scores predicted by a given *in vitro* result.

The variability of the Draize test in each phase of the study was quite striking, even though the animal tests had been carried out in a single laboratory. In Phases I and II, the variability was smallest for the least irritating chemicals, and increased as the irritancy increased. In contrast, the variability in Phase III was greatest in the middle of the irritancy range and smallest at the two ends of the scale. The Draize scores were confined to the lower end of the Draize scale (less than 46), which is the most relevant range for cosmetic formulations.

The performance of the *in vitro* tests also varied between the three phases. In general, the concordance of the *in vitro* assays was higher, and the prediction intervals were narrower, for Phase I and Phase III materials than they were for Phase II materials.

In conclusion, several features of the CTFA study are noteworthy. Firstly, the variability of the *in vivo* scores was taken into account when determining the performance of the *in vitro* methods. Secondly, regression analysis was used to determine 95% prediction intervals for the estimation of *in vivo* scores. Thirdly, the predictivity of each *in vitro* method was shown to vary according to the type of material being investigated.

The IRAG study

The Interagency Regulatory Alternatives Group (IRAG), which is made up of representatives from three US regulatory agencies (the Food and Drug Administration [FDA], the Environmental Protection Agency [EPA], and the Consumer Product Safety Commission [CPSC]), carried out a 3-year programme (1991–1994) to evaluate the performance of *in vitro* assays for eye irritation (20). The evaluation was based on existing animal and *in vitro* data, which were submitted in parallel by laboratories around the world. Over 60 data sets from 41 laboratories were received for 29 different test methods. The *in vitro* data were compared not only with the MMAS, but also with the individual tissue scores representing the damage of the cornea, conjunctiva and iris.

A set of guidelines was developed to standardise the data submissions and to facilitate their review (21). These guidelines included: general guidelines for the acceptance of data; criteria for the collection and collation of *in*

vitro data; criteria for the collection and collation of *in vivo* data (individual animal and tissue scores were requested); criteria for the review and evaluation of data; and the format to be used when reporting the summary of an evaluation (see below).

Five working groups were established, each containing 4–10 members, to review data from: organotypic models (22); chorio-allantoic membrane-based assays (23); cell function-based assays (24); cell cytotoxicity assays (25); and other assays (26). In addition, a statistical subcommittee was formed to help in the planning and analysis of the study (27). At the end of the programme, each working group published a summary of its evaluation, and presented its conclusions at an open forum (Workshop on Eye Irritation Testing, Washington, DC, USA, November 1993). Most of the reviews were based on scatterplots of paired *in vivo* and *in vitro* data and on regression analyses of the resulting relationships. The variabilities of both the *in vivo* test and the *in vitro* tests were taken into account, and were generally represented on the scatterplots by the inclusion of error bars. The reviews revealed differences in predictivity between test methods for the same types of chemicals, and between chemical types for the same test method; none of the tests showed a satisfactory performance across all chemical groups. In general, the ability to obtain strong *in vitro*–*in vivo* correlations was compromised by the variable nature of the animal test.

The IRAG study led to several conclusions: none of the *in vitro* tests, and no combination of the tests, could completely replace the animal test; alternatives to the Draize test are currently being used by industry as screens in the risk assessment process for product development; and some of the *in vitro* models have the potential to reduce animal testing, provided that they have been validated and are conducted under well-defined conditions.

The MHW/JCIA study

In 1991, the Japanese Ministry of Health and Welfare (MHW) began a study to investigate the possibility of using alternatives to the Draize test for the safety assessment of cosmetic ingredients (28). A detailed review of 16 methods led to the selection of 12 methods for inclusion in an interlaboratory validation study (29), carried out

under the auspices of the MHW and the Japanese Cosmetic Industry Association (JCIA).

The 12 methods assessed in the MHW/JCIA study were: the HET-CAM method; the HET-CAM-trypan blue staining method (CAM-TB); the RBC haemolysis method; the haemoglobin denaturation method (HD); the artificial skin models SKIN²™ (ZK1100 model) and MATREX™, cytotoxicity on normal rabbit corneal cells (CornePack™); the crystal violet staining (CVS) method using transformed rabbit corneal (SIRC) cells (SIRC-CVS); the NRU method using SIRC cells (SIRC-NRU); the reduction of 3-(4,5-dimethylthiazole-2-yl)-2,5-diphenyl tetrazolium bromide (MTT) using human cervical carcinoma (HeLa) cells (HeLa-MTT); the CVS method using Chinese hamster lung (CHL) cells (CHL-CVS); and EYTEX. A total of 27 laboratories in Japan participated in the validation study. Each method (except for MATREX and CHL-CVS) was assessed in at least five laboratories, and most laboratories assessed more than one method.

The test chemicals, comprising 38 cosmetic ingredients, were tested in three phases (9, 15 and 14 ingredients in the first, second and third phases, respectively). The samples were coded, randomised and supplied to the participating laboratories, and tested in accordance with the principles of GLP. The *in vitro* data were compared with Draize rabbit data obtained by a single laboratory in accordance with OECD Guideline 405 (4), and the variabilities of the *in vitro* and the *in vivo* data were analysed.

Interlaboratory variability, as judged by the mean CV (the coefficient of variation averaged across all chemicals), was less than 50% for all *in vitro* tests except the HET-CAM and HD tests, for which the mean CVs were greater than 50%. However, the mean CVs of these tests could be reduced to below 50% by excluding the non-irritants from the analysis. The *in vivo* data were more variable, particularly MMAS values in the range 15–50, which is important for the evaluation of cosmetic ingredients. The correlation between the *in vitro* results and the MMAS was high (Pearson's coefficient greater than 0.7) for CAM-TB, HD, SIRC-CVS, SIRC-NRU, HeLa-MTT and CHL-CVS, but it was low (0.3) for EYTEX. The Pearson's correlation coefficients for the cytotoxicity tests

exceeded 0.8 if acids, alkalis and alcohols were excluded. The MMAS scores were also grouped into five categories according to the Kay & Calandra classification scheme (29), i.e. non-irritant ($0 \leq \text{MMAS} < 0.5$), slight irritant ($0.5 \leq \text{MMAS} < 15$), mild irritant ($15 \leq \text{MMAS} < 25$), moderate irritant ($25 \leq \text{MMAS} < 50$) and severe irritant ($50 \leq \text{MMAS}$). The rank correlation between the *in vitro* results and these categories was high (Spearman's coefficient greater than 0.8) for HET-CAM and CAM-TB, and was increased if powdered substances were excluded from the analysis. In the case of the cytotoxicity tests, the rank correlation was also greater than 0.8, provided that the acids, alkalis and alcohols were excluded. In general, the Spearman's rank correlations were higher than the Pearson's correlations. In addition to comparing the *in vitro* data with the MMAS, comparisons were also made with the 24-hour weighted Draize score. The MMAS produced closer correlations than the weighted Draize score. On the basis of these results, it was concluded that none of the alternative methods could be used to test all types of test substances, and that a battery of tests would be needed to optimise the ability to predict eye irritancy.

The Way Forward

Given the experience gained in previous validation studies, the workshop participants agreed that initiatives should be taken to expedite the elimination of the Draize eye irritation test. It was decided that progress toward the short-term reduction and refinement of animal use, and the long-term replacement of the Draize test, could be achieved by four parallel activities: a) an evaluation of the benchmarking (reference standards) approach; b) a review of tiered testing strategies; c) further analyses of the data obtained in previous studies; and d) research on the mechanisms of eye irritation. Before these activities are reviewed, it is useful to consider the various purposes for which the Draize test is currently used, since this has implications for the development and validation of non-animal methods. In addition, the case for changing the Draize test protocol for the testing of solid materials is presented, since this would provide a means of refining the test in the short term.

Current uses of the Draize test

Eye irritation data are used in at least two contexts: in the hazard classification of chemicals for regulatory purposes; and in the safety assessment of ingredients and mixtures of ingredients used in a wide range of industrial, pharmaceutical and consumer products. The questions asked within each context are different, so the types of data needed for the two situations are not equivalent. This means that the development and validation of an alternative to the Draize test will be subject to different considerations, depending on the intended purpose of the alternative method.

