9.0 OTHER SCIENTIFIC REPORTS AND REVIEWS

This section contains summaries of the available data from other published or unpublished studies conducted using the ICE test method. In many of these reports, inadequate information on the substances tested (e.g., identity not specific) and/or on the results obtained from the *in vitro* or *in vivo* studies (e.g., qualitative but not quantitative ICE data; group mean but not individual *in vivo* animal scores) precluded an assessment of the performance characteristics of ICE. However, using additional data received from the authors of the reports or from alternative sources (e.g., ECVAM), the test results on some of the substances in some of these reports were used to assess the performance of ICE. The results of these analyses are provided in **Sections 6.0** and **7.0**. This section provides a summary of reports (presented in alphabetical order by lead author) and the conclusions presented by the investigators, where such information was not available. An explanation why the data presented in a report could not be used to independently assess the performance of ICE is provided. In addition, where applicable, an explanation why some data could be used as part of the performance evaluation is provided.

9.1 Reports in the Peer Reviewed Literature

9.1.1 Balls et al. (1995)

Under the auspices of the British Home Office and Directorate General XI of the European Commission, a validation study on proposed alternatives to the *in vivo* rabbit ocular toxicity test method was conducted. The goal of the evaluation was to identify at least one non-whole animal test method that could be proposed to regulatory authorities as a replacement for the currently accepted *in vivo* ocular toxicity test method. For the ICE test method, a total of 52 substances were evaluated in 60 tests in four laboratories. Four of the test substances were evaluated at two different concentrations and two substances were evaluated at three different concentrations. The ocular irritancy potentials of the test substances were ranked in terms of MMAS, which ranged from 0 to 108. The test substances evaluated in the validation study were classified as acids (4), acyl halide (1), alcohols (9), aldehyde (1), alkali (1), esters (6), heterocyclics (3), hydrocarbons (2), inorganic chemicals (4), ketones (3), organophosphate (1), pesticides (5), surfactants (6), and miscellaneous (6). *In vivo* data for 46 of the test substances, generated in compliance with OECD TG 405, were obtained from historical sources. *In vivo* rabbit eye data for 14 of the test substances were obtained from concurrent studies conducted in compliance with OECD TG 405. *In vivo* data in the report were presented as MMAS. Comparison of the ICE test results to the GHS, EPA, or EU classification systems was not conducted.

As noted in **Section 5.4.2**, neither the individual substance scores for each ICE test method endpoint nor the overall Irritation Index was included in the published report. Rather, the study reports on the correlation between each ICE test method endpoint and the MMAS for the entire set of test substances. The MMAS was chosen as the *in vivo* reference endpoint by the EC/HO working group and therefore was the single *in vivo* endpoint included in the Balls et al. (1995) evaluation. Information about the 59 substances representing a wide-range of chemical classes and irritancy ranges tested in this study can be found in **Appendix B**.

In this study, the authors first generated X/Y scatterplots to visualize the relationship between the ICE test method results and the *in vivo* MMAS values. Spearman's rank correlation test and linear regression analysis were used to compare *in vivo* MMAS values with mean corneal swelling, mean opacity score, mean fluorescein retention score, and the ICE Index score. Spearman's rank correlation coefficients and Pearson's correlation coefficients were calculated for each participating laboratory for the entire test substance set, as well as for five subsets of test substances (water-soluble substances, surfactants, solids, solutions, and liquids). The ranges of the correlation coefficients for correlations between overall classification scores and MMAS that were obtained by each of the testing laboratories are presented in **Table 9-1**.

