

SECTION II

ISOLATED CHICKEN EYE (ICE) TEST METHOD ACCURACY AND RELIABILITY REANALYSIS

[This Page Intentionally Left Blank]

904 **1.0 INTRODUCTION**

905

906 On November 1, 2004, NICEATM released draft BRDs on the current status of four *in vitro*
907 test methods for detecting ocular corrosives and severe irritants (see
908 http://iccvam.niehs.nih.gov/methods/ocudocs/ocu_brd.htm). The test methods reviewed
909 were the BCOP, the HET-CAM, the IRE, and the ICE assays. On January 11-12, 2005,
910 ICCVAM convened an Expert Panel to independently evaluate the validation status of these
911 four *in vitro* test methods for identifying ocular corrosives or severe irritants. The Expert
912 Panel Report, *Evaluation of the Current Validation Status of In Vitro Test Methods for*
913 *Identifying Ocular Corrosives and Severe Irritants*, can be obtained by contacting
914 NICEATM or electronically from <http://iccvam.niehs.nih.gov/methods/eyeirrit.htm>. Public
915 comments at the meeting revealed that additional data could be made available that had not
916 yet been provided in response to earlier requests for data. The Expert Panel subsequently
917 recommended that the additional data be requested and that a reanalysis of the accuracy and
918 reliability of each test method be conducted, to the extent possible.

919

920 In response to this recommendation, a second *FR* notice was published on February 28, 2005
921 (*FR* Vol. 70, No. 38, pp. 9661-9662; <http://iccvam.niehs.nih.gov/methods/eyeirrit.htm>)
922 requesting all available *in vitro* data on these four *in vitro* ocular irritancy test methods and
923 corresponding *in vivo* rabbit eye test method data, as well as any human exposure data (either
924 via ethical human studies or accidental exposure). The first *FR* notice requesting these data
925 had been published on March 24, 2004 (*FR* Vol. 69, No. 57, pp. 13859-13861;
926 <http://iccvam.niehs.nih.gov/methods/eyeirrit.htm>). Also, a request for relevant data was re-
927 sent directly to the primary developers or users of each test method and sent to other
928 scientists who participated in or attended the Expert Panel Meeting on January 11-12, 2005
929 and who had indicated a desire to provide additional data. No human exposure data was
930 obtained for the substances evaluated in the ICE test method, and therefore no calculations
931 could be made on the accuracy of the ICE test method for predicting human severe ocular
932 irritancy.

933

934 Other factors also necessitated a reanalysis of the accuracy of the ICE test method for
935 detecting ocular corrosives and severe irritants. First, clarification regarding the rules for
936 classification of severe irritants was obtained subsequent to the release of the four BRDs that
937 resulted in changes to the hazard classification of some of the substances used in the original
938 analysis. For the original analysis, reversibility of ocular effects for the EU and GHS hazard
939 classification systems was considered to be achieved if, by post-exposure day 21, the
940 endpoint scores fell below the threshold that resulted in a test substance being classified as a
941 severe irritant (EU [2001]; UN [2003]). The new information obtained indicated that
942 reversibility of ocular effects is achieved only when all scores reach zero by post-exposure
943 day 21. This change resulted in one substance previously classified as non-severe GHS
944 irritants now being classified as a GHS severe irritant.

945

946 Second, the chemical classes assigned to each test substance were revised to reflect a
947 standardized classification scheme (based on MeSH [<http://www.nlm.nih.gov/mesh>]) that
948 would ensure consistency in classifying substances among all *in vitro* ocular test methods
949 under consideration. This resulted in some chemicals being reclassified. The accuracy of the

950 ICE test method, by chemical class and using the GHS classification system (UN [2003]),
951 has been reanalyzed to reflect these changes.

952

953 Finally, an additional accuracy analysis was conducted. In this analysis, the accuracy of each
954 *in vitro* ocular irritancy test method for detecting ocular corrosives or severe irritants,
955 depending on whether the classification was based on the severity of the response and/or its
956 persistence to day 21 post-treatment, was determined.

957

958 For the ICE test method, the changes to the existing database that resulted from using the
959 appropriate persistence classification criteria and any new data and/or information received
960 subsequent to the release of the draft BRD are summarized in **Table II-1**. Additional ICE
961 test method data and corresponding *in vivo* rabbit eye test data were submitted by the
962 Netherlands Organisation for Applied Scientific Research (TNO) Nutrition and Food
963 Institute for the 44 substances tested in Prinsen (1996) and for an additional 50 substances
964 (Prinsen [2005]).

965

966 Also, the TNO Nutrition and Food Institute provided replicate ICE test data and the
967 corresponding *in vivo* EU hazard classification for four substances (Prinsen [2000]). The
968 efforts of Mr. Menk Prinsen and the TNO Nutrition and Food Institute in providing
969 additional data and/or information are gratefully acknowledged.

970

971 **2.0 ACCURACY OF THE ICE TEST METHOD - REANALYSIS**

972

973 The ability of the ICE test method to correctly identify ocular corrosives and severe irritants,
974 as defined by the GHS, EPA, and EU classification systems (EPA [1996]; EU [2001]; UN
975 [2003])¹, was evaluated. The three regulatory ocular hazard classification systems
976 considered during this analysis use different classification systems and decision criteria to
977 identify ocular corrosives and severe irritants based on *in vivo* rabbit eye test results. All
978 three classification systems are based on individual animal data in terms of the magnitude of
979 the response and on the extent to which induced ocular lesions fail to reverse by day 21.
980 However, there are differences among the three classification systems in regard to their
981 criteria used by NICEATM for distinguishing between a severe and a non-severe response
982 (see **Appendix A**). Thus, to evaluate the accuracy of the IRE test method for identifying
983 ocular corrosives and severe irritants, individual rabbit data collected at the different
984 observation times was needed for each substance.

985

986 The ability of the ICE test method to correctly identify ocular corrosives and severe irritants,
987 as defined by the GHS, EPA, and EU classification systems (EPA [1996]; EU [2001]; UN
988 [2003]), was evaluated using two approaches. In the first approach, the accuracy of ICE was
989 assessed separately for each *in vitro-in vivo* comparative study (i.e., publication) reviewed in
990 Sections 4.0 and 5.0 of the draft ICE BRD. In the second approach, an overall analysis of
991 ICE test method accuracy was conducted by combining results from each study, and then an

¹ For the purposes of this analysis, an ocular corrosive or severe irritant was defined as a substance that would be classified as Category 1 according to the GHS classification system (UN [2003]), as Category I according to the EPA classification system (EPA [1996]), or as R41 according to the EU classification system (EU [2001]).

992 **Table II-1. Summary of ICE Database Changes**

993

Data Source	Data Set	Number of Available Substances	Number of Acceptable Substances by Ocular Irritancy Classification System			Comments
			EPA ¹	EU ²	GHS ³	
			Cat ⁴ I/Total ⁵	R41/Total ⁴	Cat 1/Total ⁴	
Prinsen and Koëter (1993)	New ⁶	21	2/10	7/21	2/10	The decrease in the number of corrosive/severe irritants is due to the reclassification of one substance from a severe ocular irritant/corrosive to a moderate ocular irritant.
	Old ⁶	21	3/10	8/21	3/10	
Balls et al. (1995)	New	59	19/51	19/50	22/54	The decrease in the total number of usable substances is due to excluding substances from consideration due to insufficient rabbit eye test data for classification (See Appendix A).
	Old	59	20/54	21/59	22/56	
Prinsen (1996)	New	44	2/36	2/36	2/36	The <i>in vivo</i> data that corresponded to the substances tested were received, which allowed for an evaluation of all three regulatory hazard classification systems for this study (previously, the analysis of severe irritants was limited to the published EU classification for these substances). The published EU classification for four severe irritants was based only on dermal corrosivity (no rabbit eye test was performed). Therefore, these substances were excluded from the revised analysis.
	Old	44	0/29	6/44	0/29	
Prinsen (2000)	New	4	-	1/4	-	This is new information received subsequent to the original analysis. Because the corresponding <i>in vivo</i> rabbit test data were not submitted, the analysis was based on the provided EU classification only.
Prinsen (2005)	New	50	4/46	4/46	4/46	This is new information received subsequent to the original analysis. Four of these substances were classified based only on dermal corrosivity (no <i>in vivo</i> rabbit eye test was performed); these substances were excluded from the analysis.

994 ¹EPA = U.S. Environmental Protection Agency (EPA [1996]).995 ²EU = European Union (EU [2001]).996 ³GHS = Globally Harmonized System (UN [2003]).

- 997 ⁴Cat = Category.
- 998 ⁵First number (before forward slash) refers to the number of substances in each study that were classified as a severe irritant according to each classification
- 999 system (EPA, EU, and GHS). The second number (after the forward slash) refers to the number of substances that were classified, based on animal data, for
- 1000 each classification system (EPA, EU, GHS).
- 1001 ⁶New = accuracy statistics based on the revised analysis; Old = accuracy statistics based on the previous analysis included in the draft ICE BRD.

1002 overall ocular irritancy classification was assigned for each substance. When the same
1003 substance was evaluated in multiple laboratories, the overall ICE ocular irritancy
1004 classification was based on the majority of calls among all of the studies. When there was an
1005 equal number of different irritancy classifications for substances (e.g., two tests classified a
1006 substance as a nonsevere irritant and two tests classified a substance as a severe irritant), the
1007 more severe irritancy classification was used for the overall classification for the substance
1008 (severe irritant, in this case).