In the hazard identification of chemicals, the purpose of testing is to classify eye irritation potential according to classification schemes defined by regulatory authorities. Current proposals by the OECD (31, 32) recommend a tiered (stepwise) approach to hazard identification in which new chemicals can be classified as irritating to the eye on the basis of results from a non-animal test. Testing in animals is only required as a last step to confirm negative results generated by the non-animal tests applied in earlier steps. Thus, the stepwise process represents both reduction and refinement, but the *in vivo* test is not replaced.

During the development and validation of a non-animal method intended as a screen in a stepwise testing strategy, it might be sufficient that the test can place chemicals into two or more categories of eye irritation potential, without generating too many false positive results. There is less concern about the generation of false negatives because these will presumably be identified by the animal test(s) carried out in the last step of the process.

In the safety assessment of ophthalmological and cosmetic ingredients, mixtures and products, toxicologists face different requirements. In this case, the placement of test substances into broad irritation categories is often not sufficient, since it is necessary to prove the absence of adverse effects in the eye. In the past, this was accomplished by using *in vivo* endpoints such as the average eye irritation scores obtained over several days, the MMAS, the number of days required for an irritation response to clear, and the appearance of secondary lesions (for example, corneal ulcerations). All of this information was used to demonstrate that

products, particularly those used around the eye, would not cause adverse effects. Nowadays, there is an increasing reliance on non-animal methods, but if these are ever to be used as complete replacements for the Draize test, product safety toxicologists will have to be assured that the predictions obtained are reliable. It has been concluded that a longer term approach will be needed to develop mechanism-based alternatives which could serve as replacements for the Draize eye irritation test across the full range of response (33). An overview of areas of research is provided below.

Refinement of the Draize test

The Draize test could be refined by changing the way in which solid materials are treated. In the Draize test, solids are instilled as a bulk material in the conjunctival sac, where they may be held for up to 24 hours. If the substance is poorly soluble and has cytotoxic properties, the combined effect of mechanical damage and cytotoxicity can cause very severe effects. Such a high and persistent exposure does not occur in rabbits when testing liquids for eye irritation, nor does it occur when testing compounds for skin irritation (dermal exposure for 4 hours). It is a situation which is not consistent with accidental human exposure, and which cannot be mimicked by many alternatives. An example of the effects of solid entrapment is provided by sodium perborate (chemical 37 in the EC/HO study). The *in vivo* data for this substance (34, 35) show that the corneal opacity was slight or moderate an hour after exposure, and covered only a small part of the cornea (the lower part), whereas the conjunctival swelling was severe. It is not clear from the data whether solid remains were still present in the conjunctival sac at the 24-hour reading, or whether rinsing of the eye was performed. However, in two rabbits, the maximum corneal opacity (score of 4) was observed 21 days after treatment. As a result of this effect, sodium perborate is classified as R41/Category A (Appendix 1), even though it has a relatively low MMAS (score of 30) and caused low *in vitro* scores in several assays (6). Similarly, four other solid materials (captan 90 concentrate, quinacrine, and 1-naphthalene acetic acid [and its sodium salt]) which caused low or moderate *in vitro* scores in the EC/HO study, are also classified as R41/Category A (Appendix 1).

One of the disadvantages of testing solids in the standard rabbit eye test is the amount of test substance which is typically instilled into the rabbit eye. OECD Guideline 405 (4) recommends using either a volume of 0.1ml of solid (in the form of a fine, but slightly compacted, dust) or a weight of no more than 0.1g. However, because the density of solids is often much higher than 1g/ml, overdosing can occur, possibly increasing the variability in the eye effects. The disadvantages of testing solids are discussed in more detail by Walker (36), who contends that the use of a lower volume of test material, placed directly onto the cornea, gives a much better prediction of eye irritation in humans.

The benchmarking/reference standards approach

The term "reference standard" (RS) should not be confused with "positive control". A positive control is a substance which is known to give a positive response in a particular *in vitro* assay, and which is used to confirm the correct conduct of the assay. In contrast, an RS is a substance which has a known degree of toxicity *in vivo*, and which can be used *in vitro* to determine the degree of toxicity of test substances, whose effects are scaled relative to the RS. For example, if two RSs are available corresponding to known boundaries of eye irritation potential (for example, R41/R36 and R36/NI), it should be possible to classify a test substance by comparing its *in vitro* result with the *in vitro* results of the two RSs. It should also be possible to obtain a measure of confidence for the classification according to the proximity of the *in vitro* result to each boundary. Conceivably, a positive control could also act as a reference standard, if it were used both to determine the validity of an assay and to scale the toxic response of a test substance.

In industry, RSs are already widely used for making safety decisions regarding the acceptability of new formulations of existing ingredients, and for prioritising further developments. Three other roles of RSs can be foreseen: a) within companies, for the development and cross-validation of *in vitro* assays, when RSs could be used to investigate and calibrate new or existing assays by using data available in the public domain; b) in the validation of alternative methods, as a replacement for the totally blind approach which currently exists, so that substances

can be grouped into categories defined by the RSs; and c) in regulatory toxicology, for the submission of data on selected new substances to competent authorities. This would apply to substances for which the physical, chemical and other properties are known, and where it can be demonstrated that the use of the RS is toxicologically relevant.

To investigate the applicability of the RSs (benchmarking) approach in the validation and acceptance of *in vitro* tests, the ECVAM Reference Standards Working Group has been established, with the following membership: Michael Balls (ECVAM), Lesley Earl (Unilever Research, UK), Julia Fentem (ECVAM) and Richard Lewis (Zeneca CTL, UK).

Criteria for the selection, use and validation of reference standards

The ECVAM working group on RSs agreed that the following criteria should be applied to determine whether a substance is a suitable RS for use in a given *in vitro* test.

1. The RS should be readily available in a chemically pure and stable form.
2. The RS should provide reproducible results within the test system of choice.
3. The RS should be associated with *in vivo* data (preferably human) of high quality and low variability.
4. The set of RSs chosen should cover the full range of the *in vivo* toxicological endpoint, which should be clearly defined.

Having established a set of suitable RSs, the following points should be considered before they are used in testing chemicals: a) the relative toxicity of the RS and test material (if known); b) the chemistry (including structure and functional class) and physical form of the RS relative to the test substance; and c) the likely mechanisms of toxicity (if known) of the RS and test substance.

The criteria for determining whether an assay and its associated RSs are ready for prevalidation and validation are similar to the criteria applied to any alternative method.

1. The method and RSs must be well-developed and associated with good supporting data.
2. The method and RSs must be relevant to the toxicological endpoint.
3. There must be a protocol and PM covering the use of the RSs.

4. There must be evidence that the reproducibility of the RSs is adequate for the purpose.

An ECVAM study to evaluate the use of reference standards in the validation process

The ECVAM Reference Standards Working Group decided that an initial evaluation of the benchmarking approach should be made by concentrating on eye irritancy as the toxicological endpoint. There are a number of reasons for this decision: a) several validation studies for eye irritation have so far failed to find suitable alternatives, which are urgently required; b) several *in vitro* eye irritation assays are promising candidates for evaluation by the reference chemicals approach; c) there are a few groups of chemicals which could be considered as candidate RSs available in the public domain; d) good quality human exposure data could be available for some chemicals; and e) there is a large industrial community which has a wealth of experience in using RSs for assessing eye irritancy.

The details of an ECVAM-sponsored study are currently being finalised. It is envisaged that five *in vitro* methods will be included in the study, i.e. the ICE test, the BCOP test, the combined use of the HET-CAM and NRU tests, EpiOcularTM, and the RBC haemolysis test. For each method, chemicals belonging to one or more chemical groups will be tested in a single laboratory (i.e. there will be five laboratories in total). Most of the chemical groups will be defined in terms of functional class or physical form, although a mixed group will also be tested in each laboratory. The testing of chemicals will be carried out in two phases. In the first phase, each laboratory will be required to test up to five chemicals per chemical group (these will be the RSs) and to develop a PM. At this stage, the chemical identities and accompanying *in vivo* data will be supplied to the laboratories. In the second phase of testing, each laboratory will be required to repeat the testing of the RSs in Phase I and to test a further five chemicals per chemical group, which will be supplied coded. Each laboratory will be required to predict the eye irritation potential of the five test chemicals, by using the PM developed in Phase I. The reliability and relevance of each *in vitro* test, as judged by the benchmarking approach, will be assessed by independent data analysis.

