9.0 OTHER SCIENTIFIC REPORTS AND REVIEWS

9.1 Summaries of IRE Data from Published and Unpublished Studies

This section contains summaries of the available data from published or unpublished studies conducted using the IRE test method.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 IRE data, group mean but not individual *in vivo* animal scores) precluded an assessment of the performance of IRE. However, based on data received from contacting the authors or alternative sources (e.g., ECVAM), some substances included in these reports were used to assess the accuracy and reliability of IRE; these analyses are included in **Section 6.0**. This section provides a summary of reports (presented in alphabetic order by lead author) where such information was not available and the conclusions presented by the investigators. An explanation as to why the data presented in a report could not be used to independently assess the performance of IRE is provided. In addition, where applicable, an explanation as why some data could be used as part of the performance evaluation is provided.

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 IRE 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 potential 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), alkalis (1), esters (6), heterocyclics (3), hydrocarbons (2), inorganic chemicals (4), ketones (3), organophate (1), pesticides (5), surfactants (6), and miscellaneous (6). *In vivo* data for 46 of the test substances, which were generated in compliance with OECD TG 405 (OECD 1987), was 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 (OECD 1987).

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 IRE for detecting ocular corrosives and severe irritants according to the GHS (UN 2003), EPA (EPA 1996), or EU (EU 2001) classification systems. However, using data provided by ECVAM, an evaluation was conducted of the ability of the IRE 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**).

The individual scores for each IRE test method endpoint were not included in the published report in tabular form. Rather, the study reports the relationship between each IRE test

method endpoint to the MMAS in graphic form 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. A list of the 59 substances representing a wide-range of chemical classes and irritancy ranges tested in this study can be found in **Appendix B1**.

Spearman's rank correlation test and linear regression analysis were used to compare *in vivo* MMAS with irritancy in the IRE expressed as mean corneal opacity and mean corneal swelling, both measured at one and four hours. 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**.

The resulting analysis showed that overall, the IRE test method (based on the Summary Score) was not highly predictive of the MMAS (Pearson's Correlation Coefficient: 0.40 to 0.48 for the full set of test substances). Correlations with individual *in vitro* endpoints (corneal opacity and swelling) versus the MMAS also were relatively low (r = 0.25 to 0.61). Subset analyses revealed some differences among specific groups of test substances with Pearson's Correlation Coefficients ranging from 0.31 to 0.56 for water-soluble test substances, 0.10 to 0.76 for water insoluble test substances, 0.20 to 0.85 for surfactants, 0 to 0.57 for solids, 0.16 to 0.73 for solutions, and 0.11 to 0.76 for liquids.

9.1.2 Chamberlain et al. (1997)

As part of the Organotypic Models Working Group, Chamberlain et al. (1997) reviewed IRE test method data submitted to the Interagency Regulatory Alternatives Group (IRAG) on the use of isolated eyes and ocular components used to predict eye irritation potential. The protocol for the IRE test method was a modification of that described by Burton et al. (1981). A total of 107 substances were evaluated using the IRE test method. The substances represented a wide range of chemical types. The majority of substances (89) had MAS values of 30 or less (and therefore considered mild to moderate irritants) and 13 substances had MAS values ranging from 31 to 55 (and therefore considered moderate to severe irritants). The five severe irritants had MAS values equal to or greater than 55 and produced > 15% corneal swelling (**Table 9-2**). Greater than 50% of substances with MAS values between 31 and 55 (n = 13) produced corneal swelling greater than 15% in the IRE test method. When all of the substances were considered, only 38% produced > 15% corneal swelling. A Pearson's correlation coefficient of 0.50 was obtained when the IRE test results were correlated against the *in vivo* rabbit eye test results, presented as MAS scores. Consistent with some of the previous reports considered in this section, corneal opacity was not a good predictor of *in vivo* irritancy. The authors concluded that the IRE test method is suitable for screening severely irritating substances before *in vivo* animal tests are conducted, but cautioned that relying solely on organotypic methods for evidence of lack of an eye irritation hazard was not warranted at the present time.

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 (n = 59)		
IRE-Opacity, 1 Hour	0.407-0.502	0.316-0.510
IRE-Opacity, 4 Hours	0.485-0.606	0.451-0.606
IRE-Swelling, 1 Hour	0.247-0.528	0.166-0.515
IRE-Swelling, 4 Hours	0.447-0.611	0.364-0.624
IRE- Summary Score	0.399-0.483	0.473-0.603
Chemicals soluble in water (n = 30)		
IRE-Opacity, 1 Hour	0.422-0.514	0.238-0.377
IRE-Opacity, 4 Hours	0.341-0.516	0.226-0.440
IRE-Swelling, 1 Hour	0.305-0.492	0.329-0.552
IRE-Swelling, 4 Hours	0.329-0.552	0.293-0.511
IRE- Summary Score	0.471-0.560	0.311-0.426
Chemicals insoluble in water (n = 18)		
IRE-Opacity, 1 Hour	0.104-0.706	0.117-0.770
IRE-Opacity, 4 Hours	0.422-0.730	0.346-0.795
IRE-Swelling, 1 Hour	0.177-0.762	0.159-0.692
IRE-Swelling, 4 Hours	0.342-0.763	0.381-0.656
IRE- Summary Score	0.156-0.502	0.458-0.626
Surfactants (n = 12)		
IRE-Opacity, 1 Hour	0.466-0.833	0.486-0.855
IRE-Opacity, 4 Hours	0.696-0.853	0.623-0.828
IRE-Swelling, 1 Hour	0.204-0.690	0.007-0.720
IRE-Swelling, 4 Hours	0.532-0.677	0.504-0.746
IRE- Summary Score	0.513-0.666	0.613-0.839
Solids (n = 20)		
IRE-Opacity, 1 Hour	0.001-0.403	-0.056-0.373
IRE-Opacity, 4 Hours	0.231-0.564	0.130-0.534
IRE-Swelling, 1 Hour	-0.056-0.487	-0.182-0.504
IRE-Swelling, 4 Hours	0.112-0.566	-0.085-0.612
IRE- Summary Score	0.033-0.293	0.045-0.545
Solutions (n = 14)		
IRE-Opacity, 1 Hour	0.502-0.718	0.425-0.702
IRE-Opacity, 4 Hours	0.657-0.733	0.598-0.761
IRE-Swelling, 1 Hour	0.157-0.564	0.308-0.726
IRE-Swelling, 4 Hours	0.240-0.686	0.495-0.664
IRE- Summary Score	0.539-0.743	0.631-0.770
Liquids (n = 26)		
IRE-Opacity, 1 Hour	0.197-0.595	0.261-0.617
IRE-Opacity, 4 Hours	0.402-0.759	0.384-0.764
IRE-Swelling, 1 Hour	0.115-0.709	0.139-0.774
IRE-Swelling, 4 Hours	0.527-0.736	0.524-0.782
IRE- Summary Score	0.203-0.514	0.524-0.743

