7.0 HET-CAM 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 HET-CAM test method data were available for replicate eggs within individual experiments or for replicate experiments within an individual laboratory for two studies (Gilleron et al. 1996, 1997). Therefore, an evaluation of the repeatability and/or intralaboratory reproducibility of the HET-CAM test method could be conducted. Additionally, comparable HET-CAM data were available for multiple laboratories within each of three to four comparative validation studies (CEC 1991, Balls et al. 1995, Spielmann et al. 1996, and Hagino et al. 1999), which allowed for an evaluation of the interlaboratory reproducibility of the HET-CAM test method.

7.1 Selection Rationale for the Substances Used to Evaluate the Reliability of the HET-CAM Test Method

There was limited information on the rationale for substance selection used in various multilaboratory studies to evaluate the reliability of the HET-CAM test method. Most reports indicated that 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).

The selection of substances used in the CEC (1991) evaluation was based on the following criteria:

- The substances should be representative of currently used industrial chemicals and should represent a range of chemical structures.
- The substances should cover the range of eye effects from nonirritant to severe irritant.

- The *in vivo* rabbit eye studies should have been conducted in accordance with EEC criteria, and the animal data should be sufficient to allow an irritancy classification to be definitively assigned to the test substance.
- Whenever possible, the substances should have been used in previous validation studies.

As noted previously, the EC/HO validation study reported on by Balls et al. (1995) evaluated the performance and reproducibility of the HET-CAM 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). A description of the requirements for inclusion into the study was provided in **Section 3.0**.

Gilleron et al. (1996, 1997) selected substances that represented a broad spectrum of ocular irritancies, chemical classes, and chemical structures. Substances also were selected on the basis of availability of historical *in vivo* data, to avoid conducting additional tests for the validation study. Additionally, substances evaluated in the Gilleron et al. (1997) study were the same as those previously evaluated by Balls et al. (1995).

Spielmann et al. (1996) selected substances that represented a broad spectrum of ocular irritancies, chemical classes, and chemical structures. Substances also were selected on the basis of availability of historical *in vivo* data.

Hagino et al. (1999) evaluated substances that were major ingredients in cosmetic formulations and preparations. These substances included surfactants and solvents.

7.2 Analyses of Repeatability and Reproducibility

7.2.1 Quantitative Assessment of Intralaboratory Repeatability

An analysis of interlaboratory repeatability has included such approaches as:

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

Two of the reports discussed included intralaboratory repeatability data (Gilleron et al. 1996, 1997)[1]. Using these data, the consistency of HET-CAM IS(B) results obtained among identically-treated eggs within an experiment was evaluated using a CV analysis. Considering the number of replicate eggs tested in each experiment, no attempt was made to use ANOVA to determine if any individual egg score differed from any other egg scores.

7.2.1.1 Gilleron et al. (1996)

Individual egg results for 46 substances analyzed by the HET-CAM IS(B) analysis method and reported on by Gilleron et al. (1996) were received in response to a request from NICEATM. In the data provided to NICEATM, the original test results for nine of the 46

[1] Transformed data for these studies are available in **Appendices E1** and **E2**.

substances included in the 1996 publication (laurylsulfobetaine, deoxycholic acid, ethylacetoacetate, methyl isobutyl ketone, methanol, N-laurylsarcosine, promethazine hydrochloride, 2-methoxyethanol, benzethonium chloride, and imidazole) were no longer available. Since alternative HET-CAM test data generated were available for these substances, these data were provided to NICEATM. The overall replicate egg mean and median %CV values were evaluated with and without the inclusion of the data for these nine substances.

For each test substance, three different eggs were used in each of at least three replicate experiments. For this evaluation, the %CV values were determined for each endpoint (hemorrhage, lysis, coagulation) and for the overall *in vitro* IS(B) score. For each of the endpoints, there were experiments where test substances did not produce any effects (i.e., the average score of the three replicate eggs and standard deviation [SD] of the scores were both 0) (see **Appendix E3**). For the three endpoints evaluated, 69 of 146 experiments (47%) resulted in an average score and SD of zero for the hemorrhage and lysis endpoints. Additionally, 47 of 146 experiments (32%) resulted in a total average score and SD of zero for the coagulation endpoint. For the overall *in vitro* IS(B) score, 21 of 146 experiments (14%) resulted in an average score and SD of zero (**Appendix E3**). For three test substances (anthracene, ethylenediaminetetraacetic acid [EDTA] dipotassium, and iminodibenzyl), the overall IS(B) analysis method score and SD were zero for each of the eggs tested. The replicate egg repeatability %CV values for individual experiments, excluding studies where such values could not be calculated, ranged from 0.12 to 173.21 for hemorrhage, from 0.25 to 173.21 for lysis, from 0.00 to 173.21 for coagulation, and from 0.25 to 173.21 for the overall *in vitro* IS(B) score (see **Table 7-1** and **Appendix E3**).

The mean and median replicate egg repeatability %CV values for the overall *in vitro* IS(B) scores for the entire data set (last column in **Appendix E3**), excluding studies where the overall IS(B) score and SD were zero, were 32.52 and 11.49, respectively (**Table 7-1**). When the data for the nine substances noted were removed, the mean and median replicate egg repeatability %CV values for the overall IS(B) scores were 41.48 and 17.54, respectively (**Table 7-1**).

7.2.1.2 *Gilleron et al. (1997)*

Individual egg results for 60 substances evaluated by the HET-CAM IS(B) analysis method and reported on by Gilleron et al. (1997) were provided to NICEATM. Among the data, the original test results for four of the 60 substances included in the 1997 publication (Maneb, 1-napthalene acetic acid, Tween 20, and 1-napthalene acetic acid, sodium salt) were no longer available. Since alternative HET-CAM test data were available for these substances, these data were provided to NICEATM. The overall replicate egg mean and median %CV values were evaluated with and without the inclusion of these data.

Table 7-1 Intralaboratory Repeatability Results for HET-CAM Studies of Gilleron et al. (1996)

	Hemorrhage Endpoint	Lysis Endpoint	Coagulation Endpoint	Overall Irritation Score
Mean Value (SD) for All Substances[1]	1.64 (1.93)	2.68 (2.88)	3.59 (3.44)	7.92 (5.84)
Range of Values for All Substances	0.12-173.21	0.25-173.21	0.00-173.21	0.25-173.21
%CV for Substances[2]	117.56	107.52	95.69	73.74
Number of Experiments	146	146	146	146
Mean Value (SD) Excluding Nine Substances Where Original Data Was Not Available[1]	1.63 (1.90)	1.87 (2.57)	2.83 (3.25)	6.33 (5.43)
Range of Values Excluding Nine Substances Where Original Data Was Not Available	0.12-173.21	0.25-173.21	0.00-173.21	0.35-173.21
%CV Excluding Nine Substances Where Original Data Was Not Available[2]	116.13	137.49	115.07	85.84
Number of Experiments	111	111	111	111
Mean Overall *In Vitro* Score %CV for All Substances	32.52			
Median Overall *In Vitro* Score %CV for all Substances	11.49			
Mean Overall *In Vitro* Score %CV Excluding Nine Substances Where Original Data Was Not Available	41.48			
Median Overall *In Vitro* Score %CV Excluding Nine Substances Where Original Data Was Not Available	17.54			

Abbreviations: %CV = percent coefficient of variation, SD = standard deviation.

[1]Mean was calculated using the values from the "Mean for 3 Eggs" column for each endpoint and the Overall *In Vitro* Score as shown in **Appendix E3**. The SD was calculated based on the values in these individual columns.

[2]To avoid eliminating data for which the %CV (coefficient of variation) value could not be calculated (i.e., where the mean and SD both equaled 0), the %CV values were calculated using the mean and standard deviation calculated as described in footnote 1 of this table.

For each test substance, three different eggs were used in each of at least three replicate experiments. For this evaluation, the %CV values were determined for each endpoint (hemorrhage, lysis, coagulation) and for the overall *in vitro* IS(B) score. For each of the endpoints, there were experiments where test substances did not produce any effects (i.e., the average score of the three replicate eggs and standard deviation [SD] of the scores were both 0) (see **Appendix E4**). For the hemorrhage endpoint, 91 of 184 experiments (49%) resulted in an average score and SD of zero for the three replicate eggs; for the lysis endpoint, 22 of 184 experiments (12%) resulted in an average score and SD of zero; while, for the coagulation endpoint, 16 of 184 experiments (9%) resulted in an average score and SD of zero. For the overall *in vitro* IS(B) score, 6 of 184 experiments (3%) resulted in an average score and SD of zero for the three replicate eggs (**Appendix E4**). For one test substance (Maneb), the overall IS(B) analysis method score and SD were zero for each of the eggs tested. The replicate egg repeatability %CV values for individual experiments, excluding studies where such values could not be calculated, ranged from 0.23 to 173.21 for hemorrhage, from 0.00 to 173.21 for lysis, from 0.37 to 173.21 for coagulation, and from 0.13 to 173.21 for the overall *in vitro* IS(B) score (**Table 7-2** and **Appendix E4**).

The mean and median replicate egg repeatability %CV values for the overall *in vitro* IS(B) scores for the entire data set (last column in **Appendix E4**), excluding studies where such values could not be calculated, were 7.61 and 2.24, respectively (**Table 7-2**). When the data for the four substances noted were removed the mean and median replicate egg repeatability %CV values for the overall IS(B) scores were 6.99 and 2.04, respectively (**Table 7-2**).

7.2.2 Quantitative Assessment of Intralaboratory Reproducibility

Interlaboratory variability can be evaluated by assessing the CV or by using ANOVA methods. Two studies discussed in **Section 6.0** included intralaboratory reproducibility data (Gilleron et al. 1996, 1997). For both sets of studies, quantitative HET-CAM test method data were available for studies repeated three to five times in a single laboratory.

7.2.2.1 *Gilleron et al. (1996)*

Individual experimental results for 46 substances evaluated by the HET-CAM IS(B) analysis method and reported on by Gilleron et al. (1996) were received in response to a request from NICEATM. In the data provided to NICEATM, the test results for nine of the 46 substances included in the 1996 publication (laurylsulfobetaine, deoxycholic acid, ethylacetoacetate, methyl isobutyl ketone, methanol, N-laurylsarcosine, promethazine hydrochloride, 2-methoxyethanol, benzethonium chloride, and imidazole) were no longer available. Since alternative HET-CAM test data generated were available for these substances, these data were provided to NICEATM. The overall mean and median %CV values for replicate experiments were evaluated with and without the inclusion of these data.

In these studies, three different eggs were used for each experiment. Three experiments were conducted for each test substance, except for the nine substances where nonoriginal data was provided. For these substances, data for three to five experiments were provided.

Table 7-2 Intralaboratory Repeatability Results for HET-CAM Studies of Gilleron et al. (1997)

	Hemorrhage Endpoint	Lysis Endpoint	Coagulation Endpoint	Overall Irritation Score
Mean Value (SD) for All Substances[1]	1.94 (2.12)	5.60 (2.31)	6.42 (2.68)	13.96 (4.89)
Range of Values for All Substances	0.23-173.21	0.00-073.21	0.37-173.21	0.13-173.21
%CV for Substances[2]	109.10	41.24	41.78	34.99
Number of Experiments	184	184	184	184
Mean Value (SD) Excluding Four Substances Where Original Data Was Not Available[1]	2.07 (2.16)	5.75 (2.19)	6.60 (2.49)	14.42 (4.48)
Range of Values Excluding Four Substances Where Original Data Was Not Available	0.23-173.21	0.00-073.21	0.37-173.21	0.13-173.21
%CV Excluding Four Substances Where Original Data Was Not Available[2]	104.43	38.04	37.78	31.05
Number of Experiments	168	168	168	168
Mean Overall *In Vitro* Score %CV for All Substances	7.61			
Median Overall *In Vitro* Score %CV for all Substances	2.24			
Mean Overall *In Vitro* Score %CV Excluding Four Substances Where Original Data Was Not Available	6.99			
Median Overall *In Vitro* Score %CV Excluding Four Substances Where Original Data Was Not Available	2.04			

Abbreviations: %CV = percent coefficient of variation, SD = standard deviation.

[1]Mean was calculated using the values from the "Mean for 3 Eggs" column for each endpoint and the Overall *In Vitro* Score as shown in **Appendix E4**. The SD was calculated based on the values in these individual columns.

[2]To avoid eliminating data for which the %CV (coefficient of variation) value could not be calculated (i.e., where the mean and SD both equaled 0), the %CV values were calculated using the mean and standard deviation calculated as described in footnote 1 of this table.

