Regulatory Toxicology and Pharmacology 85 (2017) 132–149

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

journal homepage: www.elsevier.com/locate/yrtph

The Isolated Chicken Eye test to replace the Draize test in rabbits

Menk K. Prinsen [a, *], Coenraad F.M. Hendriksen [b], Cyrille A.M. Krul [c, d], Ruud A. Woutersen [d, e]

[a] Triskelion B.V., P.O. Box 844, 3700 AJ Zeist, The Netherlands
[b] Utrecht University/Institute for Translational Vaccinology, P.O. Box 450, 3720 AL Bilthoven, The Netherlands
[c] University of Applied Sciences Utrecht, Institute for Life Sciences & Chemistry, P.O. Box 12011, 2501 AA Utrecht, The Netherlands
[d] TNO Innovation for Life, P.O. Box 360, 3700 AJ Zeist, The Netherlands
[e] Wageningen University Research Centre, Department Toxicology, Wageningen, The Netherlands

ARTICLE INFO

Article history:
Received 14 October 2016
Accepted 27 January 2017
Available online 10 February 2017

Keywords:
Isolated Chicken Eye test
ICE
Chicken Enucleated Eye test
CEET
Alternative testing method
Eye irritation *in vitro*
Draize test

ABSTRACT

In 1944, Draize et al., published a paper entitled “Methods for the study of irritation and toxicity of substances applied topically to the skin and mucous membranes”. The Organization for Economic Co-operation and Development published their first guideline on eye irritation in 1981, using rabbits. In the early eighties the development of alternative non-animal tests to replace the Draize eye test started. The first attempts to validate alternative tests for eye irritation were considered to be relatively simple by comparing *in vitro* and *in vivo* irritation index scores. In the early nineteen-eighties, we introduced the use of isolated eyes as an alternative test for the Draize eye irritation test. What was expected to be a process of several years, however, turned out to be a decades spanning process still not fully completed. For a large part, this can be attributed to the nature of the *in vivo* test in rabbits, which is more complicated and compromised than originally believed. This paper describes, most chronologically, the development, performance, validation and application of the Isolated Eye Test and, in broader perspective, the international validation and acceptance of this alternative test by regulatory authorities and agencies.

© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Before industrialization, eye defects were mainly caused by physical trauma or by diseases caused by malnutrition, bacterial infection or parasites. In the twentieth century, when (chemical) industrialization strongly developed, it became apparent that conditions at the workplace could have distinct adverse effects on health and safety of employees. Acute and long-term exposure to a variety of industrial chemicals were responsible for a range of diseases, varying from relatively mild, non-life threatening phenomena, such as dermatitis, to incurable, lethal conditions such as cancer. After World War II, the chemical industry rapidly increased and workers became organized and more concerned with the potential risks they could encounter in the workplace. Consequently, the need for identifying health hazards and worker's protection became an important issue in most industrial countries. Moreover, people could afford more luxury products and the household and personal care industry became more and more innovative using new technologies and (chemical) ingredients. Therefore, an even larger population of people needed to be safeguarded from potential hazardous substances.

To establish the potential risk of exposure of the eyes to compounds, the Food and Drug Administration of the United States (US-FDA) adopted the Draize eye irritation test using rabbits already in 1961.

At first sight, this test is simple and straightforward and provides a useful tool for regulators. However, the controversial character of this type of animal testing became known to the general public – on 15 April 1980, Henry Spira, a Belgian-American advocate, member and founder of the Animal Rights International group bought a full-page advertisement in the New York Times, with the header: “How many rabbits does Revlon blind for beauty's sake?” – and the need to develop alternative non-animal tests became apparent. Within a year after Spira's advertisement, Revlon had

* Corresponding author. Triskelion B.V., P.O. Box 844, 3700 AJ Zeist, The Netherlands.
E-mail addresses: menk.prinsen@triskelion.nl (M.K. Prinsen), coenraad.hendriksen@intravacc.nl (C.F.M. Hendriksen), cyrille.krul@tno.nl (C.A.M. Krul), ruud.woutersen@tno.nl (R.A. Woutersen).

http://dx.doi.org/10.1016/j.yrtph.2017.01.009
0273-2300/© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

List of abbreviations

AZAN	Azocarmine & aniline
BCOP	Bovine Corneal Opacity and (fluorescein) Penetration test
CEET	Chicken Enucleated Eye test
CRO	Contract Research Organization
EC	European community
ECVAM	European Centre for the Validation of Alternative Methods
EU	European Union
EVG	Elastic Van Gieson
FDA	Food and Drug Administration
H&E	Haematoxylin & eosin
HET-CAM	Hen's Test - Chorioallantois Membrane
HO	British Home Office
ICCVAM	Interagency Coordinating Committee on the Validation of Alternative Methods
ICE	Isolated Chicken Eye test
IRAG	Interagency Regulatory Alternatives Group of the United States
IRE	Isolated Rabbit Eye test
MMAS	Modified Maximum Average Score
NEI	National Eye Institute, USA
OECD	Organisation for Economic Co-operation and Development
PAS	Periodic acid-Schiff
PM	Prediction Model
REET	Rabbit Enucleated Eye test
TG	Test Guideline
TNO-CIVO	Toegepast Natuurwetenschappelijk Onderzoek – Centraal Instituut voor Voedingsonderzoek
UN-GHS	United Nations - Globally Harmonized System

donated $750,000 to a fund to investigate alternatives to animal testing, followed by substantial donations from Avon, Bristol Meyers, Estée Lauder, Max Factor, Chanel, and Mary Kay Cosmetics. These donations led to the creation of the Centre for Alternatives to Animal Testing (http://caat.jhsph.edu).

The attempts to validate alternative tests for eye irritation in the early nineteen-eighties were considered to be relatively simple by comparing *in vitro* and *in vivo* irritation index scores. What was expected to be a process of several years, however, turned out to be a decades spanning process still not fully completed.

For a large part, this can be attributed to the nature of the *in vivo* test in rabbits, which is more complicated and compromised than originally believed.

This paper describes, most chronologically, the development, performance, validation and application of the Isolated Eye Test and, in broader perspective, the international validation and acceptance of this alternative test by regulatory authorities and agencies.

A considerable part of the paper deals with the *in vivo* Draize rabbit eye test itself, because its performance and the use of its results as the golden standard to compare the *in vitro* test, are considered to be the main obstacle for replacing one of the most controversial experimental animal tests within safety testing today.

Since the introduction of the alternative test with isolated eyes, two different isolated eye models exist for the same test, each having two different abbreviations, namely the Rabbit Enucleated Eye Test or the Isolated Rabbit Eye test (REET or IRE) and the Chicken Enucleated Eye Test or the Isolated Chicken Eye test (CEET or ICE).

2. The Draize eye irritation test

On 2 November 1944 a manuscript, entitled “Methods for the study of irritation and toxicity of substances applied topically to the skin and mucous membranes” was received for publication by the Journal of Pharmacology and Experimental Therapy. The authors of this article were John H. Draize, Geoffrey Woodard and Herbert O. Calvery from the Division of Pharmacology, Food and Drug Administration, Federal Security Agency, Washington, D.C., USA. It is more than likely that the authors never expected the kind of impact this publication would have on animal experimentation worldwide. Almost seventy years later the name Draize is still inextricably attached to two of the three most disputed toxicity tests commonly used to determine acute toxicity, i.e. the Draize eye irritation test, the Draize skin irritation test and the LD50 (lethal dose) test. The latter two tests fortunately have already been replaced by *in vitro* tests (skin irritation) or by test methods using much less animals and causing less discomfort (LD50).

The Draize eye irritation test was first adopted by the US-FDA as part of the safety evaluation of foods, drugs and cosmetics (US-Federal Register, 1961). At that time already, it was recognized that the subjective grading of ocular reactions posed a considerable problem. In order to standardize the scoring and to provide guidance to the observers, an illustrated guide was issued (FDA, 1964). Internationally, the OECD published their first guideline on eye irritation in 1981, which was subsequently adopted by the European Union (EC, 1984).

Since then several revisions of the guideline have followed, mostly not affecting the actual exposure procedure, but providing guidance for refinement and reduction of animal use and discomfort (Table 1). Examples are the exemption of testing skin corrosives and substances with pH lower than 2.0 or higher than 11.5, the use of well-validated alternatives as a screen for severe irritancy, and a tiered approach of testing (i.e. starting with one animal and continue only if non-severe irritancy is observed).

The design of the eye irritation test is actually quite simple and straightforward: a rabbit is placed on a worktable and restrained either manually or in a fixation-box. Next, the lower eye-lid is pulled out and the test substance is instilled in the conjunctival cul-de-sac formed; the upper and lower eye lids are then closed and subsequently held together for at least one second before releasing the animal. The other eye remains untreated and serves as a control.

The animal is returned to its cage and is free to remove the material. The control and test eyes are examined (without optical aid) at approximately one hour, and at approximately 24, 48, and 72 h after treatment. Ocular reactions of the test eye are judged using a scoring scale (Table 2). Residual eye effects are recorded at regular intervals, if necessary up to about 3 weeks after treatment, in order to allow the evaluation of the reversibility or irreversibility of the effects elicited. Liquids are tested in a volume of 0.1 mL and solids (ground to a fine powder) in an amount of 0.1 g or a volume of 0.1 mL. In general, 0.1 mL is the amount the conjunctival cul-de-sac can hold when the lower eye-lid is pulled out.

Despite the existence of many national guidelines on eye irritation, the exposure procedure and the scoring system for ocular

Table 1
OECD test guideline no. 405 and its revisions (procedures, interpretation results, ethics and 3 R's).

OECD TG 405	Procedure	Guidance on interpretation of results	Ethical considerations	Three R's
1981	- 0.1 mL or 0.1 g substance; - wash out only after 24 h	- Extrapolation of the results of eye irritation studies in animals to man is valid only to a limited degree. The albino rabbit is more sensitive than man to ocular irritants or corrosives in most cases. - Similar results in tests on other animal species can give more weight to extrapolation from animal studies to man. - Care should be taken in the interpretation of data to exclude irritation resulting from secondary infection.	Local anaesthetics proposed	- Three instead of six rabbits No testing of: - Strongly acidic or alkaline substances - Corrosive or severe skin irritants
1987	- 0.1 mL or 0.1 g substance; - wash out only after 24 h	Identical to 1981 Guidance	Addition of: - Animals showing severe and enduring signs of distress and pain may need to be humanely killed.	Addition of: - severe eye irritants identified in well-validated alternative studies
2002	- 0.1 mL or 0.1 g substance; - wash out after 1 h (solids)	Similar to 1981 and 1987 Guidance	Addition of: - End points for humane sacrifice - Tiered testing	Addition of: - Weight-of-the-evidence analysis on the existing relevant data - Conduct of validated and accepted *in vitro* tests; - One rabbit first
2012	- 0.1 mL or 0.1 g substance; - wash out after 1 h (solids)	Similar to 1981, 1987 and 2002 Guidance	Addition of: - Extensive directions for the use of topical anaesthetics and systemic analgesics	Addition of: - ICE test (OECD 438) - BCOP test (OECD 437)

lesions are basically identical. However, the classification systems differ considerably (Tables 3 and 4). In general, four classifications are assigned on the basis of the ocular lesions, viz. not irritating (not classified), mildly irritating, irritating and severely irritating. The EU recognizes three classifications, i.e. not classified, irritating and severely irritating (risk of serious damage to the eye). The existence of these different classification and labelling systems is not favourable for the validation of alternative test methods. Therefore,

Table 2
Draize scheme for grading of ocular lesions in the rabbit.

Tissue	Lesion	Score
Cornea Opacity-degree of density (area most dense taken for reading)	No opacity	0
	Scattered or diffuse areas, details of iris clearly visible	1[a]
	Easily discernible translucent area, details of iris slightly obscured	2
	Opalescent areas, no details of iris visible, size of pupil barely discernible	3
	Opaque, iris invisible	4
Iris	Normal	0
	Folds above normal, congestion, swelling, circumcorneal injection (any or all of these or combination of any thereof); iris still reacting to light (sluggish reaction is positive)	1[a]
	No reaction to light, haemorrhage, gross destruction (any or all of these)	2
Conjunctivae - Redness	Vessels normal	0
	Vessels definitely injected above normal	1
	More diffuse, deeper crimson red, individual vessels not easily discernible	2[a]
	Diffuse beefy red	3
Conjunctivae - Swelling	No swelling	0
	Any swelling above normal (including nictitating membrane)	1
	Swelling with lids about half closed	2[a]
	Swelling with lids about half closed to completely closed	3

[a] Lowest score considered positive according to US-EPA.

Table 3
European Union (1993[a]) classification system for eye irritation/corrosion.

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

[a] Official Journal of the European Communities, L 110 A, Volume 36, 4 May 1993.
[b] The classification is assigned if the mean tissue effect (averaged over the 24 h, 48 h and 72 h time points) exceeds the threshold value in at least two of the three animals.
[c] The classification is assigned if the mean tissue effect (averaged over the three time points and over the six animals) exceeds the threshold value.
[d] R41 is also assigned if, in at least one animal, one of the eye effects has not reversed at the end of the observation period.

