7.0 BCOP TEST METHOD RELIABILITY

An assessment of test method reliability (intralaboratory repeatability and intra- and inter-laboratory reproducibility) is an essential element of any evaluation of the performance of an alternative test method (ICCVAM 2003). Repeatability refers to the closeness of agreement between test results obtained within a single laboratory, when the procedure is performed on the same substance under identical conditions within a given time period (ICCVAM 1997, 2003). Intralaboratory reproducibility refers to the determination of the extent to which qualified personnel within the same laboratory can replicate results using a specific test protocol at different times. Interlaboratory reproducibility refers to the determination of the extent to which different laboratories can replicate results using the same protocol and test chemicals, and indicates the extent to which a test method can be transferred successfully among laboratories. A reliability assessment includes reviewing the rationale for selecting the substances used to evaluate test method reliability, a discussion of the extent to which the substances tested represent the range of possible test outcomes and the properties of the various substances for which the test method is proposed for use, and a quantitative and/or qualitative analysis of repeatability and intra- and inter-laboratory reproducibility. In addition, measures of central tendency and variation are summarized for historical control data (negative, vehicle, positive), where applicable.

Quantitative BCOP test method data were available for replicate corneas within individual experiments or for replicate experiments within an individual laboratory for four studies (Gettings et al. 1996; Southee 1998; data submission from Dr. Joseph Sina; data submission from Dr. Freddy Van Goethem). Therefore, an evaluation of the repeatability and/or intralaboratory reproducibility of the BCOP test method could be conducted. Additionally, comparable BCOP data were available for multiple laboratories within each of three comparative validation studies (Gautheron et al. 1994; Balls et al. 1995; Southee 1998), which allowed for an evaluation of the interlaboratory reproducibility of the BCOP test method.

7.1 Selection Rationale for the Substances Used to Evaluate the Reliability of the BCOP Test Method

The quality of a reliability evaluation depends on the extent to which the substances tested adequately represent the range of physicochemical characteristics and response levels that the test method must be capable of evaluating.

The rationale for substance selection used in the various intralaboratory and multilaboratory studies was previously discussed in **Section 3.0**. In brief, substances were selected for inclusion based on available *in vivo* rabbit eye data for comparison, to cover the range of ocular irritation potential, and to include substances with different physicochemical properties (e.g., solids, liquids).

As noted previously, the EC/HO validation study reported by Balls et al. (1995) evaluated the performance and reproducibility of the BCOP test method using 60 “substances” (i.e., there were 52 different substances with four substances tested at two different concentrations and

two substances tested at three concentrations, for a total of 60 possible ocular irritation outcomes). To be selected for inclusion in this study, the substances had to be single chemicals (no mixtures) available at high purity and stable when stored, and the reference *in vivo* rabbit eye data had to have been generated since 1981 according to OECD TG 405 following GLP guidelines. In addition, substances were selected to ensure an adequately diverse group of physicochemical characteristics and levels of irritancy severity. One substance (thiourea) was tested *in vitro* in the BCOP test method but, due to its excessive toxicity *in vivo*, was excluded from the comparison of *in vitro* and *in vivo* test results.

7.2 Analyses of Repeatability and Reproducibility

7.2.1 Assessment of Intralaboratory Repeatability and Reproducibility

Generally, analyses of intralaboratory reliability have included approaches such as:

- a coefficient of variation (CV) analysis - a statistical measure of the deviation of a variable from its mean (e.g., Holzhütter et al. 1996)
- analysis of variance (ANOVA) methods (e.g., Holzhütter et al. 1996; ASTM 1999)

Three of the studies discussed in **Section 6.0** included intralaboratory data (Gautheron et al. 1994, Gettings et al. 1996, and Southee 1998). For the Southee (1998) study, quantitative BCOP test method data were available for replicate corneas within individual experiments repeated two to five times for each test substance in three different laboratories. CV analyses were performed on within-experiment and between-experiment BCOP data, using the *In Vitro* Irritancy Score obtained for each test substance within each of the three testing laboratories. For the Gettings et al. (1996) study, Dr. John Harbell provided the mean permeability data obtained from three different experiments on the 25 surfactant-based formulations evaluated the CTFA Phase III study, as well as the mean permeability value for the three experiments, the standard deviation and the corresponding %CV values. In addition, Dr. Joseph Sina submitted a study of 43 substances, which included detailed BCOP data for replicate corneas. A CV analysis was conducted on the subset of substances provided by Dr. Sina that were tested using an incubation temperature of 32˚C, the temperature most commonly used in the BCOP for incubations as indicated in **Appendix A**; substances incubated at room temperature were not included in this analysis. For the Gautheron et al. (1994) study, Dr. Freddy Van Goethem provided individual cornea data collected in one of the participating laboratories (Janssen Pharmaceutica), which used six corneas per test substance. A %CV value was calculated for the opacity and permeability values and the *In Vitro* Irritancy Score for each test substance.

7.2.1.1 *Southee (1998)*

Intralaboratory Repeatability: In this study, 16 substances were evaluated in three laboratories multiple times (two to five experiments) for a total of 122 tests. Each test used three corneas. A %CV value was calculated for the opacity value, the permeability value, and the *In Vitro* Irritancy Score for each test (**Appendix E1**). **Tables 7-1**, **7-2**, and **7-3** summarize the mean and the %CV values of the *In Vitro* Irritancy Score for each test conducted in Laboratory 1, Laboratory 2, and Laboratory 3, respectively. The results for each laboratory are sorted by %CV values from lowest to highest value.

Table 7-1 Intralaboratory Repeatability of *In Vitro* Irritancy Scores for Replicate Corneas -- Laboratory 1, Southee 1998[1]

Substance	Mean *In Vitro* Irritancy Score (n = 3 corneas)	%CV	*In Vitro* Prediction
Benzalkonium chloride	138.0	0.1	Severe
NaOH (10%)	227.1	1.5	Severe
Benzalkonium chloride	137.9	1.6	Severe
Imidazole	142.0	2.1	Severe
Benzalkonium chloride	135.0	3.8	Severe
Imidazole	137.4	4.8	Severe
Imidazole	131.0	5.1	Severe
Benzalkonium chloride	195.0	5.8	Severe
4-Carboxybenzaldehyde	47.3	6.1	Moderate
Hexadecyltrimethylammonium bromide (10%)	20.0	6.3	Mild
4-Carboxybenzaldehyde	47.1	6.5	Moderate
Imidazole	145.7	8.3	Severe
Sodium lauryl sulfate (15%)	17.3	9.9	Mild
Glycerol	1.1	10.2	Mild
Butyl cellosolve	99.2	10.7	Severe
Methyl ethyl ketone	108.7	10.9	Severe
NaOH (10%)	245.0	11.7	Severe
Benzalkonium chloride	156.5	11.9	Severe
Ethanol	41.7	13.8	Moderate
Butyl cellosolve	92.8	14.0	Severe
Ethanol	31.5	14.2	Moderate
Ethanol	36.6	16.3	Moderate
Parafluoroaniline	38.3	19.6	Moderate
Methyl ethyl ketone	101.7	20.8	Severe
Ethanol	29.6	21.6	Moderate
Imidazole	112.0	22.0	Severe
Ammonium nitrate	5.9	23.4	Mild
Hexadecyltrimethylammonium bromide (10%)	23.1	25.3	Mild
Ethanol	37.6	28.6	Moderate
Triton X-100 (5%)	3.4	30.3	Mild
Parafluoroaniline	37.5	32.7	Moderate
Propyl-4-hydroxybenzoate	5.2	36.6	Mild
Triton X-100 (5%)	5.8	40.9	Mild
Propyl-4-hydroxybenzoate	3.6	44.1	Mild
Sodium lauryl sulfate (15%)	15.9	47.8	Mild
Ammonium nitrate	4.9	50.2	Mild
Glycerol	0.8	70.3	Mild
Tween 20	0.37	134.0	Mild
Tween 20	0.37	157.0	Mild
Sodium oxalate	-0.23	> 500	Mild
Sodium oxalate	-0.13	> 500	Mild
Mean %CV	**48.3**		
Median %CV	**14.2**		

[1]Substances organized by increasing %CV.

