7.0 ICE 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.

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

The quality of a reliability evaluation depends on the extent to which the substances tested adequately represent the range of physicochemical characteristics and response levels that the test method must be capable of evaluating. The only data source for conducting an assessment of ICE test method reliability was Balls et al. (1995). This study evaluated the performance and reproducibility of the ICE 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 different concentrations, for a total of 60 possible ocular irritation outcomes). To be selected for inclusion in this study, the substances had to be single chemicals (no mixtures) available at high purity and stable when stored, and the reference *in vivo* rabbit eye data must have been generated since 1981 according to OECD TG 405, following GLP guidelines. In addition, substances were selected to ensure an adequately diverse group of physicochemical characteristics and levels of irritancy severity. One substance (thiourea) was tested *in vitro* in the ICE assay but, due to its excessive toxicity *in vivo*, was excluded from the comparison of *in vitro* and *in vivo* test results (see **Section 3.1.2**).

An unpublished study (Prinsen 2000) provided data from a single laboratory that tested four substances (two surfactants and two siloxanes) in four to five separate experiments, which allowed for evaluation of intralaboratory repeatability and reproducibility. Each experiment used three eyes. One of these substances was classified as a *non-irritant* (EU classification NI), two substances were classified as *irritating to the eyes* (EU classification R36) and one was classified as *severely irritating to the eyes* (EU classification R41).

7.2 Analyses of Repeatability and Reproducibility

7.2.1 Quantitative and Qualitative Assessments of Intralaboratory Repeatability

Generally, analyses of intralaboratory repeatability have included approaches such 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 (ANOVA) methods, (e.g., Holzhütter et al. 1996; ASTM 1999).

A CV analysis was done on within-experiment data from Prinsen (2000), using scores for each endpoint (i.e., corneal thickness/swelling, corneal opacity, fluorescein retention) and the ICE Irritation Index, for each test substance (**Table 7-1**). When considering the results of this analysis, note that some test substances had a mean or a standard deviation equal to zero for some endpoints and that scores for corneal opacity and fluorescein retention have a small dynamic range (0 to 4 and 0 to 3, respectively).

Corneal thickness measurements within experiments showed %CV values ranging from 0.9 to 6.1 and corneal opacity scores showed %CV values ranging from zero to 86.6 (the highest value was obtained for the nonirritating substance). The %CV values for fluorescein retention were zero for three of the four substances and ranged from zero to 86.6 for the non-irritating substance, although this range is based on only two experiments. Finally, the %CV values for the ICE Irritation Index for the four substances ranged from -86.6 to 41.6, with the nonirritating substance exhibiting the outlying values (-86.6 and 41.6).

7.2.2 Quantitative and Qualitative Assessment of Intralaboratory Reproducibility

Generally, analyses of intralaboratory reproducibility have included approaches such 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 (ANOVA) methods, (e.g., Holzhütter et al. 1996; ASTM 1999).

The data from Prinsen (2000) was also used to do a CV analysis on between-experiment values for each endpoint (i.e., corneal thickness/swelling, corneal opacity, fluorescein retention) along with the ICE Irritation Index, for each test substance (**Table 7-2**). When considering the results of this analysis, note that scores for corneal opacity or fluorescein retention have a small dynamic range (0 to 4 and 0 to 3, respectively).

The %CV values for the corneal thickness measurement ranged from 1.8 to 6.3 and those for corneal swelling ranged from 13.9 to 138.7. The %CV values for the corneal opacity score ranged from 8.7 to 95.8. The %CV values for the fluorescein retention score ranged from zero to 141.4. Finally, the %CV values for the ICE Irritation Index ranged from 4.1 to 91.8. Note that for all endpoints considered except corneal thickness, the highest %CV values were obtained for the non-irritating substance.

Table 7-1 Intralaboratory Repeatability of ICE Test Method Endpoints – Prinsen (2000)

Substance (Experiment No.[1])	EU[2] Class[3]	CT[4] (mean[5])	CT (%CV[6])	CS[7] (mean)	CS (%CV)	CO[8] (mean)	CO (%CV)	FR[9] (mean)	FR (%CV)	Index[10] (mean)	Index (%CV)
SP-1 (1)[11]	NI	60	3.3	0.7	346.4	0.3	86.6	0.3	86.6	15	41.6
SP-1 (2)	NI	63.3	3.3	1.7	91.6	0.3	86.6	0.5	0	18.3	39.4
SP-1 (3)	NI	62.3	2.4	2.3	24.7	0.5	0	0	-	12.3	4.7
SP-1 (4)	NI	61.7	0.9	-1.3	-86.6	0	-	0	-	-1.3	-86.6
SP-1 (5)	NI	63.3	0.9	2	0	0	-	0	-	2	0
SP-4 (1)	R36	68.7	3.0	14.3	24.5	3	0	2	0	114.3	3.1
SP-4 (2)	R36	69.3	3.0	13.3	40.0	2	0	2	0	93.3	5.3
SP-4 (3)	R36	75.7	3.3	21	23.8	2.7	21.6	2	0	114.3	14.0
SP-4 (4)	R36	69.7	4.4	14	49.5	2.7	21.6	2	0	107.3	15.1
SP-5 (5)	R36	70	3.8	12.7	27.7	2	0	2	0	92.7	3.8
SU-4 (1)	R36	72	2.4	13.7	18.4	0.7	43.3	1	0	47	16.9
SU-4 (2)	R36	68.7	3.4	14	12.4	0.7	43.3	1	0	47.3	8.5
SU-4 (3)	R36	67.7	6.0	13	15.4	0.7	43.3	1	0	46.3	9.0
SU-4 (4)	R36	66.7	3.5	11	31.5	0.8	34.6	1	0	47.7	10.6
SU-4 (5)	R36	67.7	2.2	9.7	15.8	0.7	43.3	1	0	43	16.3
SU-5 (1)	R41	77.7	1.5	23	24.2	2	0	2	0	103	5.4
SU-5 (2)	R41	74.7	4.7	20.7	19.6	2	0	2	0	100.7	4.0
SU-5 (3)	R41	75.3	6.1	21	9.5	2	0	2	0	101	2.0
SU-5 (4)	R41	76.7	2.0	16.3	25.5	1.7	34.6	2	0	89.7	16.4