1009

1010 Based on the revisions made to the ICE test method database, which included the addition of
1011 46 to 50 new substances, a revised accuracy analysis has been conducted. The calculations
1012 were performed as described in Section 6.0 of the draft ICE BRD. To allow for a
1013 comparison of the results obtained in the revised analysis relative to those obtained
1014 previously, the data tables below include accuracy statistics from both analyses. However,
1015 the discussion of the results in the sections that follow relate to the revised analysis only.

1016

1017 2.1 GHS Ocular Hazard Classification System

1018

1019 The four studies (Prinsen and Koëter [1993]; Balls et al. [1995]; Prinsen [1996]; Prinsen
1020 [2005]) contained ICE test data on 171 substances, 144 of which had sufficient *in vivo* data to
1021 be assigned an ocular irritancy classification as defined by the GHS classification system
1022 (UN [2003])² (see **Appendix II-A**). Based on results from *in vivo* rabbit eye experiments,
1023 30³ of the 144 substances were classified as severe irritants (i.e., Category 1), the other 114
1024 substances were classified as nonsevere irritants (either Category 2A, 2B) or nonirritants.
1025 The 27 substances that could not be classified according to the GHS classification system due
1026 to the lack of adequate animal data are so noted in **Appendix II-A**.

1027

1028 2.1.1 Prinsen and Koëter (1993)

1029 Based on the available *in vivo* rabbit eye data, 10 of the 21 substances tested in this study
1030 could be assigned a GHS classification (**Table II-2**). The remaining 11 substances had
1031 inadequate *in vivo* data for assigning a classification according to the GHS system (UN
1032 [2003]). For the 10 substances that could be evaluated, the ICE test method has an accuracy
1033 of 80% (8/10), a sensitivity of 100% (2/2), a specificity of 75% (6/8), a false positive rate of
1034 25% (2/8), and a false negative rate of 0% (0/2).

1035

1036 2.1.2 Balls et al (1995)

1037 Based on the available *in vivo* rabbit eye data, 54 of the 59 substances tested in this study
1038 could be assigned a GHS classification (**Table II-2**). The remaining five substances had
1039 inadequate *in vivo* data for assigning a classification according to the GHS system (UN
1040 [2003]). For the 54 substances assigned a GHS classification, the ICE test method has an

² For the purpose of this accuracy analysis, *in vivo* rabbit study results were used to identify GHS Category 1 irritants (i.e., severe irritants); substances classified as GHS Category 2A and 2B irritants were identified as nonsevere irritants.

³ One chemical (benzalkonium chloride, 1%) was tested *in vivo* twice in the same laboratory. The results were discordant with respect to GHS classification. According to one test, the classification was Category 1, while results from the other test yielded a Category 2B classification. The accuracy analysis was performed with the substance classified as Category 1.

1041 **Table II-2. Evaluation of the Performance of the ICE Test Method In Predicting Ocular Corrosives and Severe Irritants**
 1042 **Compared to the *In Vivo* Rabbit Eye Test Method, as Defined by the GHS¹ Classification System, by Study and**
 1043 **Overall**
 1044

Data Source	N ²	Accuracy		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
		%	No. ³	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Prinsen and Koëter (1993) (new) ⁴	10/21	80	8/10	100	2/2	75	6/8	50/2/4	3/4	100	6/6	25	2/8	0	0/2
Prinsen and Koëter (1993) (old) ⁴	10/21	80	8/10	100	3/3	86	6/7	75	3/4	100	6/6	17	1/7	0	0/3
Balls et al. (1995) ^{5,6} (new)	54/59	69	37/54	50	11/22	81	26/32	65	11/17	70	26/37	19	6/32	50	11/22
Balls et al. (1995) ^{5,6} (old)	56/59	71	40/56	55	12/22	82	28/34	67	12/18	74	28/38	18	6/34	46	10/22
Prinsen (1996) (new)	36/44	97	35/36	50	1/2	100	34/34	100	1/1	97	34/35	0	0/34	50	1/2
Prinsen (1996) (old)	29/44	100	29/29	-	0/0	100	29/29	-	0/0	100	29/29	0	0/29	-	0/0
Prinsen (2005) (new)	46/50	89	41/46	0	0/4	98	41/42	0	0/1	91	41/45	2	1/42	100	4/4
Entire Data Set ^{6,7} (new)	144/171	83	120/144	50	15/30	92	105/114	63	15/24	88	105/120	8	9/114	50	15/30
Entire Data Set ^{6,7} (old)	92/121	82	75/92	60	15/25	90	60/67	68	15/22	86	60/70	10	7/67	40	10/25

1045 ¹GHS = Globally Harmonized System (UN [2003]).

1046 ²N = Number of substances included in this analysis/the total number of substances in the study.

1047 ³No. = Data used to calculate the percentage.

1048 ⁴New = accuracy statistics based on the revised analysis; Old = accuracy statistics based on the previous analysis included in the draft ICE BRD.

1049 ⁵One chemical (benzalkonium chloride, 1%) was tested *in vivo* twice within the same laboratory. The results were discordant with respect to GHS classification;
 1050 the analysis was performed assuming Category 1 classification.

1051 ⁶Performance calculated using the overall *in vitro* classification based on the majority and/or most severe classification among the four laboratories.

1052 ⁷Includes the data from Balls et al. (1995) using the overall *in vitro* classification based on the majority and/or most severe classification among the four
 1053 laboratories.

1054 accuracy of 69% (37/54), a sensitivity of 50% (11/22), a specificity of 81% (26/32), a false
1055 positive rate of 19% (6/32), and a false negative rate of 50% (11/22).

1056

1057 2.1.3 Prinsen (1996)

1058 Based on the *in vivo* rabbit eye data obtained subsequent to the original ICE test method
1059 analysis, 36 of the 44 substances tested in this study could be assigned a GHS classification
1060 (**Table II-2**). The remaining eight substances had inadequate *in vivo* data for assigning a
1061 classification according to the GHS system (UN [2003]). For the 36 substances that could be
1062 evaluated, the ICE test method has an accuracy of 97% (35/36), a sensitivity of 50% (1/2), a
1063 specificity of 100% (34/34), a false positive rate of 0% (0/34), and a false negative rate of
1064 50% (1/2).

1065

1066 2.1.4 Prinsen (2005)

1067 Subsequent to the original ICE test method accuracy analysis, data were submitted on 50
1068 substances. Based on the available *in vivo* rabbit eye data provided in this submission, 46 of
1069 the 50 substances tested in this study could be assigned a GHS classification (**Table II-2**).
1070 The remaining four substances had inadequate *in vivo* data for assigning a classification
1071 according to the GHS system. For the 46 substances that could be evaluated, the ICE test
1072 method has an accuracy of 89% (41/46), a sensitivity of 0% (0/4), a specificity of 98%
1073 (41/42), a false positive rate of 2% (1/42), and a false negative rate of 100% (4/4).

1074

1075 2.1.5 Entire Data Set

1076 A total of 144 substances had sufficient *in vivo* data among the four studies to perform an
1077 accuracy analysis, based on the GHS classification system (**Table II-2**). Twenty-two
1078 substances lacked sufficient *in vivo* information on which to assign a GHS classification.
1079 Based on these 144 substances, the ICE test method has an accuracy of 83% (120/144), a
1080 sensitivity of 50% (15/30), a specificity of 92% (105/114), a false positive rate of 8%
1081 (9/114), and a false negative rate of 50% (15/30).

1082

1083 2.2 **EPA Ocular Hazard Classification System**

1084

1085 The four studies (Prinsen and Koëter [1993]; Balls et al. [1995]; Prinsen [1996]; Prinsen
1086 [2005]) contained ICE test method data on 171 substances, 145 of which had sufficient *in*
1087 *vivo* data to be assigned an ocular irritancy classification according to the EPA classification
1088 system (EPA 1996)⁴ (see **Appendix II-A**). Based on results from the *in vivo* rabbit eye test,
1089 29 of these 145 substances were classified as severe irritants (i.e., Category I), while the
1090 other 116 substances were classified as nonsevere irritants or nonirritants (Categories II, III,
1091 or IV). The 26 substances that could not be classified according to the EPA classification
1092 system are so noted in **Appendix II-A**.

1093

⁴ For the purpose of this accuracy analysis, *in vivo* rabbit study results were used to identify EPA Category I irritants (i.e., severe irritants); substances classified as EPA Category II, III, or IV irritants were defined as nonsevere irritants.

1093 2.2.1 Prinsen and Koëter (1993)

1094 Based on the available *in vivo* rabbit eye data, 10 of the 21 substances tested in this study
1095 could be assigned an EPA classification (**Table II-3**). The remaining 11 substances had
1096 inadequate *in vivo* data for assigning a classification according to the EPA system (EPA
1097 1996). For the 10 substances that could be evaluated, the ICE test method has an accuracy of
1098 80% (8/10), a sensitivity of 100% (2/2), a specificity of 75% (6/8), a false positive rate of
1099 25% (2/8), and a false negative rate of 0% (0/2).

1100

1101 2.2.2 Balls et al. (1995)

1102 Based on the available *in vivo* rabbit eye data, 53 of the 59 substances tested in this study
1103 could be assigned an EPA classification (**Table II-3**). The remaining six substances had
1104 inadequate *in vivo* data for assigning a classification according to the EPA system (1996).
1105 For the 53 substances assigned an EPA classification, the ICE test method has an accuracy of
1106 72% (38/53), sensitivity of 53% (10/19), a specificity of 82% (28/34), a false positive rate of
1107 18% (6/34), and a false negative rate of 47% (9/19).