Table 7-2 Intralaboratory Repeatability of *In Vitro* Irritancy Scores for Replicate Corneas -- Laboratory 2, Southee 1998[1]

Substance	Mean *In Vitro* Irritancy Score (n = 3 corneas)	%CV	*In Vitro* Prediction
Benzalkonium chloride	157.9	2.1	Severe
NaOH (10%)	235.5	3.1	Severe
Benzalkonium chloride	150.8	4.7	Severe
Imidazole	137.6	4.9	Severe
Butyl cellosolve	111.8	4.9	Severe
NaOH (10%)	241.3	4.9	Severe
Sodium lauryl sulfate (15%)	5.4	5.9	Mild
Imidazole	134.9	7.0	Severe
4-Carboxybenzaldehyde	47.7	7.1	Moderate
Benzalkonium chloride	154.4	7.2	Severe
Imidazole	157.2	8.0	Severe
Ethanol	60.2	8.1	Severe
Propyl-4-hydroxybenzoate	7.4	8.3	Mild
Imidazole	140.1	8.5	Severe
Methyl ethyl ketone	67.8	8.5	Severe
4-Carboxybenzaldehyde	53.8	8.6	Moderate
Ethanol	54.2	9.1	Moderate
Imidazole	138.1	9.4	Severe
Benzalkonium chloride	157.2	11.5	Severe
Benzalkonium chloride	156.9	11.8	Severe
Butyl cellosolve	108.3	11.9	Severe
Ethanol	61.7	12.6	Severe
Sodium oxalate	10.3	13.5	Mild
Ethanol	54.5	15.1	Moderate
Parafluoroaniline	34.9	17.8	Moderate
Sodium oxalate	4.4	2.0	Mild
Methyl ethyl ketone	73.2	21.7	Severe
Parafluoroaniline	31.0	23.2	Moderate
Ethanol	52.7	24.3	Moderate
Ammonium nitrate	3.7	27.5	Mild
Triton X-100 (5%)	3.7	28.7	Mild
Propyl-4-hydroxybenzoate	11.2	28.7	Mild
Hexadecyltrimethylammonium bromide (10%)	34.7	35.0	Moderate
Hexadecyltrimethylammonium bromide (10%)	39.2	41.8	Moderate
Tween 20	0.3	45.8	Mild
Ammonium nitrate	3.9	46.4	Mild
Sodium lauryl sulfate (15%)	5.2	52.3	Mild
Triton X-100 (5%)	1.8	53.0	Mild
Glycerol	0.5	108.0	Mild
Glycerol	0.27	356.0	Mild
Tween 20	0.1	> 500	Mild
Mean %CV	**39.2**		
Median %CV	**11.8**		

[1]Substances organized by increasing %CV.

Table 7-3 Intralaboratory Repeatability of *In Vitro* Irritancy Scores for Replicate Corneas -- Laboratory 3, Southee 1998[1]

Substance	Mean *In Vitro* Irritancy Score (n = 3 corneas)	%CV	*In Vitro* Prediction
Ethanol	45.4	4.3	Moderate
Methyl ethyl ketone	70.3	5.1	Severe
Benzalkonium chloride	151.6	5.1	Severe
Imidazole	124.0	5.5	Severe
Benzalkonium chloride	169.7	6.0	Severe
Imidazole	128.7	6.3	Severe
Ethanol	44.4	6.7	Severe
Benzalkonium chloride	162.8	7.0	Severe
NaOH (10%)	214.8	7.2	Severe
Hexadecyltrimethylammonium bromide (10%)	31.7	7.3	Moderate
Ethanol	54.6	8.2	Moderate
Methyl ethyl ketone	73.5	8.7	Severe
4-Carboxybenzaldehyde	41.8	9.4	Moderate
NaOH (10%)	193.1	9.9	Severe
Benzalkonium chloride	163.4	9.9	Severe
4-Carboxybenzaldehyde	42.2	10.2	Moderate
Benzalkonium chloride	156.9	10.9	Severe
Propyl-4-hydroxybenzoate	6.2	11.8	Mild
Imidazole	123.4	12.0	Severe
Parafluoroaniline	22.1	12.0	Moderate
Ammonium nitrate	5.2	12.4	Mild
Parafluoroaniline	25.9	13.0	Moderate
Imidazole	140.2	13.5	Severe
Butyl cellosolve	94.9	14.5	Severe
Sodium lauryl sulfate (15%)	8.4	16.1	Mild
Ethanol	45.7	18.6	Moderate
Imidazole	139.6	18.6	Severe
Ammonium nitrate	6.7	21.6	Mild
Glycerol	0.8	21.7	Mild
Butyl cellosolve	98.2	22.0	Severe
Sodium lauryl sulfate (15%)	5.6	26.7	Mild
Sodium oxalate	4.6	28.5	Moderate
Ethanol	47.0	30.3	Severe
Sodium oxalate	2.7	33.0	Moderate
Triton X-100 (5%)	1.9	34.4	Mild
Hexadecyltrimethylammonium bromide (10%)	29.9	37.3	Moderate
Triton X-100 (5%)	3.0	37.9	Mild
Propyl-4-hydroxybenzoate	7.7	53.7	Mild
Glycerol	1.0	57.0	Mild
Tween 20	0.3	75.5	Mild
Tween 20	0.0	> 500	Mild
Mean %CV	**30.5**		
Median %CV	**12.4**		

[1]Substances organized by increasing %CV.

The ranges of %CV values for substances classified as severe irritants *in vitro* are 0.1 to 22.0 for Laboratory 1, 2.1 to 21.7 for Laboratory 2, and 5.1 to 30.3 for Laboratory 3. The within experiment mean and median %CV values for the three laboratories for all substances ranged from 30.5 to 48.3 and 11.8 to 14.2, respectively (%CV values listed as >500 were set at 500). Substances classified *in vitro* as mild irritants (i.e., *In Vitro* Irritancy Score >25) tended to have greater %CV values. The three laboratories all had at least one, but not more than two, %CV values greater than 500, which resulted from substances that had *In Vitro* irritancy Scores at or below the accepted background score of 3 to 5.

Intralaboratory Reproducibility: The between experiment %CV values of *In Vitro* Irritancy Scores for substances tested two or more times in Laboratory 1, Laboratory 2, and Laboratory 3 are presented in **Tables 7-4**, **7-5**, and **7-6**, respectively. The mean %CV values ranged from 12.6 to 14.8 for the three laboratories, while the median %CV values ranged from 6.7 to 12.4.

Table 7-4 Intralaboratory Reproducibility of Substances Tested in Multiple Experiments in Laboratory 1, Southee 1998[1]

Substance	Mean *In Vitro* Irritancy Score	No. of Exp.	%CV	*In Vitro* Prediction
Tween 20	0.37	2	0	Mild
4-Carboxybenzaldehyde	47.2	2	0.3	Moderate
Parafluoroaniline	37.9	2	1.6	Moderate
Butyl cellosolve	96	2	4.7	Severe
Methyl ethyl ketone	105	2	4.7	Severe
Ethanol	35.4	5	4.9	Moderate
NaOH (10%)	236	2	5.3	Severe
Sodium lauryl sulfate (15%)	16.6	2	6.1	Mild
Imidazole	133.7	5	9.9	Severe
Hexadecyltrimethylammonium bromide (10%)	21.6	2	10.1	Mild
Ammonium nitrate	5.4	2	13.45	Mild
Benzalkonium chloride	141.9	5	17.83	Severe
Glycerol	0.98	2	21.8	Mild
Propyl-4-hydroxybenzoate	4.4	2	25.7	Mild
Triton X-100 (5%)	4.6	2	36.7	Mild
Sodium oxalate	-0.07	2	39.3	Mild
Mean %CV	**12.6**			
Median %CV	**8.0**			

[1]Substances organized by increasing %CV.

Table 7-5 Intralaboratory Reproducibility of Substances Tested in Multiple Experiments in Laboratory 2, Southee 1998[1]

Substance	Mean *In Vitro* Irritancy Score	No. of Exp.	%CV	*In Vitro* Prediction
NaOH (10%)	238.4	2	1.7	Severe
Benzalkonium chloride	155	5	1.9	Severe
Butyl cellosolve	110	2	2.2	Severe
Sodium lauryl sulfate (15%)	5.3	2	3.1	Mild
Ammonium nitrate	3.8	2	4.3	Mild
Glycerol	0.52	2	4.5	Mild
Methyl ethyl ketone	70.5	2	5.5	Severe
Imidazole	141.6	5	6.3	Severe
Ethanol	56.7	5	7.1	Severe
4-Carboxybenzaldehyde	50.8	2	8.5	Moderate
Parafluoroaniline	32.9	2	8.5	Moderate
Hexadecyltrimethylammonium bromide (10%)	36.9	2	8.6	Moderate
Propyl-4-hydroxybenzoate	9.3	2	29.3	Mild
Tween 20	0.47	2	40.5	Mild
Triton X-100 (5%)	2.7	2	48	Mild
Sodium oxalate	7.4	2	56.4	Mild
Mean %CV			**14.8**	
Median %CV			**6.7**	

[1]Substances organized by increasing %CV.

Table 7-6 Intralaboratory Reproducibility of Substances Tested in Multiple Experiments in Laboratory 3, Southee 1998[1]

Substance	Mean *In Vitro* Irritancy Score	No. of Exp.	%CV	*In Vitro* Prediction
4-Carboxybenzaldehyde	42	2	0.57	Moderate
Butyl cellosolve	96.5	2	2.4	Severe
Methyl ethyl ketone	71.9	2	3.2	Severe
Benzalkonium chloride	161	5	4.2	Severe
Imidazole	131.2	5	6.3	Severe
NaOH (10%)	203.9	2	7.5	Severe
Ethanol	47.4	5	8.6	Moderate
Parafluoroaniline	24	2	11.3	Mild
Glycerol	0.88	2	13.4	Mild
Hexadecyltrimethylammonium bromide (10%)	33.3	2	14.2	Moderate
Propyl-4-hydroxybenzoate	7	2	15.2	Mild
Ammonium nitrate	5.9	2	18.6	Mild
Tween 20	0.4	2	23.7	Mild
Sodium lauryl sulfate (15%)	7	2	28.5	Mild
Triton X-100 (5%)	2.5	2	31.4	Mild
Sodium oxalate	3.65	2	35.6	Mild
Mean %CV			**14.0**	
Median %CV			**12.4**	

[1]Substances organized by increasing %CV.