Table 4
US-EPA (1998[a]) classification system for eye irritation.

Toxicity categories	Category I	Category II	Category III	Category IV
Eye effects	Corrosive (irreversible destruction of ocular tissue) or corneal involvement or irritation persisting for more than 21 days	Corneal involvement or irritation clearing in 8–21 days	Corneal involvement or irritation clearing in 7 days or less	Minimal effects clearing in less than 24 h

[a] Health Effects Test Guideline OPPTS870.1000, EPA 712-C-98-189, August 1998.

the implementation of the classification system of the United Nations Globally Harmonized System (UN-GHS; Table 5) in 2007 is considered to be an improvement, although there is still a difference with the system the EU applies. The UN-GHS system subdivides the irritating category (Category 2) into mild irritant (Category 2B) and irritant (Category 2A), whereas the EU only uses the category irritant (Category 2).

2.1. Awareness of alternatives for animal testing

The publication of Russell and Burch in 1959 entitled: "The Principles of Humane Experimental Technique" stood at the basis of most initiatives relating to the use and development of alternatives for animal experiments. In their publication they postulated the famous and often cited three R's: Reduction, Refinement and Replacement of animal experiments. Nowadays, the 3Rs have become a mantra for scientists and regulators in research areas involving animal experimentation. The initiatives concerning the Draize eye test mainly involved reduction of the number of animals from six to three per test and replacement by the implementation of non-animal alternatives. Certain aspects of the Draize eye test causing considerable pain and discomfort to the animal were dealt with only at a much later stage, i.e. reduction of the time for a washout of the test substance from 24 h to 1 h after instillation in 2002, and the use of systemic pain relief and topical sedation in 2012 updates of the OECD guideline 405 (Table 1).

In the early nineteen-eighties, the Netherlands Society of Toxicology (NVT) started a working group named Critical Evaluation of Toxicity Testing and in Europe, the European Research Group for Alternatives in Toxicity Testing (ERGATT) was founded to stimulate innovative toxicological research and to act as a counterpart to the John Hopkins Centre of Alternatives to Animal Testing in the USA which was founded in 1981.

For eye irritation, the policy was to select one of the most promising existing alternative methods and to focus on further development, standardization and validation in order to develop a method that would be acceptable for regulatory purposes. In addition, recommendations for a tiered approach to eye irritation testing were made, viz. testing skin irritation first, and starting the eye irritation test with one rabbit.

Because the cornea is such a highly relevant target tissue in eye irritation, it was taken as the basic principle for the development of a relevant and practical *in vitro* alternative to the animal test that had been in use as the sole test for the screening of eye irritation worldwide since the early forties of the twentieth century.

2.2. Isolated eye test method (rabbit)

In 1981, Burton published a method using isolated rabbit eyes for the *in vitro* assessment of severe eye irritants. Previously, he had discovered that the measurement of corneal thickness (swelling) by slit-lamp examination provided an objective assessment of eye irritation in the *in vivo* rabbit eye irritation test (Burton, 1972). He had examined 100 different cosmetic formulations in 600 rabbits and found not only a close relationship between the total corneal Draize score and the recorded corneal swelling, but also a relationship between corneal swelling and the conjunctival effects scored subjectively. Around that time another article on the usefulness of slit-lamp examination in the rabbit eye irritation test, including corneal thickness, was published (McDonald et al., 1973).

Between 1972 and 1981, Burton did not publish further on this subject, but it is assumed that he played with the idea of replacing the live rabbit by isolated rabbit eyes only. In his 1981 publication no further considerations for using isolated eyes were given, but a possible clue may be found in the literature reference he used for the design of the superfusion apparatus (used for maintaining the

Table 5
GHS (2007[a]) classification system for eye irritation/corrosion.

Eye effects	Category 2A[b]	Category 1[c]
Corneal opacity	≥1.0	≥3.0
Iris lesion	≥1.0	>1.5
Conjunctiva redness	≥2.0	
Conjunctiva chemosis	≥2.0	

[a] Globally Harmonised System of Classification and Labelling of Chemicals (UN-GHS) UN, New York and Geneva, 2007.
[b] All effects have to be reversible within 21 days of treatment. Subcategory of 2B: mildly irritating to the eyes, i.e. eye effects reversible within 7 days of treatment.
[c] Category 1 is also applicable if, in at least one animal, an eye effect has not reversed, or is not expected to reverse, within 21 days of treatment.

isolated eyes in good condition), which he had modified from the one described by Mishima and Kudo in 1967. Remarkably, Burton had already referred to publications by Mishima in his 1972 article, and surely have thought about the possibility of using isolated rabbit eyes in a superfusion apparatus at that time. It remains unclear why he did not pursue the use of isolated rabbit eyes sooner.

The idea of using isolated rabbit eyes was very appealing from a scientific point of view. After all one uses an *ex vivo* eye for an eye *in vivo* and, moreover, the parameters measured (corneal swelling, corneal opacity and epithelial cytoxicity by fluorescein dye) are directly comparable to the parameters measured *in vivo* (both in rabbit and in human). Therefore, Koëter and Prinsen proposed to introduce an *in vitro* eye irritation test (with isolated rabbit eyes) as a possible contribution to the reduction of experimental animal use (Koëter and Prinsen, 1985). The test method was evaluated by investigating the effects of several substances from the publication of Burton et al. (1981). The test method was further validated with 34 substances that had been investigated in the *in vivo* eye irritation test in rabbits as part of the standard toxicity testing (Koëter and Prinsen, 1985).

During the same period, several other investigators explored the use of isolated rabbit eyes as an alternative for the *in vivo* test (York et al., 1982; Price and Andrews, 1985; Jacobs and Martens, 1988, 1990).

The research and publications on isolated rabbit eyes resulted in the inclusion of this test method in the first EC Collaborative Study on the Evaluation of Alternative Methods to the Eye Irritation Test (EC, 1991). In this study, five *in vitro* cell toxicity tests, the Rabbit Enucleated Eye Test (REET) and the Hen's Egg Test Chorioallantoic Membrane (HET-CAM) were selected to undergo validation by testing 21 chemicals of different classes in at least 3 different laboratories. It was concluded that: i) The Isolated Rabbit Eye test did not misclassify many non-irritants and also had the capability to discriminate between moderate and severe effects, although irritating (R36) chemicals were underrepresented. ii) The REET produced results which were consistent across all three laboratories and generally correctly predicted the *in vivo* grade. The protocol and the method for calculating final irritancy grades (in validation studies later on called “Prediction Model”) needed harmonization before a wider interlaboratory study could be conducted. iii) The REET is nearest to the human situation and has the advantage that all types of chemicals can be investigated without the need for testing dilutions, therefore the study showed that the REET is easier to interpret than the other assays in the trial.

The REET results showed an overall correlation of 82% with the *in vivo* results, using a general classification scheme for the grading of *in vivo* and *in vitro* eye irritation (i.e. not, slight, moderate or severe irritant). Four compounds, overpredicted (slight or moderate instead of non-irritant) by the REET, were all moderate to severe irritants in the *in vivo* skin irritation test. A common physicochemical characteristic of these compounds was their hydrophobicity. A general observation in the *in vivo* rabbit eye test was that hydrophobic compounds stayed in contact with the cornea (eye) for a relatively short period of time, because they mixed poorly with the tear film on the cornea and because the nictitating membrane (third eye-lid) acted as a wiper, and rapidly removed the compound from the eye. This was in distinct contrast to the skin irritation test where the compound was kept in contact with the skin with the aid of a patch and fixative tapes for at least 4 h. The presence of a third eye-lid is an example of a condition very specific to the *in vivo* rabbit eye test influencing the exposure and cause a problem with respect to the validation of alternative methods. alternative methods are not able to mimic the presence of a third eye-lid, which is also an irrelevant condition with respect to humans. On the basis of the results with the 34 compounds, the REET was considered to be a sensitive and useful test system for the identification of eye irritants. Negative *in vitro* results should be confirmed only in case of expected regular eye contact and in a maximum of three rabbits. At that time six rabbits was the usual number in the Draize eye test in order to comply with the US guidelines, although the first OECD guideline on eye irritation (OECD, 1981) recommended the use of 3 rabbits, which could mean a considerable reduction in the use of rabbits for eye irritation testing.

The rapporteurs, however, considered the trial to be too limited to make firm conclusions and it was recommended to perform further interlaboratory trials with a larger number of laboratories and chemicals and according to the principles of GLP. An important remark in the EC report was the fact that toxicological profiles of the chemicals investigated were prepared by collecting and critically evaluating the literature data available, because it was not possible to repeat *in vivo* eye determinations for animal protection considerations. The limited availability, evaluation and appraisal of *in vivo* eye irritation data and the test method itself constitute the main cause of the exceptionally long, and not yet completed, acceptance of alternative methods to the Draize eye test. Until recently, the *in vivo* data were taken as the “Golden Standard”, which in practice meant that the *in vitro* data had to almost exactly match the variable *in vivo* result.

2.3. The use of slaughterhouse animals

Although rabbits are available as eye-donor for the isolated eye test in contract research organisations (CROs) executing routine eye- and skin irritation studies, the scientific community considered this dependency on laboratory animals a serious shortcoming. The isolated eye test would still be associated with animal experimentation and organisations not using laboratory rabbits could have difficulties in obtaining rabbit eyes. Therefore, the use of slaughterhouse animals, such as the cow, pig or chicken, as eye-donor was investigated. Firstly, the possibility to obtain eyes from abattoirs was explored and, secondly, possible practical limitations of the type of eye in relation to the experimental set-up were identified. On the basis of these explorations, the cow and pig were rejected as suitable eye-donor, although the latter was expected to be the most suitable candidate on the basis of its physiological resemblance with the human eye. The rejection of the cow eye was based on the following observations: 1) the irregular supply and variable origin of cows and, related to that, 2) the insufficient quality of the cornea in too many cases. In the case of the pig eye, the collection of suitable eyes by the investigator at the process line of the slaughterhouse appeared difficult, but doable. The most important feature of the pig's eye, also applying to the cow's eye, was unfavourable for its use in the isolated eye test, namely its corneal thickness. The pig's cornea is relatively thick (600–700 μm; Faber et al., 2008) when compared to the cornea of the rabbit (ca 400 μm; Chan et al., 1983; Li et al., 1997), but quite comparable to the thickness of the human cornea (ca 530 μm; Doughty and Zaman, 2000; Fowler et al., 2004). The cow's cornea easily doubles that thickness (900–1100 μm; OECD TG 437). Although the pig eye is better comparable to the human eye, the isolated eye test has to produce matching results with the Draize rabbit eye test in order to be accepted as an alternative. Hence, an eye (cornea) that matches closest to the rabbit eye and not to that of the human eye was needed. The chicken provided such an eye, i.e. its dimensions and corneal thickness (ca 400 μm; Fowler et al., 2004) are similar to the rabbit cornea. Another point of concern was the baseline corneal thickness of the pig which was in the upper scale of the thickness measurement device and recording of increased thickness to its full extend after treatment with moderate to severe

irritants appeared not possible. Other possibilities for measuring the corneal thickness in the higher range were not available for the Haag-Streit microscope in use. Ultrasonic pachymetry instead of the mechanical (non-invasive) measurement device was an option but the procedure required touching the cornea with a probe. In case of irritating substances, touching the cornea could result in additional damage and was therefore considered unsuitable.

The epithelium of the cornea is the first barrier against (chemical) insult and as such its thickness (number of layers of epithelial cells) is of importance. One of the most used alternative tests for eye irritation is the BCOP, which uses excised and isolated corneas of the cow. Because of its corneal thickness (not used as a parameter in the BCOP), the exposure time needed to elicit an irritation response is 10 min (for liquids) or 4 h (for solids) in the BCOP. In the ICE and REET test, only a 10-second contact period is needed to elicit a relevant irritation response. A relevant response in this case means a response comparable to that observed in the *in vivo* rabbit test. Apart from a comparable thickness to the rabbit cornea and, to a certain extent to the human cornea, the chicken cornea has another feature, which might be an advantage to predict (human) eye irritation, i.e. a well-developed Bowman's membrane (Fowler et al., 2004). This membrane, which is in between the epithelium and stroma, is also well developed in the human cornea, but poorly developed in the rabbit cornea.

With respect to the availability of eyes, attaining chicken eyes appeared relatively easy by simply collecting heads just after the sedation of the animals at the process line, transporting them to the laboratory and enucleating them from the heads within 2 h after sacrifice. This time period is sufficient to guarantee corneas meeting the acceptance criteria for testing, i.e. no or very slight opacity, no or very slight fluorescein staining, and a corneal thickness in the normal range. The daily processing of thousands of chickens at the slaughterhouse guaranteed a constant supply of suitable eyes (corneas). Since 1981 we have visited the same slaughterhouse (Nijkerkerveen, The Netherlands) for our supply and only during a period of about 6 months in 2003 could eyes not be obtained because of an Avian flu break-out (H7N7 variant) in a large part of the Netherlands. Even then, eyes could be obtained from another slaughterhouse outside the affected region.