There was insufficient information in the IRAG report to assign GHS (UN 2003), EPA (EPA 1996), and EU (EU 2001) regulatory classifications to perform an accuracy analysis in this BRD. Furthermore, as the identity of the substances considered in the IRAG analysis were kept confidential and some of the data were likely to have been generated by studies considered elsewhere in this BRD, these data were not considered further.

Table 9-2 Relationship Between MAS *In Vivo* and the Ability to Cause More Than 15% Corneal Swelling *In Vitro* (Chamberlain et al. 1997)

MAS Range	N	Materials Causing >15% Corneal Swelling	
		Number of Substances	%
0-76 (all substances	107	41	38.3
≥ 55	5	5	100
31 to 55	13	7	53.8
≤30	89	29	32.6

9.1.3 Cooper et al. (2001)

Cooper and colleagues compared the IRE test method results on seven shampoo formulations to MAS values obtained from corresponding *in vivo* rabbit eye studies. The IRE protocol was modified from Burton et al. (1981) by inclusion of the evaluation of fluorescein penetration and histopathology.

The data generated in the study suggests that the IRE test method is useful for predicting the irritant potential of shampoo formulations that, in general, tend to produce mild to moderate rather than severe irritation (**Table 9-3**). In general, there appeared to be a concentration-dependent increase in irritancy for the shampoo formulations. Based on the IRE test results, one of five full strength shampoo formulations was overpredicted and one was underpredicted, when compared to *in vivo* rabbit eye test results. These authors also suggest, as demonstrated by Jones et al. (2001), that corneal swelling often occurs in the absence of corneal opacity.

Table 9-3 Comparison of IRE Test Method Results With *In Vivo* Data (Cooper et al. 2001)

Treatment[a]	IRE Irritancy Rating	*In Vivo* Irritancy Rating (MAS)
10% A	Moderate	No Test Data
10% B	Slight/Moderate	No Test Data
10% C	Moderate	No Test Data
10% D	Moderate	No Test Data
10% E	Slight/Moderate	No Test Data
10% F	Slight/Moderate	No Test Data
10% G	Very Slight/Slight	Mild (Predicted)
100% A	Moderate	Mild (14.3)
100% B	Moderate	Moderate (30.0)
100% C	No Test Data	Extreme (59.0)
100% D	Severe	Extreme (77.0)
100% E	Mild	Moderate (Predicted)
100% F	Moderate	Moderate (Predicted)
100% G	No Test Data	Mild (Predicted)

[a] Shampoo formulations (A is base formula, B is base with 1.5% ingredient X, C is base with 3.0% ingredient X, D is base with 6.0% ingredient X, E and F are reference controls, and G is a baby shampoo).

There was insufficient information in this report to assign a GHS (UN 2003), EPA (EPA 1996), and EU (EU 2001) regulatory classification for the accuracy analysis in **Section 6.0**.

9.1.4 Gettings et al. (1996)

As part of the Phase III CTFA validation study, Gettings et al. (1996) evaluated 25 surfactant-based personal care formulations using the IRE test method. *In vitro* responses were measured using either corneal swelling in the IRE (referred to in the report as the Rabbit Enucleated Eye Test or REET I) or scored according to severity (score ranging from 0 to 3) of the REET I corneal swelling results (referred to as the Rabbit Enucleated Eye Test II). Substances with *in vitro* scores greater than 18.6 for the REET I analysis or a score greater than 1.0 for the REET II were classified as irritants. Substances that did not meet these criteria were designated nonirritants. There was no attempt to distinguish severe irritants from moderate or mild irritants. The *in vitro* data obtained in the IRE were compared to *in vivo* rabbit eye test data obtained using the Draize scoring method (Draize et al. 1944) expressed as MAS or were classified as irritant or nonirritant based on the FHSA regulatory classification (FHSA 1988). The results of these analyses are shown in **Table 9-4.**

Table 9-4 The Results of the CTFA Evaluation of *In Vitro* Alternatives to the Draize Primary Eye Irritation Test (Phase III): Surfactant-Based Formulations (Getting et al. 1996)