Review of stepwise testing strategies

Stepwise (hierarchical) testing strategies are approaches to toxicity testing in which alternative methods (structure-activity relationships, biokinetic models, physicochemical techniques and *in vitro* tests) are applied in sequence before any animal tests are carried out. In addition to providing a means of implementing the Three Rs, stepwise testing strategies optimise the use of existing knowledge and resources, and promise to improve the scientific basis of toxicity testing. The tiered approach to eye irritation testing should be particularly useful, since it seems unlikely that any single *in vitro/ex vivo* test will be capable of reproducing the complexity of the *in vivo* response.

Hierarchical testing schemes have been proposed in the literature for a variety of toxicological endpoints, including skin irritation/corrosion (37, 38), skin sensitisation (39), phototoxicity (40) and neurotoxicity (41). For the assessment of eye irritancy, there have been several proposals based on the combined use of a cytotoxicity test and an organotypic test, including: the 3T3 NRU cytotoxicity and HET-CAM tests (13); the K562 cytotoxicity and isolated rabbit eye (IRE) tests (42); and the 3T3 NRU cytotoxicity and ICE tests. The proposal made by Spielmann *et al.* (13) for the combined use of the 3T3 NRU cytotoxicity and HET-CAM tests consists of a tiered strategy for the identification of severe eye irritants (R41 chemicals), as defined by EU criteria (3). The strategy is applicable to chemicals with different solubility characteristics, for which separate PMs were derived by using the data generated in the German validation study on alternatives to the Draize test (13). The combined use of the 3T3 NRU cytotoxicity and ICE tests also appears to provide an effective means of identifying severe irritants (defined as chemicals having an MMAS > 59) according to the results of a discriminant analysis carried out on the EC/HO study chemicals, in which only three out of 40 non-severe irritants (chemicals having an MMAS < 59) would be overclassified as severe irritants (M. Liebsch, personal communication).

At the regulatory level, a tiered approach to eye irritancy/corrosivity testing is provided for in the 1987 update of OECD Guideline 405 (acute eye irritation/corrosion; 4), although no particular testing strategy is specified. A proposal for a testing strategy

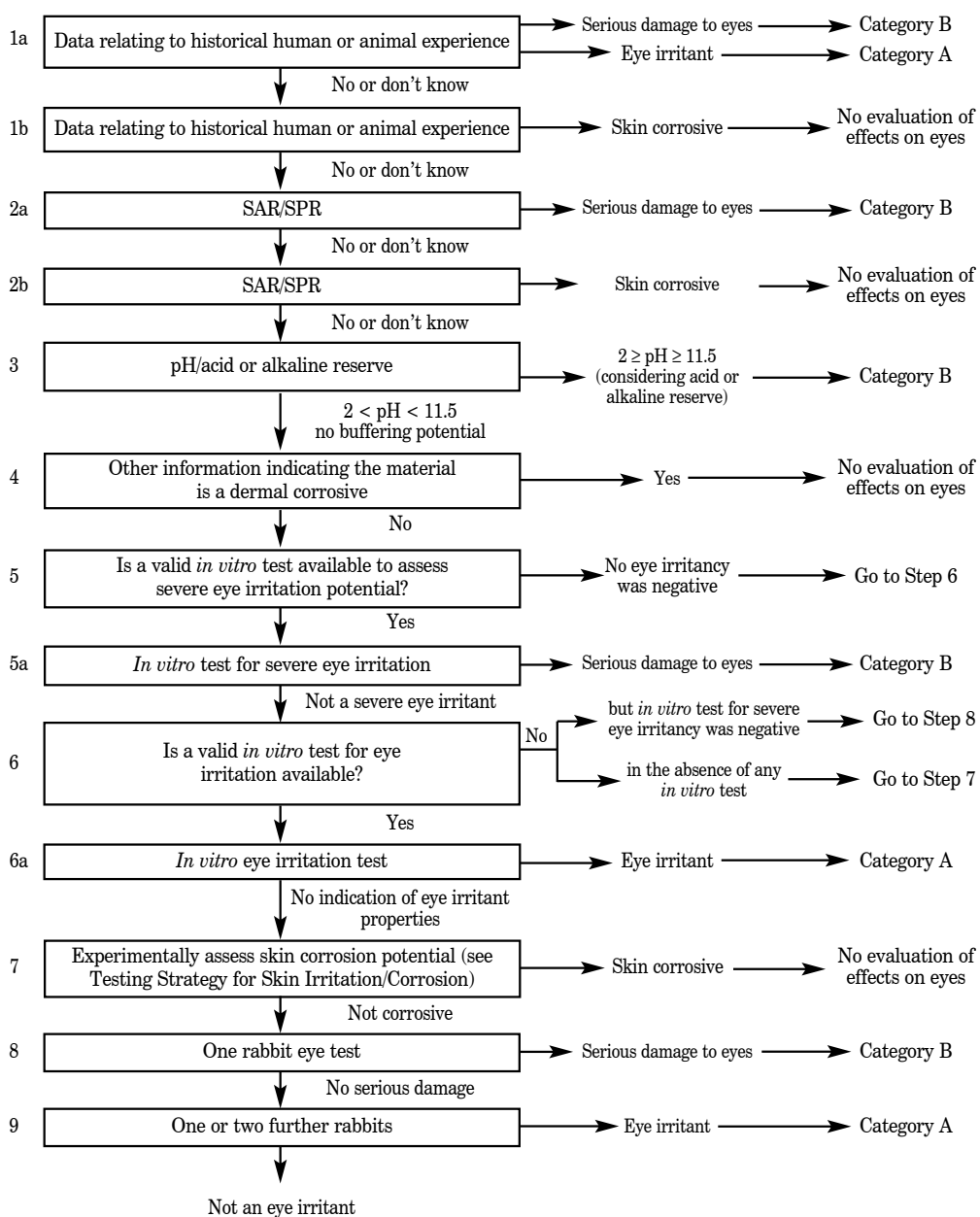
was discussed at an *OECD Workshop on Harmonization of Validation and Acceptance Criteria for Alternative Toxicological Test Methods*, held in Solna, Sweden, in January 1996 (31). Subsequently, the proposed strategy was modified by the OECD Advisory Group on Harmonization of Classification and Labelling, which incorporated a *Testing and Evaluation Strategy for Eye Irritation/Corrosion* (Figure 1) into its revised proposal for the harmonisation of hazard classification based on eye irritation/corrosion (32). Important features of the proposed OECD strategy are: a) it allows for the classification of chemicals as irritant or corrosive to the eye on the basis of validated alternative methods; b) it only permits the use of animal tests to check negative results (non-irritant and non-corrosive) generated by one or more alternative methods; and c) animal testing is refined by using a single rabbit test to detect serious damage to the eyes (in which case no further testing would be conducted) before conducting one or two additional rabbit tests to detect moderate irritancy. The results of a study carried out by ECVAM indicate that the basic design of the OECD testing strategy provides an effective means of reducing and refining the use of the Draize eye test. The report of this study (43) is intended to illustrate a general approach to the evaluation of stepwise testing strategies. The same approach has also been applied to the proposed OECD testing strategy for skin corrosion (44).

Further analysis of completed validation studies

Considerable effort has been directed toward the *post hoc* analysis of completed validation studies. For example, Menk Prinsen (TNO, Zeist, The Netherlands) has classified the EC/HO chemicals according to both the current EU classification system (3) and the harmonised classification system proposed by the OECD (32). A comparison of the two sets of classifications indicates that the OECD classification system is broadly comparable to the current EU system (Appendix 1). It is suggested that the OECD system will provide an appropriate choice of classification system during international validation studies which assess alternative methods in terms of their ability to predict eye irritation potential.

Other efforts have used the techniques of multivariate statistics to extract information

Figure 1: Proposed OECD testing and evaluation strategy for eye irritation/corrosion



Adapted from reference 53.