7.2.1.2 □ *Data from Dr. Joseph Sina (Merck)*

Intralaboratory Repeatability: In this study, 43 substances were tested in one laboratory using four corneas per test substance. A %CV value was calculated for the opacity and permeability values and the *In Vitro* Irritancy Score for each test substance (**Appendix E2**). However, only 29 of the test substances were evaluated using a protocol that incubated the corneas at 32˚C. The %CVs for the *In Vitro* Irritancy Scores of these 29 substances are shown in **Table 7-7**. The results are sorted by %CV from lowest to highest value. The ranges of %CV values for substances classified as severe irritants *in vitro* are 1.1 to 13 (n = 5). The within experiment mean and median %CV values for this study were 71 and 35%, respectively. Substances classified *in vitro* as mild irritants tended to have greater %CV values. A majority (21 of 29; 72%) of the test substances in this study were classified as mild irritants *in vitro* and, of these, 10 had *In Vitro* Irritancy Scores at or below the accepted the background score of 3 to 5, contributing to higher within experiment mean and median %CV values for this study in comparison with the Southee (1998) study, which included test substances with a greater range of irritancy.

7.2.1.3 □ *Data from Dr. John Harbell (IIVS) for Gettings et al. (1996)*

Intralaboratory Reproducibility: Dr. John Harbell provided permeability values (OD_{490}) for three replicate experiments performed in an individual laboratory for the 25 surfactant-based personal care cleaning formulations evaluated in Gettings et al. (1996). The mean permeability value of these three experiments, as well as the mean and %CV of these data also were provided. All of these data and statistics are shown in **Table 7-8**. The results are sorted by %CV from lowest to highest value. The between experiment mean and median %CV values for this study were 33.4 and 29, respectively, with a %CV range of 5% to 100%.

7.2.1.4 □ *Data from Dr. Freddy Van Goethem for Gautheron et al. 1994)*

Intralaboratory Repeatability: In this study, 52 substances were tested in 11-12 different laboratories. Dr. Freddy Van Goethem provided individual cornea data collected in one of the participating laboratories (Janssen Pharmaceutica), which used six corneas per test substance. A %CV value was calculated for the opacity and permeability values and the *In Vitro* Irritancy Score for each test substance (**Appendix E3**). The %CVs for the *In Vitro* Irritancy Scores of the 52 substances tested are shown in **Table 7-9**. The results are sorted by %CV from lowest to highest value. The ranges of %CV values for substances classified as severe irritants *in vitro* are 1.4 to 24.3 (n = 20). The within experiment mean and median %CV values for this study were 47% and 18%, respectively. Substances classified *in vitro* as mild irritants tended to have greater %CV values (ranging from 11.3% to 312.6% [n = 27]). These results were comparable to those obtained in the intralaboratory repeatability analysis of the BCOP data from Southee (1998) (see **Section 7.2.1.1**).

Table 7-7 Intralaboratory Repeatability of *In Vitro* Irritancy Scores for Replicate Corneas -- Laboratory 4 (Dr. Sina, Merck)[1]

Substance	Mean *In Vitro* Irritancy Score (n = 4 corneas)	%CV	*In Vitro* Prediction
3-Trichlorovinylaniline HCL	404	1.1	Severe
2-Amino-3,6-dimethylphenol, hydrobromide salt	150	7.1	Severe
Carbic anhydride	202	8.1	Severe
1,3-Benzenedicarboxaldehyde	29.8	10.9	Moderate
4-Bromo-2,5-dimethylphenol	131	11.7	Severe
Methyl 3-oxo-6-methoxyhexanoate	57.8	13	Severe
R-Hydroxy ester of benzoic acid compound	-12	14	Mild
Quinaldine (2-methylquinoline)	25.5	19	Moderate
Mixture of 2-chloromethyl-4,7-dimethylbenzoxazole and 2-bromomethyl dimethylbenzoxazole	18.1	19.1	Mild
Carbonitrile	21.8	21.5	Mild
Methyl boronic acid	25.1	26.6	Moderate
alpha-Pyranol, 7,7-dioxide	31.5	27.7	Mild
7-Chloroquinaldine	10.6	28.4	Mild
+-Butyl-3R-hydroxy-6-methoxyhepanoate	22.8	28.8	Mild
Cyano methylpyridine	15.5	34	Mild
Cyclic peptide	7.9	36.9	Mild
Substituted cephalosporanic acid	-4.4	40	Mild
S-Hydroxy ester of benzoic acid compound	20.8	42	Mild
t-Butyl-3-oxo-6-methoxyhexanoate	15.3	49	Mild
Aglycone; natural product	11.3	52.4	Mild
N-Acetyl-*p*-anisidine	8.38	58.7	Mild
Cyanopyridinone	-4.3	64	Mild
N-Sulfonamido hydroxyacetophenone	-5.8	117	Mild
Nitropyridinone	-3.7	124	Mild
3-Bromo-7-methyl-9-flurenone	-2.6	140	Mild
Cyclic peptide	2.7	175	Mild
Dimethyl ethylimidazo pyridine	3.36	200	Mild
tert-Butyl-6-methoxy-3-S-(2-thiophenethio) hexanoate	1.5	221	Mild
4-(2-Quinolylmethoxy)aniline	2.8	479	Mild
Mean %CV	**71**		
Median %CV	**35**		

[1]Substances organized by increasing %CV.

Table 7-8 Intralaboratory Reproducibility of Substances Tested in Multiple Experiments in Laboratory 5, Microbiological Associates[1]

Formulation	Permeability – O.D. units					
	Exp. 1	Exp. 2	Exp. 3	Mean	SD	%CV
Skin Cleaner - HZI	0.782	0.728	0.796	0.77	0.04	5
Shower Gel - HZS	1.488	1.501	1.655	1.55	0.09	6
Facial Cl Foam - HZR	0.215	0.244	0.259	0.24	0.02	9
Liquid Soap 1 - HZB	0.198	0.176	0.223	0.20	0.02	12
Shampoo 4 - HZV	0.306	0.219	0.279	0.27	0.04	17
Baby Shampoo 2 - HZF	0.505	0.342	0.427	0.42	0.08	19
Baby Shampoo 1 - HZP	0.285	0.202	0.296	0.26	0.05	20
Shampoo 3 - HZM	0.229	0.254	0.16	0.21	0.05	23
Shampoo AntiD - HZY	0.756	0.709	1.075	0.85	0.20	24
Gel Cleaner - HZE	0.186	0.15	0.246	0.19	0.05	25
Shampoo 6 - HZN	0.283	0.184	0.333	0.27	0.08	28
Liquid Soap 2 - HZW	0.356	0.21	0.417	0.35	0.10	28
Shampoo 8 - HZG	0.22	0.131	0.24	0.20	0.06	29
Foam Bath - HZL	0.625	0.976	1.136	0.91	0.26	29
Cleaning Gel – HZQ	0.214	0.114	0.165	0.16	0.05	30
Hand Soap - HZU	0.348	0.187	0.344	0.29	0.09	31
Shampoo 1 - HZC	1.193	0.612	1.067	0.96	0.31	32
Bubble bath - HZK	1.33	0.753	0.785	0.96	0.32	34
Shampoo 5 - HZD	0.318	0.15	0.225	0.23	0.08	36
Shampoo 7 - HZA	0.562	0.406	0.251	0.41	0.16	38
Shampoo 2 - HZX	0.582	0.498	1.036	0.71	0.29	41
Mild Shampoo - HZJ	0.064	0.021	0.066	0.05	0.03	51
Eye Makeup Remover - HZH	0.029	0.001	0.029	0.02	0.02	82
Polishing Scrub - HZT	0.002	0	0.002	0.001	0.00	87
Facial Cleaner - HZZ	0.008	0.004	0	0.004	0.00	100
Mean %CV	**33.4**					
Median %CV	**29.0**					

[1]Substances organized by increasing %CV.

