ESAC Shadow Review

Organotypic *in vitro* assays as screening tests to identify potential ocular corrosives and severe irritants as determined by US EPA, EU (R41) and UN GHS classifications in a tiered testing strategy, as part of a weight of evidence approach

Shadow Group members

Odile de Silva, L'Oréal, France (co-chair)
Jan van der Valk, NCA, Utrecht University, The Netherlands (co-chair)
Manfred Liebsch, ZEBET, BfR, Germany
Jon Richmond, ASPD, Home Office, UK
Dariusz Sladowski, Medical University of Warsaw, Poland

ECVAM support: Chantra Eskes

List of contents

I. Background

The US Environmental Protection Agency (EPA) formally nominated four *in vitro* ocular irritation tests methods proposed as screening tests for identifying potential ocular corrosives and severe irritants as determined by US EPA, EU (R41) and UN GHS classifications in a tiered testing strategy, as part of a weight of evidence approach. These four methods are the bovine corneal opacity & permeability (BCOP) assay, the isolated chicken eye (ICE) assay, the isolated rabbit eye (IRE) assay and the hen's egg test – chorio-allantoic membrane (HET-CAM) assay.

ICCVAM-NICEATM conducted a retrospective evaluation between the autumn of 2003 and November 2006. The final ICCVAM Test Method Evaluation Report was made available in November 2006.

ESAC established a Shadow Review Panel in order to evaluate the final conclusions and recommendations of the ICCVAM report, using a specific guidance document (Annex I).

II. General considerations

The ESAC Shadow Review Panel (ESRP) has focused its work on the strict evaluation of the report, based on the relevant BRDs and Expert Panel reviews, according to the pre-determined review criteria.

On the EU and US regulatory contexts

It needs to be stressed that the ICCVAM-NICEATM evaluation has covered only one part of the overall evaluation of the predictive capacity of the four organotypic assays which can be expected to take place in the context of the current EU regulatory requirements. In particular ICCVAM-NICEATM have focused on the capacity of these alternative test methods to predict corrosives and severely irritants, but not to evaluate and confirm lesser degrees of ocular damage.

In the context of the 7th Amendment to the EU Cosmetics Directive where the Draize eye irritation test will be banned for the safety evaluation of ingredients in March 2009, the expectation is that assays or combination of non-animal test methods will be identified and demonstrated as scientifically validated and accepted by the regulators, for the purpose of fully replacing the *in vivo* test in order to enable the safety assessment of ingredients with the same level of workforce and consumer safety as is achieved with safety evaluations based on the current *in vivo* test. Therefore, to achieve this goal it will still be necessary to further assess these four organotypic tests alongside other cellular assays for their predictive capacity to cover the whole range of the irritation response, as well as in a dose response manner, alone or in combination with other assays.

On substances versus formulations

The ICCVAM-NICEATM evaluation has covered both single substances and formulations. It is felt that the use of the tests in a tiered testing strategy is quite different for the safety evaluation of a formulation, where the *in vivo* irritation potential of the individual components is already known, and the hazard evaluation may be based on the known properties of the individual ingredients. Therefore the distinction between these different classes of test materials could be usefully made distinct.

In the context of tier testing strategies more attention should be given to the formulations composition before any further “confirmatory” *in vivo* tests are conducted following the results of the *in vitro* tests. The results of the *in vitro* test, taken together with knowledge of the individual ingredients’ *in vivo* irritation potential, could permit a sound safety evaluation without the need for a confirmatory *in vivo* test.

On the overall objective of the ICCVAM-NICEATM evaluation

Taking into account the objective of the analysis, the prevalence of severe and corrosives *in vivo* does not seem to have influenced the design of the statistical analysis.

On the criteria to determine how predictive is an assay to identify corrosives and severe irritants

It is not clearly specified which criteria have been used to establish whether the performance characteristic make an assay a satisfactory screen to identify corrosives and severe irritants. It is important to define performance criteria such as acceptable accuracy thresholds (i.e. > 70%), both to evaluate potential improvements achieved by the introduction of further modifications in the assays, and to set pass/fail criteria for other potential test systems.

On the need for different protocols and different prediction models for one assay

Although the process followed by the ICCVAM-NICEATM evaluation is very sound and clear, it is quite obvious that the option to retain only very similar protocols and only the studies where detailed *in vivo* data were available has led to a number of other studies and results which were not compliant being excluded from the evaluation of the test methods

Leaving aside the availability of detailed *in vivo* data, it can be seen that in several cases adaptations to protocols need to be made to fully optimize the test methods, as well as relevant prediction models developed, in order to accommodate the diversity of chemical classes, their physico-chemical properties and the range of irritation responses. Definitive protocols and prediction models should be clearly defined and further evaluated. It is therefore quite possible that several protocols and several prediction models are needed for a given assay.

In the case of retrospective studies, the evaluation of methods should not be hindered by the design of the studies. Rather the capacity to extract in the best way the information included in the weight of evidence should be further developed.

On the quality of in vivo data

With respect to exposure and higher sensitivity of rabbit eye compared to humans: we must accept that huge differences exist between exposure in rabbit and possible exposure in human eyes.

It should be acknowledged that the results of the *in vitro* studies are validated against those of a non-validated *in vivo* test, of which it is unknown what the false positive and negative results are compared to the human eye.

On the guidance on how to choose among several recommended tests

It is felt that a thorough guidance document needs to be prepared for each of these tests, including relevant protocols and prediction models for specific classes of substances, as well as details on which assay is recommended to be used for which purpose , based on performance, applicability, costs, skills required, feasibility (access to biological material).

This task could ideally be coordinated by ECVAM, with the help of method developers and users.