[1]No. = Number.
[2]EU = European Union (EU 2001).
[3]Class. = Classification (EU 2001).
[4]CT = Corneal thickness.
[5]Mean values calculated with scores from three eyes.
[6]%CV = % Coefficient of variation.
[7]CS = Corneal swelling.
[8]CO = Corneal opacity.
[9]FR = Fluorescein retention.
[10]Index = ICE Irritation Index (= CS x [CO x 20] + FR x 20]); No. = number.
[11]*In vivo* animal data were not provided for these substances, and therefore the EU classification that was provided by testing laboratory is presented here.

Table 7-2 Intralaboratory Reproducibility of ICE Test Method Endpoints – Prinsen (2000)

Substance (Experimental Replicates)	EU[1] Class[2]	CT[3] (mean[4])	CT (%CV[5])	CS[6] (mean)	CS (%CV)	CO[7] (mean)	CO (%CV)	FR[8] (mean)	FR (%CV)	Index[9] (mean)	Index (%CV)
SP-1 (5)[10]	NI	62.1	2.2	1.1	138.7	0.2	95.8	0.2	141.4	9.3	91.8
SP-4 (5)	R36	70.7	4.0	15.1	22.4	2.5	18.1	2	0	104.4	10.3
SU-4 (5)	R36	70.5	6.3	12.3	15.2	0.7	10.6	1	0	46.3	4.1
SU-5 (4)	R41	76.1	1.8	20.2	13.9	1.9	8.7	2	0	98.6	6.1

[1]EU = European Union (EU 2001).
[2]Class. = Classification (EU 2001).
[3]CT = Corneal thickness.
[4]Mean values calculated with scores from three eyes.
[5]%CV = % Coefficient of variation.
[6]CS = Corneal swelling.
[7]CO = Corneal opacity.
[8]FR = Fluorescein retention.
[9]Index = ICE Irritation Index (= CS x [CO x 20] + FR x 20]); No. = Number.
[10]*In vivo* animal data were not provided for these substances, and therefore the EU classification that was provided by testing laboratory is presented here.

7.2.3 Assessment of Interlaboratory Reproducibility

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

- determination of the extent of concordance among laboratories in assigning the same regulatory classification for a particular substance (e.g., Holzhütter et al. 1996)
- 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 (ANOVA) 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)

7.2.3.1 *Interlaboratory Reproducibility of Hazard Classification Based on In Vitro Irritancy Classification*

In the EC/HO study reported on by Balls et al. (1995), ICE test data for an assessment of interlaboratory reproducibility was provided for four laboratories. As described in **Section 2.0**, a categorization scheme was developed that enables the assignment of a test substance, based on its activity in the ICE assay, to an ocular irritancy category that corresponds to the EU *in vivo* rabbit ocular irritancy classification system (EU 2001). This categorization scheme was used to classify the ocular irritancy potential of the 59 substances with corresponding *in vivo* rabbit eye study data tested in the ICE assay for each of the four EC/HO participating laboratories. A similar scheme was used to classify the same 59 substances according to the EPA and GHS classification systems (EPA 1996; UN 2003) for each of the four participating laboratories. The resulting *in vitro* ocular irritation classifications were used to evaluate the extent of agreement among the laboratories.

For the Balls et al. (1995) study, 19 of the 59 substances tested were assigned an overall *in vitro* classification of corrosive/severe irritant and 40 substances were assigned an overall classification of nonsevere irritant (i.e., irritants other than severe/nonirritant). For an assessment of interlaboratory reproducibility, substances classified as corrosive/severe irritants or nonsevere irritants/nonirritants were also classified (within the GHS, EPA, and EU classification schemes [EPA 1996; EU 2001; UN 2003]) by their *in vivo* rabbit eye test results. Because the focus of this assessment is on the interlaboratory reproducibility of the ICE test method in identifying corrosives/severe irritants versus nonsevere irritants/nonirritants, considerable variability could exist among laboratories in their classification of substances as nonsevere irritants or nonirritants (e.g., three laboratories could classify a substance as a nonirritant and one laboratory could classify the same substance as a moderate irritant; for the purpose of the analysis conducted for this BRD, this would be considered 100% agreement between laboratories).

