# 6.0 HET-CAM TEST METHOD ACCURACY

## 6.1 Accuracy of the HET-CAM Test Method

A critical component of an ICCVAM evaluation of the validation status of a test method is an assessment of the accuracy of the proposed test method when compared to the current reference test method (ICCVAM 2003). This aspect of assay performance is typically evaluated by calculating:

- accuracy (concordance): the proportion of correct outcomes (positive and negative) of a test method
- sensitivity: the proportion of all positive substances that are classified as positive
- specificity: the proportion of all negative substances that are classified as negative
- positive predictivity: the proportion of correct positive responses among substances testing positive
- negative predictivity: the proportion of correct negative responses among substances testing negative
- false positive rate: the proportion of all negative substances that are falsely identified as positive
- false negative rate: the proportion of all positive substances that are falsely identified as negative

The ability of the HET-CAM test method to correctly identify ocular corrosives and severe irritants, as defined by the GHS (UN 2003), EPA (1996), and EU (2001) classification systems[1], was evaluated using two approaches. In the first approach, the performance of HET-CAM was assessed separately for each *in vitro-in vivo* comparative study (i.e., "per study" approach) reviewed in **Sections 4.0** and **5.0**. Within the "per study" analysis approach, there were two different analyses used. In the second approach, the performance of HET-CAM was assessed after pooling data across comparative studies that used the same data analysis method (i.e., IS, IS and ITC, Q-Score, or S-Score).

As mentioned above, for the "per study" accuracy analysis approach, two different types of analyses were used. In the first analysis, the HET-CAM ocular irritancy potential of each substance in each report was determined (**Appendix C**). When the same substance was evaluated in multiple laboratories within the same study (see Balls et al. 1995, Spielmann et al. 1996, and Hagino et al. 1999 in **Appendix C**), the HET-CAM ocular irritancy potential for each independent test result was determined. An overall HET-CAM ocular irritancy classification was the assigned for each substance in the study based on the majority of ocular irritancy classification calls (e.g., if two laboratories classified a substance as a nonirritant and three laboratories classified a substance as a severe irritant; the overall *in vitro* irritancy classification for the substance used in this analysis would be severe irritant). When there was an even number of different irritancy classifications for substances (e.g., two laboratories classified a substance as

[1] For the purposes of this analysis, an ocular corrosive or severe irritant is defined as a substance that would be classified as Category 1 according to the GHS classification system, Category I according to the EPA classification system, or as R41 according to the EU classification system (see **Section 1.0**).

a nonirritant and two laboratories classified a substance as a severe irritant), the more severe irritancy classification was used for the overall classification for the substance (severe irritant, in this case). Once the ocular irritancy potential classification was determined for each substance in each of the studies, the ability of the HET-CAM test method to identify ocular corrosives and severe irritants, as defined by the GHS (UN 2003), EPA (1996), and EU (2001) classification systems, was determined for each study. The overall *in vitro* and *in vivo* classifications assigned to each substance are provided in **Appendix D**.

In the second analysis used in the "per study" evaluation, each classification obtained when the same substance was evaluated in multiple laboratories was used separately to assess test method accuracy (i.e., results were not combined across multiple laboratories to develop an overall HET-CAM ocular irritancy classification). The ability of the HET-CAM test method to identify ocular corrosives and severe irritants, as defined by the three different classification systems, was then determined for reports where multiple results were available for tested substances. This analysis was applied to the CEC (1991), Balls et al. (1995), Spielmann et al. (1996), and Hagino et al. (1999) studies.

In the second approach in evaluating the accuracy of HET-CAM, results from the different studies using the same HET-CAM analysis approach were combined. As discussed in **Section 2.0**, there are several different data analysis methods that have been used (i.e., IS, IS and ITC, Q-Score, S-Score). Therefore, an accuracy assessment was conducted for each analysis method described. When the same substance was evaluated in multiple laboratories, the overall HET-CAM ocular irritancy classification was based on the majority of calls among all of the laboratories in the studies (see **Appendix C**). Once the ocular irritancy classification was determined for each substance, the ability of the HET-CAM test method to identify ocular corrosives and severe irritants, as defined by the GHS (UN 2003), EPA (1996), and EU (2001) classification systems, was determined for each analysis method (**Appendix D**). Since the test methods protocols used in different studies to generate HET-CAM test results are not identical (see **Appendix A** for comparisons of key components of test method protocols), care should be used when interpreting the results of these analyses.