Phase III Substances	REET I[a] (Percent of Control)	REET II[b] (Irritancy Score, 0-3)
Classified as Irritants by FHSA[b] Criteria		
HZQ	7.5	0.3
HZG	29.5	2.0
HZN	37.7	2.7
HZD	20.3	1.0
HZB	24.8	1.7
HZV	25.6	1.3
HZW	23.9	1.7
HZU	36.7	3.0
HZC	21.2	1.0
HZF	14.3	1.0
HZA	32.1	2.3
HZL	36.2	2.7
HZR	13.4	1.0
HZK	36.4	2.7
HZX	20.9	1.3
HZI	28.6	1.7
HZS	33.3	2.3
HZY	18.6	1.0
Classified as Non-Irritants by FHSACriteria		
HZH	7.7	0.0
HZZ	2.8	0.0
HZT	2.5	0.0
HZI	16.3	1.0
HZP	25.0	1.7
HZM	26.0	1.7
HZE	6.4	0.0

[a] Modified from Burton et al. (1981) using 20 µL test material at 10 second intervals for 1 minute. Represents percentage increase in mean corneal thickness compared to control. Score ≥ 18.6 considered irritant.

[b] Modified from Burton et al. (1981) using 20 µL test material at 10 second intervals for 1 minute. Represents a classification into one of four groups (0 to 3) based on the degree of corneal swelling. Score ≥ 1.0 considered irritant.

For the FHSA classification system (FHSA 1988) for identification of irritants, an accuracy of 80% (20/25), a sensitivity of 83% (15/18), a specificity of 71% (5/7), a false positive rate of 29% (2/7), and a false negative rate of 17% (3/18) were obtained for REET I. For the REET II test, an accuracy of 84% (21/25), a sensitivity of 94% (17/18), a specificity of 57% (4/7), a false positive rate of 43% (3/7), and a false negative rate of 6% (1/18) were obtained. The authors also calculated a separation index for each substance tested for REET I. The separation index represents the rate at which the *in vitro* endpoint (corneal swelling) and MAS do not agree. The mean of separation indices was 0.463 ± 0.026 (a standard error based on a Monte Carlo estimate of variability). A value of 1.0 indicates complete concordance with the *in vivo* outcome.

In vivo data from the Gettings et al. (1996) report were not used as provided, because FHSA classification does not include a severe irritant category. However, *in vivo* data were received from the CTFA in response to an *FR* notice that allowed for an accuracy analysis. This analysis is provided in **Section 6.0**.

9.1.5 Guerriero et al. (2004)

Guerriero and colleagues obtained data using the IRE test method protocol as described in **Section 5.1.3**. The study evaluated the response of 44 substances (30 pharmaceutical process materials, 14 ECETOC compounds) in the IRE test method. *In vitro* data were recorded as scores for corneal opacity and area, corneal swelling, scores for fluorescein intensity and area, and observations of epithelial integrity (pitting, mottling, sloughing). Test substances that produced an *in vitro* corneal opacity x area score ≥ 3, a fluorescein uptake intensity x area score ≥ 4, swelling ≥ 25, or produced corneal epithelial damage were designated as severe irritants. Test substances that did not exceed this score were classified as nonsevere irritants. Data obtained from concomitant *in vivo* rabbit eye irritation tests on these substances were classified for ocular irritancy according to the EU classification system (EU 2001). Using these multiple decision criteria, the authors correctly identified 100% (n = 15) of R41 substances. The authors concluded that use of the IRE assay supports the concept of the 3Rs (replacement, reduction, and refinement) and that the IRE assay is a valuable and practical screening tool to identify substances that are severe eye irritants.

In their 2004 report, Guerriero et al. provided a EU regulatory classification (EU 2001) for the *in vivo* data. Upon request, the authors kindly provided the individual animal *in vivo* response data, which permitted classification according to the GHS (UN 2003) and EPA (EPA 1996) classification systems. These results were used in the accuracy analysis described in **Section 6.0**.

9.1.6 Jacobs and Martens (1990)

Using an ultrasonic pachymeter, corneal swelling (expressed as a percentage) derived from the mean increase in corneal thickness produced in response to application of 34 test substances of varying irritancy levels at 4, 24, 48, and 72 hours *in vivo* was compared to that obtained in the Isolated Eye Test (IET) at two and four hours. Linear correlation between corneal swelling *in vitro* and *in vivo* tests at four hours was slight with r = 0.77. However, when test substances that produced epithelial opacity (notably acids) were omitted from the evaluation, the correlation between *in vitro* corneal swelling at two and four hours improved

to r = 0.91, when compared against the mean *in vivo* corneal swelling measured at 24, 48, and 72 hours (EU 2001). Linear correlation between mean percentage corneal opacity scores and mean corneal swelling was satisfactory with r = 0.89. In this study, a percentage increase in corneal swelling of 55% obtained in isolated rabbit eyes over two and four hours, corresponds to the limit of an irritant classification using the EEC (1984) regulatory classification system. When this criterion was applied to all of the substances excluding those that produced epithelial swelling, one false positive and no false negatives were observed.

9.1.7 Jacobs and Martens (1989)

The ultrasonic pachymeter has been shown to be more accurate than the optical pachymeter (Salz et. al. 1983; Thornton 1985) and has the advantage that it is easy to handle and transport, has rapid measuring speed, requires less operating skill, is not restricted to measurement of central corneal thickness and can be used in the presence of severe opacity (Jacobs and Martens 1988). Thirty-four chemically diverse test substances with a wide range of irritant responses were tested in the *in vivo* rabbit eye test for corneal swelling using an ultrasonic pachymeter and this data was compared to mean Draize corneal opacity, erythema, chemosis, and iritis scores. Mean corneal swelling at 24, 48, and 72 hours was determined. The eye irritation protocol described in EEC (1979) was used for the assay. Linear correlation between mean percent corneal swelling measurements and corneal opacity scores was r = 0.94. Linear correlation between mean percent corneal swelling measurements and chemosis scores were r = 0.87. Erythema scores were not linear with percent corneal swelling measurements, due to a limited erythema scale and the need for a minimum degree of erythema to be produced before corneal swelling can be measured. Mean percent corneal swelling at 24 and 72 hours using ultrasonic pachymetry were comparable to 24-hour optical pachymetry measures, while ultrasonic measures were lower than optical pachymetry measures at 72 hours. The authors suggest that addition of a quantitative and sensitive measure such as ultrasonic pachymetry to *in vivo* rabbit eye testing for ocular toxicity would reduce intra- and interlaboratory variability.