SAR = structure-activity relationship; SPR = structure-property relationship.

from validation study data sets. These techniques are particularly well-suited to the analysis of these data sets, which generally consist of many variables (physicochemical properties and biological endpoints), correlated with one another to varying degrees, for a given set of objects (chemicals or formulations). The following sections describe a number of commonly used multivariate techniques, and illustrate their application in the field of eye irritation testing. Some results of *post hoc* data analysis are also outlined above in the review of previous validation studies. At present, the data from the EC/HO and COLIPA studies are being analysed by the COLIPA Eye Irritation Task Force, to identify the outliers (substances significantly under-predicted or over-predicted by non-animal test methods) which may help to explain the outcome of these studies. The work of this task force will be reported in the near future.

Principal components analysis

PCA is a method for reducing the number of variables in a complex data set with the minimum loss of information (variance). The original variables are transformed into new variables called principal components (PCs), which are linear combinations of the original variables. The PCs are constructed in such a way that all PCs are orthogonal (uncorrelated), and the first PC accounts for the greatest proportion of the variance in the original data set, while subsequent PCs account for decreasing proportions of the remaining variance. The PCs can be interpreted in terms of their vector loadings, which are simply the coefficients in the linear combination of original variables: the greater the loading of a PC for a particular variable, the more the PC is composed of that variable.

An example of the use of PCA in a validation study is provided by Barratt *et al.* (44). PCA was used to visualise the relationship between the skin corrosivity potential of four groups of chemicals (acids, bases, electrophiles and neutral organics) and their physicochemical properties. The PCA plot for each group of chemicals showed a general separation between the corrosive chemicals and the non-corrosive chemicals, and enabled borderline chemicals to be identified. This information was used in the ECVAM validation study on alternatives to

the skin corrosivity test, to guide the selection of test chemicals, and to assist the interpretation of *in vitro* data (46).

A further illustration of the applications of PCA is provided by Lovell (47, 48). PCA was used to obtain the PCs of 18 rabbit eye tissue scores (referring to damage of the cornea, conjunctiva and iris after 24 hours, 48 hours and 72 hours) from 352 animals and 55 test substances. The first PC, which accounted for 77% of the variability in the *in vivo* data, gave approximately equal weight to the 18 tissue scores and was strongly correlated with the total Draize score (TDS) and with the MMAS, whereas the second PC, which accounted for 7% of the variability, was found to contrast damage to the cornea and iris from damage to the conjunctiva. These results indicate that the TDS and the MMAS capture most of the information about tissue damage which can be observed between the 24-hour and 72-hour time-points in the Draize test. The study also showed that the TDS (to which the corneal score contributes 80 units out of a maximum of 110) is strongly correlated with the sum of non-weighted tissues scores (to which the corneal score contributes 24 units out of a maximum of 60). It was concluded from this that the TDS and MMAS provide a suitable means of summarising the information recorded in the Draize test, despite the high weighting of the corneal score in these measures. It was also concluded that there is only limited evidence for differential responses of the different tissues (within 24–72 hours of treatment), and that alternative methods which are developed to predict specific types of tissue damage on the basis of Draize test results are unlikely to be successful.

Partial least squares

PLS analysis is similar to PCA in that it reduces the number of variables in a complex data set, but it differs in that the variables are divided into two subsets, relating to the dependent variables and the independent variables. PCA is carried out on each subset of variables, and multiple regression is used to correlate the PCs of the dependent variables with the PCs of the independent variables. The PCA and multiple regression are carried out in order to preserve as much of the variance in the dependent and independent variables as possible, while at the same time maximising the strength of the correla-

tion between the dependent variables and the independent ones. The results of PLS analysis are visualised as projections on two-dimensional maps. Variables which project furthest from the origin of the graph are the most relevant, whereas those located closest to the origin are the least relevant. Variables which project close to one another are positively correlated, whereas those which project diametrically opposite to one another are negatively correlated. Further details of the PLS method are provided by Lindberg *et al.* (49).

A study carried out by de Silva *et al.* (50) illustrates the use of PLS in establishing batteries of *in vitro* alternatives to the eye irritation test. The technique was applied to a data set consisting of 11 *in vitro* endpoints (relating to eight tests) and 27 *in vivo* endpoints for a set of 32 surfactants and surfactant-based formulations. The analyses indicated that the most predictive methods were the HET-CAM test, the BCOP assay, and the NRU-SIRC assay. The most predictive battery was composed of (in decreasing order of relevance): a) the HET-CAM test; b) the SM test; c) the BCOP assay; and d) the agarose overlay assay.

Cluster analysis

Cluster analysis (CA) is a method for visualising (and quantifying) the similarity between different objects (chemicals), or between different variables (physicochemical and toxicological endpoints). The objects or variables are placed in multi-dimensional space, so that adjacent observations can be grouped in a stepwise fashion. This results in a dendrogram in which all of the observations are grouped into one or more clusters. The similarity between observations is defined as the distance (typically, the Euclidean distance) between them. There are various types of clustering algorithm, which can be distinguished according to: a) the number of links they allow between observations; b) whether they allow links to be broken once they have been formed; and c) whether clusters are built up from individual observations, or split off from a single cluster containing all observations. Further details on cluster analysis are given by Gordon (51).

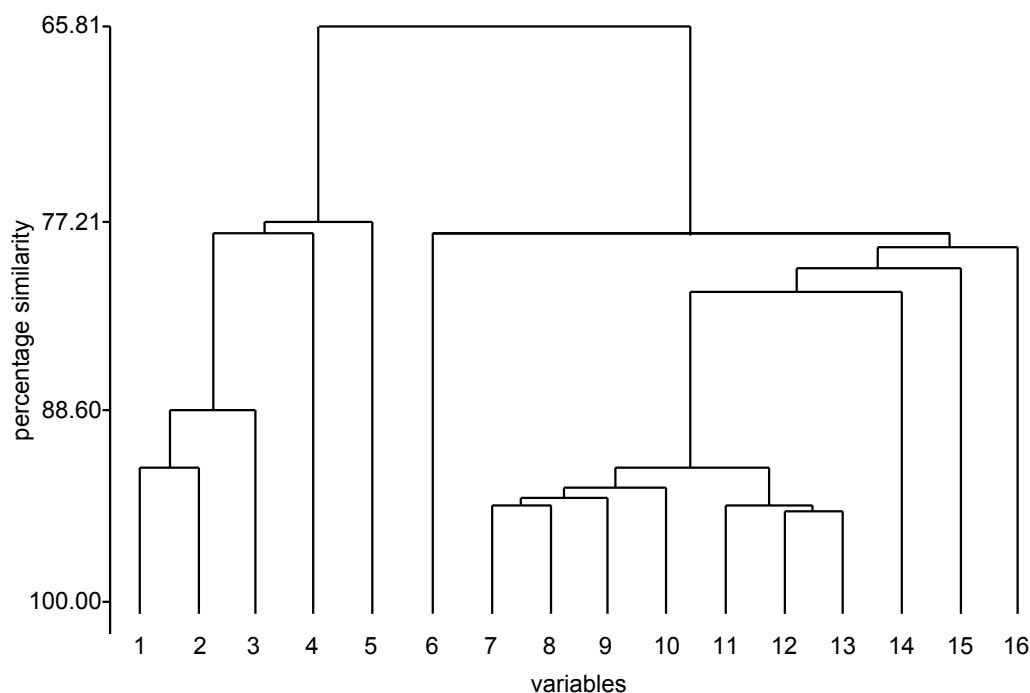
A possible use of CA would be the clustering of *in vitro* endpoints to help in the selection of tests for inclusion in a testing strategy. To illustrate this application, 16 *in*

vitro endpoints for predicting eye irritation potential were clustered on the basis of the *in vitro* scores for 43 chemicals (Figure 2). It can be seen that the cell-based assays cluster together, as do the organotypic assays, even though the distinction between the two types of assays was not fed into the clustering process. If CA were used in the design of a testing strategy, tests would be chosen from different clusters since this would maximise the amount of information provided by the strategy as a whole, i.e. the tests would be selected on the basis of their dissimilarity.