The three ocular hazard classification systems (GHS [UN 2003], EPA [1996], and EU [2001]) considered during each approach use different classification systems and decision criteria to identify ocular corrosives and severe irritants based on *in vivo* rabbit eye test results (see **Sections 1.0** and **4.0**). All three classification systems are based on individual animal response data in terms of the magnitude of the response and on the extent to which induced ocular lesions fail to reverse by day 21. Thus, to evaluate the accuracy of the HET-CAM test method for identifying ocular corrosives and severe irritants, individual rabbit data collected at the different observation times are needed for each substance. However, these data were not consistently available in the reports considered, which limited the number of test results that could be used to assess test method accuracy. Furthermore, most of the *in vivo* classifications used for the analyses presented in this section are based on the results of a single study. Unless otherwise indicated, variability in the *in vivo* classification is unknown.

### 6.1.1 GHS Classification System: HET-CAM Test Method Accuracy

#### 6.1.1.1 *Overall Test Method Accuracy*

Accuracy analyses for ocular corrosives and severe irritants, as defined by the GHS classification system (UN 2003)[2], were evaluated for the following reports: Gettings et al. (1991, 1994, 1996), Bagley et al. (1992), Vinardell and Macián (1994), Balls et al. (1995), Kojima et al. (1995), Spielmann et al. (1996), Gilleron et al. (1997), and Hagino et al. (1999). Of these reports, Balls et al. (1995), Spielmann et al. (1996), and Hagino et al. (1999) provided HET-CAM test data for substances tested in multiple laboratories.

In these studies, HET-CAM test data was provided for a total of 376 substances, 260 of which had sufficient comparative *in vivo* data that could be used to assign an ocular irritancy classification according to the GHS classification system (UN 2003). Of these 260 substances, 92 substances were classified as GHS severe irritants based on results from the *in vivo* rabbit eye test. *In vivo* and *in vitro* irritancy classifications of test substances are provided in **Appendix C** and **Appendix D**.

For one set of data (Spielmann et al. 1996), a large number of substances were available to compare the accuracy of the test method when substances were evaluated at a 10% and 100% concentration *in vitro* and 100% *in vivo*. Therefore, a comparison of the accuracy statistics of these two *in vitro* concentrations was possible. To include the additional HET-CAM test data, which were tested at 10% and 100% concentrations, appropriate data were combined with each of the Spielmann et al. (1996) data sets. These combined data sets were used to evaluate the accuracy of the IS(B) test method, when using a 10% (IS[B]-10) or 100% (IS[B]-100) concentration *in vitro*, to predict the effects produced *in vivo* at 100%. As a corollary to this evaluation, the accuracy of the IS(A) method, when substances were tested at 10% or 100% concentration *in vitro*, to predict the effects produced *in vivo* at 100% concentration also was evaluated.

Based on the data provided in the reports and when results across multiply tested substances were combined to generate a single consensus call per test substance, the HET-CAM test method has an accuracy in predicting substances classified as corrosives or severe irritants, according to the GHS classification system (UN 2003) of 41% to 83%, a sensitivity of 25% to 100%, a specificity of 9% to 100%, a false positive rate of 0% to 91%, and a false negative rate of 0% to 75%[3,4]. The performance characteristics for each report are provided in **Table 6–1**.

The performance statistic ranges for Balls et al. (1995), Spielmann et al. (1996) and Hagino et al. (1999), when results from different testing laboratories are considered separately rather than combined, are: 47% to 80% for accuracy, 27% to 87% for sensitivity, 46% to 82% for

[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] The ranges provided do not include the results obtained for Bagley et al. (1992) and Vinardell and Macián (1994); the number of chemicals evaluated (two each) was deemed to few to consider.