Table 7-9 Intralaboratory Repeatability of *In Vitro* Irritancy Scores for Replicate Corneas -- Laboratory 9 (Gautheron et al. 1994)[1]

Substance	Mean *In Vitro* Irritancy Score (n = 6 corneas)	%CV	*In Vitro* Prediction
2-Ethoxyethanol	84.4	1.4%	Severe
Cyclohexanone	141.7	5.8%	Severe
Gluconolactone	87.5	6.0%	Severe
2,4-Pentanedione	50.3	6.8%	Moderate
Promethazine hydrochloride	139.2	7.3%	Severe
Furan	50.2	7.9%	Moderate
Deoxycholic acid, sodium salt	99.6	8.0%	Severe
Benzethonium chloride	165.9	8.8%	Severe
Hexadecyltrimethylammonium bromide	69.9	9.9%	Severe
Quinacrine	57.9	10.0%	Severe
Octanol	60.9	11.2%	Severe
1-Nitropropane	16.6	11.3%	Mild
N-Lauroylsarcosine, sodium salt	62.6	11.6%	Severe
Allyl alcohol	123.3	11.7%	Severe
Butyrolactone	41.6	12.0%	Moderate
1-Phenyl-3-pyrazolidone	13.2	12.4%	Mild
Methanol	99.2	12.9%	Severe
Thiourea	151.4	13.7%	Severe
Ethanol	45.7	14.3%	Moderate
Dimethyl sulfoxide	9.4	14.4%	Mild
Ethyl acetoacetate	25.7	14.8%	Moderate
Pyridine	104.7	15.0%	Severe
2-Methoxyethanol	57.1	15.1%	Severe
Methylisobutyl ketone	19.4	15.9%	Mild
Dibenzoyl-L-tartaric acid	81.5	16.8%	Severe
Imidazole	64.3	17.3%	Severe
2-Aminophenol	13.0	19.0%	Mild
1,2,4-Trimethylbenzene	21.2	21.2%	Mild
1,2,3-Trichloropropane	91.1	22.0%	Severe
Aluminum hydroxide	9.9	23.2%	Mild
Diacetone alcohol	92.9	23.7%	Severe
Propyl-4-hydroxybenzoate	6.2	24.0%	Mild
Laurylsulfobetaine	102.4	24.3%	Severe
2,4-Dichloro-5-sulfamoylbenzoic acid	19.2	24.7%	Mild
3-Glycidoxypropyltrimethoxysilane	17.6	26.7%	Mild
Triethanolamine	3.0	34.5%	Mild
Sodium oxalate	3.2	40.9%	Mild
Triton X-155	3.1	53.3%	Mild
Tetraaminopyrimidine sulfate	2.5	54.7%	Mild
BRIJ-35	1.0	61.7%	Mild
EDTA, dipotassium salt	0.9	63.1%	Mild
Betaine monohydrate	3.5	63.7%	Mild
Magnesium carbonate	0.7	71.4%	Mild
Phenylbutazone	0.5	80.1%	Mild
Anthracene	1.4	87.4%	Mild
Petroleum ether	2.1	91.4%	Mild
Dimethylbiguanide	2.1	124.6%	Mild
Hexane	1.4	128.3%	Mild

Substance	Mean *In Vitro* Irritancy Score (n = 6 corneas)	%CV	*In Vitro* Prediction
2-Mercaptopyrimidine	-0.2	167.3%	Mild
DL-Glutamic acid	-0.2	221.3%	Mild
Iminodibenzyl	0.2	278.9%	Mild
MYRJ-45	0.5	312.6%	Mild
Mean %CV	**46.8%**		
Median %CV	**18.1%**		

[1]Substances organized by increasing %CV

7.2.2 Evaluation of Interlaboratory Reproducibility

Generally, analyses of interlaboratory variability have included approaches such as:

- the extent of concordance among laboratories in assigning the same regulatory classification for a particular substance (e.g., Holzhütter et al. 1996)
- bivariant scatter diagrams/correlation analyses for pairs of laboratories to assess the extent possibility of divergence (e.g., Holzhütter et al. 1996)
- a CV analysis (e.g., Holzhütter et al. 1996)
- analysis of variance (ANOVA) methods (e.g., Holzhütter et al. 1996; ASTM 1999)

Several of the studies discussed in **Section 6.0** included interlaboratory data for at least a subset of the substances evaluated. The ability of the BCOP test method to reproducibly identify ocular corrosives/severe irritants versus nonsevere irritants/nonirritants was evaluated using two approaches.

In the first approach, a qualitative assessment of reproducibility was conducted. In this evaluation, the individual laboratory *in vitro* ocular irritation classification for each substance was used to evaluate the extent of agreement among the participating laboratories in their ability to identify ocular corrosives/severe irritants versus nonsevere irritants/nonirritants. The reliability of BCOP was assessed separately for each study (i.e., publication) reviewed in **Sections 4.0** and **5.0**. In an alternative approach, the reliability of BCOP was assessed after combining test results across comparative studies that used the same data analysis method (i.e., use of *In Vitro* Irritancy Score). **Section 6.0** provides a further description of how data were treated for each type of analysis. Substances classified, based on BCOP data, as corrosive/severe irritants or nonsevere irritants/nonirritants were further classified by their *in vivo* rabbit eye test results, as determined within the GHS, EPA, and EU classification schemes. Because the focus of this reliability assessment is on the interlaboratory reproducibility of BCOP in identifying corrosives/severe irritants versus nonsevere irritants/nonirritants, considerable variability could exist among laboratories in their classification of substances as nonsevere irritants or nonirritants (e.g., three laboratories could classify a chemical as a nonirritant and one laboratory could classify the same chemical as an moderate irritant; for this analysis this would be considered 100% agreement between laboratories) that would not be apparent from this analysis.

In the second approach, a quantitative assessment of reproducibility was determined by calculating the CV for test substance data for which *In Vitro* Irritancy Scores were available from multiple laboratories. The reproducibility of BCOP was assessed for the studies (i.e.,

publication) reviewed in **Sections 4.0** and **5.0** where individual testing laboratory data were available.

7.2.2.1 *Interlaboratory Reproducibility of Hazard Classification Category Using the GHS Classification System*

Reliability analyses for the BCOP test method were evaluated for the following three studies: Balls et al. (1995), Gautheron et al. (1994), and Southee (1998). The agreement of classification calls among participating laboratories and the relationship to the *in vivo* classification (GHS; UN 2003) for the substances tested in each validation in each study is provided in **Table 7-10**.

For the study by Balls et al. (1995), the five participating laboratories were in 100% agreement in regard to the ocular irritancy classification for 41 (68%) of the 60 substances tested. The extent of agreement between testing laboratories was greatest for substances identified from *in vivo* rabbit eye data as corrosives or severe irritants when compared to any other combination of *in vivo* and *in vitro* results (76% of the accurately identified severe substances were shown to have 100% classification agreement among testing laboratories). Comparatively, greater disparity between individual substance classifications was observed for substances that were identified as false positives (i.e., positive *in vitro* but negative *in vivo*). For instance, 63% (36% + 27%) of the false positives exhibited less than 100% agreement in the irritancy classifications among laboratories.

For the study by Gautheron et al. (1994), there was 100% agreement in regard to the ocular irritancy classification for 35 (69%) of the 51 substances, which were tested in either 11 or 12 laboratories. Discordance in the classification results was present for substances that were correctly identified as corrosives/severe irritants and as nonsevere irritants/nonirritants.

For the study by Southee (1998), there was 100% agreement in regard to the ocular irritancy classification for 15 (94%) of the 16 substances. Discordance in the classification results was present for only one substance that was correctly identified as a nonsevere irritant/nonirritant.

7.2.2.2 *Interlaboratory Reproducibility of Hazard Classification Category Using the EPA Classification System*

Reliability analyses for the BCOP test method were evaluated for the following three studies: Balls et al. (1995), Gautheron et al. (1994), and Southee (1998). The agreement of classification calls among participating laboratories and its relationship to the *in vivo* classification (EPA 1996) for the substances tested in each validation in each study is provided in **Table 7-11**.

Table 7-10 Evaluation of the Reliability of the BCOP Test Method in Predicting Ocular Corrosives and Severe Irritants as Defined by the GHS Classification System, by Study

Report	Classification (*In Vivo/In Vitro*)[1]	No. of Testing Labs	n[2]	Substances with 100% Agreement among Labs[3]	Substances with 91-92% Agreement among Labs	Substances with 82-83% Agreement among Labs	Substances with 80% Agreement among Labs	Substances with 73% Agreement among Labs	Substances with 64-67% Agreement among Labs	Substances with 58-60% Agreement among Labs	Substances with ≤55% Agreement among Labs
Balls et al. (1995)	+/+	5	17	13 (76%)			3 (18%)			1 (6%)	
	+/-	5	5	3 (60%)			1 (20%)			1 (20%)	
	-/+	5	11	4 (36%)			4 (36%)			3 (27%)	
	-/-	5	21	16 (76%)			2 (10%)			3 (14%)	
	?/-	5	4	3 (75%)						1 (25%)	
	?/+	5	2	2 (100%)							
	Total		60	41 (68%)			10 (17%)			9 (15%)	
Gautheron et al. (1994)	+/+	11 12	5 1	3 (60%) 1 (100%)		1 (20%)					1 (20%)
	+/-	11 12	1 1	1 (100%)		1 (100%)					
	-/+	11 12	4 5	2 (50%) 2 (40%)	1 (20%)	1 (25%)		1 (25%)			2 (40%)
	-/-	11 12	15 15	12 (80%) 13 (86%)	1 (7%)	2 (13%) 1(7%)			1 (7%)		
	?/-	11 12	1 1	1(100%)				1 (100%)			
	?/+	11	2		1 (50%)				1 (50%)		
	Total		51	35 (69%)	3 (6%)	6 (12%)		2 (4%)	2 (4%)		3 (6%)
Southee (1998)	+/+	3	4	4 (100%)							
	+/-	3	3	3 (100%)							
	-/+	3	1	1 (100%)							
	-/-	3	7	6 (86%)					1 (14%)		
	?/-	3	1	1 (100%)							
	?/+	-	0								
	Total		16	15 (94%)					1 (6%)		

[1]A "+" indicates that the substance was assigned an overall classification of corrosive or a severe irritant (Category 1); a "-" indicates that the substance was assigned an overall classification of nonsevere irritant (Category 2A, 2B) or nonirritant; a "?" indicates that, due to the lack of appropriate *in vivo* data (e.g., studies were terminated too early to assess reversibility of effects; insufficient dose volume), a GHS classification could not be made. See **Section 6.1** for a description of the rules followed to classify the ocular irritancy of test substances tested multiple times *in vitro*.