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

The four participating laboratories were in 100% agreement in regard to the ocular irritancy classification (corrosive/severe irritant or nonsevere irritant/nonirritant) of 44 (75%) of the 59 substances tested. As shown in **Table 7-3:**

Table 7-3 Interlaboratory Variability of Balls et al. (1995) for Substances Classified as Ocular Corrosives/Severe Irritants or Nonsevere Irritants/Nonirritants Using the GHS Classification System

Classification (*in vivo*/*in vitro*)[1]	Number of Substances	Number of Testing Laboratories	Substances with 100% Agreement Among Laboratories (%)	Substances with 75% Agreement Among Laboratories (%)	Substances with 50% Agreement Among Laboratories (%)
+/+	11	4[2]	7 (64)	3 (27)	1 (9)
+/-	11	4	9 (82)	2 (18)	0 (0)
-/+	6	4	1 (17)	0 (0)	5 (83)
-/-	26	4	22 (85)	4 (15)	0 (0)
?/-	3	4	3 (100)	0 (0)	0 (0)
?/+	2	4	2 (100)	0 (0)	0 (0)
TOTAL	59	4[2]	44 (75)	9 (15)	6 (10)

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

[2]Scores for fluorescein retention and corneal swelling were not provided from one laboratory for one substance (trichloroacetic acid, 30%), and therefore this substance was classified based on results from only three laboratories.

- All four participating laboratories agreed on the classification of seven (64%) of the 11 substances that were GHS corrosives/severe irritants[1]. Three of the four laboratories were in agreement for the three (27%) substances with discordant *in vitro* classification results among the four participating laboratories (5% benzalkonium chloride, cyclohexanol, promethazine HCl). The discordant laboratory was never the same for these three substances. In addition, two of the four laboratories were in agreement for one (9%) substance (dibenzoyl-L-tartaric acid).
- Nine (82%) of the 11 substances classified according to the GHS based on *in vivo* rabbit eye data as corrosives/severe irritants were incorrectly classified by the four participating laboratories as nonsevere irritants (i.e., Category 2A and 2B irritants) or nonirritants. Of the two substances (18%) with discordant *in vitro* classification results among the four laboratories, three of the four laboratories were in agreement for both substances (10% cetylpyridinium bromide, 2,5-dimethylohexanediol). The discordant laboratory for these two substances was not the same laboratory

[1] As described in **Section 6.1**, the overall *in vitro* classification for each substance was determined based on the most frequent individual laboratory classification, or in the case of an even number of discordant responses, the most severe classification. For one chemical (trichloroacetic acid, 30%), scores for fluorescein retention and corneal swelling were not provided from one laboratory. Therefore, this chemical was classified based on the results from only three laboratories.

- One (17%) of the six substances (isobutanol) classified according to the GHS based on *in vivo* rabbit eye data as a nonsevere irritant/nonirritant was incorrectly classified by the four laboratories as a corrosive/severe irritant. Of the five substances (83%) with discordant *in vitro* classification results among the four laboratories, two of the four laboratories were in agreement for all five substances (ethanol, n-hexanol, isopropanol, methyl acetate, methyl ethyl ketone). The discordant laboratories for these five substances were not consistently the same two laboratories.
- All four laboratories agreed on the classification of 22 (85%) of the 26 substances classified as GHS nonsevere irritants/nonirritants. Three of the four laboratories were in agreement for the four substances (15%) with discordant classification results (n-butyl acetate, 4-carboxybenzaldehyde, dibenzyl phosphate, methyl isobutyl ketone). The discordant laboratory for three of these four substances was always the same laboratory.
- Due to the lack of appropriate *in vivo* data (e.g., studies were terminated too early to assess reversibility of effects), five (8%) of the 59 test substances could not be classified according to the GHS classification scheme. Among these five substances, all four laboratories were in agreement with the classification of three substances as nonsevere irritants/nonirritants by and two substances as corrosives/severe irritants.

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

The four participating laboratories were in 100% agreement for the ocular irritancy classification (corrosive/severe irritant or nonsevere irritant/nonirritant) of 44 (75%) of the 59 substances tested. As shown in **Table 7-4:**

- All four participating laboratories agreed on the classification of seven (70%) of the 10 substances that were EPA corrosives/severe irritants[2]. Three of the four laboratories were in agreement for the three (30%) substances with discordant *in vitro* classification results among the four participating laboratories (benzalkonium chloride, 5%, cyclohexanol, promethazine HCl). The discordant laboratory was never the same for these three substances.
- Seven (78%) of the nine substances classified according to the EPA based on *in vivo* rabbit eye data as corrosives/severe irritants were incorrectly classified by the four participating laboratories as nonsevere irritants/nonirritants. Of the two substances (22%) with discordant *in vitro* classification results among the four participating laboratories, both substances (10% cetylpyridinium bromide, 2,5-dimethylhexanediol) were incorrectly classified by three of the four laboratories. The discordant laboratory for these two substances was not the same laboratory.

[2] As described in **Section 6.1**, the overall *in vitro* classification for each substance was determined based on the most frequent individual laboratory classification, or in the case of an even number of discordant responses, the most severe classification. For one chemical (trichloroacetic acid, 30%), scores for fluorescein retention and corneal swelling were not provided from one laboratory. Therefore, this chemical was classified based on the results from only three laboratories.

Table 7-4 Interlaboratory Variability of Balls et al. (1995) for Substances Classified as Ocular Corrosives/Severe Irritants or Nonsevere Irritants/Nonirritants Using the EPA Classification System

Classification (*in vivo*/*in vitro*)[1]	Number of Substances	Number of Testing Laboratories	Substances with 100% Agreement Among Laboratories (%)	Substances with 75% Agreement Among Laboratories (%)	Substances with 50% Agreement Among Laboratories (%)
+/+	10	4[2]	7 (70)	3 (30)	0 (0)
+/-	9	4	7 (78)	2 (22)	0 (0)
-/+	6	4	1 (17)	0 (0)	5 (83)
-/-	28	4	24 (86)	4 (14)	0 (0)
?/-	3	4	3 (100)	0 (0)	0 (0)
?/+	3	4	2 (67)	0 (0)	1 (33)
TOTAL	59	4[2]	44 (75)	9 (15)	6 (10)

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

[2]Scores for fluorescein retention and corneal swelling were not provided from one laboratory for one substance (trichloroacetic acid, 30%), and therefore this substance was classified based on results from only three laboratories.