9.1.8 Jacobs and Martens (1988)

The ultrasonic pachymeter was used to measure corneal swelling (expressed as a mean percentage and standard deviation) in response to 11 substances tested in the enucleated rabbit eye test and compared to mean percentage corneal swelling results obtained on these substances in the enucleated rabbit eye test methods previously reported by Burton et al. (1981) using an optical pachymeter and by Köeter and Prinsen (1985) using an ultrasonic pachymeter. Mean percentage corneal swelling was determined 240 min after test substance application to four enucleated rabbit eyes after a 10 sec exposure to the test substance followed by saline rinse. Although the measured results were not identical, good correlation with an r-value of 0.98 was obtained by plotting a linear regression of 240-minute ultrasonic data and the optical pachymeter data from Burton et al. (1981). Standard deviations for both test methods were of the same order of magnitude, with the exception of acetone and ethanol which were higher for the ultrasonic pachymeter. Corneal opacity scores at 240 minutes compared to ultrasonic pachymetry with a Spearman's rank correlation coefficient of 0.91 ($p < 0.0005$). Using a mean epithelial damage score produced a less satisfactory correlation

(0.78; $p < 0.005$). Careful assessment of epithelial integrity in response to the applied test substances and to the ultrasonic pachymeter itself, revealed that the pachymeter did not significantly contribute to epithelial damage observed in response to the test substances. Higher values for the ultrasonic pachymeter against strongly irritating materials such as allyl alcohol, 1N sodium hydroxide, and butanol might be related to the fact that optical pachymetry units are not linear with swelling or that increased corneal opacity resulted in a concomitant decrease in refractivity of the cornea.

9.1.9 Jones et al. (2001)

Jones and colleagues published a study comparing ten shampoo formulations and seven conditioner formulations using five alternative test methods, including the IRE. The shampoos were tested at both 100% and 10% concentrations. The investigators modified the original Burton et al. (1981) IRE test method to include evaluation of fluorescein retention and evaluation of the epithelium. The investigators found generally good agreement between the irritancy ratings of the shampoo and conditioner formulations based on IRE data and their *in vivo* irritancy rating based on historical data. Eight of the 17 formulations classified as moderate irritants based on *in vivo* rabbit eye test results were either classified correctly or overpredicted, but never underpredicted (i.e., no false negatives were identified). A single severe ocular irritant formulation was correctly predicted by the IRE. However, for most test substances, corneal opacity alone was not as predictive as corneal opacity combined with corneal swelling and histology. Histology scoring appears to be responsible for some of the overpredicted classification, since a maximum number of layers lost rather than an average was used. For example, in cases where there was a wide range of responses of cell layers lost (e.g., two to seven), use of an average value instead of the maximum would have reduced the overall score. Furthermore, the conditioners tended to be overpredicted more frequently than the shampoos, perhaps because they contained predominately cationic surfactants versus the anionic and amphoteric surfactants contained in the shampoo formulations. The authors concluded that the data supports continued use of the IRE test method as an alternative to the *in vivo* rabbit eye irritation test with recognition that it can overpredict the irritancies of some formulations.

There was insufficient data provided in this report to assign GHS (UN 2003), EPA (EPA 1996), and EU (EU 2001) classifications for the tested formulations to perform an accuracy analysis in **Section 6.0**.

9.1.10 Koëter and Prinsen (1985)

A total of 34 substances were evaluated using the IRE test method and the data were compared to *in vivo* rabbit eye data obtained in the Draize test (**Table 9-5**). In this report, the test substances are indicated by code and therefore the substance names are unknown. However, physicochemical properties, including pH values, for some substances were provided. A mixture of hydrophilic (14) and hydrophobic (11) liquid substances and nine solid substances with pH values ranging from 1.8 to 13.5 were tested. In this assay, the Burton et al. (1981) protocol was modified to include fluorescein penetration and histology.

Table 9-5 Comparison of IRE *In Vitro* Irritancy Grades to *In Vivo* Rabbit Eye Test Irritancy Classifications (Koëter and Prinsen 1985)

Test Substance	Irritancy Grade	
	In Vitro[a]	***In Vivo***[b]
1	Not Irritant	Not Irritant
2	Slight	Slight
3	Slight	Slight
4	Moderate/Severe	Severe
5	Slight	Slight
6	Severe	Severe
7	Slight	Severe
8	Severe	Severe
9	Slight	Not Irritant
10	Negligible	Not Irritant
11	Not Irritant	Not Irritant
12	Not Irritant	Not Irritant
13	Moderate	Moderate
14	Slight	Not Irritant
15	Moderate	Not Irritant
16	Not Irritant	Not Irritant
17	Severe	Severe
18	Slight	Slight
19	Negligible	Not Irritant
20	Not Irritant	Not Irritant
21	Slight	Slight
22	Negligible	Not Irritant
23	Negligible	Not Irritant
24	Negligible	Not Irritant
25	Severe	Severe
26	Not Irritant	Not Irritant
27	Negligible	Not Irritant
28	Not Irritant	Not Irritant
29	Not Irritant	Not Irritant
30	Slight	Severe
31	Slight	Slight
32	Negligible	Not Irritant
33	Moderate	Not Irritant
34	Severe	Severe

[a] Based on overall Irritancy Rating

[b] Based on Draize score according to FDA guidelines (FDA 1980)

For identification of severe irritants, the accuracy was 91% (31/34), sensitivity was 63% (5/8), specificity was 93% (26/28), the false positive rate was 7% (2/28), and the false negative rate was 38% (3/8).