[4] For substances where there were two *in vivo* studies with discordant results (e.g., one study classified the substance as a Category 1 and a second study classified the substance as a Category 2A), the more severe irritancy classification was used for the accuracy analysis.

specificity, 18% to 54% for the false positive rate, 15% to 73% for the false negative rate. These performance characteristics also are provided in **Table 6-1**.

The overall performance statistics, arranged by HET-CAM data analysis method, are provided in **Table 6-2**. Based on the combined test result approach, the HET-CAM test method has an accuracy in predicting substances classified as corrosives or severe irritants, according to the GHS classification system (UN 2003), of 44% to 85%, a sensitivity of 25% to 100%, a specificity of 40% to 100%, a false positive rate of 0% to 60%, and a false negative rate of 0% to 75%

The IS(A)-100 analysis method (substances were tested *in vitro* at a concentration of 100% and compared to substances tested *in vivo* at 100%) had the highest accuracy for predicting ocular corrosives and severe irritants (85% [17/20]). It is noted that for the IS(A)-100 analysis method evaluation represents 20 substances that are mostly formulations. Comparatively, the IS(B) approach (which has a larger database and contains many individual chemicals) had the highest accuracy when 10% concentration tested *in vitro* was compared to 100% concentration tested *in vivo*. The false positive and false negative rates for this analysis method were 33% (20/61) and 30% (12/40), respectively.

6.1.1.2 *Discordant Results According to the GHS Classification System*

To evaluate discordant responses of the HET-CAM test method relative to the *in vivo* hazard classification, several accuracy sub-analyses were performed for each analysis method evaluated. These included specific classes of chemicals with sufficiently robust numbers of substances ($n \geq 5$), as well as certain properties of interest considered relevant to ocular toxicity testing (e.g., surfactants and surfactant-based formulations, pH, physical form).

*IS(A)-10 and IS(A)-100 Analysis Methods*

The overall false positive and false negative rates for the test substances evaluated are provided for two different groups: (a) substances tested at a 10% concentration, and (b) substances tested at a 100% concentration. As is shown in **Table 6-3**, the false negative rate of the IS(A) analysis method is higher when test substances are tested at a 10% concentration (75% [12/16]) when compared to 100% (0% [0/2]). However, the false positive rate of the IS(A) analysis method is lower for the 10% concentration (0% [0/8]) compared to the 100% concentration (17% [3/18]).

As indicated in **Table 6-3**, formulations were the only chemical class with a sufficient number of substances that allowed for an evaluation. Twelve out of sixteen (75%) of formulations (all of which were surfactant-based formulations and all of which were tested as liquids *in vivo*) were underpredicted by the IS(A)-10 analysis method. Comparatively, 18% (3/17) formulations (oil water formulations) evaluated by the IS(A)-100 analysis method were overpredicted. With regard to physical form for the IS(A)-100 analysis method, the false positive and false negative rates were 17% (3/18) and 0% (0/2), respectively for liquids.

Substances were more likely to be underpredicted if (a) the *in vivo* effect was based on a persistent lesion and (b) if the concentration of the test substance *in vitro* was 100% (**Table 6-3**).

**Table 6-1 Evaluation of the Performance of the HET-CAM Test Method In Predicting Ocular Corrosives and Severe Irritants Compared to the *In Vivo* Rabbit Eye Test Method, as Defined by the GHS Classification System, by Study**