The use of chicken eyes was evaluated by testing the 21 reference chemicals previously tested in the first (pilot) EC validation study of the REET (EC, 1991) which were selected to be representative of currently used industrial chemicals of different chemical classes and ranging from non-irritant to severe irritant. Describing the criteria and scoring system of the *in vitro* corneal effects together with a Prediction Model (PM) for matching the EC scheme for classification and labelling of compounds (Table 3) was an important step forward. The development of the prediction scheme was primarily a theoretical exercise based on the range of physiological responses observed in the ICE (corneal swelling, opacity and fluorescein retention of damage epithelial cells). Because the corneal effects determined in the ICE have a direct relationship with the *in vivo* response (e.g. *in vitro* opacity for prediction of *in vivo* opacity), a PM could be established based on the magnitude/range of the effects and not by using an empirically-derived mathematical algorithm to translate an *in vitro* effect to an *in vivo* effect. For instance, in the HET-CAM assay lysis of blood vessels of the chorioallantois membrane is measured as time (seconds) of first occurrence (Luepke, 1985). The number of seconds recorded cannot be translated directly to an *in vivo* effect, which is considered a serious limitation of the method. A PM could be designed only after computer calculations of data obtained for a relevant number of compounds by using the *in vivo* MMAS (Modified Maximum Average Score) as the sole parameter for the *in vivo* test. The computer calculation was based on a mathematical formula or conversion algorithm resulting in a single *in vitro* irritation index score comparable to the *in vivo* MMAS. Computer calculations were also needed with the BCOP assay, using a cornea, because the opacity is measured as a reduced light transmission value and the epithelial damage as the amount of fluorescein penetrated through the cornea, leading to an *In Vitro* Irritation Score (IVIS = mean opacity value + (15 × mean permeability OD490 value)).

The ICE followed a theoretical approach not using *in vitro* data to be compared to the *in vivo* MMAS. By knowing the ranges of the *in vitro* ICE responses (swelling, opacity and fluorescein retention), cut-off values were chosen to identify different categories of effect, viz. a non (Category I), slight (Category II), moderate (Category III) or severe (Category IV) effect. Thus after testing a compound, three categories were established, i.e. one for swelling, one for opacity and one for fluorescein retention.

The assignment of the final irritation classification to a non-irritant, (slight or moderate) irritant or severe irritant was obtained by the combination of these three categories. Again, a theoretical and weighted division of the different combinations possible was made for each final classification. For instance, at the low end of the ICE classification system the combination of I/I/I is a non-irritant and at the high end the combination of IV/IV/IV is a severe irritant. The combinations possible and respective irritation classifications are shown in Table 6.

Often the PM of other alternatives had to be revised after more compounds had been tested by the *in vitro* test or additional PM's were especially designed for certain categories of compounds. With the ICE, the criteria system for scoring the effects was never changed, while the classification system has been modified twice, i.e. once to accommodate the introduction of the UN-GHS classification system and secondly after adoption as an OECD guideline for non-irritants (OECD, 2013a).

The use of the *in vivo* MMAS in establishing the PM proved to be less ideal than thought. The EC and UN-GHS systems do not use the MMAS for determination of the irritation classification of a compound. Instead, the individual *in vivo* tissue scores are used. Evaluation of the use of the MMAS in the validation study of the European Commission and the British Home Office (EC/HO study) showed a poor correlation with these classification systems (Prinsen, 1999).

On the basis of the ICE study with 21 reference materials (Prinsen and Koëter, 1993) the conclusions were that: 1) although the ICE does not assess conjunctival damage, its sensitivity to predict ocular damage is not reduced, 2) the ICE correctly predicted the EC classifications of the 21 chemicals and 3) the ICE fitted in the previously updated EC B.5 and OECD 405 guidelines regarding acute eye irritation/corrosion now including recommendations to use alternatives for the prescreening or positive identification of strong eye irritants.

2.4. Isolated eye test method (chicken)

The use and suitability of eyes of slaughterhouse animals was first established by testing the same reference chemicals (Prinsen and Koëter, 1993) that had been tested in the EU Collaborative Study on the Evaluation of Alternative Methods to the Eye Irritation Test (EC, 1991). Although the *in vivo* data were obtained from literature this study was considered quite valuable because the *in vitro* data obtained with the slaughter eyes could be directly compared with the *in vitro* data obtained in the REET. Ideally, the *in vitro* test should be performed in parallel with the *in vivo* test, hence enabling a more direct comparison of the data.

Prinsen (1996) published the results of such a parallel testing of 44 compounds, which were considered to be a relevant cross-section of compounds (raw chemicals, finished products and

Table 6
ICE *in vitro* classification system.

General classification	UN-GHS classification[a]	Combinations of categories
Not irritating	Not classified	3 × I 2 × I, 1 × II
Slightly irritating	2B: *Mild irritant/causes eye irritation*	2 × II, 1 × I 3 × II 2 × II, 1 × III 1 × I, 1 × II, 1 × III
Moderately irritating	2A: *Irritant/causes eye irritation*	2 × III, 1 × I 2 × III, 1 × II 3 × III 1 × IV, 2 × I 1 × IV, 2 × II x IV, 2 × III 1 × IV, 1 × III, 1 × II
Severely irritating	1: *Irreversible effects on the eye/serious damage to the eye*	2 × IV,1 × I 2 × IV,1 × II 3 × IV

[a] Globally harmonised system of classification and labelling of chemicals (UN-GHS). UN, New York and Geneva, 2007.

formulations) routinely produced by the (chemical) industry. Instead of only comparing single *in vivo* and *in vitro* irritation index scores, as was the common practice in validation, the individual components used for calculation of the index score were also analysed. This was a recommendation of the United States Interagency Regulatory Alternatives Group (IRAG) made during a workshop on eye irritation testing in Washington DC, in 1993 (Scala and Springer, 1997). Fourteen different *in vivo* scores were derived from each of the 44 *in vivo* tests, covering time scores (1-hr, 24 h, 1–72 h, days to recovery), index scores (MAS and total eye score), and individual tissue scores (cornea, area of cornea involved, iris, cornea and iris combined, conjunctival redness, conjunctival swelling, conjunctival discharge, conjunctival scores combined). The *in vivo* critical scores were compared to the critical scores of the ICE test, namely the scores for corneal swelling, corneal opacity, fluorescein retention of damaged corneal epithelium and an index score (combination of the three parameters). The overall correlations found for the *in vivo* scores with the ICE *in vitro* scores were 0.90 (index score), 0.91 (corneal swelling), 0.86 (corneal opacity) and 0.82 (fluorescein retention). The correlation between the *in vivo* conjunctival scores and the ICE scores were 0.92 (index score), 0.92 (corneal swelling), 0.93 (corneal opacity) and 0.86 (fluorescein retention). These correlations substantiated the conclusion made by Burton (1972) that a relationship exists between the *in vivo* conjunctival damage and the corneal scores of the isolated eye test. Moreover, in ophthalmology the term "ocular surface" was introduced to emphasize the potential interdependence of the epithelium of the cornea and the epithelium of the conjunctivae (Thoft and Friend, 1977; Wagoner, 1997). "Subsequent clinical and research insights provided compelling evidence of the functional relationships between these two adjacent cell populations" (Thoft and Friend, 1977). Furthermore, the proven relationship and high correlation between the critical scores of the ICE test and the Draize eye test demonstrated that the test is a relevant alternative to eye irritation, and that applying regulatory irritation classification systems is "just" a matter of choosing the appropriate threshold limits belonging to the different irritation classes. This was supported by the *in vivo* and *in vitro* EC classifications obtained for the 44 compounds. Overall, it was concluded that the ICE provided a practical prescreen for the Draize rabbit eye test and that only mild to moderate irritants in the ICE, generally showing the highest sensitivity to inter- and intra-laboratory variability, should be confirmed in the rabbit eye test. Eighteen years later, the OECD adopted this conclusion. Furthermore, the "parallel" testing showed the ICE test to be robust in the sense that the practical aspects are not complicated and relatively easy to control, i.e. a saline drip is sufficient to maintain the eyes in good condition, and all compounds, regardless the physico-chemical properties, can be assayed.

The in-house repeatability of the ICE was assessed to be adequate during the Reference Standard Validation of *in vitro* tests sponsored by ECVAM (Brantom et al., 2000). Two reference compounds for the group of siloxanes (decamethylcyclopentasiloxane and cyclohexylamino-functional PDMS) and two for the group of surfactants (Triton X-500 5% and cetylpyridium bromide 6%), representing non-irritants, Category 2B and 2A, and Category 1 compounds, were tested five times each on different occasions.

The publication of the results for the first 44 compounds did not result in termination of the "parallel" testing program. The main reason for continuing the "parallel" testing was that it was considered unethical to perform any toxicity test on live animals without prior information on the reactivity of the test compound using a "relevant" biological structure such as the cornea. In those days, it was common practice to start different acute toxicity tests with a new compound almost simultaneously and, if different study directors were involved, often without consulting each other about the specific results of their studies, whereas the result of an acute irritation test would have influenced or helped their decision concerning the study design to be followed. It became apparent that the ICE fitted well in a tiered approach for acute toxicity testing. The results of the ICE provided not only information on eye irritation, but also gave information that could help to optimally design the other acute toxicity tests. If the ICE test showed severe irritancy, the *in vivo* eye irritation test was waived and the skin irritation test was initiated with one rabbit only. Important decisions for the conduct of the acute oral and dermal toxicity tests in rats could also be made on the basis of the outcome of the ICE. In most cases, these studies were started as a limit study with the highest dose level of 2000 mg/kg body weight. In case of severely irritating or corrosive compounds, the local effects on the stomach or skin could lead to severe suffering or even mortality of the animals. When the ICE test showed severe effects, dosing of high levels or high concentrations of corrosive compounds could be avoided. When the ICE showed no or negligible signs of irritation (cytotoxicity), the decision to perform the acute oral and dermal tests with the highest dose level or a high test concentration could be better justified.

It was not until 2001 that the OECD adopted the use of results of any other *in vitro* toxicity test on a compound in order to determine start levels to be used for the *in vivo* acute oral toxicity test in rats

(OECD, 2001).

For skin sensitization tests, this approach reduced the number of animals necessary for testing. At that time, the standard test for skin sensitization was the Guinea Pig Maximization test (GPMT), requiring up to 40 animals for the main test and 6–9 animals for the preliminary skin irritation test. The preliminary skin irritation test was needed to establish the appropriate (maximum tolerable) concentrations for the various phases of the study (i.e. the intradermal and topical induction and the topical challenge). Normally ranges covering concentrations from 1% up to 100% had to be investigated. When the ICE showed no or negligible irritancy, the range to be examined could be limited to only 100% and one lower concentration, which in practice meant that only 3 and not 6–9 animals were needed. At present, the GPMT is replaced by the Local Lymph Node Assay (LLNA; OECD, 2010a,b), and this guideline also mentions the use of results of any other *in vitro* toxicity test on a substance as an aid in dose selection.

In retrospect it can be concluded that the testing strategy was successful for the majority of compounds submitted for testing, i) identification of severe eye irritants without further *in vivo* testing.

ii) tiered testing of skin irritants/corrosives, iii) determination of acceptable (non-severe) dose levels of corrosive compounds in acute oral and dermal toxicity testing, and iv) reduction of the number of animals used in the preliminary irritation experiment of sensitization studies.

"Parallel" *in vivo* and *in vitro* eye irritation testing was continued with another 50 compounds, meaning that each compound was first tested in the ICE and, in case of non-severe irritancy, directly followed by an *in vivo* rabbit test, both in full compliance with the OECD principles of Good Laboratory Practice (GLP). These results were submitted to organizations dedicated to the validation of alternative non-animal tests, such as ECVAM and ICCVAM. Because the performance of the ICE was considered sufficiently established after the "parallel" testing of 94 compounds, another approach was introduced to make optimal use of the ICE at low extra costs. First a non-GLP ICE test with only one eye was carried out and depending on the outcome either a full GLP ICE test (in case of severe irritancy) or an *in vivo* rabbit eye test (in case of non-severe irritancy) was carried out. This procedure was followed until the ICE OECD guideline 438 was adopted to include the identification of non-irritants (OECD, 2002). From then on, only compounds identified by the ICE as irritating (Category 2) need to be tested in the *in vivo* eye irritation test in rabbits.

Based on the results of their microscopic examination of the corneas in the Low Volume Eye Test (LVET; Griffith et al., 1980; Griffith, 1987). Maurer et al. (2002) recommended to include histopathology of the treated corneas sampled at the end of the observation period. They suggested a direct correlation between the depth of injury in the cornea and possible recovery of eye lesions. Since measurement of recovery is not possible in the ICE (and in all other alternatives) it could be an important improvement of the method and helpful in the argumentation for absence or presence of microscopic lesions in the ICE that could predict (ir) reversibility. The outcome of this study (Schutte et al., 2009) was that, in general, the results of the ICE using the standard volume of 30 μl or mg were in line with or more conservative than the LVET in terms of classification. The level of overprediction found in the ICE was expected and considered acceptable since the ICE was developed for prediction of the Draize eye test and not the LVET.