III. Conclusions

The 2002 EU "Declaration on the acceptance of results from some alternative methods for eye irritation" reflected the current know-how. The test methods had not been successfully validated in ring-trials, but were being used in house, and as part of a tiered and hierarchical approach offered the potential of reducing animal numbers at a cost of some materials being classified as R41.

We think the ICCVAM evaluation has taken a weight of evidence approach and has been competently performed. The ICCVAM recommendations should reduce animal use (in the USA) at the expense of some materials inappropriately classified as R41.

This is our opinion on all four assays:

- With respect to the BCOP assay the ICCVAM recommendations should be endorsed by ESAC.
- With respect to the ICE assay the ICCVAM recommendations should be endorsed by ESAC.
- With respect to the IRE assay and considering the overall study, it is recommended that the Expert Panel recommendations are followed: further improvement and analysis needed before any statement is made on behalf of the ESAC.
- With respect to the HET-CAM assay and considering the overall study, it is recommended that the Expert Panel recommendations are followed: further analysis before any statement is made on behalf of the ESAC.

With regard to the readiness for regulatory purposes, the data are not structured or evaluated in a way to allow an informed conclusion to be drawn on the negative results in order for these to be used for regulatory purposes. For this purpose, a rigorously conducted evaluation would require access to and reconstruction of the original datasets. We therefore recommend that such an evaluation be conducted by ECVAM in order to assess the predictive capacity of these methods alone or in combination, for the whole range of the eye irritation response.

Other observations:

The Shadow Review Panel observed the discrepancy between the Expert Panel report and conclusions and those of the Evaluation Committee report, in particular with respect to the rejection of the IRE and HET-CAM tests as screening tests for corrosive and severe irritants. The Shadow Review Panel imputes this to the evaluation method of ICCVAM in which assays were evaluated for one unique protocol instead of considering the analysis of different protocols and prediction models developed for specific classes of chemicals. The Shadow Review Panel therefore agreed with the conclusions of the Expert Panel for the IRE and HET-CAM whereas it supported the conclusions of the Evaluation Committee for the BCOP and ICE assays.

IV. The Bovine Corneal Opacity and Permeability assay (BCOP)

1. *Consider if the ICCVAM material should allow a reasonable and expert person, for each of the test methods, to be satisfied that all of the relevant component parts of modular validation had or had not been properly addressed. The emphasis should be on 'positive reporting' - recording little against the headings where we find the required criteria have been met, and more where we have concerns they were not.*

The BCOP assay was evaluated for its ability as a screening test to identify potential ocular corrosives and severe irreversible irritant effects as determined by US EPA, EU (R41), and UN GHS classifications in a tiered-testing strategy, as part of a weight of evidence approach.

A total of 8 studies and 158 substances were used to evaluate the method. In fact, both single chemicals and formulations were studied, for which detailed *in vivo* and *in vitro* data were available. Protocols were similar but not identical.

Only part of the studies was conducted under GLP status.

The ICCVAM BRD confirms the essential components of modular validation are satisfied with respect to the BCOP assay.

Components of modular validation:

A. Test (method/model) definition

Detailed high-quality test protocols/SOPs are supplied and positive and negative controls are used.

Endpoints predicted are very clear*:* corneal opacity is measured quantitatively as the amount of light transmission through the cornea by an opacitometer*,* and permeability is assessed quantitatively by assessing changes in barrier properties of the epithelium to sodium fluorescein by a spectrophotometer. Both measurements are used to calculate an *In vitro* Irritancy Score (IVIS). This IVIS was derived empirically by the developers of the method, based on in-house results and further assay development studies.

It is recommended that histopathology is added to these measurements on a case by case basis.

Substances evaluated*:* considering that this assay is commonly used and has been subjected to several ring studies*,* the data sets used for these studies rather than a training set are discussed.

Domain of applicability: comprehensive, with identification given regarding classes of materials at greatest risk of giving false negatives and false positives.

A clear explanation of the mechanistic basis and similarity to the *in vivo* situation make the comparisons easy.

B. Within-laboratory variability

Information is provided regarding reproducibility of quantitative data, for replicate corneas, and for replicate experiments within one laboratory for three studies. Intra laboratory repeatability with IVIS was also measured.

C. Transferability

This method has been transferred successfully in numerous laboratories.

Prerequisites and experimental set-ups are well described.

D. Between-laboratory variability

Qualitative assessment*:* reproducibility of hazard classification category led to 67 to 94% of the substances being classified the same.

The substances for which there was less agreement are alcohols*,* ketones, heterocyclic compounds, as well as solvents*,* surfactants, chemical intermediates and pesticides.

E. Predictive capacity

Based on the available data base, the false negative rates for alcohols and solids range from 67 to 100%, and 42 to 50% respectively, depending on the hazard classification system. The false positive rates for alcohols, ketones and solids range from 50 to 56%, 40% and 10% respectively, depending on the hazard classification system.

When these chemical and physical classes are excluded from the data base*,* the accuracy across EPA, EU and GHS classification systems ranges from 87% to 92% and the false negative and false positive rates range from 0 to 12% and 12 to 16% respectively.

The BCOP assay is already used in Europe for in-house decision making for chemicals, mixtures, products and formulations (comparative results to benchmarks or original Gautheron classification).

EU regulators agree to accept positive BCOP results as the basis for R41 classification.

F. Applicability domain

Non-applicability is demonstrated for alcohols, ketones, heterocyclic compounds, as well as solvents, surfactants, chemical intermediates and pesticides.

G. Performance standards

Can be consulted (due consideration to improvement of the protocols and new results should be given).