[2]n indicates number of substances.

[3]Number in parentheses indicates percentage of tested chemicals.

Table 7-11 Evaluation of the Reliability of the BCOP Test Method In Predicting Ocular Corrosives and Severe Irritants as Defined by the EPA Classification System, by Study

Report	Classification (*In Vivo/In Vitro*)[1]	No. of Testing Labs	n[2]	Substances with 100% Agreement among Labs[3]	Substances with 91-92% Agreement among Labs	Substances with 82-83% Agreement among Labs	Substances with 80% Agreement among Labs	Substances with 73% Agreement among Labs	Substances with 64-67% Agreement among Labs	Substances with 58-60% Agreement among Labs	Substances with≤ 55% Agreement among Labs
Balls et al. (1995)	+/+	5	13	10 (77%)			2 (15%)			1 (8%)	
	+/-	5	5	3 (60%)			1 (20%)			1 (20%)	
	-/+	5	13	5 (38%)			5 (38%)			3 (23%)	
	-/-	5	22	15 (68%)			4 (18%)			3 (14%)	
	?/-	5	3	3 (100%)							
	?/+	5	4	4 (100%)							
	Total		60	40 (67%)			12 (20%)			8 (13%)	
Gautheron et al. (1994)	+/+	11 12	4 1	2 (50%) 1 (100%)		1 (25%)					1 (25%)
	+/-	11 12	1 1	 1 (100%)		1 (100%)					
	-/+	11 12	6 5	3 (50%) 2 (40%)	 1 (20%)	1 (17%)		1 (17%)	1 (17%)	 1 (20%)	 1 (20%)
	-/-	11 12	15 15	12 (80%) 13 (86%)	 1 (7%)	2 (13%) 1 (7%)			1 (7%)		
	?/-	11 12	1 1	 1 (100%)				1 (100%)			
	?/+	11	1		1 (100%)						
	Total		51	35 (69%)	3 (6%)	6 (12%)		2 (4%)	2 (4%)	1 (2%)	2 (4%)
Southee (1998)	+/+	3	2	2 (100%)							
	+/-	3	3	3 (100%)							
	-/+	3	2	2 (100%)							
	-/-	3	7	6 (86%)					1 (14%)		
	?/-	3	1	1 (100%)							
	?/+	3	1	1 (100%)							
	Total		16	15 (94%)					1 (6%)		

[1]A "+" indicates that the substance was assigned an overall classification of corrosive or a severe irritant (Category I); a "-" indicates that the substance was assigned an overall classification of nonsevere irritant (Category II, III) or nonirritant (category IV); a "?" indicates that, due to the lack of appropriate *in vivo* data (e.g., studies were terminated too early to assess reversibility of effects; insufficient dose volume), an EPA classification could not be made. See **Section 6.1** for a description of the rules followed to classify the ocular irritancy of test substances tested multiple times *in vitro*.

[2]n indicates number of substances.

[3]Number in parentheses indicates percentage of tested chemicals.

The participating laboratories of Balls et al. (1995) were in 100% agreement in regard to the ocular irritancy classification for 40 (67%) of the 60 substances tested. The agreement among laboratories was greatest for accurately identified corrosives/severe irritants when compared to any other combination of *in vivo* and *in vitro* results (77% of the accurately identified corrosives/severe irritants exhibited 100% classification agreement among laboratories). Comparatively, greater disparity between individual substance classifications was observed for substances that were identified as false positives. For instance, 61% (38% + 23%) of the false positives exhibited less than 100% agreement among laboratories in the irritancy classifications.

The participating laboratories of Gautheron et al. (1994) were in 100% agreement in regard to the ocular irritancy classification (corrosive/severe irritant or nonsevere irritant/nonirritant) for 35 (69%) of the 51 tested substances. Discordant results were observed for substances that were correctly identified as corrosive/severe irritant or nonsevere/irritant/nonirritant, as well as for false negatives and false positives.

For the report by Southee (1998), there was 100% agreement in regard to the ocular irritancy classification (corrosive/severe irritant or nonsevere irritant/nonirritant) for 15 (94%) of the 16 substances. Discordance in the classification results was present for only one substance that was correctly identified as a nonsevere irritant/nonirritant.

7.2.2.3 *Interlaboratory Reproducibility of Hazard Classification Category Using the EU Classification System*

Reliability analyses for the BCOP test method were evaluated for the following three studies: Balls et al. (1995), Gautheron et al. (1994), and Southee (1998). The agreement of classification calls among participating laboratories and its relationship to the *in vivo* classification (EU 2001) for the substances tested in each validation in each study is provided in **Table 7-12**.

The participating laboratories were in 100% agreement in regard to the ocular irritancy classification for 41 (68%) of the 60 substances tested by Balls et al. (1995). The extent of agreement among laboratories was greatest for accurately identified corrosives/severe irritants when compared to any other combination of *in vivo* and *in vitro* results (86% of the accurately identified corrosives/severe irritants exhibited 100% classification agreement among laboratories). Comparatively, greater disparity between individual substance classifications was observed for substances that were identified as false positives, false negatives, and those substances accurately classified as nonsevere irritants/nonirritants. For instance, 63% (36% + 27%) of the false positives, 60% (20% + 40%) of the false negatives and 25% (10% + 15%) of the correctly identified nonsevere irritants/nonirritants exhibited less than 100% agreement among laboratories in irritancy classifications.

The participating laboratories in Gautheron et al. (1994) were in 100% agreement in regard to the ocular irritancy classification for 35 (69%) of the 51 tested substances. Substances that were classified as false positives exhibited the most discordant results, with 60% (20% + 20% + 20%) of false positives exhibiting less than 100% classification agreement among laboratories.

Table 7-12 Evaluation of the Reliability of the BCOP Test Method In Predicting Ocular Corrosives and Severe Irritants (as Defined by the EU Classification System), by Study

Report	Classification (*In Vivo/In Vitro*)[1]	No. of Testing Labs	n[2]	Substances with 100% Agreement among Labs[3]	Substances with 91-92% Agreement among Labs	Substances with 82-83% Agreement among Labs	Substances with 80% Agreement among Labs	Substances with 73% Agreement among Labs	Substances with 64-67% Agreement among Labs	Substances with 58-60% Agreement among Labs	Substances with ≤55% Agreement among Labs
Balls et al. (1995)	+/+	5	14	12 (86%)			2 (14%)				
	+/-	5	5	2 (40%)			1 (20%)			2 (40%)	
	-/+	5	11	4 (36%)			4 (36%)			3 (27%)	
	-/-	5	20	15 (75%)			2 (10%)			3 (15%)	
	?/-	5	5	5 (100%)							
	?/+	5	5	3 (60%)			1 (20%)			1 (20%)	
	Total		60	41 (68%)			10 (17%)			9 (15%)	
Gautheron et al. (1994)	+/+	11 12	5 1	3 (60%) 1 (100%)		1 (20%)					1 (20%)
	+/-	11 12	1 1	 1 (100%)		1 (100%)					
	-/+	11 12	5 5	2 (40%) 2 (40%)	 1 (20%)	1 (20%)		1 (20%)	1 (20%)	 1 (20%)	 1 (20%)
	-/-	11 12	15 15	12 (80%) 13 (86%)	 1 (7%)	2 (13%) 1 (7%)			1 (7%)		
	?/-	11 12	1 1	 1 (100%)				1 (100%)			
	?/+	11	1		1 (100%)						
	Total		51	35 (69%)	3 (6%)	6 (12%)		2 (4%)	2 (4%)	1 (2%)	2 (4%)
Southee (1998)	+/+	3	4	4 (100%)							
	+/-	3	2	2 (100%)							
	-/+	3	1	1 (100%)							
	-/-	3	7	6 (86%)					1 (14%)		
	?/-	3	2	2 (100%)							
	?/+	-	0								
	Total		16	15 (94%)					1 (6%)		

[1]A "+" indicates that the substance was assigned an overall classification of corrosive or a severe irritant (Category I); a "-" indicates that the substance was assigned an overall classification of nonsevere irritant (Category II, III) or nonirritant (category IV); a "?" indicates that, due to the lack of appropriate *in vivo* data (e.g., studies were terminated too early to assess reversibility of effects; insufficient dose volume), an EPA classification could not be made. See **Section 6.1** for a description of the rules followed to classify the ocular irritancy of test substances tested multiple times *in vitro*.

[2]n indicates number of substances.

[3]Number in parentheses indicates percentage of tested chemicals.

For the study by Southee (1998), there was 100% agreement in regard to the ocular irritancy classification for 15 (94%) of the 16 substances. Discordance in the classification results was present for only one substance that was correctly identified as a nonsevere irritant/nonirritant.