1108

1109 2.2.3 Prinsen (1996)

1110 Based on the *in vivo* rabbit eye data obtained subsequent to the original ICE test method
1111 analysis, 36 of the 44 substances tested in this study could be assigned an EPA classification
1112 (**Table II-3**). The remaining eight substances had inadequate *in vivo* data for assigning a
1113 classification according to the EPA system (1996). For the 36 substances that could be
1114 evaluated, the ICE test method has an accuracy of 97% (35/36), a sensitivity of 50% (1/2), a
1115 specificity of 100% (34/34), a false positive rate of 0% (0/34), and a false negative rate of
1116 50% (1/2).

1117

1118 2.2.4 Prinsen (2005)

1119 Subsequent to the original ICE test method accuracy analysis, data were submitted on 50
1120 substances. Based on the available *in vivo* rabbit eye data provided in this submission, 46 of
1121 the 50 substances tested in this study could be assigned an EPA classification (**Table II-3**).
1122 The remaining four substances had inadequate *in vivo* data for assigning a classification
1123 according to the EPA system (1996). For the 46 substances that could be evaluated, the ICE
1124 test method has an accuracy of 89% (41/46), a sensitivity of 0% (0/4), a specificity of 98%
1125 (41/42), a false positive rate of 2% (1/42), and a false negative rate of 100% (4/4).

1126

1127 2.2.5 Entire Data Set

1128 A total of 145 substances had sufficient *in vivo* data among the four studies to perform an
1129 accuracy analysis, based on the EPA classification system (**Table II-3**). Twenty-six
1130 substances lacked sufficient *in vivo* information on which to assign an EPA classification
1131 (EPA [1996]). Based on these 145 substances, the ICE test method has an accuracy of 84%
1132 (122/145), a sensitivity of 52% (15/29), a specificity of 92% (107/116), a false positive rate
1133 of 8% (9/116) and a false negative rate of 48% (14/29).

1134 **Table II-3. Evaluation of the Performance of the ICE Test Method In Predicting Ocular Corrosives and Severe Irritants**
 1135 **Compared to the *In Vivo* Rabbit Eye Test Method, as Defined by the EPA¹ Classification System, by Study and**
 1136 **Overall**
 1137

Data Source	N ²	Accuracy		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
		%	No. ³	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Prinsen and Koëter (1993) (new) ⁴	10/21	80	8/10	100	2/2	75	6/8	50	2/4	100	6/6	25	2/8	0	0/2
Prinsen and Koëter (1993) (old) ⁴	10/21	80	8/10	100	3/3	86	6/7	75	3/4	100	6/6	17	1/6	0	0/5
Balls et al. (1995) ^{5,6} (new)	53/59	72	38/53	53	10/19	82	28/34	63	10/16	76	28/37	18	6/34	47	9/19
Balls et al. (1995) ^{5,6} (old)	54/59	72	39/54	55	11/20	82	28/34	65	11/17	76	28/37	18	6/34	45	9/20
Prinsen (1996) (new)	36/44	97	35/36	50	1/2	100	34/34	100	1/1	97	34/35	0	0/34	50	1/2
Prinsen (1996) (old)	29/44	100	29/29	-	0/0	100	29/29	-	0/0	100	29/29	0	0/29	-	0/0
Prinsen (2005) (new)	46/50	89	41/46	0	0/4	98	41/42	0	0/1	91	41/45	2	1/42	100	4/4
Entire Data Set ^{6,7} (new)	145/171	84	122/145	52	15/29	92	107/116	63	15/24	89	107/121	8	9/116	48	14/29
Entire Data Set ^{6,7} (old)	90/121	82	74/90	61	14/23	90	60/67	67	14/21	87	60/69	10	7/67	39	9/23

1138 ¹EPA = U.S. Environmental Protection Agency (EPA [1996]).

1139 ²N = Number of substances included in this analysis/the total number of substances in the study.

1140 ³Data used to calculate the percentage.

1141 ⁴New = accuracy statistics based on the revised analysis; Old = accuracy statistics based on the previous analysis included in the draft ICE BRD.

1142 ⁵One chemical (benzalkonium chloride, 1%) was tested *in vivo* twice within the same laboratory. The results were discordant with respect to EPA classification;
 1143 the analysis was performed assuming Category I classification.

1144 ⁶Performance calculated using the overall *in vitro* classification based on the majority and/or most severe classification among the four laboratories.

1145 ⁷Includes the data from Balls et al. (1995) using the overall *in vitro* classification based on the majority and/or most severe classification among the four
 1146 laboratories.

1147 2.3 EU Ocular Hazard Classification System

1148

1149 The five studies (Prinsen and Koëter [1993]; Balls et al. [1995]; Prinsen [1996]; Prinsen
1150 [2000]; Prinsen [2005]) contained ICE test method data on 175 substances, 154 of which had
1151 sufficient *in vivo* data to be assigned an ocular irritancy classification according to the EU
1152 classification system (EU [2001])⁵ (see **Appendix II-A**). Based on results from the *in vivo*
1153 rabbit eye test, 32⁶ of the 154 substances were classified as severe irritants (i.e., R41) and the
1154 other 122 substances were classified as nonsevere irritants (either R36) or nonirritants. The
1155 21 substances that could not be classified according to the EU classification system are so
1156 noted in **Appendix II-A**.

1157

1158 2.3.1 Prinsen and Koëter (1993)

1159 All 21 substances tested in this study were included in an analysis of accuracy (**Table II-4**).
1160 Based on the available *in vivo* rabbit eye data or the EU ocular irritancy classification for
1161 each substance provided in the published study (individual rabbit eye test data was not
1162 available for all of the substances), the ICE test method has an accuracy of 95% (20/21), a
1163 sensitivity of 100% (7/7), a specificity of 93% (13/14), a false positive rate of 7% (1/14), and
1164 a false negative rate of 0% (0/7).

1165

1166 2.3.2 Balls et al. (1995)

1167 Based on the available *in vivo* rabbit eye data, 50 of the 59 substances tested in this study
1168 could be assigned an EU classification (**Table II-4**). Nine substances lacked sufficient *in*
1169 *vivo* information on which to assign an EU classification (2001). For the 50 substances
1170 assigned an EU classification, the ICE test method has an accuracy of 72% (36/50),
1171 sensitivity of 53% (10/19), a specificity of 84% (26/31), a false positive rate of 16% (5/31),
1172 and a false negative rate of 47% (9/19).

1173

1174 2.3.3 Prinsen (1996)

1175 Based on the *in vivo* rabbit eye data obtained subsequent to the original ICE test method
1176 analysis, 36 of the 44 substances tested in this study could be assigned an EU classification
1177 (**Table II-4**). Eight substances lacked sufficient *in vivo* information on which to assign an
1178 EU classification (2001). For the 36 substances that could be evaluated, the ICE test method
1179 has an accuracy of 97% (35/36), a sensitivity of 50% (1/2), a specificity of 100% (34/34), a
1180 false positive rate of 0% (0/34), and a false negative rate of 50% (1/2).

1181

1182 2.3.4 Prinsen (2000)

1183 Subsequent to the original ICE test method accuracy analysis, data were submitted on four
1184 substances. The EU classifications were provided by the author for all four of these
1185 substances that were used for the accuracy analysis (**Table II-4**). For these substances, the

⁵ For the purpose of this accuracy analysis, *in vivo* rabbit study results were used to identify R41 irritants (i.e., severe irritants); substances classified as R36 were defined as nonsevere irritants.

⁶ One chemical (benzalkonium chloride, 1%) was tested *in vivo* twice in the same laboratory. The results were discordant with respect to EU classification. According to one test, the classification was R41, while results from the other test yielded an R36 classification. The accuracy analysis was performed with the substance classified as R41.

1186 **Table II-4. Evaluation of the Performance of the ICE Test Method In Predicting Ocular Corrosives and Severe Irritants**
 1187 **Compared to the *In Vivo* Rabbit Eye Test Method, as Defined by the EU¹ Classification System, by Study and**
 1188 **Overall**

Data Source	N ²	Accuracy		Sensitivity		Specificity		Positive Predictivity		Negative Predictivity		False Positive Rate		False Negative Rate	
		%	No. ³	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
Prinsen and Koëter (1993) (new) ⁴	21/21	95	20/21	100	7/7	93	13/14	88	7/8	100	13/13	7	1/14	0	0/7
Prinsen and Koëter (1993) (old) ⁴	21/21	100	21/21	100	8/8	100	13/13	100	8/8	100	13/13	0	0/13	0	0/8
Balls et al. (1995) ^{5,6} (new)	50/59	72	36/50	53	10/19	84	26/31	67	10/15	74	26/35	16	5/31	47	9/19
Balls et al. (1995) ^{5,6} (old)	59/59	73	43/59	57	12/21	82	31/38	63	12/19	78	31/40	18	7/38	43	9/21
Prinsen (1996) (new)	36/44	97	35/36	50	1/2	100	34/34	100	1/1	97	34/35	0	0/34	50	1/2
Prinsen (1996) (old)	44/44	96	42/44	100	6/6	95	36/38	75	6/8	100	36/36	5	2/38	0	0/6
Prinsen (2000) (new)	4/4	100	4/4	100	1/1	100	3/3	100	1/1	100	3/3	0	0/3	0	0/1
Prinsen (2005) (new)	46/50	89	41/46	0	0/4	98	41/42	0	0/1	91	41/45	2	1/42	100	4/4
Entire Data Set ^{6,7} (new)	154/175	87	134/154	59	19/32	94	115/122	73	19/26	90	115/128	6	7/122	41	13/32
Entire Data Set ^{6,7} (old)	121/121	85	103/121	70	26/37	92	77/84	79	26/33	88	77/88	8	7/84	30	11/37

1189 ¹EU = European Union (EU [2001]).