2. Consider if the ICCVAM processes were generally sound, that due process was followed, and due diligence exercised.

A sound process is documented and the internal evidence indicates that due process was followed.

Diligent efforts were made to obtain all relevant data (including two calls in the Federal Register) and opinion (including public participation, and an expert panel that was convened twice); these were set out accurately and objectively in the BRD; suitable expert and public comments were elicited, considered and reflected in the final documents and outputs including updated performance metrics and further analysis as hazard classification rules were clarified. The ICCVAM draft documents were therefore updated as new information became available and was considered.

The gaps and imperfections in the evidence gathered and considered seem to have been made transparent and taken into account on the analysis and outputs.

3. Consider if the ICCVAM comments and recommendations are supported by the evidence considered.

A critical and objective review of the ICCVAM documents and outputs tends to confirm that the ICCVAM comments and recommendations with respect to the BCOP assay are consistent with the evidence gathered and considered.

The available data do not support the more general conclusion that the BCOP might be used as a stand-alone full replacement for the rabbit eye safety test.

4. Advise if the ICCVAM comments and recommendations, constructed primarily in the context of the US regulatory environment, is applicable to the regulatory frameworks that apply within the EU.

In the context of this assay the evidence, analysis and outputs are structured around the US, EU and GHS categories. In the case of the BCOP assay performance is high for all 3 systems.

EU regulators are already prepared to accept positive BCOP results as the basis for R41 classification.

5. Determine if there is any additional information not considered within the ICCVAM process, but which is required to inform ESAC's deliberations and outputs.

It is worth noting that two distinct substance “groups” have been assessed*:* pure chemicals / single substances on the one hand*,* and formulations on the other*.* It is not clear how many substances of each group have in fact been used for assessing the accuracy of the method. Our assessment points to a difference on the overall number of

substances used (170 and not 147 which would account for 102 chemicals and 68 formulations?).

In view of the replacement of animal tests, it is suggested that more attention be given to the formulations composition before any further "confirmatory" *in vivo* tests are conducted. The results on the BCOP, taken together with the individual ingredients *in vivo* irritation potential, could lead to avoid a confirmatory test. In the case of single chemicals, it would be interesting to assess the corresponding level of accuracy.

Taking into account the objective of the analysis, the prevalence of severe and corrosives *in vivo* does not seem to have influenced the design of the statistical analysis.

6. Consider if our advice should be that ESAC endorses, or if (on the basis of scientific opinion and evidence) ESAC should offer modified or supplementary recommendations.

The ICCVAM recommendations are consistent with a sound analysis and consideration of the available evidence. They are applicable to testing, risk assessments, and regulatory decisions that must be undertaken to satisfy relevant EU requirements.

7. General considerations: please address any other considerations that you might have.

Both ECVAM and EU experts played a part in the ICCVAM process.

The data analysis tends to confirm that, as would be predicted from the nature of the assay, false negatives are more likely to represent an inability to identify irreversible effects rather than an inability to provoke severe damage.

The eyes used for the assay are harvested post-mortem from cattle being killed for food production.

The BCOP can be completed within the day and requires basic training and easily available equipment. It does not require any proprietary components.

The BCOP is amenable to GLP.

Although the basis of the comparative costs is given, it is surprising that the estimated costs per assay are not significantly less that the cost of the rabbit eye safety test.

It would have been interesting to see how many of the chemicals identified as R41 might have been screened out by prior testing for skin corrosivity and that the present level of use of the organotypic tests has to be taken into account in assessment of impact on animal numbers.

8. Conclusions and recommendations.

With respect to the BCOP assay the ICCVAM recommendations should be endorsed by ESAC.

9. Was the proposed (BCOP) protocol and prediction model adequate to attain the objective of the proposed test method?

Yes. See above.

10. With regard to the readiness for regulatory purposes; could the negative results be used for regulatory purposes (e.g. R36 default classification)?

The current context is one of a screening assay. The data are not structured or presented in a way to allow an informed conclusion to be drawn. The potentials risk of harm caused by decisions made on the basis of false negative results can be estimated from the data presented, but a rigorously conducted evaluation would require access to and reconstruction of the original datasets.

The ICCVAM recommendations would reduce animal use (in the USA) at the expense of some materials inappropriately classified as R41.

11. Summary

The ESAC shadow review committee concludes that with respect to the Bovine corneal opacity and permeability assay (BCOP) test ICCVAM has addressed all relevant component parts of modular validation. The process followed by ICCVAM seems sound and efforts were made to obtain and evaluate all relevant data. ICCVAM's conclusion that the BCOP could be allowed for screening substances, using a weight of evidence approach for identifying ocular corrosives and severe irritants (EU R41), is consistent with the provided evidence.

We would go further on the basis of the evidence available in the ICCVAM material: we believe the evidence provided is sufficient to confirm the EU position that a positive BCOP result (with the prediction model reviewed) should be sufficient to the formal assignment of R41 categorisation for regulatory purposes.

In cases where chemicals are over- or under-predicted, it is advised to further develop the prediction model for these classes of substances.

V. The Isolated Chicken Eye assay (ICE)

1. *Consider if the ICCVAM material should allow a reasonable and expert person, for each of the test methods, to be satisfied that all of the relevant component parts of modular validation had or had not been properly addressed. The emphasis should be on 'positive reporting' - recording little against the headings where we find the required criteria have been met, and more where we have concerns they were not.*

The ICE assay was evaluated for its ability to predict ocular corrosives and severe irreversible irritant effects as determined by US EPA 1996, EU (R41), and UN GHS classifications as part of a tiered assessment strategy.

In some cases the data available was in the form of average score data or reported classification. Archived records were not accessed.