7.2.2.4 *Common Chemical or Product Classes Among Test Substances with Discordant Interlaboratory Results*

For the Gautheron et al. (1994) study, 16 substances showed interlaboratory differences in *in vitro* classification (**Table 7-13**). Of these, nine (56%) are organic solvents, including five alcohols, one lactone, one ketone, one heterocyclic compound, and one chlorinated hydrocarbon. Four surfactants, four heterocyclic compounds (two solids and two liquids), and one acid (a solid) also showed interlaboratory differences in *in vitro* classification. Of the 10 liquid substances that produced discordant interlaboratory results in this study, nine are organic solvents.

For the Balls et al. (1995) study, 19 substances showed interlaboratory differences in *in vitro* classification (**Table 7-14**). Of these, 10 (53%) are organic solvents, including seven alcohols, one lactone, one ketone, and one ester. Two ethers, two carboxylic acids, two imides (solid), and one amidine also showed interlaboratory differences in *in vitro* classification. The ten liquid substances that produced discordant interlaboratory results in this study are all organic solvents.

7.2.2.5 *Interlaboratory Reproducibility Based on Coefficient of Variation Analysis of In Vitro Scores*

To provide a quantitative assessment of interlaboratory variability, individual laboratory BCOP test results were used to calculate a mean and CV for the *In Vitro* Irritancy Score for each substance tested in Gautheron et al. (1994), Balls et al. (1995) and Southee (1998) (**Tables 7-15, 7-16, 7-17**).

For the Gautheron et al. (1994) study, a wide range of %CV values for individual substances is evident for the *In Vitro* Irritancy Score (**Table 7-15**). The mean and median %CV values were 168% and 47%, respectively, ranging from 16.5% to 1325% for the entire set of 52 test substances. The 17 substances predicted as severe in the BCOP assay had mean and median %CV values of 36% and 17%, respectively, with a %CV range from 16.5% to 55.7%. Substances classified *in vitro* as mild irritants (i.e., *In Vitro* Irritancy Score < 25) tended to have much greater %CV values. About half (25 of 52; 48%) of the substances tested in this study were classified as mild irritants *in vitro* and, of these, 18 had *In Vitro* Irritancy Scores at or below the accepted the background score of 3 to 5, contributing to a high mean and median %CV for this study. All of the %CV values for individual substances greater than 75% (n = 17) resulted from substances that had *In Vitro* Irritancy Scores at or below the accepted background score of 3 to 5.

Table 7-13 Chemical and Product Classes of Test Substances with Discordant Interlaboratory Results in the Gautheron et al. (1994) Study

Substance	Chemical Class	Product Class	Physical Form Tested	*In Vitro* Classification (% of Labs with Classification)
Butyrolactone	Lactone; Heterocyclic	Solvent; Synthetic intermediate	Liquid	Moderate (10/12; 83%) Severe (2/12; 17%)
Deoxycholic acid, sodium salt	Alcohol; Carboxylic acid	Surfactant	10% Solution	Severe (11/12; 92%) Moderate (1/12; 8%)
Diacetone alcohol	Alcohol; Ketone	Solvent	Liquid	Moderate (8/11; 73%) Severe (3/11; 27%)
2,4-Dichloro-5-sulfamoylbenzoic acid	Amide; Organic sulfur compound	Chemical intermediate	Solid	Mild (8/12; 67%) Moderate (3/12; 25%) Severe (1/12; 8%)
Ethanol	Alcohol	Solvent	Liquid	Severe (7/11; 64%) Moderate (4/11; 36%)
Furan	Heterocyclic compound	Solvent; Chemical intermediate	Liquid	Severe (6/12; 50%) Moderate (6/12; 50%)
Hexadecyltrimethyl-ammonium bromide	Organic salt; Onium compound	Surfactant; Agricultural chemical; Germicide	Liquid	Severe (6/11; 55%) Moderate (5/11; 45%)
N-Lauroylsarcosine, sodium salt	Amide; Amino acid	Surfactant	10% Solution	Moderate (9/11; 82%) Severe (2/11; 18%)
Laurylsulfobetaine	Amine; Onium compound	Surfactant	10% Solution	Severe (10/11; 91%) Moderate (1/11; 9%)
Methanol	Alcohol	Solvent, Chemical intermediate	Liquid	Severe (8/11; 73%) Moderate (2/11; 12%) Mild (1/11; 9%)
2-Methoxyethanol	Alcohol	Solvent	Liquid	Severe (9/11; 82%) Moderate (2/11; 18%)
Octanol	Alcohol	Solvent	Liquid	Moderate (6/11; 55%) Severe (4/11; 36%) Mild (1/11; 9%)
2,4-Pentanedione	Alcohol	Solvent	Liquid	Severe (7/12; 58%) Moderate (5/12; 42%)
Promethazine hydrochloride	Amidine; Heterocyclic compound; Organic sulfur compound	Drug/therapeutic agent	Solid	Severe (9/11; 82%) Moderate (1/11; 9%) Mild (1/11; 9%)
Quinacrine	Heterocyclic compound	Drug/therapeutic agent	Solid	Moderate (5/11; 45%) Mild (4/11; 36%) Severe (2/11; 18%)
1,2,3-Trichloropropane	Hydrocarbon	Solvent	Liquid	Moderate (8/11; 73%) Severe (2/11; 18%) Mild (1/11; 9%)

Table 7-14 Chemical and Product Classes of Test Substances with Discordant Interlaboratory Results in the Balls et al. (1995) Study

Substance	Chemical Class	Product Class	Physical Form	*In Vitro* Classification (No. of Laboratories)
Butyrolactone	Lactone; Heterocycle	Solvent; Synthetic intermediate	Liquid	Severe (3/5; 60%) Moderate (2/5; 40%)
Captan 90 concentrate	Imide; Organic sulfur compound	Pesticide	Solid	Moderate (4/5; 80%) Severe (1/5; 20%)
Cetylpyridinium bromide (10%)	Heterocyclic compound; Onium compound	Surfactant, Germicide	10% Solution	Severe (4/5; 80%) Moderate (1/5; 20%)
Cyclohexanol	Alcohol	Solvent; Chemical intermediate	Liquid	Moderate (3/5; 60%) Severe (2/5; 40%)
Ethanol	Alcohol	Solvent	Liquid	Severe (4/5; 80%) Moderate (1/5; 20%)
2-Ethyl-1-hexanol	Alcohol	Solvent	Liquid	Severe (2/5; 40%) Moderate (2/5; 40%) Mild (1/5; 20%)
Fomesafen	Imide; Ether; Nitro compound	Pesticide	Solid	Severe (2/5; 40%) Mild (2/5; 40%) Moderate (1/5; 20%)
n-Hexanol	Alcohol	Solvent	Liquid	Severe (3/5; 60%) Moderate (2/5; 40%)
Isobutanol	Alcohol	Solvent	Liquid	Moderate (3/5; 60%) Severe (2/5; 40%)
Isopropanol	Alcohol	Solvent	Liquid	Severe (3/5; 60%) Moderate (2/5; 40%)
Maneb	Amine/Amidine; Organic salt	Pesticide	Solid	Severe (2/5; 40%) Mild (2/5; 40%) Moderate (1/5; 20%)
Methyl acetate	Ester	Solvent	Liquid	Moderate (4/5; 80%) Severe (1/5; 20%)
Methyl ethyl ketone	Ketone	Solvent	Liquid	Severe (4/5; 80%) Moderate (1/5; 20%)
1-Napthalene acetic acid	Carboxylic acid; Polycyclic compound;	Pesticide	Solid	Severe (4/5; 80%) Moderate (1/5; 20%)
n-Octanol	Alcohol	Solvent	Liquid	Moderate (3/5; 60%) Severe (1/5; 20%) Mild (1/5; 20%)
Sodium lauryl sulfate (15%)	Carboxylic acid (salt)	Surfactant	10% Solution	Severe (3/5; 60%) Moderate (2/5; 40%)
Trichloroacetic acid (3%)	Carboxylic acid	Herbicide; chemical intermediate	Solution	Severe (4/5; 80%) Moderate (1/5; 20%)
Triton X-100 (5%)	Ether	Surfactant	10% Solution	Severe (4/5; 80%) Moderate (1/5; 20%)
Triton X-100 (10%)	Ether	Surfactant	10% Solution	Severe (4/5; 80%) Moderate (1/5; 20%)

Table 7-15 Coefficient of Variation Analysis of the Interlaboratory Variability of the BCOP Test Method for Gautheron et al. (1994)[1]