For each of the endpoints, there were a number of experiments where the test substance did not induce any effects (i.e., the average score of the repeated experiments and SD of the scores were both 0) (see **Appendix E5**). For the overall *in vitro* IS(B) score, three of 46 experiments (7%) resulted in an average score and SD of zero for the repeated experiments (**Appendix E5**). For EDTA, the overall IS(B) analysis method score and SD were zero for all replicate experiments. The reproducibility %CV values for individual substances, excluding studies where such values could not be calculated, ranged from 2.59 to 173.21 for hemorrhage, from 1.55 to 173.21 for lysis, from 1.52 to 173.21 for coagulation, and from 6.66 to 173.21 for the overall *in vitro* IS(B) score (**Appendix E5** and **Table 7-3**).

The mean and median reproducibility %CV values for the overall *in vitro* IS(B) scores for the entire data set (last column in **Appendix E5**), excluding studies where such values could not be calculated, were 52.73 and 33.70, respectively (**Table 7-3**). When the data for the nine substances noted were removed, the mean and median reproducibility %CV values for the overall IS(B) scores were 60.66 and 39.15, respectively (**Table 7-3**).

7.2.2.2 Gilleron et al. (1997)

Individual experimental results for 60 substances evaluated by the HET-CAM IS(B) analysis method and reported on by Gilleron et al. (1997) were provided by the authors to NICEATM. Among the data, the original test results for four of the 60 substances included in the 1997 publication (Maneb, 1-napthalene acetic acid, Tween 20, and 1-napthalene acetic acid, sodium salt) were no longer available. Since alternative HET-CAM test data were available for these substances, these data were provided to NICEATM. The overall mean and median %CV values for replicate experiments were evaluated with and without the inclusion of these data.

In these studies, three different eggs were used for each experiment. Three experiments were conducted for each test substance, except for the four substances where nonoriginal data was provided. For these substances, data for three to five experiments were provided.

For each of the endpoints, there were a number of experiments where the test substance did not induce any effects (i.e., the average score of the three replicate eggs and thus the SD of the scores were both zero) (see **Appendix E6**). For the overall *in vitro* IS(B) score, none of substances resulted in an average score and SD of zero for the three replicate experiments (**Appendix E6**). The reproducibility %CV values for individual substances, excluding studies where such values could not be calculated, ranged from 0.20 to 173.21 for hemorrhage, from 0.12 to 200.00 for lysis, from 0.00 to 173.21 for coagulation, and from 0.34 to 200.00 for the overall *in vitro* IS(B) score (**Appendix E6** and **Table 7-4**).

The mean and median reproducibility %CV values for the overall *in vitro* IS(B) scores for the entire data set (last column in **Appendix E6**), excluding studies where such values could not be calculated, were 17.48 and 6.34, respectively (**Table 7-4**). When the data for the nine substances noted were removed, the mean and median reproducibility %CV values for the overall IS(B) scores were 13.49 and 5.25, respectively (**Table 7-4**).

Table 7-3 Intralaboratory Reproducibility Results for HET-CAM Studies of Gilleron et al. (1996)

	Hemorrhage Endpoint	Lysis Endpoint	Coagulation Endpoint	Overall Irritation Score
Mean Value (SD) for All Substances[1]	1.64 (2.04)	2.68 (2.96)	3.59 (3.52)	7.51 (5.28)
Range of Values for All Substances	2.59-173.21	1.55-173.21	1.52-173.21	6.66-173.21
%CV for Substances[2]	124.12	110.41	97.92	70.35
Mean Value (SD) Excluding Nine Substances Where Original Data Was Not Available[1]	1.63 (2.01)	1.87 (2.66)	2.83 (3.34)	6.33 (5.06)
Range of Values Excluding Nine Substances Where Original Data Was Not Available	2.59-173.21	1.55-173.21	4.84-173.21	14.33-173.21
%CV Excluding Nine Substances Where Original Data Was Not Available[2]	123.08	142.31	118.37	79.92
Mean Overall *In Vitro* Score %CV for All Substances	52.73			
Median Overall *In Vitro* Score %CV for all Substances	33.70			
Mean Overall *In Vitro* Score %CV Excluding Nine Substances Where Original Data Was Not Available	60.66			
Median Overall *In Vitro* Score %CV Excluding Nine Substances Where Original Data Was Not Available	39.15			

Abbreviations: %CV = percent coefficient of variation, SD = standard deviation.

[1]Mean was calculated using the values from the "Mean for 3 Eggs" column for each endpoint and the Overall *In Vitro* Score as shown in **Appendix E5**. The SD was calculated based on the values in these individual columns.

[2]To avoid eliminating data for which the %CV (coefficient of variation) value could not be calculated (i.e., where the mean and SD both equaled 0), the %CV values were calculated using the mean and standard deviation calculated as described in footnote 1 of this table.

Table 7-4 Intralaboratory Reproducibility Results for HET-CAM Studies of Gilleron et al. (1997)

	Hemorrhage Endpoint	Lysis Endpoint	Coagulation Endpoint	Overall Irritation Score
Mean Value (SD) for All Substances[1]	1.94 (2.12)	5.60 (2.31)	6.42 (2.68)	13.96 (4.89)
Range of Values for All Substances	0.20-173.20	0.12-200.00	0.00-173.21	0.34-200.00
%CV for Substances[2]	109.10	41.24	41.78	35.00
Mean Value (SD) Excluding Four Substances Where Original Data Was Not Available[1]	2.07 (2.16)	5.75 (2.18)	6.60 (2.50)	14.42 (4.48)
Range of Values Excluding Four Substances Where Original Data Was Not Available	0.20-173.21	0.12-173.21	0.00-173.21	0.34-118.75
%CV Excluding Four Substances Where Original Data Was Not Available[2]	104.43	38.04	37.78	31.05
Mean Overall *In Vitro* Score %CV for All Substances	17.48			
Median Overall *In Vitro* Score %CV for all Substances	6.34			
Mean Overall *In Vitro* Score %CV Excluding Four Substances Where Original Data Was Not Available	13.49			
Median Overall *In Vitro* Score %CV Excluding Four Substances Where Original Data Was Not Available	5.25			

Abbreviations: %CV = percent coefficient of variation, SD = standard deviation.

[1]Mean was calculated using the values from the "Mean for 3 Eggs" column for each endpoint and the Overall *In Vitro* Score as shown in **Appendix E6**. The SD was calculated based on the values in these individual columns.

[2]To avoid eliminating data for which the %CV (coefficient of variation) value could not be calculated (i.e., where the mean and SD both equaled 0), the %CV values were calculated using the mean and standard deviation calculated as described in footnote 1 of this table.

7.2.3 Quantitative and Qualitative Assessments of Interlaboratory Reproducibility

Generally, an analysis of interlaboratory variability has included such approaches as:

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

Several of the studies discussed in **Section 6.0** included interlaboratory data for at least a subset of the substances evaluated. Using this data, the ability of the HET-CAM test method to reproducibly identify ocular corrosives and severe irritants versus nonsevere irritants (i.e., moderate and slight irritant) and nonirritants were 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 HET-CAM was assessed separately for each study (i.e., publication) with multiple laboratory data (see CEC 1991, Balls et al. 1995, Spielmann et al. 1996, Hagino et al. 1999). In an alternative approach, the reliability of HET-CAM was assessed after pooling data across comparative studies that used the same data analysis method. The analysis methods where there was interlaboratory data were IS(A), IS(B)-10, IS(B)-100, Q-Score, and S-Score for the GHS and EPA classification systems. For the EU classification system, all the same HET-CAM analysis methods could be evaluated, as well as the IS(B) analysis method.

Substances classified, based on HET-CAM test 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 (UN 2003), EPA (1996), and EU (2001) classification systems.

Because the focus of this reliability assessment is on the interlaboratory reproducibility of HET-CAM 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. For example, three laboratories could classify a chemical as a nonirritant and one laboratory could classify the same chemical as a moderate irritant. Within this analysis, where a nonirritant and moderate irritant classification would be placed together, this distribution of classification calls would be considered as 100% agreement between laboratories.

In the second approach, a quantitative assessment of reproducibility was determined. CVs for test substances, where laboratory scores were available, for substances tested were reported or determined. The reproducibility of HET-CAM was assessed for studies (i.e.,

publication) reviewed in **Sections 4.0** and **5.0** where individual testing laboratory data was available (see CEC 1991, Balls et al. 1995, Spielmann et al. 1996, Hagino et al. 1999).

As discussed in **Section 2.0**, there is no standardized data collection method for HET-CAM studies and several different analysis methods have been developed (i.e., IS, Q-Score, S-Score). Therefore, the reliability assessments conducted in this section were evaluated according to each of the analysis methods described.

7.2.3.1 *Qualitative Analysis of the Interlaboratory Reproducibility of Hazard Classification Category Using the GHS Classification System*

Interlaboratory reproducibility for the HET-CAM test method was evaluated for the following reports: Balls et al. (1995), Spielmann et al. (1996) and Hagino et al. (1999). The agreement of classification calls among participating laboratories and its relationship to the GHS *in vivo* classification (UN 2003) for the substances tested in each report is provided in **Table 7-5**.

The participating laboratories were in 100% agreement in regard to the GHS ocular irritancy classification for 21 (45%) of the 47 substances tested when using the Q-Score (Balls et al. 1995). The extent of agreement between testing laboratories was greatest for substances correctly identified as GHS corrosives or severe irritants (60% [9/15] accurately identified severe substances were shown to have 100% classification agreement among testing laboratories). Comparatively, greater disparity between laboratory substance classifications was observed for false positives (i.e., positive *in vitro* but negative *in vivo*) and those substances accurately classified as nonsevere irritants. For instance, 75% (12/16) of the false positives and 58% (7/12) of the correctly identified nonsevere irritants exhibited less than 100% agreement in the GHS irritancy classifications among laboratories.

In addition to the Q-Score, Balls et al. (1995) evaluated irritancy potential for some substances using an S-Score. The participating laboratories were in 100% agreement in regard to the GHS ocular irritancy classification for 13 (68%) of the 19 tested substances. Substances that were classified as false negatives and false positives exhibited the most discordant results, with 29% (2/7) of the false negatives and 100% (2/2) of the false positives exhibiting less than 100% classification agreement between testing laboratories. There was complete agreement among testing laboratories for substances correctly classified as severe irritants or nonsevere/nonirritants, based on the GHS classification system (UN 2003).

The participating laboratories were in 100% agreement for 85 (79%) of 107 substances evaluated with the IS(B)-10 analysis method (Spielmann et al. 1996). The extent of agreement between testing laboratories was greatest for substances correctly identified as GHS nonsevere irritants or nonirritants by HET-CAM (94% [31/33]). Comparatively, greater disparity between individual substance classifications was observed for substances that were identified as false positives (56% [10/18] false positives had less than 100% concordance between testing laboratories).