- One (17%) of the six substances (isobutanol) classified according to the EPA based on *in vivo* rabbit eye data as a nonsevere irritant/nonirritant was incorrectly classified by the four participating laboratories as a corrosive/severe irritant. Of the five substances (83%) with discordant *in vitro* classification results among the four participating laboratories, all five substances (ethanol, n-hexanol, isopropanol, methyl acetate, methyl ethyl ketone) were incorrectly classified by two of the four laboratories. The discordant laboratories for these five substances were not consistently the same two laboratories.
- All four laboratories agreed on the classification of 24 (86%) of the 28 substances that were EPA nonsevere irritants/nonirritants. Three of the four laboratories were in agreement for the four substances (14%) with discordant classification results (n-butyl acetate, 4-carboxybenzaldehyde, dibenzyl phosphate, methyl isobutyl ketone). The discordant laboratory for three of these four substances was always the same laboratory.
- Due to the lack of appropriate *in vivo* data (e.g., studies were terminated too early to assess reversibility of effects), six (10%) of the 59 test substances could not be classified according to the EPA classification scheme. Among these six substances, three substances were classified as nonsevere irritants/nonirritants by all four laboratories. In addition, two substances were classified as corrosives/severe irritants by all four laboratories and one substance was classified as a corrosive/severe irritant by two of the four laboratories.

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

The participating laboratories were in 100% agreement in regard to the ocular irritancy classification (corrosive/severe irritant or nonsevere irritant/nonirritant) of 45 (76%) of the 59 substances tested. As shown in **Table 7-5:**

Table 7-5 Interlaboratory Variability of Balls et al. (1995) for Substances Classified as Ocular Corrosives/Severe Irritants or Nonsevere Irritants/Nonirritants Using the EU Classification System

Classification (*in vivo*/*in vitro*)[1]	Number of Substances	Number of Testing Laboratories	Substances with 100% Agreement Among Laboratories (%)	Substances with 75% Agreement Among Laboratories (%)	Substances with 50% Agreement Among Laboratories (%)
+/+	10	4[2]	6 (60)	3 (30)	1 (10)
+/-	9	4	7 (78)	2 (22)	0 (0)
-/+	5	4	1 (20)	0 (0)	4 (80)
-/-	26	4	23 (88)	3 (12)	0 (0)
?/-	5	4	5 (100)	0 (0)	0 (0)
?/+	4	4	3 (75)	0 (0)	1 (25)
TOTAL	59	4[3]	45 (76)	8 (14)	6 (10)

[1]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, an EU classification could not be made. See **Section 6.1** for a description of the rules followed to classify the ocular irritancy of test substances tested multiple times *in vitro.*

[2]Scores for fluorescein retention and corneal swelling were not provided from one laboratory for one substance (trichloroacetic acid, 30%), and therefore this substance was classified based on results from only three laboratories.

- All four participating laboratories agreed on the classification of six (60%) of the 12 substances that were EU corrosives/severe irritants[3]. Three of the four laboratories were in agreement for the three (30%) substances with discordant *in vitro* classification results among the four participating laboratories (5% benzalkonium chloride, cyclohexanol, promethazine HCl). The discordant laboratory was never the same for these three substances. In addition, one (10%) substance (dibenzoyl-L-tartaric acid) was correctly classified by two of the four laboratories.

[3] As described in **Section 6.1**, the overall *in vitro* classification for each substance was determined based on the most frequent individual laboratory classification, or in the case of an even number of discordant responses, the most severe classification. For one chemical (trichloroacetic acid, 30%), scores for fluorescein retention and corneal swelling were not provided from one laboratory. Therefore, this chemical was classified based on the results from only three laboratories.

- Seven (78%) of the nine substances classified according to the EU based on *in vivo* rabbit eye data as corrosives/severe irritants were incorrectly classified by the four participating laboratories as nonsevere irritants/nonirritants. Of the two substances (22%) with discordant *in vitro* classification results among the four participating laboratories, both substances (10% cetylpyridinium bromide, 2,5-dimethylhexanediol) were incorrectly classified by three of the four laboratories. The discordant laboratory for these two substances was not the same laboratory
- One (20%) of the seven substances classified according to the EU based on *in vivo* rabbit eye data as nonsevere irritants/nonirritants were incorrectly classified by the four participating laboratories as corrosives/severe irritants. Of the four substances (80%) with discordant *in vitro* classification results among the four participating laboratories, all four substances (ethanol, n-hexanol, methyl acetate, methyl ethyl ketone) were incorrectly classified by two of the four laboratories. The discordant laboratories for these five substances were not consistently the same two laboratories.
- All four laboratories agreed on the classification of 23 (88%) of the 26 substances classified as EU nonsevere irritants/nonirritants the four participating laboratories. Three of the four laboratories were in agreement for the three substances (12%) with discordant classification results (n-butyl acetate, 4-carboxybenzaldehyde, methyl isobutyl ketone). The discordant laboratory for these three substances was always the same laboratory.
- Due to the lack of appropriate *in vivo* data (e.g., studies were terminated too early to assess reversibility of effects), nine (15%) of the 59 test substances could not be classified according to the EU classification scheme. Among these nine substances, five substances were classified as nonsevere irritants/nonirritants by all four laboratories. In addition, three substances were classified as corrosives/severe irritants by all four laboratories and one substance was classified as a corrosive/severe irritant by two of the four laboratories.