Linear discriminant analysis

Linear discriminant analysis (LDA) is a method for classifying objects into two or more groups on the basis of one or more variables. It works by "plotting" the objects in one-dimensional or multi-dimensional space (depending on whether one or more variables are being used), and by constructing one (or more) linear boundaries which separate the objects into two (or more) groups. Each boundary is defined by a linear equation which contains as many terms as there are variables, and which is used in the classification of objects. In the simplest applications of LDA, there is only a single boundary between two groups. If only one variable is used, a point-like boundary results which can be used as a cut-off value for classifying the objects into the two groups. If two variables are used, the boundary can be thought of as a line, and if three or more variables are used, the boundary becomes a plane or hyperplane in multi-dimensional space. In situations where there are many potentially useful variables, stepwise LDA can be used to choose the variables which provide the best discrimination between groups. An introduction to LDA is given by McFarland & Gans (52).

LDA can be used to derive PMs for predicting the toxicological classifications of chemicals. An example is provided by Spielmann *et al.* (13), who report that the combined use of HET-CAM and the 3T3 NRU test provides a satisfactory means of distinguishing severely irritant (R41) chemicals from non-severely irritant chemicals. Stepwise LDA of ten endpoints (nine HET-CAM and one 3T3 NRU) showed that the best discrimination was achieved by using a single HET-CAM endpoint (the time taken for coagulation to occur), rather than the usual

Figure 2: Cluster analysis of 16 *in vitro* endpoints for predicting eye irritancy

1 = silicon microphysiometer; 2 = red blood cell (H_{50}); 3 = red blood cell (D_{low}); 4 = fluorescein leakage; 5 = neutral red uptake; 6 = red blood cell (D_{max}); 7 = isolated rabbit eye (opacity at 1 hour); 8 = isolated chicken eye (swelling); 9 = isolated chicken eye (opacity); 10 = isolated chicken eye (fluorescein retention); 11 = isolated rabbit eye (opacity at 4 hours); 12 = isolated rabbit eye (swelling at 1 hour); 13 = isolated rabbit eye (swelling at 4 hours); 14 = hen's egg chorio-allantoic membrane test; 15 = bovine corneal opacity/permeability; 16 = EYTEXTM.

Cluster analysis was applied to the *in vitro* data for the 60 European Commission/British Home Office study chemicals, by using the single-linkage, Euclidean distance, algorithm in Minitab 11 (Minitab, State College, PA, USA). The data were first standardised by subtraction of the mean and division by the standard deviation.

weighted combination of endpoints based on haemorrhage, lysis and coagulation. The addition of the 3T3 NRU endpoint to the model improved the identification of R41 chemicals which cause irreversible effects (according to a recent modification of the EU guideline [16], any chemical which causes an irreversible eye effect is classified as R41, regardless of the degree of that effect). On the basis of the LDA analyses, various test-

ing strategies based on the sequential application of HET-CAM and the 3T3 NRU were proposed. Depending on the solubilities of the test substance in oil and water, slightly different PMs were recommended for converting the *in vitro* test results into predictions of eye irritancy.

Research on the mechanisms of eye irritation
An international workshop on the develop-

ment of non-animal replacements for the Draize eye irritation test, organised by COLIPA, was held in Brighton, UK, on 6–8 October 1997 (33). This workshop brought together experts in basic eye research and consumer product toxicologists who conduct eye safety assessments. The expert panel concluded that there are two likely explanations for the difficulty experienced in the identification of non-animal tests which are adequately predictive for the assessment of consumer products. Firstly, the mechanistic basis of current *in vitro* methods has not yet been fully established. Secondly, the standard Draize eye irritation test may not be adequate as the basis for judging the performance of non-animal tests. It was agreed that the fastest way of replacing the Draize eye irritation test would be to undertake additional research aimed at improving our capacity to measure the eye irritation response and at providing information on the mechanisms by which chemicals cause eye irritation. A second workshop held in Brussels, Belgium, in October 1998, recommended the initiation of a research programme to: a) develop an appropriate set of reference test substances for use in the research; b) evaluate the area and depth of corneal injury as markers of eye injury; c) explore the use of early biomarkers of eye injury (for example, the release of cytokines); d) develop methods for evaluating corneal wound healing; e) develop methods for assessing the kinetics of eye injury; and f) develop methods for assessing injury to nerve cells in the cornea. A multicentre programme designed to conduct this work is now being developed and will be submitted for joint funding between interested parties.

Conclusions and Recommendations

General

1. Several reasons could explain why a number of *in vitro* tests for eye irritation have been unsuccessful in previous validation studies: a) the *in vitro* tests only partially modelled the complex *in vivo* eye irritation response; b) the protocols and PMs of the tests might have been insufficiently developed; c) the tests were judged on their ability to predict the MMAS, which is a variable measure of *in vivo* irritancy; and d) the statistical methods chosen for comparing the *in vitro* and *in vivo* results might not have been the most appropriate.
2. When developing and validating non-animal methods, it is important to clearly state the purpose of the non-animal test. Methods which are intended to be screens in a hierarchical testing scheme will be subject to different predictivity criteria than methods proposed as replacements for the *in vivo* test.
3. When assessing the predictivity of an alternative method during a validation study, a useful consideration is the best possible predictivity which can be expected on theoretical grounds. This can be estimated by carrying out computer simulations which model the effect of variability in the *in vivo* data on the strength of the *in vitro*–*in vivo* relationship.
4. During an international validation study in which the predictivity of an alternative method (or combination of alternative methods) is being evaluated in terms of its ability to classify chemicals, a suitable choice of classification scheme appears to be the proposed OECD scheme for the harmonised classification and labelling of chemicals based on eye irritation/corrosion. However, it should be remembered that classification systems do not remove variability in data; they merely reduce the apparent variability.
5. The use of RSs is widespread in industry, where they are used with great effect to make safety decisions. It is highly desirable that the knowledge which exists in industry is transferred to the regulatory bodies responsible for the hazard and risk assessment of new chemicals and formulations. It is important that agreement is reached on the chemicals which can be used as RSs. These chemicals should be well-characterised, readily available, and associated with high quality *in vivo* data (preferably obtained in accordance with international test guidelines).
6. The use of RSs could provide a new way of validating *in vitro* tests. Therefore, ECVAM intends to carry out a pilot study to evaluate the usefulness of reference chemicals in the validation of

alternatives to the Draize eye irritation test.

7. The development and implementation, in a regulatory testing framework, of appropriate testing strategies for eye irritation, which limit the use of the Draize rabbit test to the final step, are critically dependent on the availability of one or more scientifically validated *in vitro* tests for inclusion in the testing strategy. Therefore, either ways must be found to demonstrate that the *in vitro* tests currently being used in-house are indeed valid for the purposes to which they are being put, or new *in vitro* tests will need to be developed and validated.
8. Test batteries combining a cytotoxicity test with an organotypic test appear to provide an effective means of identifying irritant chemicals, at least for screening purposes. Before any combination of methods is accepted for use in a tiered assessment process, it must be adequately validated, and the acceptable rate of false positive results must be agreed.

Recommendations to the regulatory community

9. OECD Guideline 405 (acute eye irritation/corrosion) should be modified with respect to: a) the *in vivo* testing of solid materials (the solid material should be removed from the eye after treatment, to produce a more-relevant and more-reproducible exposure, and to reduce unnecessary animal suffering); and b) the use of physicochemical measurements (one or more PMs for converting measurements of pH and buffering capacity into predicted classifications of eye irritancy should be cited, along with the chemical concentration at which the pH measurements should be carried out).
10. The proposed OECD testing strategy for eye irritation/corrosion should incorporate one or more *in vitro* methods in steps 5a (screening of severe irritants) and 6a (screening of irritants), provided that these have been validated in an interlaboratory validation study.

Recommendations for further research

11. The *in vitro* test results from the EC/HO and COLIPA validation studies should

be compared with the eye irritation classifications defined by both the EU and the OECD guidelines. This will reveal whether the predictive capacity of the methods improves as a result of the new classification system.

12. The predictive abilities of several testing strategies for eye irritation should be evaluated. The evaluations should include an assessment of the validation status of the component tests and of the testing strategies as a whole.
13. Further research is needed on the quantitative structure-activity relationship modelling of eye irritancy, including the development and evaluation of models for predicting levels of irritancy (for example, R41/R36/N1).
14. A study should be conducted to examine the relationship between the pH of a chemical, its buffering capacity, and its capacity to cause tissue injury.
15. There is a need for alternative methods which are capable of modelling the persistence or reversibility of eye effects. For example, an *in vitro* assay for reversibility could be based on the release of inflammatory mediators.
16. Further research is needed to develop predictive batteries of methods based on a better understanding of the mechanisms of eye irritation. In particular, there is a need to: a) develop early markers of eye injury; b) evaluate the area and depth of corneal injury as markers of eye injury; and c) develop methods for assessing wound healing, pain and the kinetics of the eye response.