Substance	Mean *In Vitro* Irritancy Score	No. of Labs	%CV	*In Vitro* Prediction
2-Ethoxyethanol	91.3	12	16.5	Severe
2,4-Pentanedione	59.8	12	24	Severe
Allyl alcohol	156	12	27	Severe
Imidazole	87.9	12	28.5	Severe
Furan	56	12	29.4	Severe
Benzethonium chloride	133.9	11	31.7	Severe
Butyrolactone	45.6	12	32.2	Moderate
Cyclohexanone	105.6	11	33.3	Severe
2-Methoxyethanol	63.5	11	33.6	Severe
Laurylsulfobetaine	80.6	11	34	Severe
Ethyl acetoacetate	31.8	11	34.9	Moderate
Gluconolactone	76.6	11	35	Severe
Methylisobutyl ketone	19.9	11	36	Mild
Pyridine	112.8	11	38.4	Severe
Ethanol	60.7	11	39.1	Severe
3-Glycidoxypropyltrimethoxysilane	16.6	12	40	Moderate
N-Lauroylsarcosine, sodium salt	50	11	41.7	Moderate
Octanol	47.4	11	41.7	Moderate
Deoxycholic acid, sodium salt	93.5	12	43	Severe
2-Aminophenol	7	12	43.5	Mild
Hexadecyltrimethylammonium bromide	66.4	11	45.2	Severe
1-Phenyl-3-pyrazolidone	12.9	12	46.5	Mild
Dibenzoyl-L-tartaric acid	120.5	11	46.8	Severe
Dimethyl sulfoxide	11.4	11	46.9	Mild
1-Nitropropane	7.6	12	46.9	Mild
1,2,4-Trimethylbenzene	16.1	12	47	Mild
Propyl-4-hydroxybenzoate	7.9	11	48	Mild
Promethazine hydrochloride	112.4	11	49.3	Severe
1,2,3-Trichloropropane	47.5	11	50.3	Moderate
Diacetone alcohol	53.5	11	50.8	Moderate
Methanol	84.2	11	55.7	Severe
2,4-Dichloro-5-sulfamoylbenzoic acid	26.3	12	58.5	Moderate
Sodium oxalate	4.8	12	66	Mild
Quinacrine	31.1	11	74.8	Moderate
Petroleum ether	5.5	12	75.4	Mild
Dimethylbiguanide	2.9	11	82	Mild
Magnesium carbonate	3	11	83	Mild
Triethanolamine	2.2	11	101.5	Mild
Aluminum hydroxide	6.8	12	107	Mild
Tetraaminopyrimidine sulfate	6	11	107	Mild
Hexane	1.4	12	143	Mild
Iminodibenzyl	2.4	11	177.5	Mild
2-Mercaptopyrimidine	-1.25	12	208	Mild
Triton X-155	0.55	11	276	Mild
DL-Glutamic acid	0.58	12	330.6	Mild

Substance	Mean *In Vitro* Irritancy Score	No. of Labs	%CV	*In Vitro* Prediction
Anthracene	-0.33	12	430	Mild
Betaine monohydrate	0.92	12	432	Mild
MYRJ-45	-0.18	11	962	Mild
EDTA, di-potassium salt	-0.33	12	1009	Mild
BRIJ-35	-0.09	11	1280	Mild
Phenylbutazone	-0.17	12	1325	Mild
Mean %CV	**167.6 (all substances) 84 (excluding MYRJ-45, EDTA, BRIJ-35, phenylbutazone)**			
Median %CV	**46.9**			

[1] Substances organized by increasing %CV.

For the Balls et al. (1995) study, a wide range of %CV values for individual substances is evident for the *In Vitro* Irritancy Score (**Table 7-16**). The mean and median %CV values were 125% and 30.6%, respectively, ranging from 7.6% to 4511% for the entire set of 59 test substances. The 32 substances predicted as severe in the BCOP assay had mean and median %CV values of 25% and 22%, respectively, with a %CV range from 7.6% to 89.4%.

Table 7-17 presents the %CV values for the *In Vitro* Irritancy Score of individual substances tested in the Southee (1998) study. The mean and median %CV values were 32.4% and 22.8%, respectively, with a range of 7.5% to 108.8% for the entire set of test substances.

7.2.3 Additional Analyses of Interlaboratory Reproducibility

The EC Interlaboratory Study (Gautheron et al. 1994): This study found that 82.7% of the substances tested were classified the same by all laboratories when using a three-category system. In this system, substances were classified into one of the following categories: mild irritant (BCOP score [0-25], moderate irritant [25.1-55], and severe irritant [≥55.1]).

The EC/HO Validation Study (Balls et al. 1995): The study authors determined the interlaboratory correlation of BCOP results (permeability value, opacity value and *In Vitro* Irritancy Score) generated from the five laboratories that participated in the EC/HO study (**Table 7-18**). In this analysis, each laboratory was compared to each other laboratory in a pair-wise fashion for all 60 substances tested, as well as for subsets of test substances (water-soluble, water-insoluble, surfactants solids, solutions, and liquids). This analysis yielded a range of correlation coefficients for the subsets of test substances as shown in **Table 7-18** (see **Appendix F** for all correlation coefficients derived from comparing each laboratory with every other laboratory). Interlaboratory correlation coefficients for the *In Vitro* Irritancy Score generally spanned a range of 0.867 to 0.958 depending on the specific subsets of substances being evaluated. However, the correlation coefficients for the permeability value were lower (e.g., correlation coefficients BCOP – Permeability Value ranged from 0.683 to 0.906 for the full set of test substances). The correlation coefficients for the Opacity Value were slightly higher (0.898 to 0.978) than the correlation for the for the *In Vitro* Irritancy Score.

Table 7-16 Coefficient of Variation Analysis of the Interlaboratory Variability of the BCOP Test Method for Balls et al. (1995)[1]

Substance	Mean *In Vitro* Irritancy Score	No. of Labs	%CV	*In Vitro* Prediction
1-Naphthalene acetic acid, Na salt	149.2	5	7.6	Severe
Benzalkonium chloride (10%)	136.5	5	10.9	Severe
Sodium hydroxide (1%)	150	5	12.3	Severe
Cetylpyridinium bromide (6%)	71.2	5	12.7	Severe
Acetone	123	5	14	Severe
Imidazole	112.7	5	14.5	Severe
Benzalkonium chloride (5%)	128.5	5	15.6	Severe
Methyl acetate	54.9	5	17.4	Moderate
Sodium hydroxide (10%)	271.9	5	17.6	Severe
Toluene	35.6	5	18.1	Moderate
Chlorhexidine	114	5	18.3	Severe
Trichloroacetic acid (30%)	264	5	18.7	Severe
Dibenzyl phosphate	378	5	18.8	Severe
2,2-Dimethylbutanoic acid	111.9	5	19.5	Severe
Pyridine	148	5	20.1	Severe
Promethazine hydrochloride	121.4	5	20.4	Severe
Trichloroacetic acid (3%)	75.9	5	21.1	Severe
Benzalkonium chloride (1 %)	88.8	5	21.7	Severe
Parafluoraniline	30.4	5	21.7	Moderate
Methyl ethyl ketone	70.4	5	22.6	Severe
4-Carboxybenzaldehyde	78.3	5	24	Severe
Ethanol	70.6	5	24.1	Severe
Cetylpyridinium bromide (10%)	72	5	24.2	Severe
Triton X-100 (5 %)	78.3	5	24.2	Severe
Triton X-100 (10 %)	70.3	5	25.3	Severe
Isobutanol	56	5	26.1	Severe
n-Hexanol	61.9	5	27	Severe
Sodium lauryl sulfate (15 %)	63.3	5	28	Severe
Cyclohexanol	60.1	5	28.5	Severe
2,6-Dichlorobenzoyl chloride	10.4	5	30.6	Mild
Sodium lauryl sulfate (3 %)	25.8	5	30.9	Mild
Isopropanol	57.9	5	31.3	Severe
Sodium perborate	97	5	35.8	Severe
Methyl isobutyl ketone	12.6	5	36	Mild
1-Naphthalene acetic acid	78.1	5	37.4	Severe
Butyl acetate	34.6	5	38.4	Moderate
Methyl cyanoacetate	12.2	5	39.2	Mild
Ethyl acetate	32	5	40.5	Moderate
Potassium cyanate	15	5	40.9	Mild
2,5-Dimethylhexanediol	20.8	5	41.6	Mild
Benzoyl-L-tartaric acid	169.6	5	43	Severe
gamma-Butyrolactone	60.7	5	45	Severe
Tetraaminopyrimidine sulfate	15.1	5	46.3	Mild
Methylcyclopentane	2.8	5	47.8	Mild
2-Ethyl-1-hexanol	39.8	5	48.2	Moderate
Cetylpyridinium bromide (0.1%)	9.2	5	51.4	Mild

Substance	Mean *In Vitro* Irritancy Score	No. of Labs	%CV	*In Vitro* Prediction
Maneb	40.5	5	58.3	Moderate
n-Octanol	40.9	5	58.8	Moderate
Ethyl-2-methylacetoacetate	14.4	5	65.3	Mild
Ethyl trimethyl acetate	17.8	5	66.3	Mild
Ammonium nitrate	9.8	5	69.7	Mild
L-Aspartic acid	1.3	5	73.6	Mild
Captan 90 concentrate	43.8	5	75.8	Moderate
Quinacrine	1.6	5	76.9	Mild
Fomesafen	60.7	5	89.4	Severe
Sodium oxalate	14	5	143	Mild
Polyethylene glycol 400	1.1	5	145	Mild
Glycerol	0.26	5	712	Mild
Tween 20	-0.04	5	4511	Mild
Mean %CV	**125 (all test substances) 50 (excluding Tween 20)**			
Median %CV	**30.6**			

[1]Substances organized by increasing %CV.