The precise GLP status of the published data is unclear, and deviations from GLP and their significance are not reported.

The ICCVAM BRD confirms the essential components of modular validation are satisfied with respect to the ICE assay.

Components of modular validation:

A. Test (method/model) definition

A detailed high-quality test protocol/SOP is supplied.

Note – this uses fewer replicates that some previously published test methods (without significantly impairing the accuracy of the prediction model) and makes provision for positive controls.

Endpoints predicted are defined. A composite assessment is based upon corneal swelling/thickness, opacity and fluorescein retention. These are partly quantitative and partly qualitative. It is suggested that histopathology might be undertaken on a case-by-case basis for borderline cases: it is not clear how often the need for this might arise or what the associated costs might be.

Corneal swelling/thickness is found to be the most variable of the measures. ICCVAM advice is offered as to how this might be reduced.

The training set is not specifically addressed, but can be derived from the information provided. 5 peer published studies reported on 175 substances (85 were proprietary compounds), single chemicals and products/formulations, from a wide range of chemical classes and covering the whole spectrum or ocular irritancy.

The domain of applicability is comprehensive, with warning given regarding classes of materials at greatest risk of giving false negatives and false positives.

A clear explanation of the mechanistic basis, and comparative anatomy, is provided.

B. Within-laboratory variability

Relevant information is provided regarding reproducibility of data when the test is used in the same laboratory by different operators at different times and is reviewed for each of the studies reported.

C. Transferability

Reproducibility in second (and subsequent) labs is provided, and ease of transferability is discussed.

Training needs and equipment are basic.

D. Between-laboratory variability

Reproducibility in different labs at different times is reviewed by different operators.

	GHS	EPA	EU
Number of substances	59	59	59
Number of labs	4	4	4
100% agreement	44/75	44/75	45/76
75% agreement	9/15	9/15	8/14
50% agreement	6/10	6/10	6/10

E. Predictive capacity

Data are provided.

Data are incomplete with regard to supporting human and rabbit data. Where rabbit data was available, the analysis of results was generally consistent with the EU testing and classification systems.

The ICE assay is already known to be used in Europe for in-house decision making for chemicals, mixtures, products and formulations.

EU regulators have already accepted positive ICE results as the basis for R41 classification.

F. Applicability domain

Applicable chemical classes and endpoints are defined and evaluated against both liquid and solid test materials. See below.

G. Performance standards

Can be derived from the information provided.

2. Consider if the ICCVAM processes were generally sound, that due process was followed, and due diligence exercised.

A sound process is documented and the internal evidence indicates that due process was followed.

Diligent efforts were made to obtain all relevant data (including two calls in the Federal Register) and opinion (including public participation, and an expert panel that was convened twice); these were set out accurately and objectively in the BRD; suitable expert and public comments were elicited, considered and reflected in the final documents and outputs including updated performance metrics and further analysis as hazard classification rules were clarified. The ICCVAM draft documents were therefore updated as new information became available and was considered.

The gaps and imperfections in the evidence gathered and considered seem to have been made transparent and taken into account on the analysis and outputs.

3. Consider if the ICCVAM comments and recommendations are supported by the evidence considered.

A critical and objective review of the ICCVAM documents and outputs tends to confirm that the ICCVAM comments and recommendations with respect to the ICE assay are consistent with the evidence gathered and considered.

In particular the available data do not support the more general conclusion that the ICE might be used as a stand-alone full replacement for the rabbit eye safety test. Nevertheless the ICE has been shown to be suited for consideration as a partial replacement for the rabbit test as part of a tiered and hierarchical approach to detect R41 materials - subject to all negatives and perceived false positives being confirmed by other means.

4. Advise if the ICCVAM comments and recommendations, constructed primarily in the context of the US regulatory environment, is applicable to the regulatory frameworks that apply within the EU.

In the context of this assay the evidence, analysis and outputs are structured around the US, EU and GHS categories. In the case of the ICE assay performance is optimum in the context of the EU requirements.

EU regulators have already accepted positive ICE results as the basis for R41 classification. Furthermore, the ICE has to date reliably identified materials classified as positives on the basis of skin corrosivity tests.

5. Determine if there is any additional information not considered within the ICCVAM process, but which is required to inform ESAC's deliberations and outputs.

See “II General considerations”: the opinions offered there are relevant to all of these test methods.

6. Consider if our advice should be that ESAC endorses, or if (on the basis of scientific opinion and evidence) ESAC should offer modified or supplementary recommendations.

The ICCVAM recommendations are consistent with a sound analysis and consideration of the available evidence. They are applicable to testing, risk assessments, and regulatory decisions that must be undertaken to satisfy relevant EU requirements.

Further assessment and subsequent development of the test is recommended to serve as a full replacement for the animal test, as there is need to find a full replacement (either a single stand-alone test or a battery of tests) for the rabbit test. In particular the classes of chemicals that were over- or under-predicted need special attention.

7. General considerations: please address any other considerations that you might have.

- Both ECVAM and EU experts played a part in the ICCVAM process.
- In the case of the types chemicals/materials shown to most likely be over or under predicted there is a need to consider for separate advice or comment on whether and how these should be tested, and how results should be interpreted. For example with alcohols false positives of 27-50% were observed. Surfactant and solid materials tended to be under-predicted.
- The data analysis tends to confirm that, as would be predicted from the nature of the assay, false negatives are more likely to represent an inability to identify delayed or irreversible effects rather than an inability to provoke strong early changes.
- The eyes used for the assay are harvested post-mortem from chickens being killed for food production.
- The ICE can be completed in 6 hours and requires only basic training and easily available equipment. It does not require any proprietary components.
- The ICE is amenable to GLP.
- Although the basis of the comparative costs is given, the estimated cost per assay seems to be not significantly less than the cost of the rabbit eye safety test.
- It would have been interesting to see how many of the chemicals identified as R41 might have been screened out by prior testing for skin corrosivity, and that the present level of use of the Organotypic tests has to be taken into account in assessment of impact on animal numbers. Relevant data that were supplied by TNO are not discussed in the evaluation.