References

1. ECVAM (1994). ECVAM News & Views. *ATLA* **22**, 7–11.
2. Draize, J.H., Woodward, 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.
3. EC (1993). Council 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.

4. OECD (1987). *OECD Guidelines for the Testing of Chemicals, No. 405: Acute Eye Irritation/Corrosion*, 9pp. Paris, France: OECD.
5. 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.
6. 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.
7. Brantom, P.G. *et al.* (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.
8. CEC (1991). Collaborative study on the evaluation of alternative methods to the eye irritation test. Document XI/632/91-V/E/1/131/91, Part I, 54pp. Brussels, Belgium: CEC.
9. CEC (1991). Collaborative study on the evaluation of alternative methods to the eye irritation test. Document XI/632/91-V/E/1/131/91, Part II, 196pp. Brussels, Belgium: CEC.
10. 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.
11. Curren, R.D., Southee, J.A., Spielmann, H., Liebsch, M., Fentem, J.H. & Balls, M. (1995). The role of prevalidation in the development, validation and acceptance of alternative methods. ECVAM prevalidation task force report 1. *ATLA* **23**, 211–217.
12. Spielmann, H., Kalweit, S., Liebsch, M., Wirnsberger, T., Gerner, I., Bertram-Neis, E., Krauser, K., Kreiling, R., Miltenburger, G., Pape, W. & Steiling, W. (1993). Validation study of alternatives to the Draize eye irritation test in Germany: cytotoxicity testing and HET-CAM test with 136 industrial chemicals. *Toxicology in Vitro* **7**, 505–510.
13. 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. *ATLA* **24**, 741–858.
14. Künstler, K., Bartnik, F., Heitman, W., Lüpke, N.P., Sterzl, W. & Wallat, S. (1987). Abschlussbericht des BMBF-Projektes "Validierung von Ersatzmethoden für Tierversuche zur Prüfung auf Lokale Verträglichkeit". Henkel KGaA, Ressort Forschung/Universität Münster, Institut für Pharmakologie und Toxikologie, 241pp. Bonn, Germany: BMBF.
15. Spielmann, H., Gerner, I., Kalweit, S., Moog, R., Wirnsberger, T., Krauser, K., Kreiling, R., Kreuzer, H., Luepke, N-P, Miltenburger, G., Muller, N., Murmann, P., Pape, W., Siegemund, B., Spengler, J., Steiling, W. & Wiebel, F.J. (1991). Interlaboratory assessment of alternatives to the Draize eye irritation test in Germany. *Toxicology in Vitro* **5**, 539–542.
16. 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.
17. Gettings, S.D., Teal, J.J., Bagley, D.M., Demetru-lias, J.L., DiPasquale, L.C., Hintze, K.L., Rozen, M.G., Weise, S.L., Chudkowski, M., Marenus, K.D., Pape, W.J.W., Roddy, M., Schnitzinger, R., Silber, P.M., Glaza, S.M. & Kurtz, P.J. (1991). The CTFA evaluation of alternatives program: an evaluation of *in vitro* alternatives to the Draize primary eye irritation test. Phase I. Hydro-alcoholic formulations: Part 2: data analysis and biological significance. *In Vitro Toxicology* **4**, 247–288.
18. Gettings, S.D., DiPasquale, L.C., Bagley, D.M., Casterton, P.L., Chudkowski, M., Curren, R.D., Demetru-lias, J.L., Feder, P.I., Galli, C.L., Gay, R., Glaza, S.M., Hintze, K.L., Janus, J., Kurtz, P.J., Lordo, R.A., Marenus, K.D., Moral, J., Muscatiello, M.J., Pape, W.J.W., Renskers, K.J., Roddy, M.T. & Rozen, M.G. (1994). The CTFA evaluation of alternatives program: an evaluation of *in vitro* alternatives to the Draize primary eye irritation test. Phase II. Oil/water emulsions. *Food and Chemical Toxicology* **32**, 943–976.
19. Gettings, S.D., Lordo, R.A., Hintze, K.L., Bagley, D.M., Casterton, P.L., Chudkowski, M., Curren, R.D., Demetru-lias, J.L., DiPasquale, L.C., Earl, L.K., Feder, P.I., Galli, C.L., Gay, R., Glaza, S.M., Gordon, V.C., Janus, J., Kurtz, P.J., Marenus, K.D., Moral, J., Pape, W.J.W., Renskers, K.J., Rheims, L.A., Roddy, M.T., Rozen, M.G., Tedeschi, J.P. & Zyracki, J. (1996). The CTFA evaluation of alternatives program: an evaluation of *in vitro* alternatives to the Draize primary eye irritation test. Phase III. Surfactant-based formulations. *Food and Chemical Toxicology* **34**, 79–117.
20. Bradlaw, J., Gupta, K., Green, S., Hill, R. & Wilcox, N. (1997). Practical application of non-whole animal alternatives: summary of IRAG workshop on eye irritation testing. *Food and Chemical Toxicology* **35**, 175–178.
21. Scala, R.A. & Springer, J. (1997). IRAG working group 6. Guidelines for the evaluation of eye irritation alternative tests: criteria for data submission. *Food and Chemical Toxicology* **35**, 13–22.
22. Chamberlain, M., Gad, S.C., Gautheron, P. & Prinsen, M.K. (1997). IRAG working group 1. Organotypic models for the assessment/prediction of ocular irritation. *Food and Chemical Toxicology* **35**, 23–37.
23. Spielmann, H., Liebsch, M., Moldenhauer, F., Holzhütter, H.G., Bagley, D.M., Lipman, J.M., Pape, W.J., Miltenburger, H., de Silva, O., Hofer, H. & Steiling, W. (1997). IRAG working group 2. CAM-based assays. *Food and Chemical Toxicology* **35**, 39–66.
24. Botham, P., Osborne, R., Atkinson, K., Carr, G., Cottin, M. & van Buskirk, R.G. (1997). IRAG working group 3. Cell function-based assays. *Food and Chemical Toxicology* **35**, 67–77.
25. Harbell, J.W., Koontz, S.W., Lewis, R.W., Lovell, D. & Acosta, D. (1997). IRAG working group 4. Cell cytotoxicity assays. *Food and Chemical Toxicology* **35**, 79–126.
26. Curren, R.D., Sina, J.F., Feder, P., Kruszewski,