1190 ²N = Number of substances included in this analysis/the total number of substances in the study.

1191 ³Data used to calculate the percentage.

1192 ⁴New = accuracy statistics based on the revised analysis; Old = accuracy statistics based on the previous analysis included in the draft ICE BRD.

1193 ⁵One chemical (benzalkonium chloride, 1%) was tested in vivo twice within the same laboratory. The results were discordant with respect to EU classification;
 1194 the analysis was performed assuming Category 1 classification.

1195 ⁶Performance calculated using the overall *in vitro* classification based on the majority and/or most severe classification among the four laboratories.

1196 ⁷Includes the data from Balls et al. (1995) using the overall *in vitro* classification based on the majority and/or most severe classification among the four
1197 laboratories.

1198 ICE test method has an accuracy (4/4), sensitivity (1/1), and specificity (3/3) of 100%, and
1199 false positive (0/3) and false negative (0/1) rates of 0%.

1200

1201 2.3.5 Prinsen (2005)

1202 Subsequent to the original ICE test method accuracy analysis, data were submitted on 50
1203 substances. Based on the available *in vivo* rabbit eye data provided in this submission, 46 of
1204 the 50 substances tested in this study could be assigned an EPA classification (**Table II-4**).
1205 The remaining four substances had inadequate *in vivo* data for assigning a classification
1206 according to the EU system. For the 46 substances that could be evaluated, the ICE test
1207 method has an accuracy of 89% (41/46), a sensitivity of 0% (0/4), a specificity of 98%
1208 (41/42), a false positive rate of 2% (1/42), and a false negative rate of 100% (4/4).

1209

1210 2.3.6 Entire Data Set

1211 A total of 154 substances had sufficient *in vivo* data among the three studies to perform an
1212 accuracy analysis, based on the EU classification system (**Table II-4**). For these 154
1213 substances, the ICE test method has an accuracy of 87% (134/154), a sensitivity of 59%
1214 (19/32), a specificity of 94% (115/122), a false positive rate of 6% (7/122), and a false
1215 negative rate of 41% (13/32).

1216

1217 2.4 **Accuracy of the ICE Test Method for the GHS Ocular Hazard Classification** 1218 **System, by Chemical Class and Property of Interest – Reanalysis**

1219

1220 In order to further evaluate discordant responses of the ICE test method relative to the *in vivo*
1221 hazard classification, several accuracy sub-analyses were performed. These included specific
1222 classes of chemicals with sufficiently robust numbers of substances ($n \geq 5$), as well as certain
1223 properties of interest considered relevant to ocular toxicity testing (e.g., pesticides,
1224 surfactants, pH, physical form). Because the international community will soon adopt the
1225 GHS classification system for hazard labeling (UN [2003]), and considering that there were
1226 only modest differences in overall ICE test method accuracy among the three regulatory
1227 classification systems (i.e., EPA, EU, GHS), these sub-analyses focused on the GHS system
1228 only. As indicated in **Table II-5**, there were some notable trends in the performance of the
1229 ICE test method among these subgroups of substances. The chemical class of substances that
1230 was most consistently overpredicted according the GHS classification system (i.e., were false
1231 positives) by the ICE test method is alcohols. Five out the nine overpredicted substances
1232 were alcohols. The remaining chemical classes represented among the overpredicted
1233 substances were alkalis (1), esters (1), ketones (1), and one unclassified coded substance.
1234 With regard to physical form of the substances overpredicted by the ICE test method, all nine
1235 were liquids.

1236

1237 There were no chemical classes that were prominently represented among the 15 substances
1238 that were underpredicted (i.e., were false negatives) by the ICE test method according to the
1239 GHS classification system (see **Appendix II-A**). Five of the 15 substances were unclassified
1240 coded substances, and three were carboxylic acids. No other chemical classes were
1241 represented more than twice. These included heterocycles (2), onium compounds (2),
1242 polycyclics (2), alcohols (1), amines/amidines (1), imides (1), inorganic chemicals (1), and
1243 polyethers (1). However, five of the 15 unpredicted substances were labeled as surfactants,

1244 **Table II-5. False Negative and False Positive Rates of the ICE Test Method, by**
 1245 **Chemical Class and Properties of Interest, for the GHS¹ Classification**
 1246 **System**
 1247

Category	N ²	False Positive Rate ³		False Negative Rate ⁴	
		%	No. ⁵	%	No.
Overall	144	8	9/114	50	15/30
Chemical Class⁶					
Alcohol	12	50	5/10	50	1/2
Amine/Amidine	5	0	0/2	33	1/3
Carboxylic acid	10	0	0/3	43	3/7
Ester	9	13	1/8	0	0/1
Heterocycle	9	0	0/3	33	2/6
Onium compound	8	0	0/2	33	2/6
Properties of Interest					
Liquids	108	10	9/90	44	8/18
Solids	36	0	0/24	58	7/12
Pesticide	11	0	0/6	60	3/5
Surfactant – Total	21	0	0/12	56	5/9
-nonionic	4	0	0/3	100	1/1
-anionic	2	0	0/1	100	1/1
-cationic	7	0	0/1	33	2/6
pH – Total⁷	20	-	-	40	8/20
- acidic (pH < 7.0)	12	-	-	33	4/12
- basic (pH > 7.0)	8	-	-	50	4/8
Category 1 Subgroup⁸					
- Total	30	-	-	50	15/30
- 4 (CO=4 at any time)	13	-	-	39	5/13
- 3 (severity/persistence)	1	-	-	0	0/1
- 2 (severity)	6	-	-	50	3/6
- 2-4 combined ⁹	20	-	-	45	9/20
- 1 (persistence)	10	-	-	70	7/10

1248 ¹GHS =- Globally Harmonized System (UN [2003]).

1249 ²N = number of substances.

1250 ³False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*;

1251 ⁴False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro*; n =
 1252 number of substances.

1253 ⁵Data used to calculate the percentage.

1254 ⁶Chemical classes included in this table are represented by at least five substances tested in the ICE test method
 1255 and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh) as defined in **Appendix B**.

1256 ⁷Total number of GHS Category 1 substances for which pH information was obtained.

1257 ⁸NICEATM-defined subgroups assigned based on the lesions that drove classification of a GHS Category 1
 1258 substance. 1: based on lesions that are persistent; 2: based on lesions that are severe (not including CO=4); 3:
 1259 based on lesions that are severe (not including CO=4) and persistent; 4: corneal opacity (CO) = 4 at any time.

1260 ⁹Subcategories 2 to 4 combined to allow for a direct comparison of GHS Category 1 substances classified *in*
 1261 *vivo* based on some lesion severity component and those classified based on persistent lesions alone.
 1262

1262 which included anionic (1), cationic (2), and nonionic (1) surfactants (the remaining
1263 substance was coded, but described as a surfactant). Another three of the underpredicted
1264 substances were labeled as pesticides. With regard to physical form of the substances tested,
1265 eight of the fifteen underpredicted substances were liquids while seven were solids.
1266 However, considering the proportion of the total database, solids (36/144; 25%) appear more
1267 likely than liquids (108/144; 75%) to be underpredicted by the ICE test method. Similarly,
1268 among the eight underpredicted substances for which pH information was available, four
1269 were acidic (pH < 7.0) and four were basic (pH > 7.0), although basic substances appear
1270 more likely to be underpredicted (8/20; 40% vs. 12/20; 60%), given their relative
1271 proportionality in the total database. Finally, the fifteen underpredicted substances were
1272 more likely to be substances classified *in vivo* based on persistent lesions, rather than on
1273 severe lesions, as evidenced by an analysis of NICEATM-defined GHS Category 1 sub-
1274 groupings (**Table II-5**).

1275

1276 **2.5 Accuracy of the ICE Test Method for Identifying Ocular Corrosives and** 1277 **Severe Irritants – Summary of Reanalysis**

1278

1279 As detailed in **Section II-1.0** of the ICE addendum, additional or new relevant ICE test
1280 method data was received after the Expert Panel meeting on January 11 and 12, 2005 that
1281 increased the size of the comparative ICE: *in vivo* rabbit eye test database from 92 to 144
1282 substances for the GHS classification system (UN [2003]), 90 to 145 for the EPA
1283 classification system (EPA [1996]), and 121 to 154 for the EU classification system (EU
1284 [2001]). As can be seen in **Tables II-2** through **II-4**, the overall accuracy of the ICE test
1285 method changed from 82-85% (old) to 83-87% (reanalysis) depending on the classification
1286 system used), the false positive rate was reduced from 8-10% (old) to 6-8% (reanalysis),
1287 while the false negative rate was increased from 30-40% (old) to 41-50% (reanalysis).

1288

1289 Similar to the original analysis, the revised analysis indicated that alcohols are overpredicted
1290 (50% [5/10] false positive rate) in the ICE test method. Carboxylic acids were shown to have
1291 a false negative rate of 43% (3/7).