Corneal opacity was scored and corneal swelling was calculated based on the percentage increase in corneal thickness at each time point relative to a preapplication measurement, but modified with respect to the inclusion of the additional parameters -- histological assessment of the cornea and fluorescein penetration. Based upon averaging the final scores of all four *in vitro* endpoints, an overall Irritancy Rating was assigned. A comparative analysis of the IRE test results and the Draize rabbit eye test scores indicates that 28 of the 34 substances (82%) had similar irritancy ratings *in vitro* and *in vivo*. In general, the irritancy ratings were

predictive throughout the range of irritancy with a few exceptions. Two substances (6%) were underpredicted and four substances (12%) were overpredicted. Importantly, the two underpredicted substances were classified as severe ocular irritants *in vivo* on the basis of persistence of adverse effects and not the severity of the effect. The authors conclude that the IRE test method is a useful and sensitive test system for the evaluation of ocular irritation. A performance analysis on the reported data for identification of severe irritants indicated that the accuracy was 91% (31/34), sensitivity was 63% (5/8), specificity was 93% (26/28), the false positive rate was 7% (2/28) and the false negative rate was 38% (3/8).

There was insufficient information in this report to assign GHS (UN 2003), EPA (EPA 1996), and EU (EU 2001) regulatory classifications to perform an accuracy analysis in **Section 6.0**.

9.1.11 Lewis et al. (1994)

Lewis and colleagues published a report on the use of an *in vitro* test battery as a prescreen in the assessment of ocular irritancy. The authors describe a trypan blue exclusion assay using a human myeloid cell line as an initial screening test for severe irritants based on cytotoxicity. Test substances that produced < 15% cytotoxicity were tested *in vivo* using the rabbit eye test method while substances that produced > 15% cytotoxicity were tested using the IRE test method. In the IRE test method, if a substance produces less than 15% corneal swelling, one animal is tested *in vivo* since there is little likelihood of a severe irritant response. Those test substances producing greater than 15% corneal swelling are likely to be severe irritants; therefore, only one animal is tested initially using the low volume eye test in which the quantity dosed is 0.01 mL or 0.01 g. A total of 93 substances were evaluated using this tiered *in vitro* approach.

Among these 93 substances, a complex fiber formulation and a research agrochemical were classified as false negatives. Eight false positives were identified. Using nonparametric analysis, it was concluded that the majority of severe eye irritants were correctly predicted *in vitro*, with a sensitivity (ability to predict severe irritants) of 83% and a specificity (ability to identify less than severe irritants) of 90%. The authors concluded that although 10 of 11 severe eye irritants were predicted correctly using the IRE test method and 11 of 12 severe eye irritants were predicted by the trypan blue exclusion assay, the incidence of false positive responses in each of the assays still precludes their routine use as complete replacements for the *in vivo* rabbit eye test. However, the authors added that the *in vitro* battery assay approach does reduce the number of animals used and is clearly superior to reliance on skin testing data as an indicator of potential ocular effect. Using this approach, the authors report a reduction of 85% in the number of laboratory animals treated in the traditional *in vivo* rabbit eye test.

There was insufficient information in this publication to assign GHS (UN 2003), EPA (EPA 1996), or EU (EU 2001) regulatory classifications for the accuracy analysis in **Section 6.0**.

9.1.12 Price and Andrews (1985)

Price and Andrews evaluated the *in vivo* predictive accuracy of 60 substances using the IRE test method. The 60 substances included 25 industrial chemicals and 32 formulations (three

unformulated agrochemicals, 14 formulated lubricating oils and 18 formulated agrochemicals). The results were presented as a ratio of the *in vitro* prediction of irritancy with an *in vivo* classification expressed as a percentage. In this study, the Burton et al. (1981) protocol was modified to include evaluation of fluorescein penetration. Corneal thickness measurements along with evaluations of corneal appearance were recorded at regular intervals for up to five hours. Fluorescein penetration was recorded at four hours, if damage was present. Irritancy criteria for the *in vivo* eye test were based on OECD guidelines (OECD 1983). The scoring system for determination of severe irritancy *in vitro* was based on the time for corneal swelling to equal or greater than 20% (Grade IV, maximum). Lesser grades were assigned if it took longer to achieve this level of swelling (two hours, Grade III; five hours, Grade II, or less than 20% swelling in five hours, Grade I, minimal). Using these decision criteria, the results demonstrated that 10 (83%) of the 12 *in vivo* Class IV (severe) irritants and 33 (97%) of the 34 Class I (nonirritants or very mild) irritants were correctly identified by the IRE test method. For the detection of severe irritants only, a retrospective performance analysis indicated that the accuracy was 97% (58/60), sensitivity was 83% (10/12), specificity was 100% (48/48), the false positive rate was 0% (0/48) and the false negative rate was 17% (2/12).

There was insufficient information in this publication to assign GHS (UN 2003), EPA (EPA 1996), or EU (EU 2001) regulatory classifications for the accuracy analysis in **Section 6.0**.