Table 9-1 *In Vitro/In Vivo* Correlation Coefficients from Balls et al. (1995)

Index Score	Pearson's Correlation Coefficient (r)	Spearman's Correlation Coefficient (r)
Full set of test substances (58-60 depending on endpoint)		
ICE-Mean Swelling	0.433-0.567	0.372-0.510
ICE-Mean Opacity Score	0.346-00.529	0.341-0.493
ICE-Mean Fluorescein Retention	0.380-0.568	0.357-0.576
ICE Index Score	0.490-0.599	0.416-0.552
Chemicals soluble in water (29-30 depending on endpoint)		
ICE-Mean Swelling	0.417-0.572	0.294-0.509
ICE-Mean Opacity Score	0.379-0.508	0.311-0.401
ICE-Mean Fluorescein Retention	0.329-0.408	0.291-0.453
ICE Index Score	0.451-0.558	0.334-0.450
Chemicals insoluble in water (17-18 depending on endpoint		
ICE-Mean Swelling	0.539-0.751	0.501-0.680
ICE-Mean Opacity Score	0.353-0.584	0.255-0.549
ICE-Mean Fluorescein Retention	0.233-0.779	0.197-0.736
ICE Index Score	0.603-0.748	0.510-0.664
Surfactants (n = 12)		
ICE-Mean Swelling	0.428-0.889	0.350-1.811
ICE-Mean Opacity Score	0.601-0.730	0.526-0.808
ICE-Mean Fluorescein Retention	0.638-0.879	0.640-0.873
ICE Index Score	0.724-0.833	0.657-0.872
Solids (19-20 depending on endpoint		
ICE-Mean Swelling	0.331-0.545	0.160-0.464
ICE-Mean Opacity Score	0.220-0.516	0.026-0.429
ICE-Mean Fluorescein Retention	0.223-0.345	0.193-0.364
ICE Index Score	0.335-0.492	0.060-0.424
Solutions (13-14 depending on endpoint		
ICE-Mean Swelling	0.471-0.853	0.342-0.823
ICE-Mean Opacity Score	0.549-0.751	0.503-0.725
ICE-Mean Fluorescein Retention	0.672-0.833	0.705-0.824
ICE Index Score	0.692-0.777	0.617-0.761
Liquids (n = 26)		
ICE-Mean Swelling	0.484-0.703	0.511-0.725
ICE-Mean Opacity Score	0.442-0.528	0.379-0.606
ICE-Mean Fluorescein Retention	0.401-0.676	0.421-0.657
ICE Index Score	0.557-0.666	0.583-0.676

The resulting analysis showed that overall, the ICE test method (based on Index Score) was not highly predictive of the MMAS (Pearson's Correlation Coefficient: 0.49 to 0.60 for the full set of test substances). Correlations with individual *in vitro* endpoints (corneal opacity, corneal swelling, and fluorescein retention) versus the MMAS also were relatively low (r = 0.35 to 0.57). Subset analyses revealed some differences among specific groups of test substances with Pearson's correlation coefficients ranging from 0.33 to 0.56 for water-soluble test substances, 0.23 to 0.78 for water insoluble test substances, 0.43 to 0.89 for surfactants, 0.22 to 0.55 for solids, 0.47 to 0.85 for solutions, and 0.40 to 0.70 for liquids.

To evaluate interlaboratory reproducibility of the ICE test method, Spearman's rank correlation coefficients and Pearson's correlation coefficients were calculated for each pair of participating laboratories for the entire test substance set, as well as for five subsets of test substances (water-soluble substances, surfactants, solids, solutions, and liquids). This analysis has been included in **Section 7.2.3**.

Since the *in vivo* test results were expressed as MMAS, the data provided in this report could not be used to evaluate the accuracy of ICE for detecting ocular corrosives and severe irritants according to the GHS, EPA, or EU classification systems (EPA 1996; EU 2001; UN 2003). However, using data provided by ECVAM, NICEATM was able to evaluate the ability of the ICE test method to identify severe ocular irritants or corrosives, as defined by the three classification systems (**Section 6.0)**, as well as to evaluate its interlaboratory reproducibility (**Section 7.0**).