Table 7-5 Evaluation of the Reliability of the HET-CAM Test Method In Predicting Ocular Corrosives and Severe Irritants as Defined by the GHS Classification System, by Study

Report	Anal[1]	Classification (*In Vivo*/*In Vitro*)[2]	# of Labs	N[3]	Substances with 100% Agreement among Labs	Substances with 80% Agreement among Labs	Substances with 75% Agreement among Labs	Substances with 66% Agreement among Labs	Substances with 60% Agreement among Labs	Substances with ≤50% Agreement among Labs
Balls et al. (1995)	Q	+/+	2 4	4 11	3 (75%)[4] 6 (55%)	-	- 4 (36%)	-	-	1 (25%) 1 (9%)
		+/-	-	-	-	-	-	-	-	-
		-/+	4	16	4 (25%)	-	9 (56%)	-	-	3 (19%)
		-/-	2 4	1 11	1 (100%) 4 (36%)	-	- 7 (64%)	-	-	-
		?/-	2	1	1 (100%)	-	-	-	-	-
		?/+	3 4	1 2	1 (100%) 1 (50%)	-	- 1 (50%)	-	-	-
		Total	2-4	47	21 (45%)	-	21 (45%)	-	-	5 (10%)
Balls et al. (1995)	S	+/+	2	4	4 (100%)	-	-	-	-	-
		+/-	2 3 4	1 4 2	1 (100%) 2 (50%) 2 (100%)	-	-	- 2 (50%) -	-	-
		-/+	2 4	1 1	-	-	-	-	-	1 (100%) 1 (100%)
		-/-	3 4	1 2	1 (100%) 2 (100%)	-	-	-	-	-
		?/-	3	1	-	-	-	1 (100%)	-	-
		?/+	2	2	1 (50%)	-	-	-	-	1 (50%)
		Total	2-4	19	13 (68%)	-	-	3 (16%)	-	3 (16%)

Report	Anal[1]	Classification (*In Vivo*/*In Vitro*)[2]	# of Labs	N[3]	Substances with 100% Agreement among Labs	Substances with 80% Agreement among Labs	Substances with 75% Agreement among Labs	Substances with 66% Agreement among Labs	Substances with 60% Agreement among Labs	Substances with ≤50% Agreement among Labs
Spielmann et al. (1996)	IS(B) -10	+/+	2 3	18 1	16 (89%) -	- -	- -	- 1 (100%)	- -	2 (11%) -
		+/-	2 3	4 1	4 (100%) -	- -	- -	- 1 (100%)	- -	- -
		-/+	2 3	16 2	7 (44%) 1 (50%)	- -	- -	- -	- -	9 (56%) 1 (50%)
		-/-	2 3	31 2	30 (97%) 1 (50%)	- -	- -	- 1 (50%)	- -	1 (3%) -
		?/-	2 3	10 2	10 (100%) 1 (50%)	- -	- -	- 1 (50%)	- -	- -
		?/+	2 3	16 4	14 (88%) 1 (25%)	- -	- -	- 2 (50%)	- -	2 (11%) 1 (25%)
		Total		107	85 (79%)			6 (6%)		16 (15%)
Spielmann et al. (1996)	IS(B) -100	+/+	2 3	17 2	16 (94%) 1 (50%)	- -	- -	- 1 (50%)	- -	1 (6%) -
		+/-	2	2	2 (100%)	-	-	-	-	-
		-/+	2 3	27 4	20 (74%) 1 (25%)	- -	- -	- 3 (75%)	- -	7 (26%) -
		-/-	2	17	16 (94%)	-	-	-	-	1 (6%)
		?/-	2 3	6 2	6 (100%) 2 (100%)	- -	- -	- -	- -	- -
		?/+	2 3	18 4	15 (83%) 2 (50%)	- -	- -	- 2 (50%)	- -	3 (17%) -
		Total		99	81 (82%)			6 (6%)		12 (12%)
Hagino et al. (1999)	IS(A)	+/+	5	8	5 (63%)	2 (25%)	-	-	1 (12%)	-
		+/-	-	-			-	-	-	-
		-/+	5	3	3 (100%)		-	-	-	-
		-/-	5	4	1 (25%)	1 (25%)	-	-	2 (50%)	-
		?/-	-	-			-	-	-	-
		?/+	5	2	2 (100%)		-	-	-	-
		Total	2-4	17	11 (64%)	3 (18%)	-	-	3 (18%)	-

Abbreviation: GHS = Globally Harmonized System (UN 2003).
[1]Anal = analysis method used to transform the sample data into HET-CAM scores. IS(A) = method described in Luepke (1985); IS(B)-10 and IS(B)-100 = method described in Kalweit et al. (1987); Q = Q-Score, method described in Balls et al. (1995); S = S-Score, method described in Balls et al. (1995).
[2]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 or 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*.
[3]N indicates number of substances.
[4]Number in parentheses indicates percentage of tested chemicals.

For the IS(B)-100 analysis method (Spielmann et al. 1996), the participating laboratories were in 100% agreement for 81 (82%) of 99 substances evaluated. As with the IS(B)-10 analysis method, the extent of agreement between testing laboratories was greatest for substances correctly identified as GHS nonsevere irritants or nonirritants by HET-CAM (94% [16/17]). Greater disparity between laboratory substance classifications was observed for substances that were identified as false positives (32% [10/31] false positives had less than 100% concordance between testing laboratories).

For the report by Hagino et al. (1999), there was 100% agreement in regard to the GHS ocular irritancy classification for 11 (64%) of the 17 substances. Discordance in the classification results was present for substances that were correctly identified as corrosives/severe irritants and as nonsevere irritants/nonirritants. Discordance in the results obtained by different laboratories ranged from 37% (3/8) to 75% (3/4) of the substances within these two groups. Substances classified as false positives had the greatest extent of agreement among laboratories.

The overall reliability statistics, evaluated by HET-CAM data analysis method, for the IS(B)-10, S-Score and Q-Score are identical to what is shown in **Table 7-5**. For the IS(A) and IS(B)-100 analysis methods, additional data laboratory data was available for a subset of the substances tested for each analysis method. For the IS(A) analysis method, the addition of two additional test substances evaluated by Kojima et al. (1995) yielded an overall concordance pattern consistent with what was observed for the Hagino et al. (1999) data alone. For the IS(B)-100 analysis method, the additional data from different testing laboratories were obtained from Gilleron et al. (1996, 1997) and Vinardell and Macián (1994). As with the IS(A) analysis method, the addition of IS(B)-100 results from additional testing laboratories yielded a concordance pattern consistent with what was observed for Spielmann et al. (1996).

7.2.3.2 *Qualitative Analysis of the Interlaboratory Reproducibility of Hazard Classification Category Using the EPA Classification System*

Reliability analyses for the HET-CAM test method were evaluated for the following two reports: Balls et al. (1995), Spielmann et al. (1996), and Hagino et al. (1999). The agreement of classification calls among participating laboratories and its relationship to the EPA (1996) *in vivo* classification for the substances tested in each report is provided in **Table 7-6**.

The participating laboratories were in 100% agreement in regard to the EPA ocular irritancy classification for 21 (45%) of the 47 substances tested when using the Q-Score (Balls et al. 1995). The extent of agreement between testing laboratories was greatest for substances correctly identified as EPA corrosives or severe irritants (71% [10/14] of the accurately identified corrosives/severe irritants exhibited 100% classification agreement among laboratories). Comparatively, greater disparity between laboratory substance classifications was observed for false positives (i.e., positive *in vitro* but negative *in vivo*) and those substances accurately classified as nonsevere irritants. For instance, 76% (13/17) of the false positives and 58% (7/12) of the correctly identified nonsevere irritants exhibited less than 100% agreement in the EPA irritancy classifications among laboratories.

Table 7-6 Evaluation of the Reliability of the HET-CAM Test Method In Predicting Ocular Corrosives and Severe Irritants as Defined by the EPA Classification System, by Study

Report	Anal[1]	Classification (*In Vivo*/*In Vitro*)[2]	# of Labs	N[3]	Substances with 100% Agreement among Labs	Substances with 80% Agreement among Labs	Substances with 75% Agreement among Labs	Substances with 66% Agreement among Labs	Substances with 60% Agreement among Labs	Substances with 50% or Less Agreement among Labs
Balls et al. (1995)	Q	+/+	2 4	4 10	3 (75%)[4] 7 (70%)	-	- 3 (30%)	-	-	1 (25%)
		+/-	-	-	-	-	-	-	-	-
		-/+	4	17	4 (24%)	-	9 (52%)	-	-	4 (24%)
		-/-	2 4	1 11	1 (100%) 4 (36%)	-	- 7 (64%)	-	-	-
		?/-	2	1	1 (100%)	-		-	-	-
		?/+	3 4	1 2	1 (100%)	-	- 2 (50%)	-	-	-
		Total	2-4	47	21 (45%)	-	21 (45%)	-	-	5 (10%)
Balls et al. (1995)	S	+/+	2	3	3 (100%)	-	-	-	-	-
		+/-	3 4	3 2	2 (66%) 2 (100%)	-	-	1 (33%) -	-	-
		-/+	2 4	1 1	-	-	-	-	-	1 (100%) 1 (100%)
		-/-	3 4	1 2	1 (100%) 2 (100%)	-	-	-	-	-
		?/-	2 3	1 2	1 (100%)	-	-	- 2 (100%)	-	-
		?/+	2	2	1 (50%)	-	-	-	-	1 (50%)
		Total	2-4	18	12 (66%)	-	-	3 (17%)	-	3 (17%)

Report	Anal[1]	Classification (*In Vivo*/*In Vitro*)[2]	# of Labs	N[3]	Substances with 100% Agreement among Labs	Substances with 80% Agreement among Labs	Substances with 75% Agreement among Labs	Substances with 66% Agreement among Labs	Substances with 60% Agreement among Labs	Substances with 50% or Less Agreement among Labs
Spielmann et al. (1996)	IS(B)-10	+/+	2 3	9 1	8 (89%) -	- -	- -	- 1 (100%)	- -	1 (11%) -
		+/-	2	3	3 (100%)	-	-	-	-	-
		-/+	2 3	18 3	9 (50%) 1 (33%)	- -	- -	- 1 (33%)	- -	9 (50%) 1 (33%)
		-/-	2 3	31 2	31 (100%) 1 (50%)	- -	- -	- 1 (50%)	- -	- -
		?/-	2 3	10 3	10 (100%) 1 (33%)	- -	- -	- 2 (66%)	- -	- -
		?/+	2 3	21 3	19 (90%) 1 (33%)	- -	- -	- 1 (33%)	- -	2 (10%) 1 (33%)
		Total	2-3	104	84 (81%)			6 (6%)		14 (13%)
Spielmann et al. (1996)	IS(B)-100	+/+	2 3	10 1	9 (90%) 1 (100%)	- -	- -	- -	- -	1 (10%) -
		+/-	2	1	1 (100%)	-	-	-	-	-
		-/+	2 3	29 4	22 (76%) 1 (25%)	- -	- -	- 3 (75%)	- -	7 (24%) -
		-/-	2 3	17 1	16 (94%) 1 (100%)	-	-	-	-	1 (6%) -
		?/-	2 3	7 1	7 (100%) 1 (100%)	- -	- -	- -	- -	- -
		?/+	2 3	21 5	19 (90%) 2 (40%)	- -	- -	- 3 (60%)	- -	2 (10%) -
		Total	2-3	97	80 (82%)			6 (6%)		11 (11%)
Hagino et al. (1999)	IS(A)	+/+	5	7	5 (71%)	2 (29%)	-	-	-	-
		+/-	-	-	-	-	-	-	-	-
		-/+	5	4	4 (100%)	-	-	-	-	-
		-/-	5	3	1 (33%)	-	-	-	2 (66%)	-
		?/-	-	-	-	-	-	-	-	-
		?/+	5	2	1 (50%)	-	-	-	1 (50%)	-
		Total	-	16	11 (69%)	3 (27%)	-	-	3 (27%)	-

Abbreviation: EPA = U.S. Environmental Protection Agency (EPA 1996).
[1]Anal = analysis method used to transform the sample data into HET-CAM scores. IS(A) = method described in Luepke (1985); IS(B) = method described in Kalweit et al (1987); Q = Q-Score, method described in Balls et al. (1995); S = S-Score, method described in Balls et al. (1995).
[2]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 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** of the HET-CAM BRD for a description of the rules followed to classify the ocular irritancy of test substances tested multiple times *in vitro.*
[3]N indicates number of substances.
[4]Number in parentheses indicates percentage of tested chemicals.

In addition to the Q-Score, Balls et al. (1995) evaluated irritancy potential for some substances by using an S-Score. The participating laboratories were in 100% agreement in regard to the EPA ocular irritancy classification for 12 (66%) of the 18 tested substances. Substances that were classified as false negatives and false positives exhibited the most discordant results, with 20% (1/5) of false negatives and 100% (2/2) of false positives exhibiting less than 100% classification agreement among testing laboratories. There was complete agreement among testing laboratories for substances correctly classified as severe irritants or nonsevere/nonirritants, based on the EPA classification system.

The participating laboratories were in 100% agreement for 84 of the 104 (81%) substances evaluated using the IS(B)-10 analysis method (Spielmann et al. 1996). The extent of agreement between testing laboratories was greatest for substances correctly identified as EPA nonsevere irritants or nonirritants by HET-CAM (97% [32/33]). Comparatively, greater disparity between individual substance classifications was observed for substances that were identified as false positives (52% [11/21] false positive had less than 100% concordance between testing laboratories).

For the IS(B)-100 analysis method (Spielmann et al. 1996), the participating laboratories were in 100% agreement 80 (82%) of the 97 substances tested. As with the IS(B)-10 analysis method, the extent of agreement between testing laboratories was greatest for substances correctly identified as EPA nonsevere irritants or nonirritants by HET-CAM (94% [17/18]). Greater disparity between laboratory substance classifications was observed for substances that were identified as false positives (33% [10/33] false positive had less than 100% concordance between testing laboratories).