8. Conclusions and recommendations.

- With respect to the ICE assay the ICCVAM recommendations should be endorsed by ESAC.
- Furthermore ESAC may also recommend the early revision of relevant technical annexes of EU Directives.

9. Was the proposed (ICE) protocol and prediction model adequate to attain the objective of the proposed test method?

Yes. See above.

10. With regard to the readiness for regulatory purposes; could the negative results be used for regulatory purposes (e.g. R36 default classification)?

The data is not structured or presented in way to allow an informed conclusion to be drawn. The potentials risk of harm caused by decisions made on the basis of false negative results can be estimated from the data presented, but a rigorously conducted evaluation would require access to and reconstruction of the original datasets.

The ICCVAM recommendations would reduce animal use (in the USA) at the expense of some materials inappropriately classified as R41 – in terms of responsible regulation and liability this is a completely different proposition to a broader endorsement setting the scene for a proportion of R41 substances being misclassified as R36.

This is our opinion with respect all four of the test methods.

11. Additional remarks on the EU Regulatory Context

The test method evaluations took account of US, EU and GSH systems and criteria. Aggregate information is presented, as is key information coded against the different evaluation systems. In no cases does analysis of the performance of a test method when judged only against the EU requirements significantly alter the validity of the ICCVAM analysis, conclusions or recommendations.

ICE Metrics	GHS	EPA	EU
	%	%	%
Accuracy	83	84	87
Sensitivity	50	52	59
Specificity	92	92	94
+ve prediction	63	63	73
-ve prediction	88	89	90
False +ve	8	8	6
False –ve	50	48/52	41

12. Summary

The ESAC shadow review committee concludes that with respect to the Isolated Chicken Eye (ICE) test ICCVAM has addressed all relevant component parts of modular validation. The process followed by ICCVAM seems sound and efforts were made to

obtain and evaluate all relevant data. ICCVAM's conclusion that the ICE could be allowed for screening substances, using a weight of evidence approach for identifying ocular corrosives and severe irritants (EU R41), is consistent with the provided evidence.

We would go further on the basis of the evidence available in the ICCVAM material: we believe the evidence provided is sufficient to confirm the EU position that a positive ICE result (with the prediction model reviewed) should be sufficient to the formal assignment of R41 categorisation for regulatory purposes.

In cases where chemicals are over- or underpredicted, it is advised to further develop the prediction model for these classes of substances.

VI. The Isolated Rabbit Eye assay (IRE)

1. *Consider if the ICCVAM material should allow a reasonable and expert person, for each of the test methods, to be satisfied that all of the relevant component parts of modular validation had or had not been properly addressed. The emphasis should be on 'positive reporting' - recording little against the headings where we find the required criteria have been met, and more where we have concerns they were not.*

The IRE assay was evaluated for its ability as a screening test to identify potential ocular corrosives and severe irreversible irritant effects as determined by US EPA, EU (R41), and UN GHS classifications in a tiered-testing strategy, as part of a weight of evidence approach.

From literature research and submitted data, four studies (*and not 3 as stated in the November 2006 Executive Summary*) could be identified with relevant detailed *in vivo* data out of fourteen published studies and submitted data. 149 substances and formulations were evaluated.

Only few original *in vitro* data were available, however summary *in vitro* data were available and could enable *in vitro* irritancy classification.

Protocols used are similar but not identical, several differences can be noted such as evaluation of two to four endpoints and classification criteria, which does not allow for direct comparisons.

The studies have been conducted in accordance with the GLP guidelines; however data recording could not be evaluated.

The ICCVAM BRD confirms that some components of modular validation are satisfied with respect to the IRE assay.

Components of modular validation:

A. Test (method/model) definition

By the use of the target organ from the rabbit (the species used in the Draize test) it seems to be the method of choice as a replacement of the Draize test. A clear explanation of the mechanistic basis and comparative anatomy is provided.

The process of acquisition of rabbit eyes is well described. It is strongly suggested that the material should be obtained from the rabbits killed for other purposes, so the method is not requiring any specific animal sacrifice.

There is a well-defined protocol (INVITTOX n°85), however there are no details on overall scoring and the classification scheme.

Endpoints predicted are well defined. An assessment is based upon corneal swelling/thickness, opacity, fluorescein retention and epithelium impairment.

Histopathology is recommended in the INVITTOX protocol, but not used systematically in the studies; however it is not clear how this endpoint should be evaluated and how it should affect the final decision.

Corneal swelling/thickness is found to be the most variable of the measures in the case of non severe/non corrosives. The use of the optical method for the assessment of corneal swelling seems to be one of the major elements affecting the performance of the test. The error of this method is about 10%. Authors of the report present results of the excellent correlation between *in vivo* and *in vitro* corneal swelling measured by the ultrasonic method, which should drastically increase the performance of the IRE test model.

The studies which were evaluated do not have the same protocol i.e., number of endpoints measured (two to four). There are several minor modifications which could have important effects on the final performance of the test.

4 peer-reviewed published studies reported on 149 substances, single chemicals and 25 products/formulations, from a wide range of chemical classes, and covering the whole spectrum or ocular irritancy.