7.2.3.5 *Common Chemical or Product Classes Among Test Substances with Discordant Interlaboratory Results Using the GHS Classification System*

Among the 15 substances classified according to the GHS scheme that exhibited interlaboratory differences in *in vitro* classification in the Balls et al. (1995) study, four were classified as alcohols. Two of the 15 substances were classified as cationic surfactants two were classified as acetates/esters, and two were classified as ketones. Solvents was the product class appearing most frequently among these substances, with seven of the 15 substances represented. Other product classes represented by multiple substances were chemical intermediates (five substances) and synthetic flavor ingredients (four substances). In regard to physical properties, of the 15 substances with discordant results among the four laboratories, 10 were liquid (seven water soluble) and five were solid (four water insoluble).

7.2.4 Coefficient of Variation Analysis

Mean endpoint values (i.e., fluorescein retention, corneal opacity, corneal swelling) and the ICE Irritation Index for each substance were provided for each of the four laboratories

participating in the EC/HO study. As detailed in **Section 2.2.12**, the Irritation Index is derived by calculating the sum of the maximum mean scores of each of the numerical endpoints. The opacity and fluorescein retention scores are equally weighted relative to the maximum corneal swelling obtained. Historical data from the laboratory of the developer of the ICE test method indicates that the maximum swelling observed is approximately 60-80%. Therefore, the maximum opacity (score = 4) and fluorescein retention (score = 3) scores obtained for a test substance are multiplied by a factor of 20 in order to increase their weighting (Chamberlain et al. 1997).

To provide a quantitative assessment of interlaboratory variability, individual laboratory ICE test results were used to calculate a mean, standard deviation, and the %CV for corneal opacity, fluorescein retention, corneal swelling, and the irritation index for each substance tested (**Table 7-6**). Mean and median %CV values for all 59 substances were calculated to provide an assessment of overall variability. Traditionally, mean/median %CV values of less than 35% have been considered satisfactory for biologically-based test methods (Fentem et al. 1998). For ICE, a wide range of %CV values for individual substances is evident for all endpoints. The mean/median %CV values were 39%/36% (ranging from 0-159%) for fluorescein retention, 47/37% for corneal opacity (ranging from 0-159%), 77%/75% for corneal swelling (ranging from 31-159%), and 35%/32% (ranging from 10-98%) for the Irritation Index. When only severe irritants (GHS Category 1[4], based on *in vivo* data) are considered, the %CV values are lower for all endpoints, with corneal swelling (mean of 72%, median of 69%) the sole endpoint with a mean/median %CV value greater than 35%. Of the four liquid substances with a CV < 35% for corneal swelling (2,2-dimethylbutanoic acid, 2,6-dichlorobenzoyl chloride, benzalkonium chloride 5%, and cetylpyridinium bromide 10%), two were water insoluble. No solid substances had a CV < 35% for corneal swelling. It is noteworthy that some of the corneal swelling values reported in the EC/HO data are greater than 80% (**Table 7-6**), and therefore above the reported historical maximum range of 60-80%. However, different depth measuring devices may have been used by the participating laboratories to determine corneal thickness, which, unless normalized, would have contributed to the increased variability and/or the excessive values calculated for this evaluation (Prinsen M, personal communication).

Common physicochemical characteristics do not appear among the substances showing the most variable responses (defined as CV >70% for any of the endpoints). Of the 37 substances with significant variability in at least one ICE endpoint, 18 are solids (of a total of 19 solids, 12 of which are water soluble) and 19 are liquids (of a total of 40 liquids, 14 of which are water soluble). However, some chemical classes appear to predominate among the 37 substances with CV values greater than 70%; these include seven surfactants (of 12 tested), five heterocyclic compounds (of six tested), four acetate/esters (of six tested), and four acids (of six tested). Therefore, the majority of substances tested from these chemical classes exhibited increased interlaboratory variability.

---

[4] One of these substances (sodium lauryl sulfate, 15%) is classified as R36 according to EU. Two other substances (cetylpyridinium bromide, 6% and dibenzoyl-L-tartaric acid) were not classified according to EPA due to inadequate *in vivo* data with which to follow the EPA-specific classification rules. Therefore, substances classified as severe irritants according to the GHS system were used for this subanalysis in order to include the largest dataset.

Table 7-6 Coefficient of Variation Analysis of the Interlaboratory Variability of the ICE Test Method