For the report by Hagino et al. (1999), there was 100% agreement in regard to the EPA ocular irritancy classification for 11 (69%) of the 16 substances. Discordance in the classification results was observed for substances that were correctly identified as nonsevere irritants/nonirritants. Of the three correctly identified nonsevere irritants/nonirritants, two substances had less than 100% classification agreement among the laboratories. For EPA severe irritants, there was 100% laboratory agreement for 71% (5/7) of the tested substances.

The overall reliability statistics, evaluated by HET-CAM data analysis method, for the IS(B)-10, S-Score and Q-Score are identical to what is shown in **Table 7-6**. For the IS(A) and IS(B)-100 analysis methods, additional data laboratory data was available for a subset of the substances tested for each analysis method. For the IS(A) analysis method, the addition of two additional test substances evaluated by Kojima et al. (1995) yielded an overall concordance pattern consistent with what was observed for the Hagino et al. (1999) data alone. For the IS(B)-100 analysis method, the additional data from different testing laboratories were obtained from Gilleron et al. (1996, 1997) and Vinardell and Macián (1994). As with the IS(A) analysis method, the addition of IS(B)-100 results from additional testing laboratories yielded a concordance pattern consistent with what was observed for Spielmann et al. (1996).

7.2.3.3 *Qualitative Analysis of the Interlaboratory Reproducibility of Hazard Classification Category Using the EU Classification System*

Reliability analyses for the HET-CAM test method were evaluated for the following four reports: CEC (1991), Balls et al. (1995), Spielmann et al. (1996), and Hagino et al. (1999). The agreement of classification calls among participating laboratories and its relationship to the EU (2001) *in vivo* classification for the substances tested in each report is provided in **Table 7-7**.

For the CEC evaluation, the participating laboratories were in 100% agreement in regard to the EU ocular irritancy classification for 6 (23%) of the 26 substances tested when using the IS(B) analysis method. The extent of agreement among laboratories was greatest for accurately identified EU corrosives/severe irritants when compared to any other combination of *in vivo* and *in vitro* results (50% [3/6] of the identified EU 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 and those substances accurately classified as EU nonsevere irritants/nonirritants. For instance, 100% (9/9) of the false positives and 70% (7/10) of the correctly identified EU nonsevere irritants/nonirritants exhibited less than 100% agreement among laboratories in irritancy classifications.

The participating laboratories were in 100% agreement in regard to the EU ocular irritancy classification for 21 (45%) of the 47 substances tested when using the Q-Score (Balls et al. 1995). The extent of agreement between testing laboratories was greatest for substances correctly identified as EU corrosives or severe irritants (69% [9/13] of accurately identified EU corrosives/severe irritants exhibited 100% classification agreement among testing laboratories). Comparatively, greater disparity between laboratory substance classifications was observed for false positives and accurately identified EU nonsevere irritants/nonirritants. For instance, 71% (10/14) of the false positives and 58% (7/12) of the correctly identified EU nonsevere irritants/nonirritants exhibited less than 100% agreement among laboratories in irritancy classifications.

In addition to the Q-Score, Balls et al. (1995) evaluated irritancy potential for some substances using an S-Score. The participating laboratories were in 100% agreement in regard to the EU ocular irritancy classification for 12 (66%) of the 18 tested substances. Substances classified as false positives exhibited the most discordant results, with 100% (2/2) exhibiting less than 100% agreement in classification among laboratories.

The participating laboratories were in 100% agreement for 84 of the 106 (79%) substances evaluated with the IS(B)-10 analysis method (Spielmann et al. 1996). The extent of agreement between testing laboratories was greatest for substances correctly identified as EU nonsevere irritants or nonirritants by HET-CAM (93% [31/33]). Comparatively, greater disparity between individual substance classifications was observed for substances that were identified as false positives (58% [11/19] false positive had less than 100% concordance between testing laboratories).

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

Report	Anal[1]	Classification (*In Vivo*/*In Vitro*)[2]	# of Labs	N[3]	Substances with 100% Agreement among Labs	Substances with 75-99% Agreement among Labs	Substances with 50-74% Agreement among Labs	Substances with 25-49% Agreement among Labs
CEC (1991)	IS(B)	+/+	3	3	3 (100%)	-	-	-
			5	1	-	-	1 (100%)	-
			6	2	-	1 (50%)	1 (50%)	-
		+/-	7	1	-	1 (100%)	-	-
		-/+	3	3	-	-	1 (33%)	2 (66%)
			7	6	-	1 (17%)	2 (34%)	3 (51%)
		-/-	3	6	3 (50%)	-	2 (33%)	1 (17%)
			7	4	-	2 (50%)	2 (50%)	-
		?/-	-	-	-	-	-	-
		?/+	-	-	-	-	-	-
		Total	3-7	26	6 (23%)	5 (19%)	9 (35%)	6 (23%)
Balls et al. (1995)	Q	+/+	2	4	3 (75%)[4]	-	1 (25%)	-
			4	9	6 (67%)	3 (37%)		
		+/-	-	-	-	-	-	-
		-/+	4	14	4 (28%)	7 (50%)	3 (21%)	-
		-/-	2	1	1 (100%)	-	-	-
			4	11	4 (36%)	7 (63%)		
		?/-	2	1	1 (100%)	-	-	-
		?/+	3	1	1 (100%)	-	-	-
			4	6	1 (17%)	4 (67%)	1 (17%)	
		Total	2-4	47	21 (45%)	21 (45%)	5 (10%)	-

Report	Anal[1]	Classification (*In Vivo*/*In Vitro*)[2]	# of Labs	N[3]	Substances with 100% Agreement among Labs	Substances with 75-99% Agreement among Labs	Substances with 50-74% Agreement among Labs	Substances with 25-49% Agreement among Labs
Balls et al. (1995)	S	+/+	2	3	3 (100%)	-	-	
		+/-	2 3 4	1 3 2	1 (100%) 2 (66%) 2 (100%)	-	- 1 (33%)	
		-/+	2 4	1 1	-	-	1 (100%) 1 (100%)	
		-/-	3 4	1 2	1 (100%) 2 (100%)	-	-	
		?/-	3	2	-	-	2 (100%)	
		?/+	2	2	1 (50%)	-	1 (50%)	
		Total	2-4	18	12 (66%)	-	6 (34%)	
Spielmann et al. (1996)	IS(B)-10	+/+	2 3	12 1	11 (92%) -	- -	1 (8%) 1 (100%)	- -
		+/-	2	3	3 (100%)	-	-	-
		-/+	2 3	17 2	7 (41%) 1 (50%)	- -	- 1 (50%)	10 (59%) -
		-/-	2 3	31 2	30 (97%) 1 (50%)	- -	1 (3%) 1 (50%)	- -
		?/-	2 3	11 3	11 (100%) 1 (33%)	- -	- 2 (66%)	- -
		?/+	2 3	20 4	18 (90%) 1 (25%)	-	2 (10%) 2 (50%)	- 1 (25%)
		Total	2-3	106	84 (79%)		11 (10%)	11 (10%)

Report	Anal[1]	Classification (*In Vivo*/*In Vitro*)[2]	# of Labs	N[3]	Substances with 100% Agreement among Labs	Substances with 75-99% Agreement among Labs	Substances with 50-74% Agreement among Labs	Substances with 25-49% Agreement among Labs
Spielmann et al. (1996)	IS(B)-100	+/+	2	12	11 (92%)	-	1 (8%)	-
			3	1	1 (100%)	-	-	-
		+/-	2	1	1 (100%)	-	-	-
		-/+	2	28	21 (75%)	-	-	7 (25%)
			3	4	1 (25%)	-	3 (75%)	-
		-/-	2	17	16 (94%)	-	-	1 (6%)
		?/-	2	7	7 (100%)	-	-	-
			3	2	2 (100%)	-	-	-
		?/+	2	21	18 (86%)	-	-	3 (24%)
			3	2	2 (100%)	-	-	-
		Total	2-3	95	80 (84%)		4 (4%)	11 (11%)
Hagino et al. (1999)	IS(A)	+/+	5	7	5 (71%)	1 (14%)	1 (14%)	-
		+/-	-	-	-	-	-	-
		-/+	5	4	3 (75%)	1 (25%)	-	-
		-/-	5	3	1 (33%)	-	2 (66%)	-
		?/-	-	-	-	-	-	-
		?/+	5	2	2 (100%)	-	-	-
		Total	2-4	16	11 (69%)	2 (12%)	3 (19%)	-

Abbreviation: EU = European Union (EU 2001).
[1]Anal = analysis method used to transform the sample data into HET-CAM scores. IS(A) = method described in Luepke (1985); IS(B) = method described in Kalweit et al. (1987); Q = Q-Score, method described in Balls et al. (1995); S = S-Score, method described in Balls et al. (1995).
[2]A "+" indicates that the substance was assigned an overall classification of corrosive or severe irritant (Category R41); a "-" indicates that the substance was assigned an overall classification of nonsevere irritant (Category R36) or nonirritant; a "?" indicates that, due to the lack of appropriate *in vivo* data (i.e., insufficient dose volume), an EU classification could not be made. See Section 6.1 of the Draft HET-CAM BRD for a description of the rules followed to classify the ocular irritancy of test substances tested multiple times *in vitro*.
[3]N indicates number of substances.
[4]Number in parentheses indicates percentage of tested chemicals.

For the IS(B)-100 analysis method (Spielmann et al. 1996), the participating laboratories were in 100% agreement for 80 (84%) of the 95 substances tested. As with the IS(B)-10 analysis method, the extent of agreement between testing laboratories was greatest for substances correctly identified as EU nonsevere irritants or nonirritants by HET-CAM (94% [16/17]). Greater disparity between individual substance classifications was observed for substances that were identified as false positives (31% [10/32] false positive had less than 100% concordance between testing laboratories).

For the report by Hagino et al. (1999), there was 100% agreement in regard to the EU ocular irritancy classification for 11 (69%) of the 16 substances. Discordance in the classification results was observed for substances that were correctly identified as EU nonsevere irritants/nonirritants. Of three correctly identified EU nonsevere irritants/nonirritants, two substances exhibited less than 100% classification agreement among laboratories. Of the seven correctly identified EU corrosives/severe irritants, five substances (71%) produced the same classification in all five laboratories.

The overall reliability statistics, evaluated by HET-CAM data analysis method, for the IS(B), IS(B)-10, S-Score and Q-Score are identical to what is shown in **Table 7-7**. For the IS(A) and IS(B)-100 analysis methods, additional data laboratory data was available for a subset of the substances tested for each analysis method. For the IS(A) analysis method, the addition of two additional test substances evaluated by Kojima et al. (1995) yielded an overall concordance pattern consistent with what was observed for the Hagino et al. (1999) data alone. For the IS(B)-100 analysis method, the additional data from different testing laboratories were obtained from Gilleron et al. (1996, 1997) and Vinardell and Macián (1994). As with the IS(A) analysis method, the addition of the results from additional testing laboratories yielded a concordance pattern consistent with what was observed for Spielmann et al. (1996).

7.2.3.4 *Common Chemical or Product Classes Among Test Substances with Discordant Interlaboratory Results Based On Qualitative Analyses*

For each of the *in vivo* classifications systems, there were few substances that were evaluated in all reports discussed. Therefore, a direct comparison of the reliability of the analysis methods used by each report is limited. In general the ability of the HET-CAM test method to identify corrosives and severe irritants (for substances where there is repeated data to assess reproducibility and reliability) was similar between hazard classification systems evaluated. Therefore, conclusions about the HET-CAM reproducibility for one *in vivo* classification system generally apply to all classification systems, unless otherwise noted.

For the IS(A) analysis method, there were four false positive substances. The chemical classes represented by these substances included amidine, ether, carboxylic acid, amine, and alcohol. There were no chemicals or substances tested multiple times in different laboratories that were classified as false negatives to allow for an evaluation of common chemical product classes.

For the IS(B)-10 analysis method, the most common chemical classes shown to overpredicted, and where there were discordant results between testing laboratories was

alcohols. Other chemical classes, where discordant results were observed, included amines and phenols. Substances that were underpredicted tended to be underpredicted by all the testing laboratories that evaluated the substance.

For the IS(B)-100 analysis method, most of the substances evaluated produced the same response in all testing laboratories. Of the substances where there were discordant results, the chemical classes included, esters, aldehydes, and amines.