Domain of applicability: a wide range of substances and physico-chemical forms (liquids and solids), alcohols, amides, amines, carboxylic acids, esters, ethers, formulations (skin cleansers, soaps, shampoos, surfactants and solvents).

B. Within-laboratory variability

No relevant information could be provided regarding the reproducibility of data when the test is used in the same laboratory by different operators at different times. ICCVAM recommends future studies to assess the intra-laboratory reproducibility.

C. Transferability

Reproducibility in second (and subsequent) labs is provided, and ease of transferability discussed.

Training needs and equipment are basic. There are no major problems with transferability however the problems with the effect of the differences in pachymetry equipment on the final results were addressed.

D. Between-laboratory variability

Reproducibility in two different studies were reviewed.

- first study*:* n=59, corneal opacity (overall CV=43.4%, severe and corrosives CV=33.6%) corneal swelling (overall CV=49.7%, severe and corrosives CV=35.5%) were observed.
- second study, n=21, corneal opacity, corneal swelling and fluorescein penetration (overall CV=24 to 40%, severe and corrosives CV=15.4 to 35.5%) were observed.

E. Predictive capacity

Based on the available data base, the accuracy ranged from 64 to 69%, false negatives from 24 to 31%, and false positives from 35 to 40% rates across the EU, EPA and GHS classification systems.

The lack of data using all four recommended IRE endpoints made it impossible to assess performance of the assay while using the four endpoints.

Although with a variable number of endpoints, and with different classification/prediction model, this assay has been used by several laboratories, particularly within industry, with satisfaction.

It should be questioned therefore, if the poor performance achieved in this evaluation is not a reflection of the discrepancies of the studies selected rather than the lack of performance of the assay itself.

EU regulators agree to accept positive IRE results as the basis for R41 classification.

F. Applicability domain

Domain of applicability: a wide range of substances and physicochemical forms (liquids and solids), alcohols, amides, amines, carboxylic acids, esters, ethers, formulations (skin cleansers, soaps, shampoos, surfactants and solvents). Overprediction for*:* alcohols, amines, ketones as well as liquids.

G. Performance standards

Are not clearly given in the BRD, however it is possible to derive them from the data provided (due consideration to improvement of the protocols and new results should be given)

2. Consider if the ICCVAM processes were generally sound, that due process was followed, and due diligence exercised.

A sound process is documented and the internal evidence indicates that due process was followed.

Diligent efforts were made to obtain all relevant data (including two calls in the Federal Register) and opinion (including public participation, and an expert panel that was convened twice); these were set out accurately and objectively in the BRD; suitable expert and public comments were elicited, considered and reflected in the final documents and outputs including updated performance metrics and further analysis as hazard classification rules were clarified. The ICCVAM draft documents were therefore updated as new information became available and was considered.

The gaps and imperfections in the evidence gathered and considered seem to have been made transparent and taken into account on the analysis and outputs. Small imperfections (tables without units) are not decreasing the scientific value of the whole document, however the number of studies retained (4 and not 3) should be modified in the executive summary.

3. Consider if the ICCVAM comments and recommendations are supported by the evidence considered.

A critical and objective review of the ICCVAM documents and outputs tends to confirm that the ICCVAM comments and recommendations with respect to the IRE assay are consistent with the evidence gathered and considered. However, although the processes are transparent, it can be questioned whether the methodology was fully adequate to assess the performance of the assay. The selection criteria of using only studies with both detailed *in vivo* and *in vitro* data, has most probably deprived the overall exercise of relevant information.

Further evaluation should be conducted. In particular one suggestion was the use of ultrasound pachymetry to increase the accuracy of the corneal swelling measurements.

The available data do not support the more general conclusion that the IRE might be used as a stand-alone full replacement for the rabbit eye safety test.

Furthermore, it should be noted that the conclusions from the Test Method Evaluation Report differ from those stated in the Expert Panel Report.

4. Advise if the ICCVAM comments and recommendations, constructed primarily in the context of the US regulatory environment, is applicable to the regulatory frameworks that apply within the EU.

In the context of this assay the evidence, analysis and outputs are structured around the US, EU and GHS categories. In the case of the IRE assay performance is similar for all 3 systems.

EU regulators are already prepared to accept positive IRE results as the basis for R41 classification.

5. Determine if there is any additional information not considered within the ICCVAM process, but which is required to inform ESAC's deliberations and outputs.

It is worth noting that two distinct substance “groups” have been assessed: pure chemicals / single substances on the one hand, and formulations on the other. A separate analysis would have been also appropriate.

In view of the replacement of animal tests, it could be suggested that attention be given to the formulations composition before any further “confirmatory” *in vivo* tests are conducted.

Taking into account the objective of the analysis, the prevalence of severe and corrosives *in vivo* does not seem to have influenced the design of the statistical analysis.

Finally, it is nowhere indicated what the criteria defined are in order to consider an assay suitable to be used as a screening assay (i.e. where is the threshold for measuring accuracy: over 65% ? 75% ?).

6. Consider if our advice should be that ESAC endorses, or if (on the basis of scientific opinion and evidence) ESAC should offer modified or supplementary recommendations.

Considering the overall study, it is recommended that the Expert Panel recommendations are followed: further improvement needed and further analysis before any statement is made on behalf of the ESAC.

7. General considerations: please address any other considerations that you might have.

The data analysis tends to confirm that, as would be predicted from the nature of the assay, false negatives are more likely to represent an inability to identify irreversible effects rather than an inability to provoke severe damage.

High variability of the corneal thickness might be reduced modifying the measurement technique. As mentioned in the BRD, ultrasonic pachymetry might offer superior accuracy and is not affected by the changes in corneal opacity.

It also seems interesting to employ more quantitative method for the opacity assessment.