| Data Source | Anal.[1] | n[2] | Accuracy | | Sensitivity | | Specificity | | Positive Predictivity | | Negative Predictivity | | False Positive Rate | | False Negative Rate | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | % | No.[3] | % | No. | % | No. | % | No. | % | No. | % | No. | % | No. |
| **Gettings et al. (1991)** | IS(B) | 9/10 | 78 | 7/9 | 100 | 3/3 | 67 | 4/6 | 60 | 3/5 | 100 | 4/4 | 33 | 2/6 | 0 | 0/3 |
| **Gettings et al. (1994)** | IS(A) | 18/18 | 83 | 15/18 | 100 | 1/1 | 82 | 14/17 | 25 | 1/4 | 100 | 14/14 | 18 | 3/17 | 0 | 0/1 |
| **Gettings et al. (1994)** | IS(B) | 18/18 | 78 | 14/18 | 100 | 1/1 | 76 | 13/17 | 20 | 1/5 | 100 | 13/13 | 24 | 4/17 | 0 | 0/1 |
| **Gettings et al. (1996)** | IS(A) | 24/25 | 50 | 12/24 | 25 | 4/16 | 100 | 8/8 | 100 | 4/4 | 40 | 8/20 | 0 | 0/8 | 75 | 12/16 |
| **Gettings et al. (1996)** | IS(B) | 24/25 | 71 | 17/24 | 56 | 9/16 | 100 | 8/8 | 100 | 9/9 | 53 | 8/15 | 0 | 0/8 | 44 | 7/16 |
| **Bagley et al. (1992)** | IS(A) | 2/32 | 0 | 0/2 | - | - | 0 | 0/2 | 0 | 0/2 | - | - | 100 | 2/2 | - | - |
| **Vinardell and Macián (1994)** | IS(B) | 2/13 | 50 | 1/2 | - | - | 50 | 1/2 | 0 | 0/1 | 100 | 1/1 | 50 | 1/2 | - | - |
| **Balls et al. (1995)** | Q | 43/59 | 63 | 27/43 | 100 | 15/15 | 43 | 12/28 | 48 | 15/31 | 100 | 12/12 | 57 | 16/28 | 0 | 0/15 |
| | Q* | 162/177 | 62 | 101/162 | 87 | 45/52 | 51 | 56/110 | 45 | 45/99 | 88 | 56/63 | 49 | 54/110 | 13 | 7/52 |
| **Balls et al. (1995)** | S | 16/59 | 44 | 7/16 | 36 | 4/11 | 60 | 3/5 | 67 | 4/6 | 30 | 3/10 | 40 | 2/5 | 64 | 7/11 |
| | S* | 47/54 | 47 | 22/47 | 27 | 8/30 | 82 | 14/17 | 73 | 8/11 | 38 | 14/36 | 18 | 3/17 | 73 | 22/30 |
| **Kojima et al. (1995)** | IS(A) | 5/24 | 60 | 3/5 | 50 | 2/4 | 100 | 1/1 | 100 | 2/2 | 33 | 1/3 | 0 | 0/1 | 50 | 2/4 |
| **Spielmann et al. (1996)** | IS(B)-10 | 77/120 | 68 | 52/77 | 79 | 19/24 | 62 | 33/53 | 49 | 19/39 | 87 | 33/38 | 38 | 20/53 | 21 | 5/24 |
| | IS(B)-10* | 157/236 | 75 | 118/157 | 74 | 37/50 | 75 | 81/107 | 59 | 37/63 | 86 | 81/94 | 24 | 26/107 | 26 | 13/50 |
| **Spielmann et al. (1996)** | IS(B)-100 | 75/120 | 55 | 41/75 | 88 | 21/24 | 39 | 20/51 | 40 | 21/52 | 87 | 20/23 | 61 | 31/51 | 13 | 3/24 |
| | IS(B)-100* | 150/236 | 58 | 87/150 | 85 | 40/47 | 46 | 47/103 | 42 | 40/96 | 87 | 47/54 | 54 | 56/103 | 15 | 7/47 |
| **Gilleron et al. (1997)** | IS(B) | 54/60 | 41 | 22/54 | 86 | 19/22 | 9 | 3/32 | 40 | 19/48 | 50 | 3/6 | 91 | 29/32 | 14 | 3/22 |
| **Hagino et al. (1999)** | IS(A) | 15/17 | 80 | 12/15 | 100 | 8/8 | 57 | 4/7 | 73 | 8/11 | 100 | 4/4 | 43 | 3/7 | 0 | 0/8 |
| | IS(A)* | 75/85 | 67 | 50/75 | 90 | 36/40 | 40 | 14/35 | 63 | 36/57 | 78 | 14/18 | 60 | 21/35 | 10 | 4/40 |

Abbreviation: GHS = United Nations Globally Harmonized System (UN 2003).