For alcohols that were evaluated using the Q-Score analysis method and were defined as false positives by the HET-CAM test method, the extent of agreement among laboratories was 75% (i.e., three of four laboratories classified the alcohol as a severe irritant). The extent of agreement among laboratories for the classification of esters (e.g., methyl acetate, ethyl-2-methylacetoacetate), which were false positives, ranged between 50% and 75%. Compared to the Q-Score, there were not enough tested substances within each *in vivo/in vitro* combination for S-Score or the IS(A) analysis methods to draw similar conclusions for the *in vivo* classification system.

7.2.3.5 *Quantitative Analysis of the Interlaboratory Reproducibility of Hazard Classification Category*

CEC (1991): Between three and five laboratories evaluated each substance tested at 100% concentration. A subset of substances was evaluated at a concentration of 10% by three testing laboratories. Based on the two different data sets, two different evaluations were conducted[2]. For the substances tested at a 100% concentration *in vitro*, substances tested by five laboratories (excluding Laboratories #5 and #6) were assessed[3]. For the substances tested at a 10% concentration *in vitro*, substances tested by three laboratories were assessed.

Using these criteria, %CV values for 14 substances evaluated at 100% concentration and 12 substances evaluated at 10% concentration were determined. The mean and median %CV values for substances evaluated at 100% concentration *in vitro* were 31.86 and 33.04, respectively (**Table 7-8**). The mean and median %CV values for substances evaluated at 10% concentration *in vitro* were 66.29 and 60.75, respectively (**Table 7-9**).

Balls et al. (1995): This evaluation used two different analysis methods, the S-Score and Q-Score. A description of each of these analyses methods is provided in **Section 5.0**. The use of these different analysis methods was dependent upon the transparency of the test materials. For substances where the reactions on the CAM could be observed the Q-Score was calculated. Comparatively, for substances where the reactions on the CAM could not be observed the S-Score was calculated (**Appendix A** provides a description of the differences in the test method protocols used for each analysis method).

[2] Data for these analyses are available in **Appendix F1.**

[3] Data from these testing laboratories were excluded from this analysis because the study report indicated that both laboratories had difficulty in identifying lysis and thrombosis/coagulation.

Table 7-8 %CV Values for Substances Evaluated at 100% Concentration *In Vitro* Using the IS(B) Analysis Method (from CEC 1991)

Substance[1]	Conc.	Mean IS(B) Score	SD	%CV Values
2-Butoxyethyl acetate	100%	4.76	0.31	6.58
Butanol	100%	11.44	1.0	8.71
Chloroform	100%	12.8	2.43	18.98
Triacetin	100%	4.18	0.91	21.76
Glycerol	100%	9.32	2.62	28.14
Tributyltin chloride	100%	8.94	2.88	32.21
Dimethyl sulfoxide	100%	9.88	3.24	32.83
Sodium dodecyl sulfate	100%	10.02	3.33	33.25
Triethanolamine	100%	8.52	2.94	34.55
Toluene	100%	11.04	4.31	39.06
2-Methoxyethanol	100%	9.14	3.72	40.65
Mercuric chloride	100%	10.52	4.57	43.44
n-Hexane	100%	5.04	3.16	62.78
Brij 35	100%	5.58	4.18	74.90
Mean %CV of substances tested at 100%				**31.86**
Median %CV of substances tested at 100%				**33.04**
Range %CV of substances tested at 100%				**6.58-74.90**

Abbreviations: %CV = percent coefficient of variation; Conc. = concentration tested; SD = standard deviation.
[1]Substances organized by increasing %CV values.

Between two and four laboratories evaluated each substance tested in this report. For this evaluation, only substances tested by all four laboratories were assessed[4]. Using this criteria, %CV values for 40 substances evaluated using the Q-Score and five substances evaluated using the S-Score were determined. The average and median %CV values for substances evaluated with the Q-Score were 49.83 and 42.50 (range of %CVs: 15.09 to 157.25), respectively (**Table 7-10**). The average and median %CV values for substances evaluated with the S-Score were 84.42 and 71.90 (range of %CVs: 68.47 to 116.4), respectively (**Table 7-11**).

The average and median %CV values for GHS Category 1 substances (UN 2003), based on *in vivo* results, were 36.26 and 38.93 for the Q-Score and 81.53 and 81.53 for the S-Score. The average and median %CV value for EPA Category I substances (EPA 1996), based on *in vivo* results, were, 33.54 and 34.81 for the Q-Score and 81.53 and 81.53 for the S-Score.

[4] Individual laboratory data is available in **Appendix C**.

Table 7-9 %CV Values for Substances Evaluated at 10% Concentration *In Vitro* Using the IS(B) Analysis Method (from CEC 1991)

Substance[1]	Conc.	Mean IS(B) Score	SD	%CV Values
Dimethyl sulfoxide	10%	4.20	0.17	4.12
Tributyltin chloride	10%	12.13	3.11	25.61
Acetic acid	10%	14.67	5.08	34.67
Butanol	10%	10.50	5.01	47.70
Glycerol	10%	5.57	2.74	49.27
Sodium dodecyl sulfate	10%	12.53	6.79	54.15
Chloroform	10%	7.20	4.85	67.36
2-Butoxyethyl acetate	10%	2.43	2.15	88.56
Triacetin	10%	6.30	6.36	100.88
2-Methoxyethanol	10%	3.37	3.51	104.19
Triethanolamine	10%	5.07	5.46	107.86
n-Hexane	10%	4.60	5.11	111.08
Mean %CV of substances tested at 10%				**66.29**
Median %CV of substances tested at 10%				**60.75**
Range %CV of substances tested at 10%				**4.12-111.08**

Abbreviations: %CV = percent coefficient of variation; Conc. = concentration tested; SD = standard deviation.
[1]Substances organized by increasing %CV values.

Spielmann et al. (1996): Individual laboratory results on tested substances were provided in response to a request by NICEATM[5]. In the evaluation, substances were evaluated at a 10% and 100% concentration in at least two different testing laboratories. Therefore, evaluation of the reliability of the test method was conducted for each concentration tested. Additionally, in order to resolve discrepancies in results between testing laboratories, some substances were tested in one additional testing laboratory (substances are italicized in **Table 7-12**). In order to determine if the substance tested in three laboratories affected the overall %CV values, an evaluation of the overall %CV values was conducted with these substances removed.

The average and median %CV values for substances tested at 10% concentration were 60.17 and 42.65, respectively. For substances tested at 100% concentration, the average and median %CV values were lower: 35.21 and 26.22, respectively. When substances that were tested in three different testing laboratories were removed from the assessment, little change was seen in the mean and median %CV values for both concentrations tested (**Table 7-12**).

[5] Individual laboratory data is available in **Appendix C**.

Table 7-10 %CV Values for Substances Evaluated Using the Q-Score Analysis Method (from Balls et al. 1995)

Substance[1]	Conc.	GHS Category 1	EPA Category I	Mean Q-Score	SD	%CV Values
2,2-Dimethylbutanoic acid	-	-	X	12.78	1.93	15.09
Trichloroacetic acid	30%	X	X	12.32	1.89	15.35
Benzalkonium chloride	1%	X	X	4.18	0.68	16.29
Sodium hydroxide	1%	-	-	5.42	0.99	18.20
Butyl acetate	-	-	-	1.63	0.31	18.95
Methyl cyanoacetate	-	-	-	1.38	0.34	24.84
Sodium lauryl sulfate	-	-	-	2.12	0.53	25.25
Triton X-100	5%	-	-	2.25	0.61	27.14
Octanol	-	-	-	1.67	0.47	28.15
Cyclohexanol	-	X	X	4.91	1.42	29.01
Benzalkonium chloride	10%	X	X	5.59	1.72	30.68
Ethyl-2-methylacetoacetate	-	-	-	2.09	0.66	31.74
Methyl isobutyl ketone	-	-	-	1.67	0.53	31.76
Cetylpyridinium bromide	6%	X	-	2.29	0.75	32.56
Triton X-100	10%	-	-	2.32	0.82	35.62
Hexanol	-	-	-	3.88	1.45	37.40
Methyl ethyl ketone	-	-	-	4.60	1.72	37.45
Toluene	-	-	-	3.73	1.41	37.98
Sodium lauryl sulfate	15%	X	X	2.84	1.11	38.93
Cetylpyridinium bromide	10%	X	X	2.98	1.21	40.60
Parafluoraniline	-	-	-	3.55	1.57	44.31
Polyethylene glycol 400	-	-	-	1.03	0.46	44.41
Pyridine	-	X	X	8.74	3.88	44.42
Tween 20	-	X	-	0.58	0.27	45.98
Sodium hydroxide	10%	X	X	13.44	6.74	50.12
Isobutanol	-	-	-	3.82	1.98	51.99
Trichloroacetic acid	3%	-	-	10.79	5.68	52.67
Benzalkonium chloride	5%	X	X	4.76	2.61	54.87
Ethyl acetate	-	-	-	2.52	1.39	55.11
Methyl acetate	-	-	-	3.03	1.70	56.12
Ethanol	-	-	-	6.13	3.75	61.16
Acetone	-	-	-	10.75	7.41	68.95
Glycerol	-	-	-	0.79	0.56	70.83
Isopropanol	-	-	-	5.96	4.23	71.93
2,6-Dichlorobenzoyl chloride	-	-	-	5.85	4.23	72.44
2-Ethyl-1-hexanol	-	-	-	1.49	1.12	74.75
Ethyl trimethyl acetate	-	-	-	0.40	0.41	103.70
gamma-Butyrolactone	-	-	-	8.67	9.12	105.19
Cetylpyridinium bromide	0.1%	-	-	0.86	1.15	134.05
Methylcyclopentane	-	-	-	2.42	3.81	157.25

Substance[1]	Conc.	GHS Category 1	EPA Category I	Mean Q-Score	SD	%CV Values
Mean for All Substances (n=40)	**-**	**-**	**-**	**-**	**-**	**49.83**
Median for All Substances	**-**	**-**	**-**	**-**	**-**	**42.50**
Range for All Substances	**-**	**-**	**-**	**-**	**-**	**15.09-157.25**
Mean for Severe Irritants (GHS) (n=11)	**-**	**-**	**-**	**-**	**-**	**36.26**
Median for Severe Irritants	**-**	**-**	**-**	**-**	**-**	**38.93**
Range for Severe Irritants	**-**	**-**	**-**	**-**	**-**	**15.35-54.87**
Mean for Severe Irritants (EPA) (n=8)	**-**	**-**	**-**	**-**	**-**	**33.54**
Median for Severe Irritants	**-**	**-**	**-**	**-**	**-**	**34.81**
Range for Severe Irritants	**-**	**-**	**-**	**-**	**-**	**15.35-54.87**

Abbreviations: %CV = percent coefficient of variation; Conc. = concentration tested; EPA = U.S. Environmental Protection Agency (EPA 1996); GHS = Globally Harmonized System (UN 2003); SD = standard deviation.
[1]Substances organized by increasing %CV values.

Table 7-11 %CV Values for Substances Evaluated Using the S-Score Analysis Method (from Balls et al. 1995)

Substance[2]	GHS Category 1	EPA Category I	Mean S-Score	Standard Deviation	%CV
4-Carboxybenzaldehyde	-	-	4	2.83	70.71
Fomasafen	-	-	5.25	3.77	71.90
1-Napthalene acetic acid	X	X	5.75	5.44	94.59
Sodium oxalate	X	X	8	5.48	68.47
Dibenzyl phosphate	-	-	8.25	9.60	116.42
Mean for All Substances (n=5)	**-**	**-**	**-**	**-**	**84.42**
Median for All Substances	**-**	**-**	**-**	**-**	**71.90**
Range for All Substances	**-**	**-**	**-**	**-**	**68.47-116.4**
Mean for Severe Irritants (GHS) (n=2)	**-**	**-**	**-**	**-**	**81.53**
Median for Severe Irritants	**-**	**-**	**-**	**-**	**81.53**
Range for Severe Irritants	**-**	**-**	**-**	**-**	**68.47-94.59**
Mean for Severe Irritants (EPA) (n=2)	**-**	**-**	**-**	**-**	**81.53**
Median for Severe Irritants	**-**	**-**	**-**	**-**	**81.53**
Range for Severe Irritants	**-**	**-**	**-**	**-**	**68.47-94.59**

Abbreviations: %CV = percent coefficient of variation; EPA = U.S. Environmental Protection Agency (EPA 1996); GHS = Globally Harmonized System (UN 2003).
[1]Substances organized by increasing %CV values.