8. Conclusions and recommendations.

With respect to the IRE assay and considering the overall study, it is recommended that the Expert Panel recommendations are followed*:* further improvement needed and further analysis before any statement is made on behalf of the ESAC.

9. Was the proposed (IRE) protocol and prediction model adequate to attain the objective of the proposed test method?

Further improvement is needed as well as further analysis.

10. With regard to the readiness for regulatory purposes; could the negative results be used for regulatory purposes (e.g. R36 default classification)?

The data does not allow this decision.

11. Summary

The ESAC shadow review committee concludes that with respect to the Isolated Rabbit Eye test ICCVAM might not have addressed all relevant component parts of modular validation. The process followed by ICCVAM seems sound and efforts were made to obtain and evaluate all relevant data. ICCVAM's conclusion that the IRE could not be allowed for screening substances, using a weight of evidence approach for identifying ocular corrosives and severe irritants (EU R41), is consistent with the provided evidence. The shadow peer review committee does observe, however, that several endpoints (4) for the IRE are applied and evaluated, each with insufficient data provided to make a sound conclusion for each of them. Furthermore, no distinction was made between evaluations of chemicals and formulations.

The shadow review committee recommends ESAC that further development and improvements of the tests are needed before a new analysis of the IRE can be made. An ESAC statement is not appropriate at this moment.

VII. The Hen's Egg Test – Chorio-Allantoic Membrane assay (HET-CAM)

Introductory Remarks

With respect to exposure and higher sensitivity of rabbit eye compared to humans: we must accept that huge differences exist between exposure in rabbit eye, possible exposure in human eyes and exposure in the proposed *in vitro* models.

ICCVAM has provided an Expert Panel Report (March 2005) and a Test Method Evaluation Report (March 2006). Surprisingly, conclusions of both reports with respect to the HET-CAM test are to some extent contradictory.

The report below focuses on the Test Method Evaluation Report (March 2006). Comments with respect to the deviating conclusions will be made in the recommendations.

1. *Consider if the ICCVAM material should allow a reasonable and expert person, for each of the test methods, to be satisfied that all of the relevant component parts of modular validation had or had not been properly addressed. The emphasis should be on 'positive reporting' - recording little against the headings where we find the required criteria have been met, and more where we have concerns they were not.*

The HET-CAM assay was evaluated for its ability to predict ocular corrosives and severe irreversible irritant effects as determined by US EPA 1996, EU (R41), and UN GHS classifications as part of a tiered assessment strategy.

Different existing HET-CAM scores and different Prediction Models for classifying the results of the HET-CAM were evaluated. Unfortunately, only for the HET-CAM IS (B) score conclusions were drawn. Moreover, the Prediction Model (PM) for predicting the label “R41” applied to this score by ICCVAM (IS>9 = severe irritant) has not been developed for, and never been proposed for classification purposes. In contrast, this scheme was proposed to grade eye irritancy potential into four groups from non-irritant to severe irritant (Luepke, 1985, Kalweit et al 1987, 1990). Validation studies conducted with the purpose to screen chemicals that need to be labelled “R41” have never used this PM, neither during conduct of the study, nor in extensive retrospective biostatistical analyses (e.g. Spielmann et al. 1996).

The precise GLP status of some of the published data is unclear, and deviations from GLP and their significance are not reported.

The ICCVAM BRD confirms the essential components of modular validation are satisfied with respect to the HET-CAM assay.

Components of modular validation:

A. Test (method/model) definition

Protocols/SOPs: Among others, two detailed high-quality test protocols / SOP’s are supplied, one that has been specially developed for testing surfactants (INVITOX protocol No. 96) in comparison with a concurrently tested benchmark surfactant, and the other one is to be used with all kinds of test chemical classes (INVITOX protocol No. 47). Both INVITOX protocols define controls and test acceptance criteria, however updates of these protocols are needed to take into account all experiences from validation studies.

Endpoints predicted: Since the vascularised CAM has similarity with the conjunctiva of the eye, the HET-CAM is primarily predicting effects in these tissues. However, since intra- and extravasal coagulation on the CAM are evaluated, the HET-CAM is also predictive for corneal lesions. This was confirmed in the IRAG analysis of CAM based assays (Spielmann et al. 1997).

Training set: A training data set is defined in INVITOX Protocol No. 47.

The domain of applicability has not been perfectly characterised. No chemical classes are specified where the HET-CAM is particularly weak. Coloured sticky test items are mentioned to cause problems.

A clear explanation of the mechanistic basis is provided.

B. Within-laboratory variability

Relevant information is provided regarding reproducibility of data when the test is used in the same laboratory by different operators at different times and is reviewed for each of the studies reported.

C. Transferability

Reproducibility in second (and subsequent) labs is provided, and ease of transferability is discussed.

Training needs and equipment are basic. However, provision of an atlas with pictures would support standardised application of the HET-CAM.

D. Between-laboratory variability

Reproducibility obtained by different operators, in different labs and over time has thoroughly been reviewed for the IS(B) score. Although the HET-CAM score IS(B) is a weighed sum score (combining *haemorrhage*, *lysis* and *coagulation*) with similar drawbacks as the *in vivo* MAS, or MMAS regarding reproducibility, the ICCVAM analysis confirmed sufficient reproducibility between laboratories and over time:

IS(B)-10	GHS	EPA	EU
Number of substances	107	104	106
Number of labs		2-3	2-3
100% agreement	85(79%)	84(81%)	84(79%)
66% agreement (50-74)	6(6%)	6(6%)	11(10%)
50% agreement (25-49)	16(15%)	14(13%)	11(10%)

IS(B)-100	GHS	EPA	EU
Number of substances	99	97	95
Number of labs		2-3	2-3
100% agreement	81(82%)	80(82%)	80(84%)
66% agreement (50-74)	6(6%)	6(6%)	4(4%)
50% agreement (25-49)	12(12%))	11(11%)	11(11%)

E. Predictive capacity

Overall, the IS(B) score combined with the "classification system" (PM) according to Luepke (1985) comprising four categories of eye irritancy potential provided high rates of overpredictions of the label "R41" in all studies. The predictive capacity of the HET-CAM for "R41" is significantly better if adequate PM's are applied.