[1]Anal. = Data collection/analysis method used to transform the sample data into HET-CAM scores. IS(A) = Method described in Luepke (1985); IS(B), IS(B)-10, and IS(B)-100 = Method described in Kalweit et al. (1987); Q = Q-Score, Method described in Balls et al. (1995); S = S-Score, Method described in Balls et al. (1995). For those analysis methods marked with an "*", *in vitro* results across multiple testing laboratories were not pooled to develop an overall HET-CAM classification for the test substance. In these analyses, the accuracy evaluation was based on individual study results for substances evaluated in multiple laboratories. Additional information on this approach is provided in **Section 6.1**
[2]n = Number of substances included in this analysis/the total number of substances in the study.
[3]The data on which the percentage calculation is based.

**Table 6-2 Evaluation of the Performance of the HET-CAM Test Method In Predicting Ocular Corrosives and Severe Irritants Compared to the *In Vivo* Rabbit Eye Test Method, as Defined by the GHS Classification System, by HET-CAM Analysis Method**

| Analysis Method[1] | n[2] | Accuracy | | Sensitivity | | Specificity | | Positive Predictivity | | Negative Predictivity | | False Positive Rate | | False Negative Rate | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | % | No.[3] | % | No. | % | No. | % | No. | % | No. | % | No. | % | No. |
| IS(A)-10[4] | 24 | 50 | 12/24 | 25 | 4/16 | 100 | 8/8 | 100 | 4/4 | 40 | 8/20 | 0 | 0/8 | 75 | 12/16 |
| IS(A)-100[4] | 20 | 85 | 17/20 | 100 | 2/2 | 83 | 15/18 | 40 | 2/5 | 100 | 15/15 | 17 | 3/18 | 0 | 0/2 |
| IS(A) | 60 | 63 | 38/60 | 48 | 13/27 | 76 | 25/33 | 62 | 13/21 | 64 | 25/39 | 24 | 8/33 | 52 | 14/27 |
| IS(B)-10[4] | 101 | 68 | 69/101 | 70 | 28/40 | 67 | 41/61 | 58 | 28/48 | 77 | 41/53 | 33 | 20/61 | 30 | 12/40 |
| IS(B)-100[4] | 138 | 54 | 75/138 | 87 | 34/39 | 41 | 41/99 | 37 | 34/92 | 89 | 41/46 | 59 | 58/99 | 13 | 5/39 |
| IS(B)[5] | 106 | 58 | 61/106 | 79 | 33/42 | 44 | 28/64 | 48 | 33/69 | 76 | 28/37 | 56 | 36/64 | 21 | 9/42 |
| Q-Score | 43 | 63 | 27/43 | 100 | 15/15 | 43 | 12/28 | 48 | 15/31 | 100 | 12/12 | 57 | 16/28 | 0 | 0/15 |
| S-Score | 16 | 44 | 7/16 | 36 | 4/11 | 60 | 3/5 | 67 | 4/6 | 30 | 3/10 | 40 | 2/5 | 64 | 7/11 |

Abbreviation: GHS = Globally Harmonized System (UN 2003)
[1]IS(A), IS(A)-10, IS(A)-100 = Method described in Luepke (1985); IS(B), IS(B)-10, IS(B)-100 = Method described in Kalweit et al. (1987); Q = Q-Score, Method described in Balls et al. (1995); S = S-Score, Method described in Balls et al. (1995).
[2]n = Number of substances evaluated in each study.
[3]The data on which the percentage calculation is based.
[4]The analysis compares the ability of the specified concentration tested *in vitro* (IS(A)-10 represents the 10% concentration tested *in vitro*) to predict the effect produced by the undiluted test substance tested *in vivo*.
[5]This analysis excludes substances evaluated in Spielmann et al. (1996).