Table 7-12 %CV Values for Substances Evaluated Using IS(B) Analysis Method (from Spielmann et al. 1996)

Substance Name[1]	CASRN	Mean IS(B)-10 Score	IS(B)-10 SD	%CV for IS(B)-10	Mean IS(B)-100 Score	IS(B)-100 SD	%CV for IS(B)-100
7-Acetoxyheptanal		1.55	2.19	141.42	10.95	8.56	78.14
n-Acetyl-methionine	1115-47-5	9.85	5.30	53.84	-	-	-
Ambuphylline	5634-34-4	13.25	3.61	27.22	14.85	2.90	19.52
4-Amino-5-methoxy-2-methylbenzenesulfonic acid	6471-78-9	9.80	4.34	44.29	12.17	3.20	26.31
Anisole	100-66-3	3.65	5.16	141.42	18.80	0.42	2.26
B 25		0.00	0.00	-	0.00	0.00	-
n-Butanal	123-72-8	3.95	3.89	98.46	19.20	1.56	8.10
n-Butanol	71-36-3	13.95	6.15	44.10	16.60	5.09	30.67
Butyl carbamate	592-35-8	6.80	5.93	87.21	12.67	1.93	15.27
Caffeine sodium benzoate	8000-95-1	6.37	1.66	26.11	13.10	5.31	40.52
Caffeine sodium salicylate	8002-85-5	8.60	1.70	19.73	17.40	1.98	11.38
Camphen	79-92-5	6.00	5.66	94.28	-	-	-
Cerium-2-ethylhexanoate	24593-34-8	7.40	0.71	9.56	17.18	2.93	17.09
1-Chloroctane-8-ol		5.55	1.77	31.85	16.50	3.11	18.86
3-Cyclohexene-1-methanol	1679-51-2	10.95	1.20	10.98	18.95	0.07	0.37
DC 8		0.00	0.00	-	2.50	3.54	141.42
1,4-Dibutoxybenzene	104-36-9	2.10	2.97	141.42	-	-	-
Diepoxid 126	2386-87-0	5.50	3.38	61.42	10.53	4.82	45.78
2,5-Dimethylhexanediol	110-03-2	6.65	3.61	54.23	13.85	3.89	28.08
3,6-Dimethyloctanol		0.15	0.21	141.42	4.30	0.00	0.00
4,4-Dimethyl-3-oxo-pentanenitrile	59997-51-2	4.95	0.92	18.57	6.20	0.71	11.40
1-(2,6-Dimethylphenoxy)-2-propanone	53012-41-2	7.42	9.99	134.67	11.80	7.60	64.42
Diphocars		14.70	5.09	34.63	15.10	3.96	26.22
1,2-Dodecanediol	1119-87-5	5.48	5.75	104.84	3.20	1.27	39.77
DTPA Pentasodium salt	140-01-2	15.58	0.11	0.73	19.65	0.35	1.80
Ede 140		1.70	2.40	141.42	2.30	3.25	141.42
1,2-Epoxydodecane	2855-19-8	2.05	2.90	141.42	4.95	5.02	101.42
Ethiosan		1.90	2.69	141.42	-	-	-
Ethyl butanal	97-96-1	1.80	2.55	141.42	18.05	0.92	5.09

Substance Name[1]	CASRN	Mean IS(B)-10 Score	IS(B)-10 SD	%CV for IS(B)-10	Mean IS(B)-100 Score	IS(B)-100 SD	%CV for IS(B)-100
Gadopentetic acid dimeglumine salt	86050-77-3	4.70	2.40	51.15	5.70	3.54	62.03
Genomoll	115-96-8	9.30	0.14	1.52	10.75	1.20	11.18
C12/C14-Glucoside		9.57	1.01	10.57	16.50	0.20	1.21
L-Glutamic acid hydrochloride	138-15-8	12.95	1.77	13.65	13.45	2.47	18.40
Glycediol		0.90	1.27	141.42	2.04	2.06	101.21
Granuform	30525-89-4	1.45	2.05	141.42	0.00	0.00	#DIV/0!
Hexahydrofarnesyl-acetone	502-69-2	1.75	0.78	44.45	6.10	2.69	44.05
Hexamethylenetetramine	100-97-0	5.05	1.06	21.00	11.15	0.07	0.63
1,2,6-Hexanetriol	106-69-4	7.90	5.09	64.45	17.05	2.47	14.52
Hnol		0.40	0.57	141.42	4.05	2.76	68.09
Hoe MBF		0.00	0.00	-	0.18	0.25	141.42
Hydo 98		11.65	1.77	15.17	-	-	-
2-Hydroxyethyl imino disodium acetate	135-37-5	11.15	3.18	28.54	13.25	3.18	24.01
2-Hydroxyisobutyric acid	594-61-6	12.85	2.90	22.56	13.45	3.04	22.61
Hypo 20		3.60	5.09	141.42	6.51	3.38	51.92
Hypo 36		4.10	0.14	3.45	12.95	4.17	32.22
Hypo 45		5.17	5.15	99.62	8.33	3.76	45.16
Hypo 54		4.15	0.21	5.11	4.15	0.07	1.70
Hyton		15.25	2.47	16.23	18.40	0.28	1.54
Iminodiacetic acid	142-73-4	8.25	7.43	90.01	6.85	5.98	87.23
Isobornyl acetate	125-12-2	2.90	1.70	58.52	6.35	2.47	38.97
Isobutanal	78-84-2	1.05	1.48	141.42	19.70	0.42	2.15
Isodecylglucoside		13.55	5.16	38.10	14.35	5.16	35.97
Isononylaldehyde	35127-50-5	0.00	0.00	-	7.25	3.89	53.64
alpha-Ketoglutaric acid	328-50-7	18.95	0.21	1.12	19.75	0.07	0.36
alpha-Lactid	4511-42-6	8.60	6.08	70.66	3.90	2.75	70.55
L-Lysine Monohydrate	39665-12-8	9.13	1.24	13.56	13.65	4.60	33.67
3-Mercapto-1,2,4-triazole	3179-31-5	11.30	9.90	87.61	-	-	-
m-Methoxybenzaldehyde	591-31-1	3.15	1.34	42.65	12.65	1.48	11.74
Methyl acetate	79-20-9	4.35	0.07	1.63	17.95	2.62	14.58
Methylpentynol	77-75-8	13.85	2.19	15.83	16.50	5.09	30.86

Substance Name[1]	CASRN	Mean IS(B)-10 Score	IS(B)-10 SD	%CV for IS(B)-10	Mean IS(B)-100 Score	IS(B)-100 SD	%CV for IS(B)-100
N-(2-Methylphenyl)-imidodi-carbonimidic diamide	93-69-6	17.40	0.42	2.44	-	-	-
2-Methyl-1-propanol	78-83-1	17.80	0.14	0.79	19.80	0.85	4.29
Methyltriglycol	112-35-6	4.50	0.57	12.57	14.75	3.18	21.57
Methyltriglycol	112-35-6	7.00	5.66	80.81	16.60	5.37	32.37
Napt		3.10	1.70	54.74	8.00	3.25	40.66
Nitro-bis-octylamide		0.85	1.20	141.42	4.05	3.46	85.55
Olak		17.50	1.98	11.31	18.25	1.77	9.69
Ölesulf		16.85	0.07	0.42	19.25	0.49	2.57
Phenylephrine hydrochloride	61-76-7	9.85	1.77	17.95	19.10	1.13	5.92
Phenylthiourea	103-85-5	2.00	2.83	141.42	1.55	2.19	141.42
Phosphonat A		6.70	0.14	2.11	6.80	4.67	68.63
Acefyllin piperazinate	18833-13-1	7.13	9.95	139.49	12.97	3.45	26.63
PO 2		2.15	3.04	141.42	0.15	0.21	141.42
Polyethylene glycol butyl ether	9004-77-7	13.30	3.39	25.52	19.25	0.07	0.37
Polyethylene glycol dimethyl ether	24991-55-7	2.05	2.90	141.42	13.70	8.63	62.97
Polyethylene glycol	25322-68-3	0.50	0.71	141.42	7.15	0.78	10.88
Polyhexamethylene guanidine		10.10	1.27	12.60	15.05	0.64	4.23
Polysolvan	7397-62-8	16.15	0.49	3.06	17.65	2.47	14.02
Potassium cyanate	590-28-3	17.30	2.12	12.26	17.65	2.47	14.02
Potassium hexacyanoferrate II	14459-95-1	16.50	1.84	11.14	11.75	7.71	65.60
Potassium hexacyanoferrate III	13756-66-2	5.23	1.45	27.74	6.08	0.53	8.73
2-Pseudojonon		5.75	4.17	72.56	5.70	2.26	39.70
RK Blau		2.00	2.83	141.42	-	-	-
Sacyclo		1.70	2.40	141.42	3.85	0.78	20.20
Sept		7.00	4.24	60.61	17.85	2.76	15.45
Trimethoxypropylsilane	1067-25-0	3.80	0.14	3.72	9.10	6.51	71.49
Trimethoxyoctylsilane	3069-40-7	5.00	4.10	82.02	9.20	1.13	12.30
Silan 165	29055-11-6	0.35	0.49	141.42	5.65	2.19	38.80
Silan 167	41453-78-5	1.40	1.84	131.32	3.50	1.70	48.49
Silan 253	18784-74-2	3.00	0.00	0.00	12.30	3.39	27.59
Sodium bisulfite	7631-90-5	13.30	0.85	6.38	18.40	2.26	12.30

Substance Name[1]	CASRN	Mean IS(B)-10 Score	IS(B)-10 SD	%CV for IS(B)-10	Mean IS(B)-100 Score	IS(B)-100 SD	%CV for IS(B)-100
Sodium sulfite	7757-83-7	12.25	1.34	10.97	14.20	2.69	18.92
Sodium cyanate	917-61-3	12.65	3.04	24.04	9.45	1.77	18.71
Sodium disilicate	13870-28-5	20.20	0.71	3.50	17.40	1.13	6.50
Sodium hydrogen sulfate	7681-38-1	17.75	1.48	8.37	18.65	0.78	4.17
Sodium lauryl ether sulfate	3088-31-1	14.10	5.09	36.11	18.45	0.78	4.22
Sodium monochloroacetate	3926-62-3	3.75	5.30	141.42	13.45	3.75	27.86
Sodiumpyrosulfite	7681-57-4	14.87	2.41	16.22	14.60	3.05	20.90
4-((2-Sulfatoethyl)sulfonyl)-aniline	2494-89-5	19.05	1.48	7.79	-	-	-
TA 01946 Alkylsilan		8.80	1.70	19.28	13.10	4.38	33.47
Theophylline sodium acetate	8002-89-9	9.40	5.66	60.18	-	-	-
Tocla		16.30	4.81	29.50	16.95	4.88	28.78
Triisooctylamine	25549-16-0	0.40	0.57	141.42	9.05	7.14	78.91
2,2,3-Trimethyl-3-cyclo-pentene-1-acetaldehyde	4501-58-0	2.60	0.42	16.32	12.20	3.54	28.98
Trioxane	110-88-3	11.33	2.93	25.91	17.90	0.14	0.79
Wessalith Slurry		6.57	4.86	74.00	9.90	8.20	82.85
Xanthinol nicotinate	437-74-1	7.65	5.16	67.48	13.20	5.94	45.00
Mean %CV Value				**60.17**			**35.21**
Median %CV Value				**42.65**			**26.22**
Range %CVs				**0-141.42**			**0-141.42**
Mean %CV Value (Minus Substances Tested in 3 Laboratories)				**58.07**			**34.62**
Median %CV Value (Minus Substances Tested in 3 Laboratories)				**31.85**			**21.57**
Range %CVs (Minus Substances Tested in 3 Laboratories)				**0-141.42**			**0-141.42**

Abbreviations: CV = coefficient of variation; CASRN = Chemical Abstract Service Registry Number.
[1]Italicized substances represent chemicals that were tested in three testing laboratories. Data for these substances were removed to determine their impact on the calculated %CV values for this data set.

Hagino et al. (1999) and Ohno et al. (1999): The Japanese Ministry of Health and Welfare evaluated the HET-CAM test method in five different laboratories as part of a validation effort to assess alternative ocular irritation test method. Nine, 15, and 14 cosmetic ingredients were evaluated in the first, second, and third steps of the validation study, respectively. These studies used the IS(A) analysis method to assess potential irritancy classifications. Average individual laboratory results and standard deviations for tested substances were reported in Hagino et al. (1999). **Appendix F2** provides the average IS(A) values for each testing laboratory for each substance evaluated in this validation effort.