Substance Name	FR (mean)	FR (%CV)	CO (mean)	CO (%CV)	CS (mean)	CS (%CV)	Index (mean)	Index (%CV)
1-Naphthalene acetic acid[1]	***1.3***	***40.0***	***1.0***	***5.1***	***21.3***	***88.0***	***65.4***	***28.1***
1-Naphthalene acetic acid, Na salt	***3.0***	***0.0***	***2.8***	***8.7***	***69.2***	***52.4***	***185.3***	***19.4***
2,2-Dimethylbutanoic acid	***3.0***	***0.0***	***2.7***	***10.0***	***54.1***	***32.1***	***167.2***	***12.7***
2,5-Dimethylhexanediol	***2.1***	***33.1***	***1.7***	***56.5***	***23.5***	***115.0***	***98.7***	***57.7***
2,6-Dichlorobenzoyl chloride	2.0	11.6	1.1	37.1	18.2	30.8	80.2	14.3
2-Ethyl-1-hexanol	1.8	54.7	2.0	17.8	42.7	49.3	117.6	10.3
4-Carboxybenzaldehyde	1.2	45.6	1.3	95.2	26.4	159.4	76.2	91.0
Acetone	1.8	47.8	1.1	50.2	20.0	97.4	77.7	33.2
Ammonium nitrate	1.6	26.2	1.1	42.2	16.8	101.8	70.5	37.4
Benzalkonium chloride (1 %)	***1.9***	***29.2***	***1.9***	***55.9***	***21.6***	***80.2***	***97.5***	***23.6***
Benzalkonium chloride (10%)	***3.0***	***0.0***	***2.4***	***17.5***	***53.6***	***51.5***	***161.8***	***16.7***
Benzalkonium chloride (5%)	***2.5***	***40.0***	***2.4***	***20.4***	***45.2***	***35.7***	***143.2***	***25.2***
Dibenzoyl-L-tartaric acid	***1.7***	***51.3***	***2.3***	***29.5***	***25.2***	***132.2***	***105.7***	***38.3***
Captan 90 concentrate	***0.3***	***158.7***	***0.9***	***41.6***	***17.0***	***63.5***	***41.8***	***52.6***
Cetylpyridinium bromide (0.1%)	0.8	21.6	0.5	115.5	12.0	65.6	38.9	35.5
Cetylpyridinium bromide (10%)	***2.3***	***25.2***	***1.9***	***43.4***	***28.0***	***34.5***	***113.0***	***16.1***
Cetylpyridinium bromide (6%)	***2.4***	***20.9***	***1.3***	***46.1***	***28.6***	***52.0***	***98.1***	***21.0***

Substance Name	FR (mean)	FR (%CV)	CO (mean)	CO (%CV)	CS (mean)	CS (%CV)	Index (mean)	Index (%CV)
Chlorhexidine	***3.0***	***0.0***	***3.8***	***13.3***	***78.4***	***80.3***	***218.8***	***26.1***
Cyclohexanol	***2.8***	***14.3***	***2.3***	***10.6***	***52.2***	***66.9***	***154.3***	***24.1***
Dibenzyl phosphate	1.9	34.9	1.4	68.8	23.0	54.5	88.5	38.4
Ethanol	2.3	17.0	2.6	13.3	43.8	46.3	142.4	16.1
Ethyl acetate	2.2	26.1	2.1	7.2	36.5	72.2	121.5	20.4
Ethyl trimethyl acetate	1.1	78.3	0.7	121.0	11.6	114.0	46.9	98.3
Ethyl-2-methylacetoacetate	0.8	37.8	0.5	93.4	5.1	118.5	30.1	43.3
Fomesafen	0.7	67.0	0.7	63.1	5.9	61.1	35.1	42.3
Gammabutyrolactone	1.6	42.3	1.7	18.6	22.2	74.5	89.8	20.6
Glycerol	1.1	75.4	0.6	76.6	13.4	81.4	47.9	58.8
Imidazole	***3.0***	***0.0***	***3.1***	***20.1***	***99.8***	***87.6***	***222.2***	***38.8***
Isobutanol	3.0	3.4	2.4	12.3	61.4	40.7	168.8	16.1
Isopropanol	2.0	48.1	1.8	55.0	35.5	76.2	112.0	34.9
L-aspartic acid	1.7	28.0	1.3	44.5	21.0	113.6	82.0	50.3
Maneb	0.5	115.5	1.0	70.7	12.6	108.9	42.7	86.0
Methyl acetate	1.9	47.0	2.5	16.9	38.3	95.3	126.0	26.1
Methyl cyanoacetate	0.5	67.6	0.6	47.8	16.0	117.6	38.9	63.3
Methyl ethyl ketone	2.7	17.6	2.4	18.2	42.4	91.6	143.4	34.0

Substance Name	FR (mean)	FR (%CV)	CO (mean)	CO (%CV)	CS (mean)	CS (%CV)	Index (mean)	Index (%CV)
Methyl isobutyl ketone	2.4	20.4	2.3	20.9	31.1	72.9	125.1	32.6
Methylcyclopentane	0.6	81.6	0.4	66.7	7.5	131.9	27.0	43.0
n-Butyl acetate	1.1	50.0	2.1	18.6	25.7	53.6	89.8	14.4
n-Hexanol	2.3	60.9	2.3	32.2	46.8	62.6	137.2	23.5
n-Octanol	1.6	27.5	1.7	35.2	45.1	96.0	112.2	43.1
Parafluoraniline	3.0	0.0	2.1	4.9	46.6	46.7	147.4	14.4
Polyethylene glycol 400	1.2	65.7	0.6	76.6	14.1	61.8	49.7	31.7
Potassium cyanate	1.0	59.9	0.9	82.9	17.2	53.8	55.1	45.0
Promethazine HCl	***2.7***	***17.8***	***2.4***	***29.7***	***56.9***	***101.1***	***157.9***	***44.5***
Pyridine	***3.0***	***0.0***	***2.6***	***18.2***	***60.9***	***50.3***	***173.4***	***22.8***
Quinacrine	***1.1***	***67.2***	***0.8***	***97.2***	***8.6***	***44.2***	***47.0***	***65.3***
Sodium hydroxide (1%)	1.5	51.8	1.9	46.4	33.2	50.7	100.3	36.6
Sodium hydroxide (10%)	***3.0***	***0.0***	***3.6***	***12.3***	***111.6***	***66.6***	***243.9***	***29.9***
Sodium lauryl sulfate (3 %)	0.8	66.7	0.3	158.7	15.4	109.4	36.5	40.0
Sodium lauryl sulfate (15 %)	***1.1***	***41.6***	***0.7***	***63.4***	***15.4***	***77.1***	***49.8***	***31.8***
Sodium oxalate	***0.6***	***56.1***	***0.3***	***118.6***	***8.8***	***116.3***	***26.9***	***20.1***
Sodium perborate	***0.8***	***62.6***	***0.7***	***35.0***	***12.1***	***72.1***	***41.2***	***29.4***
Tetraaminopyrimidine sulfate	1.2	20.4	1.4	34.8	13.7	84.6	65.2	28.8