For industry, the ICE test was considered useful for several purposes, such as 1) EU/GHS classification and labelling of powder and liquid household cleaning products, 2) screening of candidate formulations, and 3) weight-of-evidence approach by determining the profile of new cleaning product formulations against benchmark products. A definite conclusion on the usefulness of histopathology in the ICE could not be made, but the data showed that assessment of the histopathological lesions in the various parts of the cornea was possible, enabling the application of the "Depth-of-Injury" theory of Maurer et al. (2002). The question remained if this theory established on the basis of an inflammatory process in the rabbit's eye (consisting of initial ocular injury and subsequent repair over days/weeks) correlated with the irritation process or damage in the ICE.

2.5. Optimizing histopathology in the Isolated Chicken Eye test

For the histopathological observations in the ICE, a quite basic and routine procedure of processing the eye/cornea is used. The eye is preserved in formalin at the end of the study, i.e. 4 h after the 10-second exposure, and further processed into a paraffin block from which slides of the longitudinal section of the corneal centre are prepared. This area was considered appropriate since the application of the test compound was made at the centre of the cornea and in general the slit-lamp observation showed confluent, homogenous corneal effects. In case of non-homogenous effects, such as focal spots with more severe opacity not present in the central area, these parts were also examined. The choice of the staining appeared an important issue. In first instance, the most common staining by Haematoxylin & Eosin (H&E) was used, which was later replaced by the Periodic Acid-Schiff (PAS) staining, which provided a much better discrimination of the different layers of the cornea. The three major layers of the cornea, i.e. epithelium, stroma and endothelium were well visible with both stainings. The other structures, such as basement membrane and Bowman's membrane (between epithelium and stroma) and Descemet's membrane (between stroma and endothelium), were not that visible with H&E, whereas PAS provided much better results. The integrity of the membranes was considered to play a role in the injury and recovery process of the cornea, and the visibility of these membranes by microscope was considered important for an adequate histopathological assessment by the pathologist. Therefore, other staining methods specifically targeting collagen-rich membranes were tested on corneas treated with compounds representing a non-irritant, two irritants and a severe irritant (corrosive). The microscopic examination focused on the basement and Bowman's membrane and not on Descemet's membrane, because damage to this membrane adjacent to the endothelium was considered to result in severe, irreversible effects and not borderline effects. Of the five stainings selected, i.e. H&E, PAS, Trichrome, Azocarmine & Aniline (AZAN) and Elastic Van Gieson (EVG), PAS was clearly superior with respect to visibility of the membranes and the quality of the morphology of the various corneal structures. Moreover, the histopathological examinations provided interesting facts and insights. After severe corrosive damage to the cornea by sodium hydroxide, observed macroscopically through slit-lamp observation as maximum swelling and very severe opacity of the cornea, the basement and the Bowman's membranes appeared undamaged while effects were seen in the underlying stroma. Does this mean that the functionality of the membranes was not compromised or is light microscopy not able to detect such damage of the membranes? This observation led to the conclusion that depth of injury is not the only factor determining the seriousness of corneal injury. Personal communications with an ophthalmologist of the University Medical Centre of Utrecht, specialized in the cornea, and publications (Wagoner, 1997; Terry and Khosla-Gupta, 2002; Rama et al., 2010) indicated that in the clinic, emphasis is put on corneal opacity and corneal stem cell survival after (chemical) injury (Table 7). With severe stem cell damage, recovery of the corneal damage by re-epithelialization of migrating stem cells from the corneal limbus is not possible. In that case, the recovery process

Table 7
Roper-Hall (1965) classification of chemical burns of the human eye.

Grade	Findings	Prognosis
I	Corneal epithelial damage; no limbal ischemia	Good
II	Cornea hazy but iris details seen; ischemia less than $^1/_3$ of limbus	Good
III	Total loss of corneal epithelium; stromal haze blurring iris details; ischemia at $^1/_3$ to ½ of limbus	Guarded
IV	Cornea opaque, obscuring view of iris or pupil; ischemia at more than ½ of limbus	Poor

will result in complete, irreversible conjunctivalization of the ocular surface. The stem cells are rather superficially located in the limbal region of the eye and as such involved in early contact with a topically applied compound. Therefore, the possibility of screening the viability of stem cells after chemical injury could be of value. However, explorations to stain stem cells of the chicken cornea by p63 immunostaining appeared unsuccessful.

Although the assessment of reversibility or irreversibility in the ICE is undoubtedly of value, especially for household and personal care companies, the EU/GHS classification and labelling of severe eye irritants does not make a distinction between these two categories; they are all classified as Category 1: “irreversible effects on the eye/serious damage to the eye”. The use of additional histopathology in the ICE over the past ten years has proven that it is mainly confirmative of the results already obtained by the slit-lamp observation and, in some instances, can be used to support decision making in borderline cases between irritating and severely irritating compounds. In general, the histopathology results provided no reason for altering the irritation classification.

In literature on the cornea, an interesting observation was that vacuolation of epithelial cells can occur in the human cornea after chemical insult, but it does not occur in the rabbit cornea (Grant, 1986). In the ICE, especially with detergent products, vacuolation of epithelial cells was regularly observed by the pathologist. ICE studies commissioned by the International Association for Soaps, Detergents and Maintenance Products (AISE) showed that vacuolation of epithelial cells, and in particular its location in the epithelial layer (top, mid or bottom region), may be reason for an upgrade of the classification to Category 1 in case of borderline severe corneal effects (Cazelle et al., 2014).

2.6. Other alternatives

In the early nineteen-nineties another alternative method using corneas was developed, namely the bovine corneal opacity test (BCOP; Gautheron et al., 1992). They used bovine corneas, not *in situ*, but excised from the eye-ball and clamped inside a chamber. At first sight the method appears quite similar to the Isolated Chicken Eye (ICE) test, i.e. using corneas and measuring opacity and fluorescein penetration, but the differences are remarkable. Corneal opacity is measured quantitatively as the amount of light transmission through the cornea. Permeability is measured quantitatively as the amount of sodium fluorescein dye that passes across the full thickness of the cornea, as detected in the medium in the posterior chamber (OECD, 2013a). An empirically-derived formula is used to calculate an *in vitro* irritancy score (IVIS = mean opacity value + (15 × mean permeability OD490 value)).

Other alternatives using reconstructed (human) corneal tissue, the so-called 3D models, such as the SkinEthic Human Corneal Epithelium test and the EpiOcular™ test were developed in the late 20th early 21st century and were also validated in several studies. A drawback of these corneal models is that only the epithelial layer of the cornea is reconstituted which might pose a problem in discriminating irritating from moderately/severely irritating substances.

At present, only the ICE and the BCOP tests are officially adopted by the OECD for Identifying.

i) Chemicals Inducing Serious Eye Damage (OECD, 2009a,b) and ii) Chemicals Not Requiring Classification for Eye Irritation or Serious Eye Damage (OECD, 2013a). The Fluorescein Leakage test has been adopted by the OECD for Identifying Ocular Corrosives and Severe Irritants (OECD, 2012), but with specific limitations, i.e. only applicable to water soluble chemicals and excluding strong acids and bases, cell fixatives and highly volatile chemicals.

The BCOP, the ICE test and 7 other test systems were considered to be the most promising alternatives to be further validated and in 1992 the British Home Office (HO) and the Directorate General XI of the European Commission (EC) commissioned a validation study on alternatives to the Draize eye irritation test, to be known as the EC/HO validation study. The first priority was given to evaluate the possibility of identification of substances severely irritating to the eye, while also evaluating the methods for predicting the irritants and non-irritants (Balls et al., 1995). The methods selected, their principle, expression of results together with the pros and cons are presented in Table 8.

2.7. International validation of the ICE (EC/HO study)

With the introduction of alternative methods for the Draize Eye Test in the mid-eighties, a certain wild-growth of alternative methods occurred and internationally (mostly) in Western Europe clusters of specific methods could be seen. Roughly, the IRE/ICE was developed and practiced mainly by the UK and the Netherlands, the Bovine and Corneal Opacity and Permeability test (BCOP) by France, the Hen's Egg Test Chorioallantoic Membrane (HET-CAM) by Germany and the Neutral Red Uptake test (NRU) by several European countries. With the increasing interest in these methods and the need for regulatory acceptance it became apparent that a formal validation of the alternative methods was needed.

A first (pilot) validation of several alternative methods, including HET-CAM, REET and NRU was commissioned by the EU in 1987 (the Collaborative study on the evaluation of alternative methods to the eye irritation test, 1991), which was the basis of one of the largest validation programs held in the early nineties, known as the EC/HO study. In this comprehensive study, organized by the European Commission (EC) and the British Home Office (HO), nine methods including the ICE (Table 6) were each carried out by four laboratories testing 60 chemicals which represented different chemical classes and irritation potential (Balls et al., 1995).

The chemicals were selected from the ECETOC database and were considered to have reliable *in vivo* eye irritation data. Basically that meant; the tests having been performed under GLP conditions and in compliance with OECD TG 405 (OECD, 1981). No other assessment with respect to the quality of the individual data was made. The outcome of the study was quite disappointing; none of the methods were capable of identifying the eye irritation potential of the compounds (maximum overall correlation ranged from 0.34 to 0.55). Breaking up the compounds into different categories such as liquids, solids, surfactants, non-irritants, severe irritants etc. did not improve the correlation significantly, although the group of surfactants showed the best results across all methods.

One of the reasons for this disappointing result was considered

Table 8
Alternative *In vitro* tests for eye irritancy considered most promising (based on EC/HO study).

Alternative	Principle	Expression of results	Pro's	Con's
Red blood cell haemolysis test	Leakage of haemoglobin (H) from red blood cells and denaturation (D)	H_{50} and D values equivalent to MMAS[a]	- relatively simple set-up - relatively simple performance	- single index score - no direct relation with ocular tissue - no reversibility - testing of non-soluble substances
Neutral red uptake test	Inhibition of neutral red uptake (NRU) into mouse 3T3 cells	NRU_{50} values equivalent to MMAS	- relatively simple set-up - relatively simple performance	- single index score - no direct relation with ocular tissue - no reversibility - extreme PH, non-soluble substances
Fluorescein leakage test	Fluorescein leakage (FL) by damage to the tight junctions of Madin-Darby canine kidney cells	FL_{20} score equivalent to MMAS	- relatively simple set-up - relatively simple performance	- single index score - no direct relation with ocular tissue - no reversibility - viscous materials, extreme PH, non-soluble substances
EYTEX method	Turbidity of reagent	EYTEX Draize equivalent (EDE) score equivalent to MMAS	- relatively simple set-up - relatively simple performance	- single index score - no direct relation with ocular tissue - no reversibility - testing of solids, coloured samples, surfactants, water-solubles - interference/inhibition with matrix
HET-CAM method	Haemorrhage, lysis and coagulation in the chorioallantoic membrane of embryonated chicken eggs	Reaction time for occurrence of haemorrhages, lysis and coagulation within 5 min combined into a Q score equivalent to MMAS	- relatively simple set-up - relatively simple performance	- single index score - no direct relation with ocular tissue - limited testing of solids and sticky materials - use of live embryo - subjective scoring - no reversibility
Silicon microphysiometer test	Reduction in the metabolic acidification rate of L929 fibroblasts	MRD_5) values equivalent to MMAS	- assesses functional cell changes	- single index score - no direct relation with ocular tissue - very limited testing of substances (37–48%) - laborious - complex and expensive system
Bovine corneal opacity/permeability test	Changes in opacity and in permeability of isolated bovine corneas	In vitro irritancy score (IVIS) equivalent to MMAS	- highly standardized - "human" parameters - ocular tissue - eyes relatively easy attainable - objective scoring	- single index score - no direct observation (black box) - cornea excised - thick cornea compared to rabbit/human - laborious - no reversibility - no conjunctival damage - testing of solids, coloured substances
Isolated chicken eye test Isolated rabbit eye test	Corneal swelling, corneal opacity and fluorescein staining of damaged epithelial cells of the cornea	Degree of severity (categories) for each endpoint separately and combination of the three categories into regulatory classification	- eyes easy attainable - relatively simple set-up - relatively simple performance - ocular tissue - "human" parameters - slit-lamp microscopical assessment - objective scoring corneal swelling - direct translation to human ocular damage - all substances can be assayed neat	- no reversibility - no conjunctival damage - subjective scoring opacity, fluorescein retention - experienced observer
Draize rabbit eye test	Corneal opacity, iritis and conjunctival damage of one eye treated in the conjunctival sac	Degree of severity for each endpoint separately and classification on the basis of the most affected tissue (degree and/or persistency)	- simple set-up - simple performance - rabbits easily attainable - large eye - in vivo response including recovery	- unrealistic exposure area (inside eye-lid) - undefined exposure time (seconds to 24 h) - subjective scoring - experienced observer - animal behaviour influencing eye effects - unrealistic assessment of recovery (no aftercare)