The interlaboratory reproducibility was evaluated by comparing the mean %CV values. The evaluation showed that for the chemicals evaluated, the mean %CV values were: 50.2 for the ten substances evaluated in the first phase of the validation study, 114.0 for the 44 substances evaluated in the second phase of the validation study, and 39.2 for the 42 substances evaluated in the third phase of the validation study. The mean %CV value for all 96 substances (when the three phases were pooled) was 74.6. The investigators proposed that the relatively high %CV was caused by variations of the results of nonirritants, which had low *in vitro* scores. When nonirritants were removed from the analysis, the mean %CV value was 45.8 (n=68).

The average and median %CV for substances classified as GHS Category 1 (UN 2003) for the substances described in Hagino et al. (1999)[6], which described the third validation phase, were 24.4 and 27.0, respectively (see **Table 7-13**)[7]. The average and median %CV for substances classified as EPA Category I (EPA 1996) were 25.86 and 26.43, respectively (see **Table 7-13**).

Table 7-13 %CV Values for Substances Evaluated Using the IS(A) Analysis Method (from Hagino et al. 1999)

Substance[1]	Conc.	GHS Category 1	EPA Category I	Mean IS(A) Score	SD	%CV
Acetic acid	10%	X	X	17.35	1.34	7.73
Stearyltrimethylammonium chloride	10%	X	X	13.60	3.00	22.08
Potassium laurate	10%	X	X	15.32	4.00	26.18
Domiphen bromide	10%	X	X	14.05	3.71	26.43
Butanol	10%	X		9.70	2.69	27.72
di-(2-Ethylhexyl) sodium sulfosuccinate	10%	X	X	9.45	2.62	27.78
Cetyltrimethylammonium bromide	10%	X	X	14.15	4.46	31.55
Lactic acid	100%	X	X	14	5.50	39.26
Mean for Severe Irritants (GHS) (n=8)						**26.09**
Median for Severe Irritants						**27.08**
Range for Severe Irritants						**7.73-39.26**
Mean for Severe Irritants (EPA) (n=7)						**25.86**
Median for Severe Irritants						**26.43**
Range for Severe Irritants						**7.73-39.26**

Abbreviations: %CV = percent coefficient of variation; Conc. = concentration tested; EPA = U.S. Environmental Protection Agency (EPA 1996); GHS = Globally Harmonized System (UN 2003).
[1]Substances organized by increasing %CV values.

[6] Percent CV values were not determined for the other phases, because average data were not provided in literature references.
[7] Individual laboratory data is available in **Appendix C**.

7.2.4 Additional Analyses of Interlaboratory Reproducibility

7.2.4.1 *Balls et al. (1995)*

The investigators of this study presented interlaboratory correlation coefficients between each pair wise combination of laboratories that were involved in the testing phase of the validation study. For example, interlaboratory correlation coefficients of the *in vitro* data for all the tested substances were developed for Laboratory A when compared to Laboratory B, C, D, and E. Summary of the interlaboratory correlation coefficients calculated in this analysis are provided in **Table 7-14** (see **Appendix G** for all correlation coefficients derived from comparing each laboratory with every other laboratory).

Table 7-14 Interlaboratory Correlation Coefficients in Balls et al. (1995)

Index Score	Interlaboratory Pearson's Correlation (r) of the *In Vitro* Data
Full set of test substances (11-49 depending on endpoint)	
HET-CAM Q-Score	0.473-0.790
HET-CAM S-Score	-0.171-0.808
HET-CAM Q-Score, with cut-off at 2	0.449-0.814
HET-CAM S-Score, with cut-off at 2	-0.316- -0.043
Chemicals soluble in water (5-25 depending on endpoint)	
HET-CAM Q-Score	0.355-0.711
HET-CAM S-Score	0.420-0.949
HET-CAM Q-Score, with cut-off at 2	0.470-0.927
HET-CAM S-Score, with cut-off at 2	Not Evaluated
Chemicals insoluble in water (4-12 depending on endpoint)	
HET-CAM Q-Score	0.580-0.944
HET-CAM S-Score	-0.910-0.852
HET-CAM Q-Score, with cut-off at 2	0.562-0.816
HET-CAM S-Score, with cut-off at 2	Not Evaluated
Surfactants (12)	
HET-CAM Q-Score	0.438-0.876
HET-CAM S-Score	Not Evaluated
HET-CAM Q-Score, with cut-off at 2	0.420-0.966
HET-CAM S-Score, with cut-off at 2	Not Evaluated
Solids (7-17 depending on endpoint)	
HET-CAM Q-Score	0.500
HET-CAM S-Score	-0.171-0.808
HET-CAM Q-Score, with cut-off at 2	0.985
HET-CAM S-Score, with cut-off at 2	Not Evaluated
Solutions (14 depending on endpoint)	
HET-CAM Q-Score	0.712-0.880
HET-CAM S-Score	Not Evaluated
HET-CAM Q-Score, with cut-off at 2	0.590-0.974
HET-CAM S-Score, with cut-off at 2	Not Evaluated
Liquids (26)	
HET-CAM Q-Score	0.221-0.755
HET-CAM S-Score	Not Evaluated
HET-CAM Q-Score, with cut-off at 2	0.591-0.771
HET-CAM S-Score, with cut-off at 2	Not Evaluated

For some of the endpoints, the range in correlation coefficients was rather large (e.g., correlation coefficients for chemicals insoluble in water ranged for the HET-CAM S-score endpoint from -0.910 to 0.852). There also were a large number of negative correlation coefficients noted. Review of the results did not indicate that there was one specific laboratory that yielded consistently high or low correlation coefficients.

7.2.4.2 *Blein et al. (1991)*

The investigators assessed the intralaboratory reproducibility with four substances (propylene glycol, Tween 20, SDS, and benzalkonium chloride). There was no rationale provided for the selection of these substances. The report indicated that the reproducibility of results for each substance was good within each laboratory (data not provided). Interlaboratory reproducibility evaluations were conducted with the same four substances and results with diluted and undiluted substances were examined. This analysis indicated that there were no significant differences ($p = 0.055$) in HET-CAM scores between the laboratories when diluted products were evaluated. However, there was a significant difference ($p = 0.01$) in HET-CAM scores when undiluted products were evaluated.[8]

7.2.4.3 *Doucet et al. (1999)*

Comparative screening of 40 cosmetic formulations was conducted to assess the usefulness of *in vitro* ocular toxicity test methods, including the HET-CAM test method, when compared to the currently accepted *in vivo* rabbit eye test method. The formulations were classified as skin care products (10), sunscreen products (10), surfactant based products (10), and alcoholic products (10). In this study, the *in vitro* scores (calculated as IS) were calculated as described in Luepke (1985). A substance with an IS value greater than five was defined as an irritant. Comparative *in vivo* results (calculated as MMAS) were calculated from concurrently conducted studies run according to the method described by Draize et al. (1944). A substance with an MMAS value greater than 15 was defined as an irritant.

An intralaboratory evaluation was conducted with one of the surfactant-based products (not identified) classified as an irritant. Twenty HET-CAM experiments with this substance were conducted; a %CV of 10 was obtained.

7.2.4.4 *Spielmann et al. (1991)*

Starting in 1988, a national validation study on two alternative ocular toxicity test methods was initiated by ZEBET. In this report, 27 coded substances that represented a variety of chemical and toxicological properties were evaluated in 12 laboratories to assess intralaboratory and interlaboratory reproducibility of the test method. All but four substances were evaluated at 10% concentrations; these four remaining substances were evaluated at concentrations ranging from 0.5% to 100%. The lowest concentration required to produce a slight reaction on the CAM also was determined. In this report, the *in vitro* scores (calculated as IS) were calculated as described in Kalweit et al. (1990). The irritation classification scheme used in the evaluation was performed according to Luepke (1985). The *in vivo* results (classified into irritation categories per an investigator defined classification

[8] In the report, the authors refer to a table (Table 2) that contains the results of the interlaboratory analysis. However, the table is not shown in the report. Therefore, review of the results by NICEATM could not be conducted.

system) were obtained from historical results for studies that were conducted as described by Draize et al. (1944).

For the analysis presented in this report, the classifications for each laboratory for each substance were determined. The irritation classification made by a majority of the laboratories was determined to be the *in vitro* classification call for the substance. The investigators then stated that if 75% of all the laboratories determined a correct classification (i.e., *in vitro* classification was the same as the *in vivo* classification), then the overall call by the testing laboratories was "correct."

A review of the data presented in the report indicates that nine of 12 substances that were classified as a corrosive or severe irritant based on the Draize test result were correctly classified by a majority of the testing laboratories when using the HET-CAM test method. For eight of the correctly identified severe irritants, between 80% and 100% of the testing laboratories classified the test substances as a strong irritant.[9] For the remaining three severe irritants, two were classified as inconclusive and one substance was classified as a negative (nonsevere irritant; i.e., false negative) by 90% of the testing laboratories.

There were 15 substances classified as a nonirritant, slight irritant, or moderate irritant based on a Draize test result. Of these substances, seven were correctly identified as nonsevere irritant substances by at least 75% of the testing laboratories. Of the remaining eight substances, five were identified as false positives and three were classified as inconclusive (i.e., a majority [75%] of the testing laboratories did not classify the test substances as an irritant or nonsevere irritant). The concordance between testing laboratories for the false positives ranged from 75% to 91% of the laboratories.

7.2.4.5 *Spielmann et al. (1993)*

Starting in 1988, a national validation study on two alternative ocular toxicity test methods was initiated by ZEBET. In this second report, 136 coded substances that represented a variety of chemical and toxicological properties were discussed. The substances tested were evaluated at 10% concentrations. The lowest concentration required to produce a slight reaction on the CAM also was determined. The studies were conducted in two laboratories with experience in the test method. In this report, the *in vitro* scores (calculated as IS) were calculated as described by Kalweit et al. (1990). The irritation classification scheme used in the evaluation used both the IS and ITC values. The *in vivo* scores (classified per the EU classification scheme [EU 1992]) were obtained from historical results that were conducted as described by Draize et al. (1944) and were conducted in accordance with GLP standards.

Of the 136 substances tested, 46 were classified as severe irritants (R41) based on *in vivo* studies. Of these 46 R41 substances, both laboratories correctly identified 22 of the substances (48%) as severe irritants. For the remaining 24 substances, 15 were classified as nonirritant or moderately irritant by both laboratories, two were classified as nonirritant or moderately irritant by one laboratory and irritant by the other, four were classified as irritant

[9] For one substances (SDS) it is stated, "since even the low concentration of 1% led to predominately high scores, the substance was classified as a correct positive." (Spielmann et al. 1991)

(R36) by both laboratories, one was classified as an irritant by one laboratory and severely irritant by the other, and two were identified as inconclusive.

The remaining 90 substances were classified as nonsevere irritants and nonirritants based *in vivo* results (10 substances were R36 and 80 were nonirritant). Of these substances, both laboratories classified 65 of the substances (72%) as nonsevere irritants (R36) and nonirritants.

7.3 Historical Positive and Negative Control Data

7.3.1 Historical Positive Control Data

Historical positive control data were obtained from two sources, in response to a request from NICEATM. For one set of data, positive control substances were dimethyl formamide (DMF) and imidazole. Studies were conducted with and without the use of a TSA. For a second set of data SDS and NaOH were used. For the negative control studies a TSA was not used.

7.3.1.1 *Positive Control Studies Using DMF and Imidazole*

Positive control studies were conducted with imidazole and DMF (see **Appendix H1-H2**). With the DMF studies that were conducted with the TSA, the hemorrhage endpoint was evaluated inside and outside the TSA. Of note, the time of development of the hemorrhage endpoint inside the TSA was lower than the time to development of the hemorrhage endpoint outside the TSA (**Table 7-15**). Two proposed reasons for the difference in time to development, according to Dr. Vanparys (submitter of the data), are (1) the vessels outside the TSA may open more easily than those under the TSA, or (2) once the liquid is applied it accumulates around the edge of the TSA rather than between the TSA and CAM.