Substance Name	FR (mean)	FR (%CV)	CO (mean)	CO (%CV)	CS (mean)	CS (%CV)	Index (mean)	Index (%CV)
Toluene	1.4	29.1	1.6	32.1	26.6	87.6	86.6	31.6
Trichloroacetic acid (3%)	2.0	22.4	1.9	27.9	26.4	38.6	104.4	15.5
Trichloroacetic acid (30%)	***3.0***	***0.0***	***4.0***	***0.0***	***92.5***	***92.5***	***226.3***	***23.2***
Triton X-100 (10 %)	1.7	42.1	0.8	61.2	17.6	50.6	66.8	35.0
Triton X-100 (5 %)	1.3	35.6	0.7	145.1	22.8	81.4	62.4	42.8
Tween 20	1.2	81.7	0.6	76.6	11.7	110.0	47.9	39.9
Mean for All Substances	**1.8**	**38.8**	**1.6**	**46.8**	**32.4**	**77.2**	**100.5**	**34.8**
Median for All Substances	**1.7**	**35.6**	**1.7**	**37.1**	**25.2**	**74.5**	**89.8**	**31.8**
Range for All Substances	**0.3-3.0**	**0-158.7**	**0.3-4.0**	**0-158.7**	**5.1-111.6**	**30.8-159.4**	**26.9-243.9**	**10.3-98.3**
Mean for Severe Irritants (GHS)	**2.2**	**29.9**	**2.1**	**34.2**	**44.8**	**72.4**	**129.0**	**30.3**
Median for Severe Irritants	**2.5**	**23.0**	**2.3**	**25.0**	**36.9**	**69.5**	**128.1**	**25.6**
Range for Severe Irritants	**0.3-3.0**	**0-158.7**	**0.3-4.0**	**0-118.6**	**8.6-111.6**	**32.2-132.2**	**26.9-243.9**	**12.7-65.3**

FR = Fluorescein retention; CO = Corneal opacity; CS = Corneal swelling; Index = ICE Irritation Index; %CV = Coefficient of variation expressed as a percentage

[1]Test substances listed in bolded italics are classified *in vivo* as severe irritants (Category 1) according to GHS.

7.2.5 Additional Analysis of Interlaboratory Reproducibility

In the EC/HO validation study, Balls et al. (1995) determined the interlaboratory correlation between ICE test method endpoint data generated by each laboratory for all 60 substances tested, as well as for subsets of test substances (water-soluble, water-insoluble, surfactants, solids, solutions, and liquids). This analysis yielded a range of correlation coefficients for the subsets of test substances as shown in **Table 7-7** (see **Appendix E** for all correlation coefficients derived from comparing each laboratory with every other laboratory). Interlaboratory correlation coefficients generally spanned a range of 0.6 to 0.9 depending on the specific subsets of substances being evaluated. However, the range in correlation coefficients for some endpoints was larger (e.g., correlation coefficients for ICE-Mean Swelling ranged from 0.210 to 0.757 when testing substances that are insoluble in water).

Review of the mean *in vitro* data from this study indicates that wide ranges of corneal swelling values were recorded for the five insoluble test substances that were classified as ocular corrosives/severe irritants. For all five substances, the same laboratory produced the highest values, with mean corneal swelling percentages ranging from 1.5 to 6 times greater than the next highest mean corneal swelling value for the same substance tested by the other three laboratories. In addition, of the 14 remaining ocular corrosives/severe irritants (soluble and surfactant combined), a considerably higher value was reported for corneal swelling by the same laboratory for 12 substances. This trend was also apparent for nonsevere irritants/nonirritants.

Although the interlaboratory variability for fluorescein retention or corneal opacity was not as pronounced for the insoluble ocular corrosives/severe irritants and could not be associated with a single laboratory, the ranges of correlation coefficients for these endpoints are also relatively high. Therefore, the apparently large interlaboratory variability noted among these substances cannot be attributed to a single laboratory or to a single endpoint.

7.3 Historical Positive and Negative Control Data

As noted in **Section 2.0**, concurrent positive control substances have not been employed in the ICE test method, and therefore, an evaluation of historical positive control data is not possible.

At least one eye is traditionally included in each ICE study as a negative/vehicle control (isotonic saline). Individual eye data that could be used to perform a CV analysis on between-experiment values for each of the test method endpoints (i.e., corneal thickness/swelling, corneal opacity, fluorescein retention) along with the ICE Irritation Index for each test substance were obtained from negative control eyes. This analysis revealed that responses in the negative control eye remain relatively consistent (**Table 7-8**).