[a] MMAS = Modified maximum average score.

to be the use of the MMAS as the sole parameter derived from the *in vivo* data. This score, ranging from 0 to 110, is an average of the maximum individual tissue scores of the animals recorded for a compound. By using certain cut-off values for the MMAS, a compound is classified as a non-irritant (score 0–25), irritant (score 25–59) or severe irritant (score >59). Assessment of the entire process of ocular inflammation by a single index instead of using the *in vivo* data to its full potential appeared rather inadequate. For instance, the classification system used in the EC does not use a single irritation index score, but is based on individual tissue scores (i.e. for cornea, iris and conjunctivae separately) and/or the (ir) reversibility of these effects within 21 days (Table 3). In 1998, the

OECD published a proposal for the harmonization of hazard classification based on eye irritation/corrosion, which is comparable to the EC classification system, because it also uses the individual tissue scores separately (Table 5). However, slightly lower thresholds were used for classification as an irritant (Category 2) or severe irritant (Category 1), and additionally recovery of eye effects within 7 days were used to discriminate between a mild irritant (2B) and an irritant (2A). Because the MMAS is not used for regulatory classification, the impact of the EC and proposed OECD classification criteria on the irritating potential of the compounds tested in the EC/HO study was investigated. First of all, it was established that applying the two classification systems to the EC/HO compounds resulted in classifications that were comparable between the two systems, i.e. R36 compounds were also Category 2A/2B compounds, and R41 compounds were also Category 1 compounds. Subsequently, it was demonstrated that the MMAS cut-offs of 25 and 59 belonging to, respectively, irritant and severe irritant, were not appropriate for classification according to these two regulatory systems. Eight compounds with an MMAS lower than 59 were in fact severe irritants (R41/Category 1) and 4 compounds with a MMAS higher than 59 and 3 compounds with a MMAS lower than 25 were irritants (R36/Category 2). One of the reasons was that an MMAS could be lower than 59 during the study while ocular effects persisted until day 21 which according to the classification system is reason for R41/Category 1 classification. The recommendation was that validation of alternatives would benefit from the use of classifications based on the proposed OECD harmonized system (later on adopted as the EU-CLP and UN/GHS classification systems). After this publication on the role of the MMAS in 1999, it was not until 2004 that a new initiative was undertaken to (re-)validate four of the alternative methods which were considered the most promising in the EC/HO study, especially for the screening of severe irritants.

2.8. International validation of the ICE (ICCVAM study)

The validation was an initiative of ICCVAM and NICEATM (National Toxicology Program (NTP) Interagency Centre for the Evaluation of Alternative Toxicological Methods) of the USA in collaboration with ECVAM of the EC and the methods selected were the ICE, the HET–CAM, the IRE (formerly called REET) test and the BCOP. An independent Expert Panel was established for each alternative method to determine the validation status of these methods. After a public request for data on the four methods in 2004, a public meeting was held in January 2005 at the National Institutes of Health (NIH), Bethesda, USA to assess the current validation status of the *in vitro* test methods proposed for identifying compounds that may cause serious eye damage and to develop recommendations for further validation. During that meeting, the ICE method was presented. The attitude of the Panel members towards the validation procedure in general was rather disappointing. The focus was on the belief that the poor correlation with the Draize Eye Test was due to shortcomings in the practical performance of the *in vitro* methods which should be improved, rather than critically addressing the validity and shortcomings of the Draize eye test itself and its consequences for the validation of *in vitro* alternatives. Emphasis was again put on the statistical evaluation of the *in vivo* and *in vitro* data (Prinsen, 2005).

Another public request for *in vitro* data was made by ICCVAM in February 2005. ICCVAM received data of in total174 compounds, i.e. previously tested by i) Prinsen and Koëter (1993; 21 compounds).

ii) Prinsen (1996, 2005, dataset of 94 compounds), and iii) Balls et al. (1995; 59 compounds), which were compiled in a background review document (ICCVAM, 2006a). This time the MMAS was no longer used for reanalysis of the accuracy and reliability of the ICE, but regulatory classification criteria (UN-GHS, EC and EPA) were applied as was recommended previously (Prinsen, 1999). ICCVAM and its Ocular Toxicity Working Group summarized the Expert Panel evaluation, the revised analyses, the public comments, and the comments of the Scientific Advisory Committee on Alternative Toxicological Methods in a final Test Method Evaluation Report (ICCVAM, 2006b).

The conclusion was that "*there are sufficient data to substantiate the use of the BCOP or the ICE test methods, with certain limitations, as screening tests to identify compounds as ocular corrosives and severe irritants in a tiered-testing strategy, using a weight-of-evidence approach, for regulatory hazard classification purposes*". The limitations of the ICE were the testing of alcohols based upon the false positive rate and of solids and surfactants based upon the false negative rates. The limitations for the BCOP were the testing of alcohols and ketones based upon the false positive rates and of solids based upon the false negative rate.

Although the acceptance as a screen for ocular corrosives and severe irritants was a success for the ICE, the way the data were used by ICCVAM for reanalyses and for adjustments in the experimental procedures of the ICE was debatable. It was acknowledged that the *in vivo* rabbit eye test was subject to variability, but the *in vivo* data were taken as absolutely accurate (the "Golden Standard") in predicting the eye irritation potential of a compound.

The *in vivo* irritation classifications were assigned if the study performance met the criteria set by ICCVAM. One of the criteria (assessment of full reversibility of any eye effect) was the reason that the database (ICCVAM, 2006a) contained many gaps compared to the original data submitted, whereas sufficient *in vivo* data were available for classification. For example, two compounds (2,2-dimethyl butanoic acid and p-fluoroaniline) identified in the EC/HO validation study as severe irritants (R41) on the basis of the individual *in vivo* data (ECETOC, 1998) were rejected with the remark "study criteria not met". The two compounds were correctly identified as R41 by the ICE and by most other *in vitro* methods participating in the validation study. The *in vivo* classification was based on sound scientific judgment, but ICCVAM decided to exclude the compounds, because a 21-day observation period was not completed. The OECD/EC guidelines (at the time of testing) specified that the observation period should be long enough to evaluate the reversibility or irreversibility of the lesions. The six rabbits treated with 2,2-dimethyl butanoic acid still showed slight to severe corneal opacity and neovascularization of the cornea at 14 days after treatment. It was considered evident that these lesions would not have cleared within a 21-day observation period. Thus, the 14-day observation period applied was in agreement with the guidelines and should not have been a reason for discarding the results by ICCVAM. The same applied for p-fluoroaniline, which caused moderate to severe corneal opacity and iritis score 2 (highest score possible; no reaction to light, haemorrhage, gross destruction). The test was terminated on day 3, which is also in agreement with the guidelines which state that animals may be humanely sacrificed if the severity of the effects is considered too high.

Similar cases occurred in the "parallel" data set of the 94 compounds provided by Prinsen (2005). These cases also mainly concerned *in vivo* eye irritation studies that were terminated earlier than 21 days after treatment because of the severity of eye effects or compounds lacking an *in vivo* eye irritation study because of proven skin corrosivity in the *in vivo* rabbit skin irritation test (10 compounds), that was performed immediately after the ICE had shown severe eye irritancy. In full agreement with the guidelines, the *in vivo* eye irritation test in rabbits was waived in these ten cases. The individual *in vivo* skin irritation data and the ICE data were provided to ICCVAM, but the 10 compounds remained excluded

from the analyses because of "classification assigned on the basis of skin corrosion assay" (SC). Remarkably, ICCVAM claimed to apply the criteria for classification according to the EC (1993) and UN (2007), whereas both guidelines unambiguously state "corrosive to skin" as one of the criteria for classifying a compound as severely eye irritating.

Another remarkable conclusion in the ICCVAM document was the underperformance of surfactants by the ICE. This conclusion was primarily based on the 6 surfactants examined in the EC/HO validation study. The correlation percentages for surfactants in the EC/HO study tested by the four participants that performed the ICE were 72, 82, 83 and 76%, compared to an overall mean correlation percentage of 54%. In general, the chemical class of surfactants was best predicted by each of the nine *in vitro* methods participating in the EC/HO study. The fact the ICE has been employed by P&G for more than ten years for their product development (Schutte et al., 2009), the majority of which contains surfactants covering the whole spectrum of eye irritancy, is in contrast with ICCVAM's conclusion. More recently, member companies of the International Association for Soaps, Detergents and Maintenance Products (AISE) increasingly use the ICE for their (surfactant-containing) products (Cazelle et al., 2014).

With respect to the ICE study performance the ICCVAM expert panel identified two major issues: 1) the variability in swelling percentages obtained by the four laboratories performing the ICE, and 2) the use of only one control eye per experiment.

2.8.1. Variability in swelling percentages (high CVs)

The variation in swelling percentage was caused by the use of different pachymeters with different slit width settings. This was already intensively discussed by the Management Team of the EC/HO study, but was considered of no concern because the *in vivo* MMAS was compared to the critical scores of the ICE (i.e. the max. mean swelling%, max. mean opacity score and mean fluorescein score) and not to the regulatory irritation classifications. ICCVAM overlooked this fact and decided to apply the TNO ICE system for categorizing effects to the other three participating labs as well. This was a valid approach for the opacity and the fluorescein scores because the scoring is exactly the same for the four labs, but it could not be used integrally for the swelling %. The misconception of the variability in corneal swelling by the ICCVAM expert panel led to incorrect conclusions and recommendations. For example, centering lights needed to be installed on the optical pachymeter to improve the determination of corneal thickness by ensuring consistent central corneal thickness measurements across laboratories. The purpose of these lights in human ophthalmology is to guide the patient's eye to a fixed point and thus perform the reading at the centre of the cornea. This is used because the subjects often (involuntarily) move their eyes making the (central) corneal thickness assessment difficult and variable. The chicken eye is isolated and fixed, so there is no movement at all. Therefore, the corneal thickness can be measured in a very accurate and reproducible way at the centre of the cornea without any additional aid.

2.8.2. Use of control eyes

ICCVAM decided to increase the number of negative control eyes from the usual one per experiment to three, because three was the accepted minimum of replicates in *in vitro* testing in general. The use of only one negative control eye has been employed and approved by all users of the isolated eye test (both with rabbit eyes and with chicken eyes) since the introduction of the method by Burton in 1981 and during the EC/HO study. The use of only one negative control eye is justified since the effects of the cornea treated with the test material are not assessed or evaluated in any way against the effects of the control eye. This is possible because, prior to testing, the quality and suitability of each cornea can be accurately assessed and, moreover, each cornea provides its own pre-dose baseline thickness/opacity/fluorescein control values. Furthermore, all compounds (liquids, pastes and solids) are tested neat and, therefore, effects of solvents need not to be examined. The purpose of the negative control eye is only to demonstrate the appropriateness of the general conditions in the chambers of the superfusion apparatus, i.e. the saline drip onto the cornea and chamber temperature, necessary to maintain corneas in the proper condition during the 6-hour test period. All ICE experiments used for the ICCVAM BRD were performed with one negative control, representing 354 independent test runs or replicates. These negative controls never showed any unusual effect during the 6-hour test period and adequately demonstrated the appropriateness of using only one negative control for the purpose of monitoring the general conditions of the test system. Other alternatives like the BCOP need to use three or more control corneas because pre-dose corneal observations, to assess their suitability for testing, are not possible. The BCOP control values at each observation time point are needed for subtraction from the values of the test corneas.

Following a public request for comments on the draft ICCVAM Test Method Evaluation Report: "Evaluation of the Current Validation Status of *In Vitro* Test Methods for Identifying Ocular Corrosives and Severe Irritants" (Federal Register, 2005a) and the "In Vitro Ocular Toxicity Draft Background Review Document (BRD)" on the ICE test method (Federal Register, 2005b), the discussion concentrated on the exclusion of the ICE data of skin corrosive compounds and other data gaps in the BRD. Concern was expressed on the decision to pool the data of the various ICE (validation) studies for analyses without analysing the individual studies separately. *In vivo* and *in vitro* data obtained in "parallel" are of a higher quality than ICE data compared to *in vivo* data obtained from literature, because in the latter case it is obtained by different observers, with different batches of the compound, under different laboratory conditions and often with only summarizing data reported. Furthermore, reservations were made about the handling of the *in vivo* data, i.e. that no lessons were learned from previous validation studies, which made clear that another approach for validation was needed. The disadvantages and shortcomings of the *in vivo* rabbit eye test and their implications on the test results should be addressed first before starting the validation process.