9.1.2 Chamberlain et al. (1997)

This report describes a retrospective study of various alternative ocular irritation toxicity test methods that was conducted by the U.S. Interagency Regulatory Alternatives Group (IRAG). In response to a request by IRAG to the scientific community, one ICE test method submission was received for consideration. For reasons of confidentiality, information (substances tested, sponsors) submitted to the working group was not provided in the report. The report indicated that the ICE test method protocol used by Prinsen and Koëter (1993) was used to generate ICE test method data in this study. ICE test method data on 20 substances were provided. These substances included industrial chemicals, pesticides, detergents, commercial formulations, and foodstuffs. The 20 substances included 12 liquids, six solids, one gel, and one paste. The number of substances in each chemical class and other physicochemical characteristics (e.g., pH) were provided. Since the confidential data reviewed by IRAG may have overlapped with data provided in the reports already reviewed, this evaluation was not included in the main sections of this BRD.

In vivo rabbit eye reference data were provided for 15 of the 20 substances. The remaining five substances were found to be severe irritants in the ICE test method and therefore not evaluated *in vivo*. The *in vivo* ocular MAS values for the 15 tested compounds ranged from 0 to 68. The *in vivo* rabbit ocular tests were stated to have been conducted according to OECD TG 405. The protocol used to generate the *in vivo* reference data and information on the number of substances that were identified as non-irritants, irritants, and severe irritants are not provided in the published report. However, the *in vitro* ICE and *in vivo* rabbit ocular irritation data reportedly met the guidelines developed by a separate IRAG working group for

acceptance and evaluation of data submitted for comparing *in vitro* and *in vivo* test results (Scala and Springer 1997). This guideline provides general requirements for data acceptance, criteria for acceptable *in vitro* and *in vivo* data, and criteria for the consistent review and evaluation of data. According to this guideline, GLP compliant data are assigned greater significance, but submitted data need not be collected in compliance with these guidelines. It is unknown if these data were obtained from studies conducted in compliance with GLP guidelines. The original study data has not been made available.

Individual *in vitro* endpoint scores, ICE Index scores, individual animal results, or *in vivo* MAS scores were not provided in the report. However, the *in vivo*/*in vitro* correlation between the ICE Index scores and 10 different *in vivo* endpoints (**Table 9-2**) were calculated using Pearson's correlation coefficients.

Table 9-2 *In Vitro*/*In Vivo* Correlations in Chamerberlain et al. (1997)

In Vivo Endpoint	*In Vitro*/*In Vivo* Correlation (Pearson's Correlation Coefficient; r)
MAS	0.94
Total Opacity Score	0.94
Total Area Score	0.89
Total Iris Score	0.96
Total Redness Score	0.95
Total Swelling Score	0.93
Total Score for Discharge of the Conjunctivae	0.97
Number Days to Recover Score	0.96
Total Score for All the Effects of the Conjunctivae	0.97
Total Score for All the Effects of the Cornea	0.92

MAS = Maximum Average Score

Based on the Pearson's correlation coefficients, the ICE Index Score and the *in vivo* MAS values for the 15 test substances evaluated by the IRAG working group were highly correlated (r = 0.94). Correlations with other *in vivo* endpoints (corneal opacity, swelling, etc.) also were relatively high (r = 0.89 to 0.97). No other assessments of accuracy (e.g., concordance, sensitivity, specificity, false negative and false positive rates) were conducted and could not be evaluated since original data were not provided.

9.1.3 Prinsen (1996)

The author used a similar statistical approach to that of Balls et al. (1995) to calculate Pearson's correlation coefficient. However, this study included a comparison of the ICE Irritation Index and its individual components to 14 different *in vivo* scores (including MAS). A correlation analysis of the Irritation Index Score and the *in vivo* MAS for the 39 test substances evaluated *in vitro* by Prinsen (1996) resulted in a Pearson's correlation coefficient (r) of 0.91 (**Table 9-3**), a much higher correlation than that reported by Balls et al. (1995).