- F.H., Osborne, R. & Régnier, J.F. (1997). IRAG working group 5. Other assays. *Food and Chemical Toxicology* **35**, 127–158.
27. Feder, P., Carr, G., Holzhütter, H.G., Lovell, D. & Springer, J. (1997). Statistical planning and analysis considerations in the evaluation of *in vitro* alternatives to whole animal use for eye irritation testing. *Food and Chemical Toxicology* **35**, 167–174.
28. Ohno, Y., Kaneko, T., Kobayashi, T., Inoue, T., Kuroiwa, Y., Yoshida, T., Momma, J., Hayashi, M., Akiyama, J., Atsumi, T., Chiba, K., Endo, T., Fujii, A., Kakishima, H., Kojima, H., Masamoto, K., Masuda, M., Matsukawa, S., Ohkoshi, K., Okada, J., Sakamoto, K., Takano, K. & Takanaka A. (1994). First-phase validation of the *in vitro* eye irritation tests for cosmetic ingredients. *In Vitro Toxicology* **7**, 89–94.
29. Ohno, Y., Kaneko, T., Kobayashi, T., Inoue, T., Kuroiwa, Y., Yoshida, T., Momma, J., Hayashi, M., Akiyama, J., Atsumi, T., Chiba, K., Endo, T., Fujii, A., Kakishima, H., Kojima, H., Masamoto, K., Masuda, M., Matsukawa, S., Ohkoshi, K., Okada, J., Sakamoto, K., Takano, K. & Takanaka A. (1995). First-phase interlaboratory validation of the *in vitro* eye irritation tests for cosmetic ingredients. I. Overview, organization and results of the validation study. *AATEX* **3**, 123–136.
30. Kay, J.H. & Calandra, I.C. (1962). Interpretation of eye irritation tests. *Journal of the Society of Cosmetic Chemists* **3**, 281–289.
31. OECD (1996). *Final Report of the OECD Workshop on Harmonization of Validation and Acceptance Criteria for Alternative Toxicological Test Methods*, 62pp. Paris, France: OECD.
32. OECD (1998). *Revised Proposal for the Harmonization of Hazard Classification Based on Eye Irritation/Corrosion*. ENV/MC/CHEM/HCL(98)5, 10pp. Paris, France: OECD.
33. Bruner, L.H., de Silva, O., Earl, L.K., Easty, D.L., Pape, W. & Spielmann, H. (1998). Report on the COLIPA workshop on mechanisms of eye irritation. *ATLA* **26**, 811–820.
34. ECETOC (1992). *ECETOC Technical Report No. 48, Eye Irritation: Reference Chemicals Data Bank*, 169pp. Brussels, Belgium: ECETOC.
35. ECETOC (1998). *ECETOC Technical Report No. 48, Eye Irritation: Reference Chemicals Data Bank, Second Edition*, 236pp. Brussels, Belgium: ECETOC.
36. Walker, A.P. (1985). A more realistic animal technique for predicting human eye response. *Food and Chemical Toxicology* **23**, 175–178.
37. 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.
38. Botham, P.A., Earl, L.K., Fentem, J.H., Roguet, R. & van de Sandt, J.J.M. (1998). Alternative methods for skin irritation testing: the current status. ECVAM skin irritation task force report 1. *ATLA* **26**, 195–211.
39. 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 and Chemical Toxicology* **33**, 1051–1057.
40. Spielmann, H., Lovell, W.W., Hölzle, E., Johnson, B.E., Maurer, T., Miranda, M.A., Pape, W.J.W., Sapora, O. & Sladowski, D. (1994). *In vitro* phototoxicity testing. The report and recommendations of ECVAM workshop 2. *ATLA* **22**, 314–348.
41. Atterwill, C.K., Bruinink, A., Drejer, J., Duarte, E., Abdulla, E.M., Meredith, C., Nicotera, P., Regan, C., Rodríguez-Farré, E., Simpson, M.G., Smith, R., Veronesi, B., Vijverberg, H., Walum, E. & Williams, D.C. (1994). *In vitro* neurotoxicity testing. The report and recommendations of ECVAM workshop 3. *ATLA* **22**, 350–362.
42. Lewis, R.W., McCall, J.C. & Botham, P.A. (1994). Use of an *in vitro* test battery as a prescreen in the assessment of ocular irritancy. *Toxicology in Vitro* **8**, 75–81.
43. Worth, A.P. & Fentem, J.H. (1999). A general approach for evaluating stepwise testing strategies. *ATLA* **27**, 161–177.
44. 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. *ATLA* **26**, 709–720.
45. 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. I. Selection and distribution of the test chemicals. *Toxicology in Vitro* **12**, 471–482.
46. 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. II. Results and evaluation by the management team. *Toxicology in Vitro* **12**, 483–524.
47. Lovell, D.P. (1996). Principal component analysis of Draize eye irritation tissue scores from 72 samples of 55 chemicals in the ECETOC data bank. *Toxicology in Vitro* **11**, 295–304.
48. Lovell, D.P. (1996). Use of principal component analysis for Draize eye irritation tissue scores. In *Animal Alternatives, Welfare and Ethics* (ed. L.F.M. van Zutphen & M. Balls), pp. 737–746. Amsterdam, The Netherlands: Elsevier.
49. Lindberg, W., Perrson, J. & Wold, S. (1983). Partial least squares method for spectrofluorimetric analysis of mixtures of humic acid and lignin sulfonate. *Analytical Chemistry* **55**, 643–648.
50. de Silva, O., Cottin, M., Dami, N., Roguet, R., Catroux, P., Toufic, A., Sicard, C., Dossou, K.G., Gerner, I., Schlede, E., Spielmann, H., Gupta, K.C. & Hill, R.N. (1997). Evaluation of eye irritation potential: statistical analysis and tier testing strategies. *Food and Chemical Toxicology* **35**, 159–164.
51. Gordon, A.E. (1981). *Classification: Methods for the Exploratory Analysis of Multivariate Data*. New York, USA: Chapman & Hall.
52. McFarland, J.W. & Gans, D.J. (1990). Linear discriminant analysis and cluster significance analysis. In *Comprehensive Medicinal Chemistry, Vol. 4, Quantitative Drug Design* (ed. C.A. Ramsden), pp. 667–689. Oxford, UK: Pergamon.
53. OECD (1998). *Revised Proposal for the Harmonization of Validation and Acceptance Criteria for Alternative Toxicological Test Methods*. Paris, France: OECD.

Appendix 1

An Evaluation of the OECD Proposal for the Harmonised Classification of Eye Irritants and Corrosives

Menk K. Prinsen

TNO Nutrition and Food Research Institute, Division of Toxicology, 3700 AZ Zeist, The Netherlands

Summary — Classifications of eye irritation/corrosion were assigned to 59 of the chemicals used in the European Commission/British Home Office (EC/HO) validation study by applying both European Union (EU) criteria and the harmonised criteria proposed by the OECD. It was found that the application of the two classification systems to the 59 chemicals resulted in comparable classifications: all of the chemicals classified as R36 or R41 according to EU criteria were classified as Category B or Category A, respectively, according to the proposed OECD criteria. Only two of the 59 chemicals, ethanol and ethyl-2-methylacetoacetate, were classified differently by the two systems; they were unclassified by the EU system, and were classified as Category B by the OECD system. It is concluded that: a) the proposed OECD classification system is broadly equivalent to the EU classification system; and b) future validation studies on alternatives to the Draize test would benefit from the application of the OECD classification system to the test chemicals.

Introduction

The failure of recent validation studies (1, 2) to find a suitable replacement to the Draize test is partly a result of the statistical method chosen for evaluating the performance of these methods was assessed by correlating the *in vitro* test scores with the Modified Maximum Average Draize Test Score (MMAS), which is problematic because the MMAS shows considerable variability, particularly in the middle of the irritancy range. The use of the MMAS as the *in vivo* endpoint in an international validation study could also be regarded as inappropriate on the grounds that most regulatory systems (for example, the European Union [EU], US Environmental Protection Agency, US Food and Drug Administration, and Canadian workplace systems) do not classify chemicals on the basis of their MMAS values, but according to their effects in individual tissues of the eye (conjunctiva, cornea and iris), taking

into account the recovery from or irreversibility of these effects. However, it would be difficult to accommodate all of the different classification systems when conducting a validation study. Fortunately, this should soon be unnecessary, since an OECD proposal for the global harmonisation of criteria for the classification of eye irritants (3, 4) is in its final phase of acceptance. The aim of this study was to examine the effect of applying the OECD criteria to chemicals which have already been classified according to EU criteria (5).

Materials and Methods

The animal data for 59 of the European Commission/British Home Office (EC/HO) chemicals were taken from the ECETOC reference chemicals data bank (6). These data were used to classify the 59 chemicals according to both EU criteria (5) and the proposed OECD criteria (3, 4), which are summarised in Tables I and II, respectively. Chemicals which

Table I: European Union classification system for eye irritation/corrosion

Effect	R36 (irritating to eyes)		R41 (risk of serious damage to eyes) ^a	
	Three animals ^b	Six animals ^c	Three animals ^b	Six animals ^c
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
Conjunctival redness	≥ 2.5	≥ 2.5		
Conjunctival chemosis	≥ 2.0	≥ 2.0		

^aA classification of R41 is also assigned if one of the eye effects has not reversed at the end of the observation period in at least one animal.

^bThe classification is assigned if the mean tissue effect (averaged over the 24-hour, 48-hour and 72-hour periods) is greater than or equal to the threshold value in at least two of the three animals. In this study, the same criteria were applied if four rabbits were used.

^cThe classification is assigned if the mean tissue effect (averaged over the three periods and over the six animals) is greater than or equal to the threshold value.