1292

1293 The total database for surfactants was increased from 13 to 21 substances. However, given
1294 the stability of the false negative rate (old analysis: 57% [4/7]; new analysis 56% [5/9]), these
1295 substances still appear to be underpredicted by the ICE test method. With the additional
1296 data, it was now possible to evaluate the accuracy of the ICE test method for pesticides.
1297 While the false positive rate for these substances was 0% (0/6), the false negative rate (60%
1298 [3/5]) suggests that these substances may be underpredicted by the ICE test method.

1299

1300 As noted in **Section II-2.4**, eight of the fifteen underpredicted substances were liquids while
1301 seven were solids. However, considering that the total number of solids (36) in the database
1302 is much smaller than the number of liquids (108), solids appear more likely to be
1303 underpredicted (58%) than liquids (44%) by the ICE test method. In comparison to the
1304 original analysis, the false negative rate of solid substances changed from 55% (6/11) to 58%
1305 (7/12). However, the false negative rate for liquids was increased in the revised analysis
1306 from 29% (4/14) to 44% (8/18).

1307

1307 Using the expanded database, an analysis was conducted of the ability of the ICE test method
 1308 to identify ocular corrosives and severe irritants, depending on the nature of the *in vivo* ocular
 1309 lesions (i.e., severity and/or persistence) responsible for classification of a substance as an
 1310 ocular corrosive/severe irritant. As indicated in **Table II-5**, the fifteen underpredicted
 1311 substances were more likely to be substances classified *in vivo* based on persistent lesions
 1312 (false negative rate = 70% [7/10]), rather than on severe lesions (false negative rate = 45%
 1313 [9/20]).

1314

1315 A new analysis not included originally was an evaluation of accuracy related to acidic or
 1316 basic pH. Among the eight underpredicted substances for which pH information was
 1317 available, four were acidic (pH < 7.0) and four were basic (pH > 7.0). Again, basic
 1318 substances (8) occupy a smaller proportion of the total database than acidic substances (12),
 1319 and were more often underpredicted (50% vs. 33%). However, it is noted that pH
 1320 information was obtained for only 20 of the 30 total Category 1 substances.

1321

1322 **Table II-6** provides a summary of the revised analysis of the overall performance of the ICE
 1323 test method defined by the GHS classification system (UN [2003]). As noted from this
 1324 analysis, the false positive substances were mild to moderate ocular irritants (i.e., GHS
 1325 Category 2A or 2B). No nonirritating substances were classified as severe irritants.
 1326 However, the mild irritants (Category 2B; n = 1/12) were less likely to be overpredicted as
 1327 severe irritants/ocular corrosives than the moderate irritants (Category 2A, n = 8/23). The
 1328 false negative substances were predominantly confined to those classified, based on ICE test
 1329 results as Category 2A (n=4) or Category 2B (n=10), although one false negative substance
 1330 was classified as a nonirritant.

1331

1332

1333 **Table II-6. Overall Accuracy of the ICE Test Method in the Predicting the Irritancy**
 1334 **of a Substance as Defined by the GHS¹ Classification System**

1335

		<i>In Vitro</i> Classification				
		1	2A	2B	Nonirritant	TOTAL
<i>In Vivo</i> Classification ²	1	15	4	10	1	30
	2A	8	9	4	2	23
	2B	1	2	8	1	12
	Nonirritant	0	6	22	51	79
	TOTAL	24	21	44	55	144

1336

¹GHS = Globally Harmonized System (UN [2003]).

1337

²Twenty-seven substances included in **Appendix II-A** had insufficient data with which to assign a precise GHS classification and therefore were not included in this table.

1338

1339

1340

1341 Compared to the overall underprediction rate of the ICE test method (15/30; 50%), the
 1342 underprediction rate for pesticides is 60% (3/5), for surfactants is 56% (5/9), and for solids is

1343 58% (7/12). Compared to the overall overprediction rate of the ICE test method (8%; 9/114),
1344 the overprediction rate of the ICE test method for alcohols is 50% (5/10).

1345

1346 **3.0 RELIABILITY OF THE ICE TEST METHOD - REANALYSIS**

1347

1348 An assessment of test method reliability (intralaboratory repeatability and intra- and inter-
1349 laboratory reproducibility) is an essential element of any evaluation of the performance of an
1350 alternative test method (ICCVAM [2003]). Repeatability refers to the closeness of
1351 agreement between test results obtained within a single laboratory when the procedure is
1352 performed on the same substance under identical conditions within a given time period
1353 (ICCVAM [1997, 2003]). Intralaboratory reproducibility refers to the determination of the
1354 extent to which qualified personnel within the same laboratory can replicate results using a
1355 specific test protocol at different times. Interlaboratory reproducibility refers to the
1356 determination of the extent to which different laboratories can replicate results using the
1357 same protocol and test chemicals, and indicates the extent to which a test method can be
1358 transferred successfully among laboratories. A reliability assessment includes reviewing the
1359 rationale for selecting the substances used to evaluate test method reliability, a discussion of
1360 the extent to which the substances tested represent the range of possible test outcomes, the
1361 properties of the various substances for which the test method is proposed for use, and a
1362 quantitative and/or qualitative analysis of repeatability and intra- and inter-laboratory
1363 reproducibility. In addition, measures of central tendency and variation are summarized for
1364 historical control data (negative, vehicle, positive), where applicable.

1365

1366 **3.1 Substances Used to Re-evaluate the Reliability of the ICE Test Method**

1367

1368 While intralaboratory repeatability and reproducibility were not originally evaluated due to a
1369 lack of appropriate data, subsequent to the original analysis, additional data were received for
1370 four substances (two surfactants and two siloxanes). This unpublished study (Prinsen
1371 [2000]) provided data from a single laboratory, which tested each of these substances in four
1372 to five separate experiments, and therefore allowed for such an evaluation. The only source
1373 of data for conducting an assessment of ICE test method interlaboratory reproducibility was
1374 Balls et al. (1995). This study evaluated the performance and reproducibility of the ICE test
1375 method using 60 substances (i.e., there were 52 different substances with four substances
1376 tested at two different concentrations and two substances tested at three different
1377 concentrations, for a total of 60 possible ocular irritation outcomes). One substance
1378 (thiourea) was tested *in vitro* in the ICE assay but, due to its excessive toxicity *in vivo*, was
1379 excluded from the comparison of *in vitro* and *in vivo* test results.

1380

1381 **3.2 Reanalysis of ICE Test Method Intralaboratory Repeatability**

1382

1383 Generally, analyses of intralaboratory repeatability have included approaches such as:
1384 • a CV analysis, which is a statistical measure of the deviation of a variable
1385 from its mean (e.g., Holzhütter et al. [1996])
1386 • ANOVA methods (e.g., Holzhütter et al. [1996]; ASTM [1999]).

1387

1387 Due to the lack of available ICE test data for replicate enucleated chicken eyes within
1388 individual experiments and for experiments conducted on the same substance under identical
1389 conditions, an evaluation of the intralaboratory repeatability of the ICE test method could not
1390 previously be conducted. As noted above, additional data were received for four substances
1391 from a single laboratory, which tested each of these substances in four to five separate
1392 experiments. Each experiment used three eyes. A CV analysis was performed on within-
1393 experiment ICE test method data, using scores for each of the test method endpoints (i.e.,
1394 corneal thickness/swelling, corneal opacity, fluorescein retention) along with the ICE
1395 Irritation Index for each test substance (**Table II-7**). These CV values are not very
1396 informative given the nature of the data (0 means and standard deviations for some test
1397 substances, limited ranges of possible values for corneal opacity or fluorescein retention).
1398 However, the analysis of intralaboratory repeatability indicates that the corneal thickness
1399 measurement was generally repeatable when results were compared within experiments, as
1400 evidenced by the range of %CV values (0.9 to 6.1). The other endpoints evaluated produced
1401 somewhat more variable responses, most prominent with the nonirritating substance (SP-1).
1402 However, this could be an exaggeration of variability given the relatively small values that
1403 were produced from the nonirritating substance relative to the irritating and corrosive
1404 substances (i.e., corneal swelling values of 2, 0, and 3 yield a much higher % CV than values
1405 of 11, 14, and 18, but may not be indicative of truly increased variability). A similar
1406 discussion can also be applied to the variability among the qualitative endpoints (i.e., corneal
1407 opacity and fluorescein retention) given the small dynamic range of their scores (0-4 or 0-3,
1408 respectively).

1409

1410 **3.3 Reanalysis of ICE Test Method Intralaboratory Reproducibility**

1411

1412 Generally, analyses of intralaboratory reproducibility have included approaches such as:

1413

- a CV analysis, which is a statistical measure of the deviation of a variable
1414 from its mean (e.g., Holzhütter et al. [1996])
- ANOVA methods (e.g., Holzhütter et al. [1996]; ASTM [1999]).

1415

1416

1417 Due to the lack of available ICE test data for experiments conducted multiple times on the
1418 same substance in the same laboratory, an evaluation of ICE test method intralaboratory
1419 reproducibility could not previously be conducted. However, the data from Prinsen (2000)
1420 could also be used to perform a CV analysis on between-experiment values for each of the
1421 test method endpoints (i.e., corneal thickness/swelling, corneal opacity, fluorescein retention)
1422 along with the ICE Irritation Index for each test substance (**Table II-8**). Results similar to
1423 those obtained from an analysis of intralaboratory repeatability were obtained from this
1424 analysis. The corneal thickness measurement was again generally reproducible (%CV = 1.8
1425 to 6.3), but the %CV values for the remaining endpoints had a much larger range (e.g.,
1426 corneal swelling %CV = 13.9 to 138.7). However, if the nonirritating substance is removed,
1427 the range of %CV values is reduced (e.g., corneal swelling %CV = 13.9 to 22.4).