9.1.13 Whittle et al. (1992)

In an interlaboratory trial of the IRE test method, Whittle and colleagues studied the ocular effect of 27 substances (17 liquids and 10 solids) representing a variety of chemicals and surfactants using the IRE test method. A modification of the IRE protocol described by Burton et al. (1981) was used that included an assessment of fluorescein retention and an evaluation of epithelial cell erosion. For two laboratories, the exposure duration (ten seconds) was the same as that proposed by Burton; in the third laboratory, the exposure duration was increased to one minute. The two laboratories that used the ten-second exposure protocol were able to separate severe/moderate from the mild eye irritants. *In vivo* irritancy was rated as severe, moderate/severe, moderate, slight/moderate, or slight, using in-house historical data on the *in vivo* rabbit eye test.

For the majority of test substances, evaluation of corneal swelling with a ten second exposure was a better indicator of irritancy than corneal opacity. For example, for the 17 liquid substances tested, all seven moderate to severe irritants induced corneal swelling of greater than 11% in both laboratories. However, corneal opacity was induced by only two of the seven-moderate/severe substances in both laboratories and by another substance in only one of the two laboratories. For the ten solid substances tested, corneal swelling was >12.5% for the three moderate to severe irritants in both laboratories, while corneal opacity was induced by two of three moderate to severe irritants and only in one of two laboratories. Evaluation of results from the 60-second exposure did not appear to provide additional benefit in identifying severe irritants. The investigators concluded that the IRE test method was useful for separating moderate to severe eye irritants from the milder eye irritants. However, it was also clear from the study that corneal opacity alone was not predictive of mild/moderate or moderate irritants using a ten-second exposure. The consistency of rating of irritancy

between laboratories was considered excellent for liquids, but was less impressive for solid materials.

There was insufficient information in this study to conduct an accuracy analysis as described in **Section 6.0**.

9.1.14 York et al. (1982)

York and colleagues published a report describing preliminary findings of an *in vitro* test for the assessment of eye irritancy in consumer products. A modification of the Burton et al. (1981) protocol was used in which evaluation of fluorescein penetration and histopathology were included. Eleven test substances with a span of irritancy ranging from no effect to very severe ocular damage (as reported in literature) were evaluated. The authors compared their *in vitro* irritancy ratings (mild to severe) to an *in vivo* Irritancy Grade (1-10; 10 being the most severe) described by Carpenter and Smyth (1946). Of 10 substances graded using the Carpenter and Smyth scale, three substances rated severe *in vitro* had *in vivo* grades of 10, 9 and 8, respectively, and were correctly predicted. Allyl alcohol was rated moderate/severe *in vitro*, assigned a five (moderate) on the *in vivo* scale, and therefore overpredicted. Toluene was underpredicted *in vitro* as negligible/slight whereas it had a scale of 7 (moderate/severe) *in vivo*. Overall, the authors conclude that the IRE test method is a valid model to use as a screening procedure for strong irritants.

There was insufficient information in this report to assign a GHS (UN 2003), EPA (EPA 1996), or EU (EU 2001) classification for the accuracy analysis in **Section 6.0**.

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 IRE test method data and *in vivo* reference data, was published on March 24, 2004. In addition, authors of published IRE studies were contacted to request original IRE data and *in vivo* reference data. In response to the *FR* notice, Guido Jacobs of the Institute for Hygeine and Epidemiology (Brussels, Belgium) and Dan Marsman of Proctor and Gamble (P&G; Cincinnati, Ohio) submitted reports of IRE test method data and *in vivo* rabbit eye test data.

9.2.1 Jacobs and Martens (January 1987)

Twenty-one substances were tested in the *in vivo* rabbit eye test (EEC 1979) and results were obtained for erythema, edema, corneal opacity, iritis, pain response, damage of the corneal epithelium, healing, and corneal swelling. This *in vivo* data was compared to the enucleated eye test of Burton et al. (1981) using the same set of substances. Mean percentage corneal swelling was determined in three rabbits over 24, 48, and 72 hours. Mean percentage corneal swelling in the enucleated eye test was obtained over 0.5, 1, 2, 4, and 5 hours. *In vitro* corneal swelling with various *in vivo* endpoint results correlated with corneal opacity ($r = 0.92$), erythema ($r = 0.91$), and percent fluorescein retention ($r = 0.94$). Correlation between mean percentage corneal swelling at four hours and the mean calculated over all observation times (24, 48, and 72 hours) was not as good ($r = 0.82$). Erythema appeared to be the most

sensitive indicator of ocular damage, and some degree of erythema was required before corneal opacity or chemosis were triggered. No correlation between pain response and production of ocular lesions was found. Test substances could be divided into two groups, one in which corneal swelling was increasing at five hours and one in which it had reached a maximum level by five hours. When *in vivo* clinical observations are considered (i.e., corneal opacity, erythema, chemosis, and iritis scores), the first group represents moderate to severe ocular irritants, whereas the latter group represents mild to moderate ocular irritants. The authors concluded that the enucleated eye test is a valid screening method for ocular irritation, although eye irritation classification cannot be based on the results of percentage corneal swelling alone or based on evaluation of a relatively small set of test substances.