3. The Draize eye test and *in vitro* alternatives; a left-handed marriage?

Recognizing, understanding, and the correct appreciation of the *in vivo* test conditions and their effect on the results are crucial for the evaluation of the *in vitro* results obtained by the alternative test method. The *in vitro* test conditions can usually be fully controlled in contrast to the conditions of testing in the *in vivo* rabbit test. The *in vivo* results are used as the "Golden Standard" for all comparisons with the alternative tests. However, the shortcomings of this test should not be ignored, especially if one is aware of the nature and extend of these shortcomings. Up to now, in compliance of the rabbit test with the current guidelines was the criterion used in validation studies for the validity and acceptance of the *in vivo* data. Once that was established, the *in vivo* results are treated as the absolute truth without any room for interpretation or expert judgment. This means that a compound is classified as a non-irritant, an irritant or a severe irritant based on the *in vivo* test without any kind of nuance or deliberation. There is general agreement that the *in vivo* test in rabbits is far from perfect, but the implications of the inconsistencies of the *in vivo* test for the validation of the alternatives were never taken into account.

In fact, over the years an increasingly rigid attitude towards

questioning the value of the *in vivo* eye irritation data can be noted. This can be considered the root of the problems encountered since the very first validation took place in 1989.

There is a common believe and acknowledgement that the *in vivo* rabbit eye test produces variable results due to the fact that different labs and observers are involved and data had been obtained over a very long period of time. However, this is only a small part of the problem. There are more serious reasons to consider the *in vivo* data in a critical way.

There are several important issues that play a crucial role in the outcome of the *in vivo* test:

3.1. The kind of exposure

By instillation of the compound in the conjunctival cul-de-sac, closing the eye-lids for one second and then releasing the rabbit, the exposure is undefined. It can be anything from minutes (liquids) to 24 h (solids). No washing out of remnants from the conjunctival sac was allowed before 24 h after treatment. Only after modification of the OECD guideline in 2002 was a wash-out allowed after one hour. Especially with poorly soluble/dissolving powders the results can be devastating if the powder is present for one hour, let alone for up to 24 h. It should be noted that the ICCVAM validation was mainly with *in vivo* data obtained before 2002. Remarkably, Draize only mentions the testing of liquids, solutions and ointments and not the testing of solids. In fact, in his 14-page publication (Draize et al., 1944), only one sentence is dedicated to the actual test procedure for eye irritation testing. Overall, eye irritation was dealt with in a rather limited way when compared to his discussion on dermal toxicology and skin sensitization. One wonders what would have happened if Draize had extended his eye irritation investigations to the testing of solids.

These undefined exposure conditions are in contrast to the basic principles of toxicity testing. Moreover, this kind of exposure condition by placing such a large amount of compound in a retracted eye-lid will hardly occur in humans.

A well-defined and standardized exposure in toxicity testing is one of the pillars of hazard and risk assessment. For instance, in the acute skin irritation in rabbits, a semi-occlusive exposure of 0.5 mL–6.25 cm^2 of skin is applied for 4 h under a patch and fixative tape. These are standardized conditions and remarkably the skin irritation test in rabbits has been fully replaced by alternatives since 2010 (OECD, 2004; OECD, 2010a,b).

3.2. Behaviour of the animals

The behaviour of rabbits after treatment may also differ considerably. After treatment the animal is immediately released and is placed in its home cage where it can move freely. Usually, they start grooming and/or scratching. One rabbit out of a group of 3 treated may do this excessively, while on the other end of the behavioural spectrum another animal may freeze and not react at all. Again these variations in behaviour add considerably to the variability of the results.

3.3. Treatment of the eye post-exposure

When significant irritation occurs in an early stage, the treatment of the eye post-exposure highly determines the outcome for classification. The observation times after treatment are essentially the only moments that the animals are handled outside the cage. In case of a moderate eye irritant those time points are normally 1 h, 24 h, 48 h, 72 h, 7 days, 14 days and 21 days. In between, the animals are not handled except for a cage-side observation once a day. The enclosure of solid materials up to 24 h in the conjunctival cul-de-sac, sometimes in combination with mechanical damage, can have a devastating effect on the eye. In the case of poorly water-soluble solids with distinct cytotoxic properties, the entrapped solid can rapidly cause a considerable and increasing swelling of the conjunctivae, making it very difficult for the animal to remove the material. If, at the 1-hour observation, the lower eye-lid is not pulled away far enough by the observer, a bulk of test material deeply hidden in the conjunctival cul-de-sac may remain unnoticed. In most cases this continuous exposure for the next 24 h results in a complete closure of the eye-lids by the abundant production of colloidal discharge which often forms a sealing crust. Upon opening the sealed eyelids, purulent discharge, and other inflammatory debris are released. If the animal (treated eye) is not receiving special care of the eye an otherwise irritating compound can easily become a severe one.

The swelling of the conjunctivae can be such that removal of the remains of the test compound is hardly possible. In the majority of these cases, the eye is permanently damaged or can only be saved by applying special care, such as regular daily cleaning and rinsing of the eye and eye-lids, often including cutting off the eye-lashes to prevent further sealing. This special care is not common practise in the Draize eye test nor is it mentioned in the guidelines, whereas it can certainly relieve the discomfort and pain experienced by the rabbit considerably. In general, keeping the eye-lids open is essential for the recovery process, otherwise the enclosed inflammatory exudate will further damage the cornea. If no further extensive remedial treatment is given to the animal, the exposure conditions described can lead to an opacity score of 3 or 4 instead of the initial score of 1 or 2. In these cases, recovery from these injuries has little or no relevance for man. As with the exposure conditions, these kind of circumstances are not representative for the human situation. After accidental exposure, one will seek “immediate” care in case of ocular damage, and the victim will usually receive medical treatment, if required. The unrealistic exposure conditions in the Draize eye test impelled P&G to develop the *in vivo* rabbit Low Volume Eye Test (LVET) for their products (Griffith, 1987). For instance, the testing of a dish wash detergent tab would result in dramatic ocular effects in the standard *in vivo* Draize eye test, because the tab is ground to a fine powder and instilled as a bulk in the conjunctival cul-de-sac of the rabbit and remains there for at least one hour (before 2002 up to 24 h). Nobody would consider this as real exposure circumstances, nor will it occur in real life. In the ICE, the exaggerated test conditions can be mimicked by leaving the powder on the cornea for up to 60 s instead of the standard 10 s, but what relevance does it have? The LVET was designed to mimic the possible human exposure, by exposure to a tenth of the usual amount directly onto the rabbit's cornea in the Draize eye test, and was extensively used for household care products. Now P&G uses the ICE for their purposes because the test also mimics the possible human exposure more closely. In general, it was astonishing that both ECVAM and ICCVAM urged that the ICE method needed to be modified in order to mimic the extreme exposure conditions of the *in vivo* Draize eye test, rather than modifying the exposure (to solids) in the *in vivo* Draize eye test.

Another phenomenon that occasionally occurred in the Draize eye test is the development of a secondary infection following the eye effects caused by the compound (initial infection). In the past, the hygiene standards in the laboratories were not as high as currently, and the treated eye could be infected by the scratching/grooming of the animal with its paws. In addition to the inflammation caused by a compound, the eye is more vulnerable to microbiological infection, causing initial mild to moderate effects during the first days after exposure developing into more severe and prolonged effects during the 21 days observation period. An

interesting example of such an event can be found in one of the compounds tested in the EC/HO study, which was also used in the ICCVAM validation of the ICE.

1-Napthaleneacetic acid was tested in six rabbits of which one rabbit showed very unusual persisting and increasing effects after day 7, compared to the eye effects observed in the other five rabbits (ECETOC, 1998). The effects of the five rabbits followed a pattern which is normally expected for the initially slight to moderate eye effects, i.e. gradually decreasing in severity after day 3 followed by a complete recovery on day 14 or day 21. In the sixth rabbit, a similar pattern was observed until day 7, but thereafter the slight opacity observed increased to a moderate opacity on day 10 and finally a very severe opacity on days 14 and 21. This difference in the pattern of the eye effects is remarkable and most probably caused by a secondary infection in the eye of the animal. Based on the result in this rabbit, the compound was classified as Category 1, whereas the initial effects (24–72 h) in the 6 rabbits would lead to a Category 2(A) classification. The ICE test also classified the compound as Category 2.

It is remarkable that the OECD guideline 405 of 1981 already stated that "Care should be taken in the interpretation of data to exclude irritation resulting from secondary infection". However, this issue was neither addressed in the EC/HO validation nor in the ICCVAM validation.

3.4. Observation/grading of eye effects

In the early days of validation of alternatives for eye irritation it was recognized that the variability could be high in the Draize eye test, and this was considered to be caused by subjective scoring by different observers and by interlaboratory variability (Weil and Scala, 1971; Lordo et al., 1999; Ohno et al., 1999). Unfortunately, the publications by Lordo and Ohno did not include the individual *in vivo* rabbit eye data which might have provided more insight in the underlying causes of the variation. Subjective scoring is indeed part of the problem but a large part of the variation presently ascribed to subjective scoring might in fact be caused by differences in animal behaviour, differences in exposure times, and absence (or presence) of post-treatment care. For instance in the study of Weil and Scala, in one laboratory ethanol 95% caused a combined score (all tissues combined) of 2, 9, 22, 15, 38 and 110 in the 6 rabbits at the 72-hour observation time point. A score of 110 is the maximum score possible. Amongst the 24 laboratories the median score for ethanol 95% ranged from 0 to 42. This cannot be explained by subjective scoring only. The subjective nature of the observation definitely plays a role with compounds causing effects near the thresholds for classification (not classified/irritant and irritant/severe irritant).

First of all there is the grading/scoring of the effects itself. The Atlas of eye effects of the FDA (1964) already gave rise to debate. For example the redness of the conjunctivae of the eye no. 6 of Plate 1 (Fig. 1) is stated to be score 2 (moderate redness: more diffuse, deeper crimson red, individual vessels not easily discernible; see also Table 2 of the Introduction). Based on the grading of eye effects of all compounds tested at TNO since 1981, it should be the maximum score of 3 (severe redness: diffuse beefy red), because a more intense redness cannot be observed. Eye no. 3 of Plate 1 is presented as a case of redness score 1 (slight redness: vessels definitely injected above normal), whereas this would be a good example of score 2 for redness. Eye no. 2 of Plate 1 is more representative of a redness score 1 than of a normal eye (redness: vessels normal). The other plates of the Atlas contain more examples of grading that are considered questionable and subject to debate.

Another issue concerning the subjective scoring is the decision the observer has to make in certain cases where the score of one animal at one time point can make the difference between, for instance, not classified and irritant (Category 2). For EU-CLP and UN-GHS classification, the individual tissue scores of each animal is first averaged over the 24–72 h time points and next the average score of the two rabbits showing the highest score for a specific tissue determines classification or not. The threshold score for redness or for swelling of the conjunctivae is an overall average of 2.0 for classifying as a Category 2 compound. Table 9 shows a theoretical case where one of the 6 scores can make the difference between classifying or not. The score of 1.0 in Table 9 is assigned to animal 2 at the 72 h time point, but can theoretically be at any of the 6 different places.

3.5. Appreciation of the in vivo data

With the knowledge of the factors influencing the *in vivo* results the "black or white" approach applied by ICCVAM can hardly be defended. Weil and Scala (1971) even concluded that the eye irritation test in rabbits as published by the Federal agencies of the US should not be recommended as standard procedure in any new regulations. However, the Draize eye test has been used practically unchanged until now.

Balls and Fentem (1993), concluded: *"It is very rare for any allowance to be made for the variability of the animal data, which are thus given a status which they do not deserve. They wrongly become the "true" values which the non-animal tests must struggle to reproduce. Also, insufficient allowance is made for the doubt which must be placed on values which fall within the barrier zones on both sides of category cut-off points. This is particular worrying when Cooper two-by-two way plots are used as a basis for establishing the sensitivity, specificity, predictivity and concordance of in vitro test data"*. Bruner et al. (1996), concluded after computer simulations that even if the alternative methods were perfectly reproducibly (if their coefficients of variation were 0), the variability in the Draize scores alone would restrict the Pearson's correlation coefficients to the range 0.65–0.80 when the Draize scores are between 0 and 40, which are typical for (mild) irritants.

4. ICE OECD test guideline 438

One of the conclusions in the ICCVAM test method evaluation report (ICCVAM, 2006b) was: "*There are sufficient data to support the use of the ICE test method, in appropriate circumstances and with certain limitations, as a screening test to identify compounds as ocular corrosives and severe irritants (i.e., EPA Category I, UN GHS Category 1, EU R41) in a tiered-testing strategy, as part of a weight-of-evidence approach. The identified limitations for this method are based on the false negative and false positive rates that are observed for certain chemical and physical classes. Based on the available database, the false negative rates for alcohols, surfactants and solids range from 33% (1/3) to 50% (1/2), 44% (4/9) to 57% (4/7), and 46% (6/13) to 70% (7/10), depending on the hazard classification system (EC, UN-GHS or EPA) used. Additionally, the false positive rates for alcohols range from 27% (3/11) to 50% (5/10)*". Two of the alcohols in the data base were ethanol and butanol. Both caused severe irritancy in the ICE (and BCOP) but were Category 2 according to the *in vivo* data of ECETOC (1998). As discussed previously the *in vivo* data of Weil and Scala for ethanol (and also for butanol) showed very high inter- and intra-laboratory variations, making the ICCVAM conclusion on performance of the ICE with respect to alcohols questionable. Also the surfactant examined in the study of Weil and Scala showed high inter- and intralaboratory variations.