1428 **Table II-7. Intralaboratory Repeatability of ICE Test Method Endpoints – Prinsen (2000)**

1429

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

1430 ¹No. = Number.1431 ²EU = European Union (EU [2001]).1432 ³Class. = Classification (EU [2001]).

- 1433 ⁴CT = Corneal thickness.
1434 ⁵Mean values calculated with scores from three eyes.
1435 ⁶%CV = % coefficient of variation.
1436 ⁷CS = Corneal swelling.
1437 ⁸CO = Corneal opacity.
1438 ⁹FR = fluorescein retention.
1439 ¹⁰Index = ICE Irritation Index (= CS x [CO x 20] + FR x 20); No. = number.
1440 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.

1441 **Table II-8. Intralaboratory Reproducibility of ICE Test Method Endpoints – Prinsen (2000)**

1442

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

1443 ¹EU = European Union (EU [2001]).1444 ²Class. = Classification (EU [2001]).1445 ³CT = Corneal thickness.1446 ⁴Mean values calculated with scores from three eyes.1447 ⁵%CV = % coefficient of variation.1448 ⁶CS = Corneal swelling.1449 ⁷CO = Corneal opacity.1450 ⁸FR = fluorescein retention.1451 ⁹Index = ICE Irritation Index (= CS x [CO x 20] + FR x 20); No. = number.1452 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.

1453 3.4 Reanalysis of ICE Test Method Interlaboratory Reproducibility

1454

1455 Generally, analyses of interlaboratory variability have included approaches such as:

1456

- 1457 • the extent of concordance among laboratories in assigning the same regulatory
1458 classification for a particular substance (e.g., Holzhütter et al. [1996])
- 1459 • a CV analysis, which is a statistical measure of the deviation of a variable
1460 from its mean (e.g., Holzhütter et al. [1996])
- 1461 • ANOVA methods (e.g., Holzhütter et al. [1996]; ASTM [1999])
- 1462 • bivariate scatter diagrams/correlation analyses for pairs of laboratories to
1463 assess the extent possibility of divergence (e.g., Holzhütter et al. [1996])

1463

1464 In the EC/HO study reported by Balls et al. (1995), ICE test data for an assessment of
1465 interlaboratory reproducibility was provided for four laboratories. While the draft BRD
1466 contained the same analysis as detailed below, new information regarding *in vivo*
1467 classification of substances according to the three regulatory classification schemes was
1468 provided, which resulted in changes to the classification of some substances. Therefore, a
1469 revised analysis was conducted to reflect the updated classifications. As previously stated in
1470 the draft ICE BRD, 19 of the 59 substances tested in this study were assigned an overall *in*
1471 *vitro* classification of corrosive/severe irritant and 40 substances were assigned an overall
1472 classification of nonsevere irritant (i.e., irritants other than severe or nonirritant). For an
1473 assessment of interlaboratory reproducibility, substances classified as corrosive/severe
1474 irritants or nonsevere irritants/nonirritants were further classified within the GHS, EPA, and
1475 EU classification schemes (EPA [1996]; EU [2001]; UN [2003]) by their *in vivo* rabbit eye
1476 test results. Because the focus of this assessment is on the interlaboratory reproducibility of
1477 the ICE test method in identifying corrosives/severe irritants versus nonsevere
1478 irritants/nonirritants, considerable variability could exist among laboratories in their
1479 classification of substances as nonsevere irritants or nonirritants (e.g., three laboratories
1480 could classify a substance as a nonirritant and one laboratory could classify the same
1481 substance as a moderate irritant; for the purpose of the analysis, this would be considered
1482 100% agreement between laboratories).

1483

1484 3.4.1 Qualitative Reanalysis of Interlaboratory Reproducibility

1485

1485 3.4.1.1 *GHS Ocular Hazard Classification System*

1486

1486 The four participating laboratories were in 100% agreement in regard to the ocular irritancy
1487 classification (corrosive/severe irritant or nonsevere irritant/nonirritant) of 44 (75%) of the 59
1488 substances tested. As shown in **Table II-9**:

1489

- 1489 • All four participating laboratories agreed on the classification of seven (64%)
1490 of the 11 substances that were GHS corrosives/severe irritants⁷. Three of the
1491 four laboratories were in agreement for the three (27%) substances with
1492 discordant *in vitro* classification results among the four participating

⁷ 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.

1493 laboratories for three substances (5% benzalkonium chloride, cyclohexanol,
1494 promethazine HCl). The discordant laboratory was never the same for these
1495 three substances. In addition, two of the four laboratories were in agreement
1496 for one (9%) substance (dibenzoyl-L-tartaric acid).

- 1497 Nine (82%) of the 11 substances classified according to the GHS based on *in*
1498 *vivo* rabbit eye data as corrosives/severe irritants were incorrectly classified by
1499 the four participating laboratories as nonsevere irritants (i.e., Category 2A and
1500 2B irritants) or nonirritants. Of the two substances (18%) with discordant *in*
1501 *vitro* classification results among the four laboratories, three of the four
1502 laboratories were in agreement for both substances (10% cetylpyridinium
1503 bromide, 2,5-dimethylhexanediol). The discordant laboratory for these two
1504 substances was not the same laboratory.
- 1505 One (17%) of the six substances (isobutanol) classified according to the GHS
1506 based on *in vivo* rabbit eye data as a nonsevere irritant/nonirritant was
1507 incorrectly classified by the four laboratories as a corrosive/severe irritant. Of
1508 the five substances (83%) with discordant *in vitro* classification results among
1509 the four laboratories, two of the four laboratories were in agreement for all
1510 five substances (ethanol, n-hexanol, isopropanol, methyl acetate, methyl ethyl
1511 ketone). The discordant laboratories for these five substances were not
1512 consistently the same two laboratories.
- 1513 All four laboratories agreed on the classification of 22 (85%) of the 26
1514 substances classified as GHS nonsevere irritants/nonirritants (UN [2003]).
1515 Three of the four laboratories were in agreement for the four substances (15%)
1516 with discordant classification results (n-butyl acetate, 4-carboxybenzaldehyde,
1517 dibenzyl phosphate, methyl isobutyl ketone). The discordant laboratory for
1518 three of these four substances was always the same laboratory.
- 1519 Due to the lack of appropriate *in vivo* data (e.g., studies were terminated too
1520 early to assess reversibility of effects), five (8%) of the 59 test substances
1521 could not be classified according to the GHS classification scheme (UN
1522 [2003]). Among these five substances, all four laboratories were in agreement
1523 with the classification of three substances as nonsevere irritants/nonirritants
1524 and two substances as corrosive/severe irritants.

1525

1525 **Table II-9. Interlaboratory Variability of Balls et al. (1995) for Substances Classified**
 1526 **as Ocular Corrosives/Severe Irritants or Nonsevere Irritants/Nonirritants**
 1527 **Using the GHS¹ Classification System**
 1528

Classification (<i>in vivo</i> / <i>in vitro</i>) ²	Data Set	Number of Substances	Number of Testing Labs ³	Substances with 100% Agreement Among Labs	Substances with 75% Agreement Among Labs	Substances with 50% Agreement Among Labs
++	New ⁴	11	4 ³	7 (64%)	3 (27%)	1 (9%)
	Old ⁴	12	4 ³	8 (67%)	3 (25%)	1 (8%)
+/-	New	11	4	9 (82%)	2 (18%)	0 (0%)
	Old	10	4	8 (80%)	2 (20%)	0 (0%)
-/+	New	6	4	1 (17%)	0 (0%)	5 (83%)
	Old	6	4	1 (17%)	0 (0%)	5 (82%)
--	New	26	4	22 (85%)	4 (15%)	0 (0%)
	Old	28	4	24 (86%)	4 (14%)	0 (0%)
?/-	New	3	4	3 (100%)	0 (0%)	0 (0%)
	Old	2	4	2 (100%)	0 (0%)	0 (0%)
?/+	New	2	4	2 (100%)	0 (0%)	0 (0%)
	Old	1	4	1 (100%)	0 (0%)	0 (0%)
TOTAL	New	59	4 ³	44 (75%)	9 (15%)	6 (10%)
	Old	59	4 ³	44 (75%)	9 (15%)	6 (10%)

1529 ¹GHS = Globally Harmonized System (UN [2003]).

1530 ²A “+” indicates that the substance was assigned an overall classification of corrosive or a severe irritant
 1531 (Category 1); a “-” indicates that the substance was assigned an overall classification of nonsevere irritant
 1532 (Category 2A, 2B) or nonirritant; a “?” indicates that, due to the lack of appropriate *in vivo* data (e.g., studies
 1533 were terminated too early to assess reversibility of effects), a GHS classification could not be made. See
 1534 **Section II-2.0** for a description of the rules followed to classify the ocular irritancy of test substances tested
 1535 multiple times *in vitro*.

1536 ³Scores for fluorescein retention and corneal swelling were not provided from one laboratory for one substance
 1537 (trichloroacetic acid, 30%), and therefore this substance was classified based on results from only three
 1538 laboratories.

1539 ⁴New = accuracy statistics based on the revised analysis; Old = accuracy statistics based on the previous
 1540 analysis included in the draft ICE BRD.