**Table 6-3 False Positive and False Negative Rates of the IS(A)-10 and IS(A)-100 Analysis Methods, by Chemical Class and Properties of Interest, for the GHS Classification System**

| Category | n[1] | False Positive Rate[2] | | False Negative Rate[2] | |
|---|---|---|---|---|---|
| | | % | No. | % | No. |
| **Overall IS(A)-10** | 24 | 0 | 0/8 | 75 | 12/16 |
| **Overall IS(A)-100** | 20 | 17 | 3/18 | 0 | 0/2 |
| ***Chemical Class[3]-IS(A)-10*** | | | | | |
| **Formulation** | 24 | 0 | 0/8 | 75 | 12/16 |
| ***Chemical Class[3]-IS(A)-100*** | | | | | |
| **Formulation** | 18 | 18 | 3/17 | 0 | 0/1 |
| ***Properties of Interest*** | | | | | |
| **Physical Form: IS(A)-10 Liquids** | 24 | 0 | 0/8 | 75 | 12/16 |
| **Physical Form: IS(A)-100** | | | | | |
| **Liquids** | 20 | 17 | 3/18 | 0 | 0/2 |
| **Solids** | 0 | - | - | - | - |
| **Surfactant-Based Formulations – IS(A)-10** | 24 | 0 | 0/8 | 75 | 12/16 |
| **Category 1 Subgroup-IS(A)-10[4]** | | | | | |
| **- Total** | 2 | - | - | 0 | 0/2 |
| **- 4 (CO=4 at any time)** | 1 | - | - | 0 | 0/1 |
| **- 3 (severity/persistence)** | 0 | - | - | - | - |
| **- 2 (severity)** | 0 | - | - | - | - |
| **- 2-4 combined[5]** | 1 | - | - | 0 | 0/1 |
| **- 1 (persistence)** | 1 | - | - | 0 | 0/1 |
| **Category 1 Subgroup-IS(A)-100[4]** | | | | | |
| **- Total** | 16 | - | - | 75 | 12/16 |
| **- 4 (CO=4 at any time)** | 0 | - | - | - | - |
| **- 3 (severity/persistence)** | 0 | - | - | - | - |
| **- 2 (severity)** | 0 | - | - | - | - |
| **- 2-4 combined[5]** | 0 | - | - | - | - |
| **- 1 (persistence)** | 16 | - | - | 75 | 12/16 |

Abbreviations: CO = corneal opacity, GHS = Globally Harmonized System (UN 2003).

[1] n = number of substances

[2] False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*; False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro*.

[3] Chemical classes included in this table are represented by at least five substances evaluated by the analysis method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh).

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

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

*IS(B)-10 and IS(B)-100 Analysis Methods*

Due to the various concentrations of test substances evaluated in this test method, different permutations of these sub-analyses are provided for comparative purposes. The overall false positive and false negative rates for the test substances evaluated are provided for two different groups: (a) substances tested at a 10% concentration in the entire database and (b) substances tested at a 100% concentration in the entire database. As is shown in **Table 6-4**, the false negative rate of the IS(B) analysis method is higher when test substances are tested at a 10% concentration (30%, 12/40) when compared to 100% (13%, 5/39). However, the false positive rate of the IS(B) analysis method is lower for the 10% concentration (33%, 20/61) compared to the 100% concentration (59%, 58/99).

As indicated in **Table 6-4**, there were some trends in the performance of the HET-CAM test method among subgroups of the tested substances. The chemical class of substances that was consistently overpredicted (i.e., were false positives) by both analysis methods is alcohols. Eight out of a total of 16 (89% [8/9]) and 14 out of a total of 24 alcohols (88% [14/16]) were overpredicted by the IS(B)-10 and IS(B)-100 analysis methods, respectively. Additional chemical classes that were overpredicted by both analysis methods were ethers, amines, organic salts, and heterocyclic compounds. Formulations appeared to have the lowest false positive rates for both analysis methods (0% [0/8] and 26% [6/23]). The chemical classes that were underpredicted by both the IS(B)-10 and IS(B)-100 analysis methods were amines and ethers. Generally, the false negative and false positive rates for the same chemical class were higher for the IS(B)-100 analysis method when compared to the IS(B)-10 analysis method.