9.2.2 Jacobs and Martens (May 1987)

An ultrasonic pachymeter was used to measure the percentage corneal swelling using the enucleated eye technique described by Köeter and Prinsen (1985) on the irritancy of 11 test substances reported by Burton et al. (1981) using optical pachymetry. Pachymetry data from one enucleated rabbit eye per test substance at 240 min was compared to the same substance tested in three enucleated rabbit eyes performed after 5, 30, 60, 120, 240, and 300 min. Using the 240 min readings on the four rabbit eyes evaluated with the ultrasonic pachymeter, a good correlation of $r = 0.98$ with the optical data was obtained. In addition, corneal swelling correlated well with corneal opacity scores at 240 min from Burton et al. (1981) with a Spearman rank correlation coefficient of $r = 0.91$ ($p = <0.0005$). Disadvantages of the optical pachymeter include changing refractive index by stromal swelling and a nonlinear correlation between actual and apparent (as viewed by the angle of the optical glass plate) corneal thickness. The study reported that the ultrasonic technique was a considerable improvement over the optical technique in: 1) simplicity of use, 2) short measuring time with ability to measure multiple eyes at each time point, 3) 10-fold increase in resolution, 4) wider range of corneal swelling is covered, since measurement is not hampered by corneal opacity, 5) measurement possible at all sites on corneal surface, 6) subjective aspects of optical pachymeter are not an issue with the probe tip of the ultrasonic instrument. One potential issue is damage to the epithelium by contact with the probe tip, although no adverse effects were observed in the study.

9.2.3 Proctor and Gamble (P&G) Submission from Drs. Daniel Marsman and Karen Acuff

9.2.3.1 *Summary of P&G Confocal Ocular Test Method*

The method of evaluation and scoring of the ocular toxicity of test substances used by P&G is substantially different from that used by many other investigators. The major difference is that confocal microscopy is used to determine the depth of corneal injury in addition to the area of involvement using a low volume eye test (LVET). This published methodology has been applied mainly to the testing of surfactant-based products (Jester et al. 1996; Maurer et al. 1996, 1997, 1998; Jester et al. 1998).

P&G has optimized this experimental methodology for use in the IRE (referred to as the *Ex Vivo* Rabbit Eye Test (ExRET) by P&G. P&G developed a Depth of Injury (DOI) method of evaluating the area and depth of corneal injury that is particularly important in evaluating an ocular response to surfactant-based substances. This measurement is obtained by staining the

eyes with Syto 10®, a fluorescent nucleic acid stain that penetrates cell membranes and labels all cells. Dead Red® is a cell-impermeant nucleic acid stain that labels only cells with compromised membranes. Measurement of the depth of corneal penetration is based on the depth at which no further staining of dead cells (as evidenced by dead cell staining) is observed and only live cells are present. The Normalized Depth of Injury (NDI) is the lone endpoint in the ExRET test method and is measured after 30-second exposure to the test substance using measurements in five regions of the cornea (center and four corresponding quadrants). The NDI is calculated as the mean of these five regions of the cornea divided by the overall corneal thickness (measured as the distance between the endothelial membrane and the basement membrane). The NDI is expressed as a percentage and is calculated by dividing the measured depth of injury by the overall corneal thickness and multiplying by 100. The experimental mean of NDI values for five eyes is expressed as a percentage. The final reported value is the average NDI obtained in three separate experiments. Liquid test substances are generally tested neat or may be diluted in water. One rabbit eye is treated for 30 seconds with 10 µL of D-MEM without phenol red containing 0.3% AlbuMax and 1% Dextran as a negative control. Five rabbit eyes are treated for 30 seconds with 10 µL of test substance. Two rabbit eyes are treated with the positive control for 30 seconds. Assays are conducted at room temperature and the eyes are rinsed with phosphate-buffered saline (PBS) to remove the test substance. Each test substance is tested three times for a total of 15 eyes per test substance, six eyes for the positive control and three for the negative control. A valid negative control response has an NDI = 0, and the NDI of the positive control should be within two standard deviations of the historical mean positive control.

9.2.3.2 *P&G Data*

P&G submitted data from the ExRET. Irritancy data obtained in the ExRET assay was compared to *in vivo* rabbit eye data obtained using confocal microscopy *in vivo*. In some studies, ExRET irritancy data was compared to data obtained using either conventional histopathology of LVET-treated tissues or a standard LVET *in vivo* rabbit model. Products tested included surfactants, general chemicals, surfactant-based dishwashing products and bleach-containing laundry additive products. Summarized NDI measurements and/or histopathology with predicted irritancy categories were presented in tabular and graphical form for each test substance. The data provided allowed for the development of an ExRET prediction model that contains: 1) a definition of the specific purposes for which the test was conducted; 2) definition of all possible results that may be obtained; 3) an algorithm to convert each test result into a prediction of the toxic effect of interest; and 4) the probability of the accuracy of the prediction for three irritancy categories (slight, mild/moderate, or severe).

The irritancy of anionic, nonionic and cationic surfactants as determined by confocal microscopy *in vivo* and *in vitro* is shown in **Table 9-6**. There is a general agreement in the assigned irritancy classification between *in vivo* data and the ExRET test method. The irritancy ratings assigned to three anionic, three nonionic and four cationic surfactants (including two severe irritants) by *in vivo* and ExRET test methods were in agreement. For the set of ten general substances (**Table 9-7**) tested *in vivo*, three (8% sodium hydroxide, 12% hydrogen peroxide and 15% hydrogen peroxide) were classified as severe irritants. Of these, all three were underpredicted as mild/moderate irritants by the ExRET test method.

Three substances (cyclohexanol, p-fluoroaniline, and formaldehyde) were overpredicted *in vitro*. In **Table 9-8**, the irritancy results from the LVET test method, conventional histopathology, and the ExRET test method *in vitro* are compared. The LVET irritancy ratings for three products, LDL659, LDL298, and LDL645 were based on MAS of 45.9, 50.3, and 53 and ratings of moderate, moderate and severe were assigned, respectively. Using histopathology, a level of mild/moderate was assigned to all three formulations, which was an underprediction when compared to LVET.