F. Applicability domain

Applicable chemical classes and endpoints are defined and evaluated against both liquid, solid test materials and formulations. The analysis shows that the HET-CAM is weak in correctly predicting the *in vivo* eye irritation potential of alcohols. Thus, for testing alcohols, a special modification of the HET-CAM should be developed, or results from concurrently tested benchmark alcohols (with well established *in vivo* irritancy potential) be included in the evaluation.

G. Performance standards

Performance Standards have not been defined by ICCVAM. However, since the HET-CAM is used in Europe in a regulatory context, Performance Standards should be defined, as they can be derived from the information provided.

2. Consider if the ICCVAM processes were generally sound, that due process was followed, and due diligence exercised.

A sound process is documented and the internal evidence indicates that due process was followed.

Diligent efforts were made to obtain all relevant data (including two calls in the Federal Register) and opinion (including public participation, and an expert panel that was convened twice); these were set out accurately and objectively in the BRD; suitable expert and public comments were elicited, considered and reflected in the final documents and outputs including updated performance

metrics and further analysis as hazard classification rules were clarified. The ICCVAM draft documents were therefore updated as new information became available and was considered.

The gaps and imperfections in the evidence gathered and considered seem to have been made transparent and taken into account on the analysis and outputs.

3. Consider if the ICCVAM comments and recommendations are supported by the evidence considered.

A critical and objective review of the ICCVAM documents and outputs tends to confirm that the ICCVAM comments and recommendations with respect to the HET-CAM assay are consistent with the evidence gathered and considered.

In particular the available data do not support the more general conclusion that the HET-CAM might be used as a stand-alone full replacement for the rabbit eye safety test.

On the other hand, the HET-CAM has been shown to be suited for consideration as a partial replacement for the rabbit test as part of a tiered and hierarchical approach to detect R41 materials - subject to all negatives and perceived false positives being confirmed by other means.

Furthermore, the conclusions from the Test Method Evaluation Report deviate from those stated in the Expert Panel Report. Whereas the first concludes that the HET-CAM is not optimal for the identification of ocular corrosives and severe irritants, the Expert Panel agreed that the HET-CAM with the IS(B) analysis method is useful for identification of these substances, and that other Prediction Models than the IS(B) score and the “classification Scheme” of Luepke may be more appropriate for the purpose of identification of “R41”.

4. Advise if the ICCVAM comments and recommendations, constructed primarily in the context of the US regulatory environment, is applicable to the regulatory frameworks that apply within the EU.

In the context of this assay the evidence, analysis and outputs are structured around the US, EU and GHS categories. In the case of the HET-CAM assay performance is optimum in the context of the EU requirements.

EU regulators have already accepted positive HET-CAM results as the basis for R41 classification.

5. Determine if there is any additional information not considered within the ICCVAM process, but which is required to inform ESAC's deliberations and outputs.

As stated above, since the “classification scheme” of Luepke comprising four arbitrary groups of eye irritancy, has not been developed for identification of R41

chemicals, further analyses by ECVAM are highly recommended and should be considered.

6. Consider if our advice should be that ESAC endorses, or if (on the basis of scientific opinion and evidence) ESAC should offer modified or supplementary recommendations.

For the reasons given above, it is not advised that ESAC endorses the ICCVAM recommendations. A revised analysis that takes into account an appropriate PM is recommended.

7. General considerations: please address any other considerations that you might have.

- Both ECVAM and EU experts played a part in the ICCVAM process.
- In the case of the types chemicals/materials shown to most likely to be over or under-predicted there is a need to consider for separate advice or comment on whether and how these should be tested, and how results should be interpreted.
- The data analysis tends to confirm that, as would be predicted from the nature of the assay, false negatives are more likely to represent an inability to identify delayed or irreversible effects rather than an inability to provoke strong early changes.
- The HET-CAM is amenable to GLP.
- The costs per assay are estimated less than the cost of the rabbit eye safety test.

8. Conclusions and recommendations.

- With respect to the HET CAM the ICCVAM recommendations should currently not be endorsed by ESAC.
- Further analyses by ECVAM applying an appropriate PM specially developed to identify R41 (as e.g. described in Spielmann et al. 1996) are considered necessary.
- From the ICCVAM analysis it became apparent that clear guidance is needed on which of the different HET-CAM test protocols and which of the various prediction models (developed for specific purposes) is recommended to be used for the identification of eye corrosives and severe eye irritants.

9. Summary

The ESAC shadow review committee concludes that with respect to the HET-CAM ICCVAM might not have addressed all relevant component parts of modular validation. The process followed by ICCVAM seems sound and efforts were made to obtain and evaluate all relevant data. ICCVAM's conclusion that the HET-CAM could not be allowed

for screening substances, using a weight of evidence approach for identifying ocular corrosives and severe irritants (EU R41), is not consistent with the provided evidence because the Prediction Model applied was not developed for identification of severe eye irritants and corrosives.

The shadow review committee recommends ESAC that further analyses applying adequate prediction models are conducted. An ESAC statement is not appropriate at this moment.