1543 3.4.1.2 EPA Ocular Hazard Classification System

1544 The four participating laboratories were in 100% agreement for the ocular irritancy
 1545 classification (corrosive/severe irritant or nonsevere irritant/nonirritant) of 44 (75%) of the 59
 1546 substances tested. As shown in **Table II-10**:

- 1547 • All four participating laboratories agreed on the classification of seven (70%)
 1548 of the 10 substances that were EPA corrosives/severe irritants⁸. Three of the
 1549 four laboratories were in agreement for the three (30%) substances with

⁸ As described in **Section II-2.0**, 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.

1550 discordant *in vitro* classification results among the four participating
1551 laboratories (benzalkonium chloride, 5%, cyclohexanol, promethazine HCl).
1552 The discordant laboratory was never the same for these three substances.
1553 • Seven (78%) of the nine substances classified according to the EPA based on
1554 *in vivo* rabbit eye data as corrosives/severe irritants were incorrectly classified
1555 by the four participating laboratories as nonsevere irritants/nonirritants. Of
1556 the two substances (22%) with discordant *in vitro* classification results among
1557 the four participating laboratories, both substances (10% cetylpyridinium
1558 bromide, 2,5-dimethylhexanediol) were incorrectly classified by three of the
1559 four laboratories. The discordant laboratory for these two substances was not
1560 the same laboratory.
1561 • One (17%) of the six substances (isobutanol) classified according to the EPA
1562 based on *in vivo* rabbit eye data as a nonsevere irritant/nonirritant was
1563 incorrectly classified by the four participating laboratories as a
1564 corrosive/severe irritant. Of the five substances (83%) with discordant *in vitro*
1565 classification results among the four participating laboratories, all five
1566 substances (ethanol, n-hexanol, isopropanol, methyl acetate, methyl ethyl
1567 ketone) were incorrectly classified by two of the four laboratories. The
1568 discordant laboratories for these five substances were not consistently the
1569 same two laboratories.
1570 • All four laboratories agreed on the classification of 24 (86%) of the 28
1571 substances that were EPA nonsevere irritants/nonirritants. Three of the four
1572 laboratories were in agreement for the four substances (14%) with discordant
1573 classification results (n-butyl acetate, 4-carboxybenzaldehyde, dibenzyl
1574 phosphate, methyl isobutyl ketone). The discordant laboratory for three of
1575 these four substances was always the same laboratory.
1576 • Due to the lack of appropriate *in vivo* data (e.g., studies were terminated too
1577 early to assess reversibility of effects), six (10%) of the 59 test substances
1578 could not be classified according to the EPA classification scheme. Among
1579 these six substances, three substances were classified as nonsevere
1580 irritants/nonirritants by all four laboratories. In addition, two substances were
1581 classified as a corrosive/severe irritant by all four laboratories and one
1582 substance was classified as a corrosive/severe irritant by two of the four
1583 laboratories.
1584

1584 **Table II-10. Interlaboratory Variability of Balls et al. (1995) for Substances Classified**
 1585 **as Ocular Corrosives/Severe Irritants or Nonsevere Irritants/Nonirritants**
 1586 **Using the EPA¹ Classification System**
 1587

Classification (<i>in vivo</i> / <i>in vitro</i>) ²	Data Set	Number of Substances	Number of Testing Labs ³	Substances with 100% Agreement Among Labs	Substances with 75% Agreement Among Labs	Substances with 50% Agreement Among Labs
+ / +	New ⁴	10	4 ³	7 (70%)	3 (30%)	0 (0%)
	Old ⁴	11	4 ³	8 (73%)	3 (27%)	0 (0%)
+ / -	New	9	4	7 (78%)	2 (22%)	0 (0%)
	Old	9	4	7 (78%)	2 (22%)	0 (0%)
- / +	New	6	4	1 (17%)	0 (0%)	5 (83%)
	Old	6	4	1 (17%)	0 (0%)	5 (83%)
- / -	New	28	4	24 (86%)	4 (14%)	0 (0%)
	Old	28	4	24 (86%)	4 (14%)	0 (0%)
? / -	New	3	4	3 (100%)	0 (0%)	0 (0%)
	Old	3	4	3 (100%)	0 (0%)	0 (0%)
? / +	New	3	4	2 (67%)	0 (0%)	1 (33%)
	Old	2	4	1 (50%)	0 (0%)	1 (50%)
TOTAL	New	59	4 ³	44 (75%)	9 (15%)	6 (10%)
	Old ²	59	4 ³	44 (75%)	9 (15%)	6 (10%)

¹EPA = U.S. Environmental Protection Agency (EPA [1996]).

²A “+” 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 II-2.0** for a description of the rules followed to classify the ocular irritancy of test substances tested multiple times *in vitro*.

³Scores 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.

⁴New = accuracy statistics based on the revised analysis; Old = accuracy statistics based on the previous analysis included in the draft ICE BRD.

3.4.1.3 EU Ocular Hazard 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 II-11**:

- All four participating laboratories agreed on the classification of six (60%) of the 10 substances that were EU (2001) corrosives/severe irritants⁹. Three of

⁹ As described in **Section II-2.0**, 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.

1608 the four laboratories were in agreement for the three (30%) substances with
1609 discordant *in vitro* classification results among the four participating
1610 laboratories (5% benzalkonium chloride, cyclohexanol, promethazine HCl).
1611 The discordant laboratory was never the same for these three substances. In
1612 addition, one (10%) substance (dibenzoyl-L-tartaric acid) was correctly
1613 classified by two of the four laboratories.

- 1614 • Seven (78%) of the nine substances classified according to the EU (2001)
1615 based on *in vivo* rabbit eye data as corrosives/severe irritants were incorrectly
1616 classified by the four participating laboratories as nonsevere
1617 irritants/nonirritants. Of the two substances (22%) with discordant *in vitro*
1618 classification results among the four participating laboratories, both
1619 substances (10% cetylpyridinium bromide, 2,5-dimethylhexanediol) were
1620 incorrectly classified by three of the four laboratories. The discordant
1621 laboratory for these two substances was not the same laboratory.
- 1622 • One (20%) of the five substances classified according to the EU (2001) based
1623 on *in vivo* rabbit eye data as a nonsevere irritant/nonirritant was incorrectly
1624 classified by the four participating laboratories as a corrosive/severe irritant.
1625 Of the four substances (80%) with discordant *in vitro* classification results
1626 among the four participating laboratories, all four substances (ethanol, n-
1627 hexanol, methyl acetate, methyl ethyl ketone) were incorrectly classified by
1628 two of the four laboratories. The discordant laboratories for these five
1629 substances were not consistently the same two laboratories.
- 1630 • All four laboratories agreed on the classification of 23 (88%) of the 26
1631 substances classified as EU (2001) nonsevere irritants/nonirritants the four
1632 participating laboratories. Three of the four laboratories were in agreement
1633 for the three substances (12%) with discordant classification results (n-butyl
1634 acetate, 4-carboxybenzaldehyde, methyl isobutyl ketone). The discordant
1635 laboratory for these three substances was always the same laboratory.

1636

1637 3.4.2 Quantitative Reanalysis of Interlaboratory Reproducibility

1638 As detailed in the draft BRD, to provide a quantitative assessment of interlaboratory
1639 variability, individual laboratory ICE test results were used to calculate a mean, standard
1640 deviation, and the %CV for corneal opacity, fluorescein retention, corneal swelling, and the
1641 irritation index for each of the 59 substances tested in the Balls et al. (1995) study. Mean and
1642 median %CV values were calculated to provide an assessment of overall variability. This
1643 analysis was not affected by the information received subsequent to the release of the draft
1644 BRD on November 1, 2004, and therefore is not presented here.

1645

1646 3.4.3 Additional Reanalyses of Interlaboratory Reproducibility

1647 The draft BRD also contains a description of the analysis performed by Balls et al. (1995) in
1648 which they determined the interlaboratory correlation between ICE test method endpoint data
1649 generated by each laboratory for all substances tested, as well as for subsets of test
1650 substances (water-soluble, water-insoluble, surfactants, solids, solutions, and liquids). This
1651 analysis was not affected by the information received subsequent to the release of the draft
1652 BRD on November 1, 2004, and therefore is not presented here.

1653

1653 **Table II-11. Interlaboratory Variability of Balls et al. (1995) for Substances Classified**
 1654 **as Ocular Corrosives/Severe Irritants or Nonsevere Irritants/Nonirritants**
 1655 **Using the EU¹ Classification System**
 1656

Classification (<i>in vivo</i> / <i>in vitro</i>) ²	Data Set	Number of Substances	Number of Testing Labs ³	Substances with 100% Agreement Among Labs	Substances with 75% Agreement Among Labs	Substances with 50% Agreement Among Labs
+ / +	New ⁴	10	4 ³	6 (60%)	3 (30%)	1 (10%)
	Old ⁴	12	4 ³	9 (67%)	3 (25%)	1 (8%)
+ / -	New	9	4	7 (78%)	2 (22%)	0 (0%)
	Old	9	4	7 (78%)	2 (22%)	0 (0%)
- / +	New	5	4	1 (20%)	0 (0%)	4 (80%)
	Old	7	4	2 (29%)	0 (0%)	5 (71%)
- / -	New	26	4	23 (88%)	3 (12%)	0 (0%)
	Old	31	4	28 (90%)	3 (10%)	0 (0%)
? / -	New	5	4	5 (100%)	0 (0%)	0 (0%)
	Old	0	4	0 (0%)	0 (0%)	0 (0%)
? / +	New	4	4	3 (75%)	0 (0%)	1 (25%)
	Old	0	4	0 (0%)	0 (0%)	0 (0%)
TOTAL	New	59	4 ³	45 (76%)	8 (14%)	6 (10%)
	Old	59	4 ³	45 (76%)	8 (14%)	6 (10%)

1657 ¹EU = European Union (EU [2001]).