Table 7-17 Coefficient of Variation Analysis of the Interlaboratory Variability of the BCOP Test Method for Southee (1998)[1]

Substance	Mean *In Vitro* Irritancy Score	No. of Labs	%CV	*In Vitro* Prediction
Butyl cellosolve	100.9	3	7.5	Severe
Benzalkonium chloride	160	3	8.5	Severe
NaOH (10%)	226	3	8.6	Severe
Imidazole	136.9	3	9.1	Severe
4-Carboxybenzaldehyde	46.7	3	9.5	Moderate
Parafluoroaniline	32.1	3	19.1	Moderate
Methyl ethyl ketone	82.5	3	21.6	Severe
Ethanol	48.7	3	22.1	Moderate
Ammonium nitrate	5.03	3	23.4	Mild
Hexadecyltrimethylammonium bromide (10%)	29.3	3	27.1	Moderate
Glycerol	0.72	3	33.5	Mild
Propyl-4-hydroxybenzoate	6.9	3	37.7	Mild
Triton X-100 (5%)	3.3	3	44.8	Mild
Sodium lauryl sulfate (15%)	9.7	3	57.1	Mild
Tween 20	0.23	3	79.8	Mild
Sodium oxalate	3.6	3	108.8	Mild
Mean %CV	**32.4**			
Median %CV	**22.8**			

[1]Substances organized by increasing %CV.

Table 7-18 Interlaboratory Correlation Ranges Determined for Various Subsets of Tested Substances in Balls et al. (1995)

BCOP Test Method Value	Interlaboratory Pearson's Correlation Coefficient (r) of the *In Vitro* Data
Full set of test substances[1] (60)	
BCOP - Permeability Value	0.683-0.906
BCOP - Opacity Value	0.898-0.978
BCOP - *In Vitro* Irritancy Score	0.867-0.958
Chemicals soluble in water (30)	
BCOP - Permeability Value	0.521-0.880
BCOP - Opacity Value	0.927-0.971
BCOP - *In Vitro* Irritancy Score	0.855-0.952
Chemicals insoluble in water (18)	
BCOP - Permeability Value	0.688-0.963
BCOP - Opacity Value	0.896-0.991
BCOP - *In Vitro* Irritancy Score	0.898-0.981
Surfactants (12)	
BCOP - Permeability Value	0.766-0.966
BCOP - Opacity Value	0.947-0.995
BCOP - *In Vitro* Irritancy Score	0.914-0.989
Solids (20)	
BCOP - Permeability Value	0.563-0.934
BCOP - Opacity Value	0.903-0.977
BCOP - *In Vitro* Irritancy Score	0.852-0.960
Solutions (14)	
BCOP - Permeability Value	0.731-0.933
BCOP - Opacity Value	0.955-0.989
BCOP - *In Vitro* Irritancy Score	0.914-0.980
Liquids (26)	
BCOP - Permeability Value	0.612-0.893
BCOP - Opacity Value	0.913-0.967
BCOP - *In Vitro* Irritancy Score	0.851-0.956

[1]As noted in **Section 3.0**, one substance (thiourea) was tested *in vitro* in the BCOP assay but, due to its excessive toxicity *in vivo*, it was excluded from the comparison of *in vitro* and *in vivo* test results, and thus excluded from the evaluation in **Section 7.2.1**. However, *in vitro* data for this substance was included in the original Balls et al. (1995) analysis.

7.3 Historical Positive and Negative Control Data

An example of historical data for positive controls was provided by IIVS (current as of July 22, 2004), as shown in **Table 7-19**.

Table 7-19 Historical Positive Control Data for the BCOP Assay

Positive Control	Opacity	OD_{490}	*In Vitro* Score
Ethanol (10 min exposure)			
Mean (n = 632)	31.2	1.422	52.7
SD	4.8	0.345	6.4
CV	15.3%	24.3%	12.1%
Upper and lower limits[1]	21.7 – 40.7	0.742 – 2.112	39.9 – 65.4
Imidazole (4 hour exposure)			
Mean (n = 125)	76.4	1.768	103.0
SD	18.4	0.488	16.6
CV	24.1%	27.6%	16.2%
Upper and lower limits*	39.7 – 113.2	0.792 – 2.745	69.7 – 136.2

Abbreviations: CV = Coefficient of variation; n = Number of tests; SD = Standard deviation.
[1]The upper and lower limits are the upper and lower 95% confidence limits (+/- 2 SDs) around the mean.

7.4 Summary

A quantitative assessment of intralaboratory data (*In Vitro* Irritancy Scores) from three studies (Southee 1998; Dr. Sina's submission; Dr. Van Goethem's submission) indicates the extent of intralaboratory repeatability of the BCOP test method for substances predicted as severe eye irritants. For the 16 substances evaluated in the Southee (1998) study, the median %CV for *In Vitro* Irritancy Scores for replicate corneas ranged from 11.8 to 14.2 for the three laboratories. For the 29 substances evaluated by Dr. Sina, the within experiment mean and median %CV values for *In Vitro* Irritancy Scores were 71 and 35, respectively. The dataset provided by Dr. Sina included 10 substances with low *In Vitro* Irritancy Scores around the background range of the assay (< 3.5), contributing to the increased variability of this dataset. However, the range of %CV values for the five substances predicted as severe irritants (*In Vitro* Scores >55.1) in this study is 1.1 to 13. For the 52 substances evaluated by Dr. Van Goethem in the Gautheron et al. (1994) study, the median %CV for *In Vitro* Irritancy Scores for replicate corneas was 18.1%, comparable to the results obtained with the data from Southee (1998).

A quantitative assessment of intralaboratory data (*In Vitro* Irritancy Scores) from two studies (Gettings et al. 1996; Southee 1998) indicates the extent of intralaboratory reproducibility of the BCOP test method for substances predicted as severe eye irritants. For the Gettings et al. (1996) study, the between experiment (n = 3) mean and median %CV values for permeability values were 33.4 and 29.0, respectively, for 25 surfactant-based personal care cleaning formulations. For the Southee (1998) study, the mean %CV values for *In Vitro* Irritancy Scores for the 16 substances tested two or more times in Laboratory 1, Laboratory 2, and Laboratory 3 ranged from 12.6 to 14.8 for the three laboratories, while the median %CV values ranged from 6.7 to 12.4.

A qualitative assessment of the data provided for multiple laboratories in three studies (Gautheron et al. 1994; Balls et al. 1995; Southee 1998) indicates the extent of interlaboratory reproducibility. In an assessment of interlaboratory reproducibility of hazard classification (EPA, EU, or GHS), the five participating laboratories for the Balls et al.

(1995) study were in 100% agreement in regard to the ocular irritancy classification for 40 to 41 (67% to 68%) of the 60 substances tested *in vitro* in the study, depending on the classification system used. The extent of agreement between testing laboratories was greatest for substances identified from *in vivo* rabbit eye data as corrosives or severe irritants when compared to any other combination of *in vivo* and *in vitro* results (76% to 86% of the accurately identified severe substances were shown to have 100% classification agreement among testing laboratories). For the study by Gautheron et al. (1994), regardless of the classification system used, there was 100% agreement in regard to the ocular irritancy classification for 35 (69%) of the 51 substances, which were tested in either 11 or 12 laboratories. For the study by Southee (1998), there was 100% agreement in regard to the ocular irritancy classification for 15 (94%) of the 16 substances, regardless of the classification system used. Substances with less than complete agreement in the testing laboratories include those representing such chemical classes as alcohols, ketones, and heterocyclic compounds, and such product classes as surfactants, organic solvents, chemical intermediates, detergents, and pesticides.

A quantitative evaluation of interlaboratory reproducibility was conducted for three studies (Gautheron et al. 1994; Balls et al. 1995; Southee 1998) by performing a %CV analysis of *In Vitro* Irritancy Scores obtained for substances tested in multiple laboratories. For the Gautheron et al. (1994) study, the 17 substances predicted as severe in the BCOP assay had mean and median %CV values of 36% and 17%, respectively, for results obtained in either 11 or 12 laboratories. For the Balls et al. (1995) study, the 32 substances predicted as severe in the BCOP assay had mean and median %CV values of 25% and 22%, respectively, for results obtained in five laboratories. For the Southee (1998) study, the mean and median %CV values for the *In Vitro* Irritancy Scores of the 16 substances were 32.4% and 22.8%, respectively, for three laboratories.

Balls et al. (1995) also determined the interlaboratory correlation between BCOP test method endpoint data generated by each laboratory for all 60 substances tested[1], as well as for various subsets of test substances (water-soluble, water-insoluble, surfactants, solids, solutions, and liquids). This analysis yielded a range of correlation coefficients for the subsets of test substances. Interlaboratory correlation coefficients for the *In Vitro* Irritancy Score generally spanned a range of 0.867 to 0.958 depending on the specific subsets of substances being evaluated. However, the correlation coefficients for the permeability value were lower (e.g., correlation coefficients BCOP – Permeability Value ranged from 0.683 to 0.906 for the full set of test substances). The correlation coefficients for the Opacity Value were higher (0.898 to 0.978) than the correlation for the *In Vitro* Irritancy Score.

[1] In some analyses of the Balls et al. (1995) validation results, 59 substances were considered. In other analyses, 60 substances were considered. The difference in the total number of substances is due to the exclusion of one substance, thiourea, in some analyses due to its excessive *in vivo* toxicity.

[This Page Intentionally Left Blank]