In 2009, on the basis of the ICCVAM evaluation report the ICE and BCOP were adopted as an OECD Test Guideline (TG 438 and TG 437, respectively) for the screening of severe eye irritants. The false

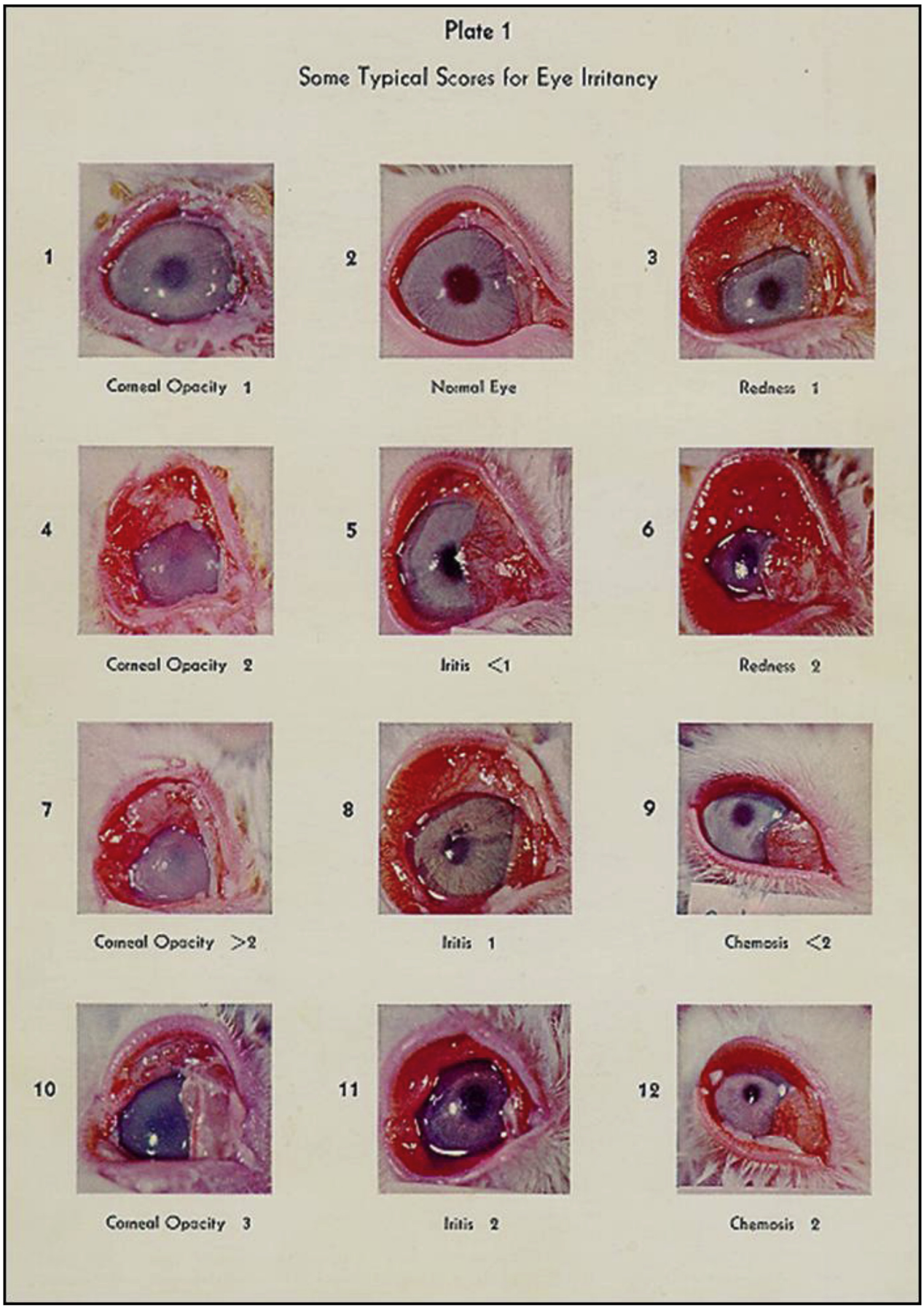

Fig. 1. FDA guidance on scoring of ocular lesions; Plate 1 (FDA, 1964).

Table 9
Examples of conjunctival scores (redness or swelling); Category 2 (first table) versus not classified (second table).

Time point 24 Hours	
Animal 1	2.0
Animal 2	2.0
Time point 48 Hours	
Animal 1	2.0
Animal 2	2.0
Time point 72 Hours	
Animal 1	2.0
Animal 2	2.0
Time point 4 Hours	
Animal 1	2.0
Animal 2	2.0
Time point 48 Hours	
Animal 1	2.0
Animal 2	2.0
Time point 72 Hours	
Animal 1	2.0
Animal 2	1.0

negative rates for identifying severe irritants, i.e. compounds identified by the ICE as not severely irritating, were not considered critical since these compounds are still to be tested in the *in vivo* rabbit eye test. In the OECD test guideline, the limitations with respect to the screening of surfactants, alcohols and solids were also mentioned. Specifically it was mentioned that: "*The current validation database did not allow for an adequate evaluation of some chemical or product classes (*e.g. *formulations). However, investigators could consider using this test method for testing all types of compounds (including formulations), whereby a positive result could be accepted as indicative of an ocular corrosive or severe irritant response. However, positive results obtained with alcohols should be interpreted cautiously due to risk of over-prediction*". The specific mentioning of formulations is remarkable because the ICE is used more frequently for formulations than for pure compounds. Moreover, the Draize eye test also does not make any distinction between the testing of pure compounds and formulations. In general, one should realize that at that time the eye irritation potential of almost all, if not all, compounds had been determined for regulatory purposes in one single type of eye (test), i.e. the rabbit eye (test).

A follow-up evaluation of the usefulness and limitations of alternatives for identifying mild/moderate and non-irritant chemicals was made by NICEATM/ICCVAM, in collaboration with ECVAM and JaCVAM. In early 2011, a proposal for updating the BCOP TG 437 for the identification of chemicals not requiring a classification for eye irritation was submitted to the OECD by means of a Standard Project Submission Form (SPSF). The BCOP database comprised 196 compounds of which 89 were non-irritants, and these data were used to draft a Streamlined Summary Document (SSD). The BCOP was considered appropriate because the percentage of false negatives for non-irritants was 0%. However, the percentage of false positives was 69%. The ICE test was not proposed for such an update, because the review panel maintained the original recommendation to use the ICE only for classification of ocular corrosives and severe irritants. Specific objections against the use of the ICE for chemicals not requiring a classification was the fact that two compounds of the "parallel" dataset (coded TNO-28 and TNO-94) identified as non-irritants by the ICE turned out to be severe irritants in the *in vivo* rabbit eye test. It was argued that the review panel had not studied the nature of the effects of these two false negative substances in detail. TNO-94 was an anti-fouling paint for the shipping industry, a specific type of product, which produced reversible irritating eye effects in two out of three rabbits. In the third rabbit an unusual effect occurred, i.e. adherence of the paint to the cornea which was reason to humanely sacrifice the animal on day 1. Anti-fouling paints are designed to be very durable which may explain the findings in this rabbit. Whether or not this peculiar effect is relevant for humans, excluding (anti-fouling) paints from ICE testing would have no major consequences for the applicability of the method for screening of non-irritants in general. TNO-28 caused no corneal or iris effects; only conjunctival effects were observed. The conjunctival effects observed with this compound were below the threshold for classification as an eye irritant. Eye effects had cleared completely in one rabbit after 72 h and in another rabbit after 7 days. The third rabbit showed increased conjunctival effects at 48 h after treatment and on day 14 moderate redness and slight swelling of the conjunctivae were still observed. Importantly, a white ocular discharge was also observed which was a sign of secondary infection. The fact that the same effects were observed at 21 days after treatment supported this assumption. One week later, the eye effects in this rabbit had cleared completely.

Overall, the ICE "parallel" data set provided by TNO showed a false negative rate for non-irritants of 6% and a false positive rate of 1%. Therefore, the OECD was asked to reconsider the applicability of the ICE for the purpose of identifying non-irritants. During an OECD expert meeting (6–7 December 2012, Paris), the limitations of the *in vivo* Draize rabbit eye irritation test and their implications for validation purposes were recognized and summarized in the document (OECD, 2013b) as follows:

1. *"The in vivo rabbit eye irritation/corrosion test has no standardized exposure regimen. Therefore, the duration of exposure of the test substance with the rabbit eyes remains unknown and can vary from a few minutes to several hours. In addition, for solids and sticky chemicals it is unclear how much of the compound (solid, paste or liquid) stays in contact with the eye;*
2. *The limited reproducibility of the Draize rabbit eye test method;*
3. *The subjectivity in the allocation of the rabbit ocular tissue scores;*
4. *The type of exposure which does not reflect a potential human accidental exposure;*
5. *The differences in physiology and sensitivity to tested chemicals between rabbit and human eyes".*

The re-evaluation of the ICE ICCVAM dataset showed that individual *in vitro* and *in vivo* classifications of a number of compounds need further considerations. Discrepancies were found in the final *in vivo* and *in vitro* classifications for a number of compounds that had an impact on the final number of false negative compounds. After re-evaluation, the ICE test method had an overall accuracy of 82%, a false positive rate of 33%, and a false negative rate of 1% (instead of 6%) for non-irritants, when compared to the *in vivo* data classified according to the UN-GHS. If anti-fouling organic solvent containing paints were excluded from the database, the accuracy of the ICE test method was 83%, the false positive rate 33% and the false negative rate 0%.

In September 2013, the OECD TG 438 for the ICE was officially adopted including the identification of non-irritants (in general about 80% of the chemicals tested are non-irritants). This was a huge success for the ICE and for the 3Rs in general, but it still meant that compounds not identified by the ICE as non-irritant or severe irritant have to be tested in the *in vivo* rabbit eye test.

5. Lessons learned and considerations

- For eye irritation with a lack of human data, the combined *in vitro/in vivo* (parallel) testing instead of using *in vivo* rabbit eye irritation data from literature only provided an more relevant setting for i) developing and validating the alternative method, ii) introducing the method to industry and regulatory authorities, iii) getting insight in, and critically address the pros and cons of both the *in vitro* and *in vivo* test system.
- A meaningful validation of an *in vitro* alternative model in the middle range of irritancy (Category 2 classification) cannot be reached with the current *in vivo* rabbit eye irritation data set due to the large variability.
- The procedure to select or to accept models suitable as an *in vitro* alternative to eye irritation should be more critical than in the past. Alternatives should have a direct relation to (human) ocular irritancy and be developed on the basis of the mechanistic principles of (human) ocular inflammation, instead of matching Draize eye test results only.
- It should be realized that the existing *in vivo* rabbit eye irritation data does not reflect the inflammatory and recovery processes in humans. Therefore, the data of the Draize eye test are not suitable for the development of *in vitro* models for eye irritation focusing on discrimination between severe, but reversible or severe, but irreversible eye effects.
- The *in vivo* eye irritation test in rabbits should no longer be allowed, and should be completely replaced by alternative methods, for instance the ICE.
- The household and personal care industry that share their strategy to fulfil regulatory demands without the use of the *in vivo* animal test with other (chemical) industries and regulatory authorities, should be rewarded, e.g. by providing a 'safe harbour' to evaluate data derived from alternative methods (Schiffelers et al., 2014).

6. Conclusions

Alternatives to the Draize eye irritation test should preferably make use of *ex vivo* eyes, eye tissue, or eye tissue equivalents in order to measure, both qualitatively and quantitatively parameters similar or identical to those in the clinic. The appropriateness of the presently available alternatives, based on the above mentioned criteria and in order of suitability, is as follows:

1. Models using intact isolated eyes with slit-lamp microscope observations and histopathology, e.g. the ICE or IRE.
2. Models using excised corneas with light transmission and fluorescein measurements and histopathology, e.g. the BCOP (Bovine) or PCOP (Porcine).
3. 2D or 3D human corneal epithelium reconstruction models, which have as major disadvantage the lack of the different membranes of the cornea, corneal stroma and corneal endothelium.

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Transparency document

Transparency document related to this article can be found online at http://dx.doi.org/10.1016/j.yrtph.2017.01.009.

References

Balls, M., Fentem, J.H., 1993. On Recognizing and Overcoming Barriers to the Acceptance of Alternative Methods. Alternative Methods in Toxicology and the Life Sciences 11, Part 11 a. Mary Ann Liebert, Inc., publishers, New York.

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. Toxicol. Vitro 9, 871–929.

Brantom, P.G., Cassidy, S., Eigler, D., Esdaile, D., Fentem, J.H., Liebsch, M., McPherson, J., Pfannenbecker, U., Prinsen, M.K., 2000. TNO BIBRA Report No. 2/3387/00. Preliminary Evaluation of the Application of Reference Standards in the Prevalidation and Validation of *in Vitro* Tests for Eye Irritation, vol. 2. December 2000.