Table 7-15 Comparison of Means and Standard Deviations for Positive Controls Tested With and Without Test Substance Applicator

Positive Control	N[1]	Hemorrhage[2] (mean ± SD[4])	Lysis[2] (mean ± SD)	Coagulation[2] (mean ± SD)	*In Vitro* Score[3] (mean ± SD)
DMF: With TSA[4]	69	0.02 ± 0.17	6.93 ± 0.03	8.82 ± 15.77	15.77 ± 0.19
DMF: With TSA[4]	10	3.36 ± 0.32	6.54 ± 0.19	8.81 ± 0.04	18.71 ± 0.38
DMF: Without TSA	2	4.00 ± 0.13	6.84 ± 0.05	8.76 ± 0.08	19.60 ± 0.15
Imidazole: Without TSA	15	4.50 ± 0.39	6.84 ± 0.08	8.66 ± 0.17	20.00 ± 0.45

Abbreviations: DMF = dimethylformamide; SD = standard deviation; TSA = test substance applicator (as described in Gilleron et al. 1996, 1997)
[1]N = number of tests.
[2]Mean values of time until development of identified endpoint.
[3]*In Vitro* irritation score calculated as IS(B).
[4]Hemorrhage endpoint in studies described in the first row were evaluated inside the TSA, while hemorrhage endpoint in studies described in the second row were evaluated outside the TSA.

Using the data provided, the intralaboratory reproducibility of the positive controls was evaluated. For the positive control imidazole, the %CV values were calculated for each endpoint as well as for the overall IS(B) score. The range of %CV values was 0.12 to 18.97

for the hemorrhage endpoint, 0.34 to 1.20 for the lysis endpoint, and 0.20 to 2.11 for the coagulation endpoint. The range of %CV values for the overall IS(B) score was 0.12 to 1.58. The average and median %CV values for the overall IS(B) score were 0.97 and 0.50, respectively (**Table 7-16**).

Table 7-16 Intralaboratory Reproducibility Results for Evaluation of Imidazole as a Positive Control

	Hemorrhage Endpoint	Lysis Endpoint	Coagulation Endpoint	Overall Irritation Score
Mean (SD)	4.5 (0.39)	6.84 (0.08)	8.66 (0.17)	20.00 (0.45)
Range of %CV	0.12 – 18.97	0.34-1.20	0.20-2.11	0.12-1.58
Overall %CV	8.6	1.10	1.99	2.23
Mean Total Score %CV	0.97			
Median Total Score %CV	0.50			

Abbreviations: %CV = percent coefficient of variation; SD = standard deviation.

For the positive control DMF, the data where hemorrhages develop inside the TSA was evaluated. The range of %CV values was 0.00 to 1.27 for the lysis endpoint and 0.00 to 1.76 for the coagulation endpoint. For the hemorrhage endpoint, a single test produced a result other than zero for the mean and the tested eggs and the standard deviation; the %CV value for the single test was 173.94. The range of %CV values for the overall IS(B) score was 0.04 to 14.07. The average and median %CV values for the overall IS(B) score were 0.59 and 0.29, respectively (**Table 7-17**).

Table 7-17 Intralaboratory Reproducibility Results for Evaluation of DMF as a Positive Control

	Hemorrhage Endpoint	Lysis Endpoint	Coagulation Endpoint	Overall Irritation Score
Mean (SD)	0.02 (0.17)	6.93 (0.03)	8.82 (0.09)	15.77 (0.19)
Range[1] of %CV values	173.94[1]	0.00-1.27	0.00-1.76	0.04-14.07
Overall %CV	850	0.49	1.05	1.20
Mean Total Score %CV	0.59			
Median Total Score %CV	0.29			

Abbreviations: %CV = percent coefficient of variation; SD = standard deviation/
[1]Range is representative of a single value since CV values for other experiments could not be calculated, since mean and SD values were zero.

7.3.1.2 *Positive Control Studies Using SDS and NaOH*

HET-CAM studies using 1% SDS and 0.1 N NaOH were provided in response to a request from NICEATM. Additional information on these data, as well as an alternative analysis conducted, is provided in **Appendix H3**. Using the mean values determined for these studies, the overall irritation score calculated (according to the method of Kalweit et al. 1987, 1990) for these substances classified them as irritants (**Table 7-18**).

Table 7-18 Means and Standard Deviations of Positive Control Test Substances

Positive Control	Hemorrhage[1] (mean ± SD[2])	Lysis[1] (mean ± SD)	Coagulation[1] (mean ± SD)
1% SDS (n=377)	14.69 ± 5.36	35.18 ± 17.15	---[2]
0.1 N NaOH (n=336)	8.96 ± 4.96	35.60 ± 24.71	48.04 ± 34.56

Abbreviations: NaOH = sodium hydroxide; SD = standard deviation; SDS = sodium dodecyl sulfate.
[1]Mean values of time until development of identified endpoint.
[2]It was indicated that 1% SDS does not produce coagulation in the CAM after application. However, in the studies conducted coagulation was identified in a single study. In these evaluations, the non-existing data was calculated with an arbitrary value of "0." Therefore, the calculation of a mean value for the coagulation endpoint was not meaningful.

7.3.2 Historical Negative Control Data

HET-CAM studies using 0.9% NaCl as a negative control were provided in response to a request from NICEATM. Studies were conducted with and without the use of a TSA (see **Appendix I**). The use of a TSA was described in Gilleron et al. (1996, 1997) (see **Section 2.2.4.3**).

Over 90 tests with 0.9% sodium chloride (NaCl) using the TSA and three tests with 0.9% NaCl without using TSA were provided. As shown in **Table 7-19**, time to development of endpoints and the overall irritation scores calculated were consistent and classified as nonirritants for all tests.

Table 7-19 Comparison of Means and Standard Deviations of 0.9% NaCl With and Without Use of the Test Substance Applicator

0.9% NaCl	N[1]	Hemorrhage[2] (mean ± SD)	Lysis[2] (mean ± SD)	Coagulation[2] (mean ± SD)	*In Vitro* Score[3] (mean ± SD)
With TSA	92	0 ± 0	0 ± 0	0 ± 0	0 ± 0
Without TSA	3	0 ± 0	0 ± 0	0 ± 0	0 ± 0

Abbreviations: NaCl = sodium chloride; SD = standard deviation; TSA = test substance applicator.
[1]N = number of tests
[2]Mean values of time until development of identified endpoint.
[3]*In Vitro* irritation score calculated as IS(B).

7.4 Summary

The analysis of intralaboratory repeatability was evaluated using data from two different publications (Gilleron et al. 1996, 1997) for the IS(B) analysis method. In both studies, the hemorrhage endpoint had a high %CV value (104-117). Additionally, the %CV values for the coagulation endpoint were the lowest of the three endpoints evaluated in the HET-CAM test method. However, the actual values were quite disparate between the two studies (e.g., Gilleron et al. 1996 coagulation %CV = 95.69, Gilleron et al. 1997 coagulation %CV = 41.78). The difference in the numbers may be due to several factors including test substances evaluated and differences in the test method protocols used between the two studies. The calculated variability for the endpoints and the overall test method may be exaggerated because of the relatively small values that are obtained from each of the

endpoints (5 for hemorrhage, 7 for lysis, and 9 for coagulation). Similar results were obtained from the analysis of intralaboratory reproducibility. The overall irritation score was generally reproducible (%CV values of 53 and 17.5 for the two studies evaluated).

A qualitative assessment of the data provided for multiple laboratories in three to four studies indicates the extent of interlaboratory reproducibility. Given the relatively homogeneous performance of the HET-CAM test method among the three classification systems, the discussions for the individual studies and analysis methods encompasses all three hazard classification systems, unless otherwise indicated.

In an assessment of interlaboratory reproducibility of hazard classification (EPA, EU, or GHS), the two to four participating laboratories for the Balls et al. (1995) study were in 100% agreement in regard to the ocular irritancy classification for 21 (45%) of the 47 substances analyzed using the Q-Score analysis method. The extent of agreement between testing laboratories for the Q-Score analysis method was greatest for substances correctly identified as corrosives or severe irritants when compared to any other combination of *in vivo* and *in vitro* results (60% to 71% [9/15 to 10/14] of the accurately identified severe substances were shown to have 100% classification agreement among testing laboratories, depending on the classification system). Comparatively, participating laboratories were in 100% agreement for 12 to 13 (66% to 68%) of the 18 to 19 substances analyzed using the S-Score analysis method, depending on the classification system used.

For the IS(B)-10 analysis methods (Spielmann et al. 1996), the participating laboratories were in 100% agreement for 84 to 85 (79% to 81%) of 104 to 107 substances evaluated. The extent of agreement between testing laboratories was greatest for substances correctly identified as GHS nonsevere irritants or nonirritants by HET-CAM (94% to 97% [31/33 to 32/33]). Comparatively, greater disparity between individual substance classifications, for all hazard classifications, was observed for substances that were identified as false positives (52% to 58% false positive had less than 100% concordance between testing laboratories).

For the IS(B)-100 analysis method, the participating laboratories were in 100% agreement for 80 to 81 (82% to 84%) of the 95 to 99 substances evaluated. As with the IS(B)-10 analysis method, the extent of agreement between testing laboratories was greatest for substances correctly identified as GHS nonsevere irritants or nonirritants by HET-CAM and greater disparity between individual substance classifications was observed for substances that were identified as false positives.

For the report by Hagino et al. (1999), there was 100% agreement in regard to the GHS ocular irritancy classification for 11 (64% to 69%) of the 16 to 17 substances evaluated in five laboratories. Discordance in the classification results was present for substances that were correctly identified as corrosives/severe irritants and as nonsevere irritants/nonirritants. Substances classified as false positives had the greatest extent of agreement among laboratories.

Sufficient *in vivo* information for the CEC (1991) study was only available to assess the interlaboratory reproducibility performance for the EU classification system. For the CEC

evaluation, the participating laboratories were in 100% agreement in regard to the EU ocular irritancy classification for 6 (23%) of the 26 substances tested. The extent of agreement among laboratories was greatest for accurately identified EU corrosives/severe irritants when compared to any other combination of *in vivo* and *in vitro* results (50% [3/6] of the identified EU 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 and those substances accurately classified as EU nonsevere irritants/nonirritants.

The overall reliability statistics, arranged by HET-CAM data analysis method, for the IS(B), IS(B)-10, S-Score and Q-Score are identical to what is shown in **Table 7-5, 7-6,** and **7-7**. For the IS(A) and IS(B)-100 analysis methods, additional data laboratory data was available for a subset of the substances tested for each analysis method. For both of these analysis methods, the addition of the results from additional testing laboratories yielded a concordance pattern consistent with what was observed for Hagino et al. (199) and Spielmann et al. (1996).

Substances with less than complete agreement in the testing laboratories depended upon the analysis method evaluated. For the IS(A) analysis method, chemical classes included amidine, ether, carboxylic acid, amine, and alcohol. For the IS(B)-10 analysis method, the most common chemical classes shown to be overpredicted, and where there were discordant results between testing laboratories, was alcohols. For the Q-Score analysis method, alcohols were shown to produce discordant results between testing laboratories.

A quantitative evaluation of interlaboratory reproducibility was conducted for four studies (CEC 1991; Balls et al. 1995; Spielmann et al. 1996; Hagino et al. 1999) by performing a %CV analysis of *in vitro* scores obtained for substances tested in multiple laboratories. For CEC (1991), two different evaluations were conducted based on the concentration tested *in vitro*. For 14 substances evaluated at 100% concentration, the mean and median %CV values were 31.86 and 33.04, respectively. For 12 substances evaluated at 10% concentration, the mean and median %CV values were 34.6 and 33.1, respectively. For the Balls et al. (1995) study, the average and median %CV values for substances evaluated with the Q-Score were 49.83 and 42.50, respectively. The average and median %CV values for the substances evaluated with the S-Score were 84.42 and 71.90, respectively. For the substances evaluated in Spielmann et al. (1996), the average and median %CV values for substances tested at 10% concentration were 60.17 and 42.65, respectively. For substances tested at 100% concentration in Spielmann et al. (1996), the average and median %CV values were lower: 35.21 and 26.22, respectively. When substances that were tested in three different testing laboratories were removed from the assessment, little change was seen in the mean and median %CV values for both concentrations tested. For Hagino et al. (1999), the average and median %CV for substances classified as GHS Category 1 (UN 2003) were 24.4 and 27.0, respectively. The average and median %CV for substances classified as EPA Category I (EPA [1996]) were 23.86 and 26.0, respectively.

Finally, historical positive and negative control data were provided by two different sources. The negative control substance evaluated was 0.9% NaCl. The positive control substances were DMF, imidazole, 1% SDS, and 0.1 N NaOH. The studies showed that, between

experiments, the results for all control substances were reproducible. Additionally, studies indicated that all control substances consistently produced appropriate responses (e.g., negative control consistently produced a response that would be classified as nonirritant and positive controls consistently produced a response that would be classified as severe irritant).

[This Page Intentionally Left Blank]