1658 ²A “+” indicates that the substance was assigned an overall classification of corrosive or severe irritant
 1659 (Category R41); a “-” indicates that the substance was assigned an overall classification of nonsevere irritant
 1660 (Category R36) or nonirritant; a “?” indicates that, due to the lack of appropriate *in vivo* data, an EU
 1661 classification could not be made. See **Section II-2.0** for a description of the rules followed to classify the ocular
 1662 irritancy of test substances tested multiple times *in vitro*.

1663 ³Scores for fluorescein retention and corneal swelling were not provided from one laboratory for one substance
 1664 (trichloroacetic acid, 30%), and therefore this substance was classified based on results from only three
 1665 laboratories.

1666 ⁴New = accuracy statistics based on the revised analysis; Old = accuracy statistics based on the previous
 1667 analysis included in the draft ICE BRD.
 1668
 1669

1670 3.5 ICE Test Method Historical Positive and Negative Control Data - Reanalysis

1671
 1672 Concurrent positive control substances have not been employed in the ICE test method, and
 1673 therefore, an evaluation of historical positive control data is not possible. One eye is
 1674 traditionally included in each study as a negative/vehicle controls (isotonic saline). However,
 1675 irritancy data for this control eye were not available for inclusion in the original analysis.
 1676 Subsequent to the original analysis, individual eye data were obtained from negative control
 1677 eyes that could be used to perform a CV analysis on between-experiment values for each of
 1678 the test method endpoints (i.e., corneal thickness/swelling, corneal opacity, fluorescein
 1679 retention) along with the ICE Irritation Index for each test substance (**Table II-12**). This
 1680 analysis revealed that responses in the negative control eye remain relatively consistent.
 1681

1681
1682
1683**Table II-12. 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

1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696¹No. = Number.²Max = Maximum.³Index = ICE Irritation Index (= CS x [CO x 20] + FR x 20).⁴Isotonic saline.⁵SD = Standard deviation.⁶CV = coefficient of variation (%CV = [standard deviation/mean] x 100); FR = fluorescein retention**3.6 Reliability of the ICE Test Method for Identifying Ocular Corrosives and Severe Irritants – Summary of Reanalysis**

Previously, an evaluation of the intralaboratory repeatability and reproducibility of the ICE test method could not be conducted. However, subsequent to the original reliability analysis

1697 (see draft ICE BRD, November 1, 2004), replicate data received allowed for a quantitative
1698 analysis of intralaboratory repeatability and reproducibility of ICE test method endpoints.
1699

1700 The range of %CV values for the corneal thickness measurement, when results were
1701 compared within experiments, was from 0.9 to 6.1. The other endpoints evaluated produced
1702 ranges of %CV values that were larger, with variability most prominent with the nonirritating
1703 substance (SP-1). However, this could be an exaggeration of variability given the relatively
1704 small values that were produced from the nonirritating substance relative to the irritating and
1705 corrosive substances (i.e., corneal swelling values of 2, 0, and 3 yield a higher % CV than
1706 values of 11, 14, and 18). A similar discussion can also be applied to the variability among
1707 the qualitative endpoints (i.e., corneal opacity and fluorescein retention) given the small
1708 dynamic range of their scores (0-4 or 0-3, respectively).
1709

1710 The range of %CV values for the corneal thickness measurement, when results were
1711 compared across labs, was from 1.8 to 6.3. The %CV values for the remaining endpoints had
1712 a larger range (e.g., corneal swelling %CV = 13.9 to 138.7). However, if the nonirritating
1713 substance is removed, the range of %CV values is reduced (e.g., corneal swelling %CV =
1714 13.9 to 22.4).
1715

1716 The previous analysis also included an evaluation of interlaboratory reproducibility using
1717 both qualitative and quantitative approaches. While the quantitative analysis was unaffected
1718 by the new information that was received, the qualitative analysis (correct classification as an
1719 ocular corrosive/severe irritant or as a non-corrosive/non-severe irritant) of the individual
1720 laboratory test results obtained for the EC/HO validation study (Balls et al., [1995])
1721 mandated that this analysis be repeated. However, the results obtained in the revised analysis
1722 were not different from the original analysis (see **Tables II-9 to II-11**).
1723

1724 **4.0. REFERENCES**

1725

1726 ASTM. 1999. Standard practice for conducting an interlaboratory study to determine the
1727 precision of a test methods. ASTM E691-92. In: Annual Book of ASTM Standards.
1728 Philadelphia, PA: American Society for Testing and Materials.
1729

1730 Balls M, Botham PA, Bruner LH, Spielmann H. 1995. The EC/HO international validation
1731 study on alternatives to the Draize eye irritation test. Toxicol In Vitro 9:871-929.
1732

1733 EPA. 1996. Label Review Manual: 2nd Ed. EPA737-B-96-001. Washington, DC: U.S.
1734 Environmental Protection Agency.
1735

1736 EU. 2001. Commission Directive 2001/59/EC of 6 August 2001 adapting to technical
1737 progress for the 28th time Council Directive 67/548/EEC on the approximation of the laws,
1738 regulations and administrative provisions relating to the classification, packaging and
1739 labeling of dangerous substances. Official Journal of the European Communities L255:1-333.
1740

1741 Holzhütter HG, Archer G, Dami N, Lovell DP, Saltelli A, Sjöström M. 1996.
1742 Recommendations of the application of biostatistical methods during the development and

- 1743 validation of alternative toxicological methods. ECVAM Biostatistics Task Force Report 1.
1744 ATLA 24:511-530.
- 1745
1746 ICCVAM. 2003. ICCVAM Guidelines for the Nomination and Submission of New, Revised,
1747 and Alternative Test Methods. NIH Publication No: 03-4508. Research Triangle Park, NC:
1748 National Toxicology Program. Available:
1749 <http://iccvam.niehs.nih.gov/docs/guidelines/subguide.htm> [accessed 12 July 2005].
1750
- 1751 ICCVAM. 1997. Validation and Regulatory Acceptance of Toxicological Methods: A Report
1752 of the *ad hoc* Interagency Coordinating Committee on the Validation of Alternative Methods.
1753 NIH Publication No: 97-3981. Research Triangle Park, NC: National Toxicology Program.
1754 Available: <http://iccvam.niehs.nih.gov/docs/guidelines/validate/pdf> [accessed 12 July 2005].
1755
- 1756 MeSH Medical Subject Heading. Available: <http://www.nlm.nih.gov/mesh> [accessed 12 July
1757 2005].
1758
- 1759 NICEATM. 2004. Draft Background Review Document. Current Status of *In Vitro* Test
1760 Methods for Identifying Ocular Corrosives and Severe Irritants: The Isolated Chicken Eye
1761 (ICE) Test Method. Available: http://iccvam.niehs.nih.gov/methods/ocudocs/ocu_brd.htm
1762 [accessed 12 July 2005].
1763
- 1764 NIEHS. 2005. Second request for data on chemicals evaluated by *in vitro* or *in vivo* ocular
1765 irritancy test methods. Fed Reg 70:9661-9662. Available:
1766 <http://iccvam.niehs.nih.gov/methods/eyeirrit.htm> [accessed 12 July 2005].
1767
- 1768 NIEHS. 2004. Request for public comment on the nomination for ocular toxicity test
1769 methods and related activities and request for data on chemicals evaluated by *in vitro* or *in*
1770 *vivo* ocular irritancy test methods. Fed Reg 69:13859-13861. Available:
1771 <http://iccvam.niehs.nih.gov/methods/eyeirrit.htm> [accessed 12 July 2005].
1772
- 1773 Prinsen MK. 2005. *In vitro* and *in vivo* data for 94 substances tested in the isolated chicken
1774 eye test. Unpublished data provided directly to NICEATM by MK Prinsen, TNO Nutrition
1775 and Food Research Institute.
1776
- 1777 Prinsen MK. 2000. Chicken enucleated eye test with reference surfactants and siloxane
1778 polymers in Phase II of the reference standard validation project; an *ex vivo* alternative to the
1779 Draize eye irritation test with rabbits. TNO Report V99.521b. Unpublished report
1780 provided directly to NICEATM by M Prinsen, TNO Nutrition and Food Research Institute.
1781
- 1782 Prinsen MK. 1996. The chicken enucleated eye test (CEET): A practical (pre)screen for the
1783 assessment of eye irritation/corrosion potential of test materials. Food Chem Toxicol 34:291-
1784 296.
1785
- 1786 Prinsen MK, Koëter BWM. 1993. Justification of the enucleated eye test with eyes of
1787 slaughterhouse animals as an alternative to the Draize eye irritation test with rabbits. Food

- 1788 Chem Toxicol 31:69-76.
1789
1790 UN. 2003. Globally Harmonized System of Classification and Labelling of Chemicals
1791 (GHS). New York & Geneva: United Nations Publications. Available:
1792 http://www.unece.org/trans/danger/publi/ghs/ghs_rev00/00files_ehtml. [accessed 12 July
1793 2005].