Table II: Proposed OECD classification system for eye irritation/corrosion

Effect	Category B ^a	Category A ^b
Corneal opacity	≥ 1.0	≥ 3.0
Iris lesion	≥ 1.0	> 1.5
Conjunctival redness	≥ 2.0	
Conjunctival chemosis	≥ 2.0	

^aAll 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.

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

Table III: Classifications of eye irritation/corrosion for 59 chemicals obtained by applying European Union (EU) and OECD criteria

No. Test chemical	<i>In vivo</i> data		Classification ^a	
	Reference ^b	MMAS	EU	OECD
1. Sodium hydroxide (10%)	82	108	R41	A
2. Benzalkonium chloride (10%)	186	108	R41	A
3. Trichloroacetic acid (30%)	36	106	R41	A
4. Cetylpyridium bromide (10%)	193	90	R41	A
5. Cetylpyridium bromide (6%)	191	86	R41	A
6. Benzalkonium chloride (5%)	184	84	R41	A
7. Captan 90 concentrate	170	83	R41	A
8. Chlorhexidine	231	82	R41	A
9. Cyclohexanol	77	80	R41	A
10. Quinacrine	230	82	R41	A
11. Promethazine hydrochloride	229	72	R41	A
12. Parafluoroaniline	105	70	R41	A
13. Triton X-100 (10%)	207	69	R36	B
14. Acetone	157	66	R36	B
15. Hexanol	74	65	R36	B
16. 1-Naphthalene acetic acid, sodium salt	168	64	R41	A
17. Sodium oxalate	147	61	R41	A
18. Isobutanol	70	60	R36	B
19. Imidazole	124	59	R41	A
20. Sodium lauryl sulphate (15%)	176	59	R36	B
21. 2-Ethyl-1-hexanol	68	51	R36	B
22. 4-Carboxybenzaldehyde	79	50	R36	B
23. Methyl ethyl ketone	155	50	R36	B
24. Pyridine	123	48	R41	A
25. 1-Naphthalene acetic acid	166	47	R41	A
26. Benzalkonium chloride (1%)	180/182	34/56	R41	A
27. 2,2-Dimethylbutanoic acid	34	45	R41	A
28. γ -Butyrolactone	228	43	R36	B
29. Thiourea	— ^c	— ^c	— ^c	— ^c
30. Octanol	66	41	R36	B
31. Methyl acetate	28	40	R36	B
32. L-Aspartic acid	33	37	R36	B
33. Benzoyl-L-tartaric acid	32	37	R41	A
34. Triton X-100 (5%)	203/205	32/34	NI/R36	NI/B
35. Potassium cyanate	146	31	R36	B

Table III: continued

No. Test chemical	<i>In vivo</i> data		Classification ^a	
	Reference ^b	MMAS	EU	OECD
36. Isopropanol	64	30	R36	B
37. Sodium perborate	144	30	R41	A
38. Dibenzyl phosphate	161	30	R36	B
39. 2,5-Dimethylhexanediol	63	28	R41	A
40. Methyl cyanoacetate	27	28	R36	B
41. Sodium hydroxide (1%)	80	26	R36	B
42. Ethanol	62	24	NI	B
43. 2,6-Dichlorobenzoyl chloride	49	24	R36	B
44. Ammonium nitrate	143	18	R36	B
45. Ethyl-2-methylacetoacetate	26	18	NI	B
46. Sodium lauryl sulphate (3%)	174	16	NI	NI
47. Ethyl acetate	24	15	NI	NI
48. Maneb	164	14	R36	B
49. Fomesafen (acid form)	163	14	NI	NI
50. Tetraaminopyrimidine sulphate	122	10	NI	NI
51. Toluene	101	9	NI	NI
52. Butyl acetate	20	8	NI	NI
53. Trichloroacetic acid (3%)	30	7	NI	NI
54. Methyl isobutyl ketone	149	5	NI	NI
55. Tween 20	201	4	NI	NI
56. Ethyl trimethyl acetate	18	4	NI	NI
57. Methylcyclopentane	138	4	NI	NI
58. Cetylpyridinium bromide (0.1%)	187	3	NI	NI
59. Glycerol	56	2	NI	NI
60. Polyethylene glycol 400	195	0	NI	NI

^aNI = non-irritant; R36/category B = irritating to eyes; R41/category A = risk of serious damage to eyes.

^bData from reference 6.

^cAcutely toxic, and therefore discarded from study.

MMAS = Modified Maximum Average Draize Test Score.

Bold type denotes different classifications obtained by applying the EU and OECD systems.

could not be classified as irritant or corrosive to the eye were classified as non-irritant (NI).

In addition, the 59 chemicals were divided into three groups: MMAS ≤ 25; 25 < MMAS ≤ 59; and MMAS > 59. These cut-off values

were arbitrarily chosen as means of classifying chemicals into three groups. This three-fold categorisation of chemicals was compared with the EU classification of chemicals (NI/R36/R41).

Results

The EU and OECD classifications for the 59 chemicals, ordered in terms of decreasing MMAS, are given in Table III.

R41 classifications

Twenty two of the 59 chemicals were classified as R41, of which sodium hydroxide (10%; chemical 1) and benzalkonium chloride (10%; chemical 2) had the highest MMAS values (108), whereas 2,5-dimethylhexanediol (chemical 39) had the lowest MMAS (28). The R41 classification for the latter chemical is due to the persistence of eye effects in at least one rabbit.

R36 classifications

Twenty chemicals were classified as R36, of which the non-ionic surfactant Triton X-100 (10%; chemical 13) had the highest MMAS (69) and the pesticide Maneb (chemical 48) had the lowest MMAS (14).

NI classifications

Seventeen chemicals were classified as NI; Triton X-100 (5%; chemical 34) had the highest MMAS (32) and polyethylene glycol 400 (chemical 60) had the lowest MMAS (0).

Harmonised OECD classifications

With the exception of two chemicals, all of the chemicals classified as R41 were also classified as Category A, and all of the chemicals classified as R36 were also classified as Category B. The two exceptions, ethanol and

ethyl-2-methylacetoacetate, were classified as NI on the basis of EU criteria, but as Category B on the basis of the harmonised OECD criteria. The subcategory B1 (mildly irritating) was not taken into consideration. The difference in the classification of ethanol under the two systems is interesting, given that ethanol is used as the positive control in the bovine corneal opacity/permeability (BCOP) assay.

Application of MMAS cut-offs

The comparison between the EU classifications and the classifications obtained by applying the MMAS cut-offs of 25 and 59 is summarised in Table IV. The results show that the cut-off values cannot be used to classify chemicals reliably.

Conclusions

It is concluded that: a) the EU and proposed OECD systems for the classification of eye irritants/corrosives are broadly equivalent; b) the MMAS cut-offs of 25 and 59 are not appropriate for classifying chemicals according to the two systems; and c) future validation studies on alternatives to the Draize test would benefit from the use of classifications based on the proposed OECD harmonised system.

Acknowledgement

The author gratefully acknowledges Ingrid Gerner (BgVV, Berlin, Germany) for her

Table IV: Classification of eye irritants on the basis of Modified Maximum Average Draize Test Score (MMAS)

	European Union classification			Total
	NI	R36	R41	
MMAS ≤ 25	16	3	0	19
25 < MMAS ≤ 59	1	13	8	22
MMAS > 59	0	4	14	18
Total	17	20	22	59

NI = non-irritant; R36 = irritating to eyes; R41 = risk of serious damage of eyes.

constructive comments regarding this work.

References

1. 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.
2. Brantom, P.G. *et al.* (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.
3. OECD (1996). *Final Report of the OECD Workshop on Harmonization of Validation and Acceptance Criteria for Alternative Toxicological Test Methods*, 62pp. Paris, France: OECD.
4. OECD (1998). *Revised Proposal for the Harmonization of Hazard Classification Based on Eye Irritation/Corrosion*. ENV/MC/CHEM/HCL(98)5, 10pp. Paris, France: OECD.
5. 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.
6. ECETOC (1998). *ECETOC Technical Report No. 48, Eye Irritation: Reference Chemicals Data Bank*, 2nd edn, 236pp. Brussels, Belgium: ECETOC.