Table 9-3. *In Vitro*/*In Vivo* Correlation Coefficient from Prinsen (1996)

In Vivo Endpoint	*In Vitro*/*In Vivo* correlation (Pearson's Correlation Coefficient; r)
MAS	0.91
Total Opacity Score	0.87
Total Area Score	0.86
Total Iris Score	0.92
Total Redness Score	0.88
Total Swelling Score	0.90
Total Score for Discharge of the Conjunctivae	0.92
Number Days to Recover Score	0.88
Total Score for All the Effects of the Conjunctivae	0.92

MAS = Maximum Average Score

Correlations with the remaining 13 individual *in vivo* endpoints were also relatively high (r = 0.86 to 0.92), as were the correlations of individual *in vitro* endpoints (corneal swelling, opacity, and fluorescein retention) to the MAS (0.83 to 0.92). A list of the substances tested in this study is provided in **Appendix C**.

The data also showed that all of the substances defined as corrosive were classified as having a risk of causing serious eye damage (EU classification R41 [EU 2001]) by the ICE test method. However, because the MAS is not used for regulatory classification, this evaluation was not included in the main sections of this BRD.

9.2 Data Received in Response to the ICCVAM *Federal Register* Notice or from Study Authors

An *FR* notice (Vol. 69, No. 57, pp. 13859-13861; available at http://iccvam.niehs.nih.gov/methods/eyeirrit.htm), requesting original ICE test method data and *in vivo* reference data, was published on March 24, 2004. In addition, authors of published ICE studies were contacted to request original ICE data and *in vivo* reference data. In response to the *FR* notice, Procter and Gamble submitted ICE test method data and *in vivo* rabbit eye test data. Original data for the Prinsen and Koëter (1993) and the Prinsen (1996) studies could not be obtained.

9.2.1 Procter and Gamble (P&G) Submission from Drs. Daniel Marsman and Karen Acuff

On behalf of P&G, Drs. Daniel Marsman and Karen Acuff submitted sets of ICE test method data from studies performed to evaluate the ability of the ICE test method to discriminate between chemicals and benchmark proprietary formulations representing several different consumer laundry and cleaning products with varying eye irritation potential. The report notes that the ICE test method studies were conducted at TNO Nutrition and Food Institute, and provides the TNO protocol. This ICE test method protocol is the same as that used to

generate the proposed standardized protocol. ICE test method data on 28 substances were provided. These substances included surfactant raw materials, light duty dishwashing liquids, heavy-duty liquid laundry detergents, bleach containing laundry additives, and fabric enhancers. All of the formulations tested were liquid or surfactant solutions. The quantitative composition of each formulation was provided using both generalized and specific chemical information (e.g., total nonionic surfactants = 5 to 10%, sodium xylene sulfonate = 1 to 4%). The number of substances in each chemical class and other physicochemical characteristics (e.g., pH) were provided. The *in vivo* reference data used to compare the ICE test method was obtained using the low volume eye test (LVET) or available human data. The LVET varies from the traditional *in vivo* rabbit eye test by using an application of only 10 μL of a test substance, rather than the traditional 100 μL, in an attempt to reduce the amount of pain and suffering potentially experienced by test animals. The MAS and days to clearing (DTC) were provided for each test substance along with the EU classification. No individual animal data was provided for any of the test substances. Because reference data were not generated with the standard *in vivo* protocol, this evaluation was not included in the main sections of this BRD.

Mean maximum *in vitro* endpoint scores (along with the time of occurrence) were provided in the report, along with histopathology findings (when performed), and the predicted EU classification, but no individual eye scores were included. No statistical analysis was performed. Rather, a simple numerical assessment of the extent to which the ICE test method accurately predicted the *in vivo* classification was performed. **Table 9-4** provides the comparative results of each test substance. Several of the test substances were not assigned an EC classification based on the *in vivo* test, but rather were included in the evaluation based on accidental human exposure.