Table 9-6 Irritancy of Surfactant-Based Products Using P&G *In Vivo* and ExRET *In Vitro* Confocal Microscopy Test Methods

Test Substance	Conc (%)	Irritancy Rating (Confocal Microscopy Test Method)	
		In Vivo (n)[1]	ExRET *In Vitro* (n)
Anionic Surfactants			
Sodium lauryl sulfate	5	Slight (24)	Slight (50)
Sodium linear alkyl benzene sulfonate	35	Mild/Mod (43)	Mild/Mod (75)
Sodium alkyl ethoxylate sulfate	42.75	Mild/Mod (20)	Mild/Mod (90)
Nonionic Surfactants			
Polyoxyethylene glycol monoalkyl ether	100	Slight[2]	Slight (75)
Polyoxyethylene sorbitan	100	Slight[2]	Slight (75)
Alkyl E7(avg)ethoxylate	99	Mild/Mod (23)	Mild/Mod (50)
Cationic Surfactants			
3-Isotridecyloxypropyl-bis(polyoxyethylene) ammonium chloride	100	Slight (24)	Slight (75)
3-Decyloxypropyl-bis(polyoxyethylene) amine	100	Mild/Mod (6)	Mild/Mod (75)
Alkylbenzyldimethylammonium chloride	100	Severe (5)	Severe (40)
Cetyltrimethylammonium chloride	100	Severe (15)	Severe (45)
Cetyltrimethylammonium chloride	75	NT	Severe (25)
Cetyltrimethylammonium chloride	50	NT	Mild/Mod (25)
Cetyltrimethylammonium chloride	25	NT	Severe (25)
Cetyltrimethylammonium chloride	10	NT	Severe (50)

[1]Represents the total number of eyes used.
[2]n value was not available at time of submission.
NT = Not tested; Conc = Concentration; Mod = Moderate

Table 9-7 Irritancy of General Chemicals Using P&G *In Vivo* and ExRET *In Vitro* Confocal Microscopy Test Methods

Test Substance	Conc (%)	Irritancy Rating (Confocal Microscopy Test Method)	
		In Vivo (n)[1]	ExRET *In Vitro* (n)
Acid			
Acetic acid	3	Slight (26)	Mild/Mod (75)
Acetic acid	10	Mild/Mod (32)	Mild/Mod (75)
Alkali			
Sodium hydroxide	2	Slight (26)	Mild/Mod (75)
Sodium hydroxide	8	Severe (20)	Severe (75)
Bleach			
Sodium perborate monohydrate		Slight (26)	Mild/Mod (75)
Sodium hypochlorite		Slight (26)	Mild/Mod (75)
Hydrogen peroxide	6	NA	Slight (25)
Hydrogen peroxide	10	NA	Slight (75)
Hydrogen peroxide	12	NA	Severe (25)
Hydrogen peroxide	15	NA	Severe (75)
Alcohol			
Cyclohexanol		Severe (31)	Mild/Mod (75)
Aromatic amine			
p-Fluoroaniline		Severe (33)	Mild/Mod (75)
Ketone			
Acetone		Slight (55)	Slight (55)
Aldehyde			
Formaldehyde Old	37	Severe (24)	Slight (75)
Formaldehyde New[2]	37	NT	Mild/Mod (25)

[1]Represents the total number of eyes tested.
[2]Includes zone of dead cells in calculation of NDI
NA - Data was not available at time of submission.
NT = Not tested; Conc = Concentration; Mod = Moderate

Table 9-8 Irritancy of Surfactant-Based Liquid Dishwashing Formulations Using LVET and Histopathology *In Vivo* and P&G ExRET Confocal Microscopy *In Vitro* Test Method

Product Name	*In Vivo*		*In Vitro*
	LVET MAS/DTC[1]	Histopathology[2]	Ex RET (n^3)
LDL659	Moderate	Mild/Mod	Slight (75)
LDL298	Severe	Mild/Mod	Mild/Mod (75)
LDL645	Severe	Mild/Mod	Mild/Mod (75)

[1]Maximum Average Score (MAS) and Days to Clear (DTC). LDL659 had a MAS of 45.9 clearing in 7 days. LDL298 and LDL645 had MAS values of 50.3 and 53, respectively, and cleared in 21 days.
[2]Conventional histopathology
[3]Represents the total number of eyes tested (usually multiples of 15 eyes/test article from three experiments).
LVET = Low volume eye test; Mod = Moderate

However, the histopathology ratings *in vivo* were in agreement for two of the three formulations and one of the three (LDL659) was underpredicted by the ExRET test method. For bleach-containing laundry additives using the same battery of test methods (**Table 9-9**), two of the four test substances (Peroxi694 and Peroxi695) were underpredicted *in vitro*. Another substance (Hypo686) was overpredicted. In general, the ExRET test method appears to be optimized for evaluation of surfactant-based chemicals, but was not optimized for evaluation of test substances from general chemical classes or from other formulation-based product classes.

Table 9-9 Irritancy of Bleach-Containing Laundry Additive Products Using LVET *In Vivo* and P&G ExRET *In Vitro* Confocal Microscopy Test Methods

Product Name	*In Vivo*		*In Vitro*
	LVET MAS/DTC[1]	Histopathology[2]	Ex RET (n^3)
Peroxi694	Moderate	Mild/Moderate	Slight (75)
Peroxi695	Moderate	Mild/Moderate	Mild/Moderate (75)
Hypo686	Severe	Mild/Moderate	Mild/Moderate (95)
Hypo580	Severe	Mild/Moderate	Mild/Moderate (90)

[1]Maximum Average Score (MAS) and Days to Clear (DTC).
[2]Conventional histopathology
[3]Represents the total number of eyes tested (usually multiples of 15 eyes/test article from three experiments).
LVET = Low volume eye test

[This Page Intentionally Left Blank]