Bruner, L.H., Carr, G.J., Chamberlain, M., Curren, R.D., 1996. Validation of alternative methods for toxicity testing. Toxicol. Vitro 10 (4), 479–501.

Burton, A.B.G., 1972. A method for the objective assessment of eye irritation. Fd. Cosm. Toxicol. 10, 209–217.

Burton, A.B.G., York, M., Lawrence, R.S., 1981. The in vitro assessment of severe eye irritants. Fd. Cosm. Toxicol. 19, 471–480.

Cazelle, E., Eskes, C., Hermann, M., Jones, P., McNamee, P., Prinsen, M., Taylor, H., Wijnands, M.V.W., 2014. Suitability of histopathology as an additional endpoint to the isolated chicken eye test for classification of non-extreme pH detergent and cleaning products. Toxicol. Vitro 28, 657–666.

Chan, T., Payor, S., Holden, B.A., 1983. Corneal thickness profiles in rabbits using an ultrasonic pachometer. Investigative Ophthalmol. Vis. Sci. 24 (10), 1408–1410.

Doughty, M.J., Zaman, M.L., 2000. Human corneal thickness and its impact on intraocular pressure measures: a review and meta-analysis approach. Surv. Ophthal 44 (5), 367–408.

Draize, J.H., Woodard, G., Calvery, H.O., 1944. Methods for study of irritation and toxicity of substances applied topically to the skin and mucous membranes. J. Pharmacol. Exp. Ther. 82, 377–390.

EC, 1984. Directive 67/548 (6th adaptation); Annex V Part B: methods for the determination of toxicity B.5. Acute toxicity, eye irritation. Official J. Eur. Community 27 (L251), 109.

EC, 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 and part II. European Commission, Brussels.

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 J. Eur. Communities (L110A), 1–86.

ECETOC, 1998. Eye Irritation – Reference Chemicals Data Bank. Technical Report No. 48(2). European Centre for Ecotoxicology and Toxicology of Chemicals, Brussels.

Faber, C., Scherfig, E., Prause, J.U., Sørensen, K.E., 2008. Corneal thickness in pigs measured by ultrasound pachymetry in vivo. Scand. J. Lab. Anim. Sci. 35, 39–43.

FDA, 1964. Illustrated Guide for Grading Eye Irritation by Hazardous Substances. Washington, D.C. 20204, USA.

Volume 70, No. 53, Monday, March 21 Federal Register notice 13513, 2005a. National Toxicology Program; National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM); Availability of Expert Panel Report on the Evaluation of the Current Validation Status of in Vitro Test Methods for Identifying Ocular Corrosives and Severe Irritants.

Volume 70, No. 142, Tuesday, July 26 Federal Register notice 43149, 2005b. National Toxicology Program (NTP); NTP Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM); Availability of Revised Analyses and Proposed Reference Substances for in Vitro Test Methods for Identifying Ocular Corrosives and Severe Irritants.

Fowler, W.C., Chang, D.H., Roberts, B.C., Zarovnaya, E.L., Proia, A.D., 2004. A new paradigm for corneal wound healing research: the white leghorn chicken (Gallus gallus domesticus). Curr. Eye Res. 28 (No. 4), 241–250.

Gautheron, P., Dukic, M., Alix, D., Sina, J.F., 1992. Bovine corneal opacity and permeability test: an in vitro assay of ocular irritancy. Fund. Appl. Toxicol 18, 442–449.

Grant, W.M., 1986. Corneal Epithelial Vacuoles. Toxicology of the Eye, third ed., vol. 1. Charles C. Thomas Publishers, Springfield, Illinois, USA, p. 9.

Griffith, J.F., 1987. The low volume eye irritation test. Soap Cosm. Chem. Spec. 63 (4), 32–63.

Griffith, J.F., Nixon, G.A., Bruce, R.D., Reer, P.J., Bannan, E.A., 1980. Dose-response studies with chemical irritants in the albino rabbit eye as a basis for selecting optimum testing conditions for predicting hazard to the human eye. Toxicol. Appl. Pharmacol. 55, 501–513.

ICCVAM, 2006a. Current Status of in Vitro Test Methods for Identifying Ocular Corrosives and Severe Irritants: Isolated Chicken Eye Test Method. NIH. Publication No: 06–4513.

ICCVAM, 2006b. In Vitro Ocular Toxicity Test Methods for Identifying Severe Irritants and Corrosives. NIH. Publication No: 07–4517.

Jacobs, G., Martens, M., 1988. The enucleated eye test: a comparison of the use of ultrasonicand optic pachometers. Toxicol. Vitro 2 (4), 253–256.

Jacobs, G., Martens, M., 1990. Quantification of eye irritation based upon in vitro changes of corneal thickness. ATLA 17, 255–262.

Koëter, H.B.W.M., Prinsen, M.K., 1985. Comparison of in Vivo and in Vitro Eye Irritation Test Systems: a Study with 34 Substances. Alternative Methods in Toxicology 3, Chapter A9. Mary Ann Liebert, Inc., publishers, New York.
Li, H.F., Petroll, W.M., Møller-Pedersen, T., Maurer, J.K., Cavanagh, H.D., Jester, J.V., 1997. Epithelial and corneal thickness measurements by in vivo confocal microscopy through focusing (CMTF). Curr. Eye Res. 16 (3), 214–221.
Lordo, R.A., Feder, P.I., Gettings, S.D., 1999. Comparing and evaluating alternative (in vitro) tests on their ability to predict the Draize maximum average score. Toxicol. Vitro 13 (1), 45–72.
Luepke, N., 1985. Hen's egg chorioallantoic membrane test for irritation potential. Fd. Chem. Toxicol. 23, 287–291.
Maurer, J.K., Parker, R.D., Jester, J.V., 2002. Extent of corneal injury as the mechanistic basis for ocular irritation: key findings and recommendations for the development of alternative assays. Regul. Toxic. Pharmacol. 36, 106–117.
McDonald, T.O., Baldwin, H.A., Beasley, C.H., 1973. Slit-lamp examination of experimental animal eyes. J. Soc. Cosmet. Chem. 24, 163–180.
Mishima, S., Kudo, T., 1967. In vitro incubation of rabbit cornea. Investig. Ophthalmol. 6 (4), 329–339.
OECD, 1981. OECD Guideline for Testing of Chemicals No. 405: Acute Eye Irritation/corrosion, Adopted 12 May 1981. Organisation for Economic Co-operation and Development, Paris.
OECD, 2001. OECD Guidelines for the Testing of Chemicals No. 423: Acute Oral Toxicity (Acute Toxic Class method, Paris, France).
OECD, 2002. OECD Guideline for Testing of Chemicals No. 405: Acute Eye Irritation/corrosion, Adopted 24 April 2002. Organisation for Economic Co-operation and Development, Paris.
OECD, 2004. OECD guideline for testing of chemicals No. 431. In: In Vitro Skin Corrosion: Human Skin Model Test, Adopted 13 April 2004. Organisation for Economic Co-operation and Development, Paris.
OECD, 2009a. OECD Guideline for Testing of Chemicals No. 437: Bovine Corneal Opacity and Permeability Test Method for Identifying Ocular Corrosives and Severe Irritants, Adopted 7 September 2009. Organisation for Economic Co-operation and Development, Paris.
OECD, 2009b. OECD Guideline for Testing of Chemicals No. 438: Isolated Chicken Eye Test Method for Identifying Ocular Corrosives and Severe Irritants, Adopted 7 September 2009. Organisation for Economic Co-operation and Development, Paris.
OECD, 2010a. OECD Guidelines for the Testing of Chemicals No. 429: Skin Sensitisation; Local Lymph Node Assay, Paris, France.
OECD, 2010b. OECD Guideline for Testing of Chemicals No. 439: in Vitro Skin Irritation - Reconstructed Human Epidermis Test Method, Adopted 22 July 2010. Organisation for Economic Co-operation and Development, Paris.
OECD, 2012. OECD Guideline for Testing of Chemicals No. 460: Fluorescein Leakage Test Method for Identifying Ocular Corrosives and Severe Irritants, Adopted 2 October 2012. Organisation for Economic Co-operation and Development, Paris.
OECD, 2013a. OECD Guideline for Testing of Chemicals No. 438: Isolated Chicken Eye Test Method for Identifying I) Chemicals Inducing Serious Eye Damage and Ii) Chemicals Not Requiring Classification for Eye Irritation or Serious Eye Damage, Adopted 26 July 2013. Organisation for Economic Co-operation and Development, Paris.
OECD, 2013b. Draft Streamlined Summary Document (SSD) for the Isolated Chicken Eye (ICE) Test Method. ENV/JM/TG/RD(2013)19.
Ohno, Y., Kaneko, T., Inoue, T., Morikawa, Y., Yoshida, T., Fuji, A., Masuda, M., Ohno, T., Hayashi, M., Momma, J., Uchiyama, T., Chiba, K., Ikeda, N., Imanashi, Y., Itakagaki, H., 1999. Interlaboratory validation of the in vitro eye irritation tests for cosmetic ingredients. (1) Overview of the validation study and draize scores for the evaluation of the tests. Toxicol. Vitro 13 (1), 73–98.
Price, J.B., Andrews, I.J., 1985. The in vitro assessment of eye irritation using isolated eyes. Fd. Cosm. Toxicol. 23 (2), 313–315.
Prinsen, M.K., 1996. The chicken enucleated eye test (CEET): a practical (pre)screen for the assessment of eye irritation/corrosion potential of test materials. Fd. Chem. Toxicol. 34 (3), 291–296.
Prinsen, M.K., 1999. An evaluation of the OECD proposal for the harmonised classification of eye irritants and corrosives. Report of ECVAM workshop 34, eye irritation testing: the way forward, appendix 1. ATLA 27, 72–77.
Prinsen, M.K., 2005. In Vitro and in Vivo Data for 94 Substances Tested in the Isolated Chicken Eye Test. Unpublished Data provided Directly to NICEATM by M.K. Prinsen. TNO Nutrition and Food Research Institute.
Prinsen, M.K., Koëter, H.B.W.M., 1993. Justification of the Enucleated Eye Test with eyes of slaughterhouse animals as an alternative to the Draize eye irritation test with rabbits. Fd. Chem. Toxicol. 31 (1), 69–76.
Rama, P., Matuska, S., Paganoni, G., Spinelli, A., De Luca, M., Pellegrini, G., 2010. Limbal stem-cell therapy and long-term corneal regeneration. New Eng. J. Med. 363, 147–155.
Roper-Hall, M.J., 1965. Thermal and chemical burns of the eye. TransOphthalmol. Soc. U.K. 85, 631–646.
Russell, W.M.S., Burch, R.L., 1959. The Principles of Humane Experimental Technique. South Mimms, Potters Bar, Herts: Universities Federation for Animal Welfare.
Scala, R.A., Springer, J., 1997. IRAG working group 6; guidelines for the evaluation of eye irritation alternative tests: criteria for data submission. Fd. Chem. Toxicol. 35, 13–22.
Schiffelers, M.J., Blaauboer, B.J., Bakker, W.E., Hendriksen, C.F., Koëter, H.B., Krul, C., 2014. Regulatory acceptance and use of 3R models for pharmaceuticals and chemicals: expert opinions on the state of affairs and the way forward. Regul. Toxicol. Pharmacol. 69, 41–48.
Schutte, K., Prinsen, M.K., McNamee, P.M., Roggeband, R., 2009. The isolated chicken eye test as a suitable in vitro method for determining the eye irritation potential of household cleaning products. Regul. Toxicol. Pharmacol. 54 (3), 272–281.
Terry, K.T., Khosla-Gupta, B.A., 2002. Chemical and Thermal Injuries to the Ocular Surface. Ocular Surface Disease Medical and Surgical Management, Part III. Springer-Verlag New York, Inc, pp. 100–112.
Thoft, R.A., Friend, J., 1977. Biochemical transformation of regenerating ocular surface epithelium. Investigative Ophthalmol. Vis. Sci. 16, 14–20.
United Nations-Economic Commission for Europe (UN/ECE), 2007. Globally Harmonised System of Classification and Labelling of Chemicals (GHS). UN, New York and Geneva. Available at: http://www.unece.org/trans/danger/publi/ghs/ghs_rev02/02files_e.html.
US-Federal Register, 1961. Title 21 Food and Drugs, Part 191 Hazardous Substances, § 191.12 Test for Eye Irritants.
Wagoner, M.D., 1997. Chemical injuries of the eye: current concepts pathophysiology and therapy. Surv. Ophthal 41 (4), 367–408.
Weil, C.S., Scala, R.A., 1971. Study of intra- and interlaboratory variability in the results of rabbit eye and skin irritation tests. Toxicol. Appl. Pharmacol. 19, 276–360.
York, M., Lawrence, R.S., Gibson, G.B., 1982. An in vitro test for the assessment of eye irritancy in consumer products – preliminary findings. Int. J. Cosmet. Sci. 4, 223–234.