As demonstrated in **Table 9-4**, the ICE test method was able to accurately discriminate between surfactant raw materials (Adogen 444 [50%] and benzalkonium chloride [10%]) classified as severe irritants, by correctly determining the EU classification (based on LVET results) for both substances. However, results with various formulations were somewhat less predictive. With regard to light duty dishwashing liquids, the ICE test method accurately predicted the EU classification of three of the four test substances evaluated. The one remaining test substance reportedly would have been correctly predicted if histopathological findings had been included in the evaluation.

For the heavy-duty liquid laundry detergents, only one test substance was assigned an *in vivo* EU classification, which was marginally predicted by the ICE test method (i.e., AISE B5 was classified by the LVET as a borderline NI/R36 and the ICE test method classified it as an NI). The remaining six test substances were included based on mild, reversible effects noted in humans. The ICE test method predicted that all would be either NI or NI/R36.

For the bleach-containing laundry additives, the ICE test method correctly predicted the EU classification of only one of the five substances tested, underpredicting one severe irritant, and overpredicting three nonirritants. However, P&G stated that some of these errors might be corrected by including histopathology.

Table 9-4 EU Classification of P&G Consumer Laundry/Cleaning Products Based on the LVET and the ICE Test Methods

Test Material	*In Vivo* Classification (EU)	*In Vitro* Classification (EU)
Adogen 444 (50%)	R41	R41
Benzalkonium chloride (10%)	R41	R41
LDL645	R41	R36(R41[1])
Peroxi695*	R41	R36
LDL298*	R36/41	NI
Neodol 45-7	R36	R36
Peroxi694*	R36	R36
FE1828	NP	R36
FE2586	NP	NI(R36[6])
FE2587	NP	NI
FE2588	NP	R36
FE2589	NP	NI
FE2592*	NP	NI
HDL1813[4]	NP	NI
HDL1814[4]	NP	NI
HDL1815*[4]	NP	NI/R36
HDL2209[4]	NP	NI/R36
HDL2591*[4]	NP	NI
HDL809*[4]	NP	NI
AISE B5*[3]	NI/R36	NI
AISE C16[2]	NI/R36	NI
LDL659	NI/R36	R36
5% Sodium lauryl sulfate	NI	NI
FE2590[6]	NI	R41
Hypo580[6]	NI	R41
Hypo686[6]	NI	R41
Peroxi696[6]	NI	R36

Abbreviations: FE = Fabric enhancer; HDL = Heavy duty liquid laundry detergent; Hypo = Hypochlorite-containing bleach; LDL = Light duty dishwashing liquid; LVET = Low Volume Eye Test; Peroxi = Hydrogen peroxide-containing bleach; NP = Not provided

[1]Classification could be upgraded to R41 based on histopathology

[2]Formulation administered to 10 human volunteers. Corneal and conjunctival effects were observed that cleared within 24 hours

[3]Formulation administered to 10 human volunteers. Corneal and conjunctival effects were observed that cleared within 48 hours

[4]Corneal effects following accidental exposures to the human eye cleared within 1-2 days, with an occasional case taking up to 2 weeks to clear.

[5]Classification could be upgraded to R36 based on histopathology.

[6]Designated as a benchmark formulation for the particular category.

Finally, the ICE test method overpredicted the classification of the one fabric enhancer for which such data were provided. The authors state that these test substances are non- to very slightly irritating to the eye. However, the basis for this statement is not provided.

Therefore, the ICE test method provided variable results in this study when compared to the classification based on the LVET, particularly with respect to consumer formulations.

Although a total of 27 substances were tested in the ICE test method, only 15 were presented with *in vivo* EU classifications. Of these 15 test substances, the ICE test method accurately predicted the EC classification of eight. In addition, eight of these 15 test substances were designated as benchmark formulations for their respective category, of which the ICE test method accurately predicted the EU classification of only two. However, P&G commented on several occasions that the predictivity of the ICE test method could be enhanced if histopathological findings were included in the evaluation.