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

Index Score	Interlaboratory Pearson's Correlation Coefficient (r) of the *In Vitro* Data
Full set of test substances[1] (58-60 depending on endpoint)	
ICE-Mean Swelling	0.627-0.750
ICE-Mean Opacity Score	0.679-0.759
ICE-Mean Fluorescein Retention	0.525-0.768
ICE Index Score	0.759-0.801
Chemicals soluble in water (29-30 depending on endpoint)	
ICE-Mean Swelling	0.691-0.808
ICE-Mean Opacity Score	0.771-0.847
ICE-Mean Fluorescein Retention	0.676-0.692
ICE Index Score	0.858-0.881
Chemicals insoluble in water (17-18 depending on endpoint)	
ICE-Mean Swelling	0.210-0.757
ICE-Mean Opacity Score	0.414-0.851
ICE-Mean Fluorescein Retention	0.371-0.847
ICE Index Score	0.569-0.905
Surfactants (12)	
ICE-Mean Swelling	0.392-0.920
ICE-Mean Opacity Score	0.438-0.759
ICE-Mean Fluorescein Retention	0.546-0.950
ICE Index Score	0.724-0.854
Solids (19-20 depending on endpoint)	
ICE-Mean Swelling	0.722-0.869
ICE-Mean Opacity Score	0.595-0.868
ICE-Mean Fluorescein Retention	0.674-0.869
ICE- Index Score	0.752-0.883
Solutions (13-14 depending on endpoint)	
ICE-Mean Swelling	0.539-0.889
ICE-Mean Opacity Score	0.717-0.907
ICE-Mean Fluorescein Retention	0.543-0.901
ICE- Index Score	0.464-0.914
Liquids (26)	
ICE-Mean Swelling	0.461-0.779
ICE-Mean Opacity Score	0.692-0.770
ICE-Mean Fluorescein Retention	0.394-0.748
ICE Index Score	0.745-0.856

[1]As noted in **Section 7.1**, one substance (thiourea) was tested *in vitro* in the ICE assay but, due to its excessive toxicity *in vivo*, excluded from the comparison of *in vitro* and *in vivo* test results, and thus excluded from the evaluation in **Section 7.2.1**. However, *in vitro* data for this substance was included in the original Balls et al. (1995) analysis.

Table 7-8 Intralaboratory Reproducibility of ICE Test Method Endpoints – Negative Control (Isotonic Saline) Data

Substance (Experiment No.[1])	Max[2] Corneal Thickness	Max Corneal Swelling (%)	Max Corneal Opacity	Max Fluorescein Retention	Irritation Index[3]
Negative Control[4] (1)	63	0	0	0	0
Negative Control (2)	61	-2	0	0	-2
Negative Control (3)	63	-2	0	0	-2
Negative Control (4)	60	0	0	0	0
Negative Control (5)	62	0	0	0	0
Negative Control (6)	61	-2	0	0	-2
Negative Control (7)	62	0	0	0	0
Negative Control (8)	65	0	0	0	0
Negative Control (9)	62	-2	0	0	-2
Negative Control (10)	62	0	0	0	0
Negative Control (11)	64	2	0	0	2
Negative Control (12)	61	0	0	0	0
Negative Control (13)	64	0	0	0	0
Negative Control (14)	64	0	0	0	0
Negative Control (15)	67	2	0	0	2
Negative Control (16)	60	2	0	0	2
Mean	62.6	-0.1	0	0	-0.1
SD[5]	1.9	1.4	0	0	1.4
%CV[6]	3.0	-1088.1	-	-	-1088.1

[1]No. = Number.
[2]Max = Maximum.
[3]Index = ICE Irritation Index (= CS x [CO x 20] + FR x 20]).
[4]Isotonic saline.
[5]SD = Standard deviation.
[6]CV = Coefficient of variation (%CV = [Standard deviation/Mean] x 100); FR = Fluorescein retention

7.4 Summary of Results

The range of %CV values for the corneal thickness measurement, when results were compared within experiments, was 0.9 to 6.1. The other endpoints evaluated produced ranges of %CV values that were larger, with variability most prominent with the nonirritating substance (SP-1). However, this could be an exaggeration of variability given the relatively small values that were produced from the nonirritating substance relative to the irritating and

corrosive substances (i.e., corneal swelling values of 2, 0, and 3 yield a higher % CV than values of 11, 14, and 18). A similar discussion can also be applied to the variability among the qualitative endpoints (i.e., corneal opacity and fluorescein retention) given the small dynamic range of their scores (0-4 or 0-3, respectively).

The range of %CV values for the corneal thickness measurement, when results were compared across labs, was from 1.8 to 6.3. The %CV values for the remaining endpoints had a larger range (e.g., corneal swelling %CV = 13.9 to 138.7). However, if the nonirritating substance is removed, the range of %CV values is reduced (e.g., corneal swelling %CV = 13.9 to 22.4).

A qualitative assessment of the data for the EC/HO validation study (Balls et al. 1995) revealed that all four laboratories were in 100% agreement on the classification of 60% to 70% of substances classified as corrosives/severe irritants, 85% to 88% of substances classified as nonsevere irritants/nonirritants and 75% to 76% of all 59 substances considered in the study, regardless of the system used to classify the substances, when using the ICE test method.

A quantitative assessment of the data for the EC/HO validation study (Balls et al. 1995) was also done by conducting a %CV analysis for each endpoint and for the ICE Irritation Index, for each substance tested. For all substances tested, the mean/median %CV for the ICE Irritation Index was 34.8%/31.8% and 30.3%/25.6% when only substances classified as severe irritants according to the GHS classification system were considered. Historically, mean/median %CV values of <35% have been considered as satisfactory for interlaboratory reproducibility (Fentem et al. 1998).

Also, in the EC/HO validation study (Balls et al. 1995) determined interlaboratory correlation between ICE test method endpoints and the ICE Irritation Index for all substances tested and for various subsets. For all substances, the correlation coefficient for the ICE Irritation Index ranged from 0.759 to 0.801.

Analysis of the responses of negative control eyes in 16 different experiments revealed that responses were relatively consistent.

[This Page Intentionally Left Blank]