With regard to physical form of the substances overpredicted by the IS(B)-10 analysis method, the false positive and false negative rates were 19% (3/16) and 37% (7/19), respectively for liquids and 58% (11/19) and 13% (1/8) for solids. For the IS(B)-100 analysis method, the false positive and false negative rates were 65% (33/51) and 0% (0/9), respectively for liquids and 67% (16/24) and 24% (4/17) for solids.

Information regarding the pH of test substances was only available for a subset of the substances tested at a concentration of 10% or 100% using the IS(B) analysis method. Among the substances that were tested at a 10% concentration, two out of a total of 35 test substances were underpredicted (false negative rate: 13%; 2/16). Among these two, both were acidic (pH < 7.0). For substances tested at a 100% concentration, two out of 35 test substances were underpredicted. Of these substances, one was acidic (pH < 7.0) and one was basic (pH > 7.0). For substances that were overpredicted, basic substances were more overpredicted than acidic substances when tested at a 10% concentration *in vitro* (false positive rate of basic substances = 80% [4/5] vs. false positive rate of acidic substances: 50% [7/14]) (see **Table 6-4**).

Finally, substances were more likely to be underpredicted if (a) the *in vivo* effect was based on a persistent lesion, and (b) if the concentration of the test substance *in vitro* was 10% (**Table 6-4**).

**Table 6-4 False Positive and False Negative Rates of the IS(B)-10 and IS(B)-100 Analysis Methods, by Chemical Class and Properties of Interest, for the GHS Classification System**

| Category | n[1] | False Positive Rate[2] | | False Negative Rate[2] | |
|---|---|---|---|---|---|
| | | % | No. | % | No. |
| **Overall IS(B)-10** | 101 | 33 | 20/61 | 30 | 12/40 |
| **Overall IS(B)-100** | 138 | 59 | 58/99 | 13 | 5/39 |
| ***Chemical Class[3]-IS(B)-10*** | | | | | |
| **Alcohols** | 16 | 89 | 8/9 | 25 | 2/7 |
| **Aldehyde** | 5 | 0 | 0/4 | 100 | 1/1 |
| **Amine** | 7 | 60 | 3/5 | 50 | 1/2 |
| **Ether** | 14 | 50 | 5/10 | 50 | 2/4 |
| **Formulation** | 24 | 0 | 0/8 | 44 | 7/16 |
| **Heterocyclic compound** | 7 | 86 | 6/7 | - | 0/0 |
| **Organic salt** | 7 | 57 | 4/7 | - | 0/0 |
| ***Chemical Class[3]-IS(B)-100*** | | | | | |
| **Alcohols** | 24 | 88 | 14/16 | 13 | 1/8 |
| **Aldehyde** | 6 | 80 | 4/5 | 0 | 0/1 |
| **Amine** | 9 | 83 | 5/6 | 33 | 1/3 |
| **Carboxylic acid/Carboxylic acid salt** | 11 | 60 | 3/5 | 17 | 1/6 |
| **Ester** | 12 | 90 | 9/10 | 0 | 0/2 |
| **Ether** | 16 | 50 | 6/12 | 25 | 1/4 |
| **Formulation** | 27 | 26 | 6/23 | 0 | 0/4 |
| **Heterocyclic compound** | 12 | 78 | 7/9 | 33 | 1/3 |
| **Inorganic salt** | 5 | 100 | 2/2 | 0 | 0/3 |
| **Ketone** | 6 | 67 | 4/6 | - | 0/0 |
| **Organic salt** | 9 | 86 | 6/7 | 0 | 0/2 |
| ***Properties of Interest*** | | | | | |
| **Physical Form: IS(B)-10** | | | | | |
| **Liquids/Solutions** | 35 | 19 | 3/16 | 37 | 7/19 |
| **Solids** | 27 | 58 | 11/19 | 13 | 1/8 |
| **Unknown** | 39 | 23 | 6/26 | 31 | 4/13 |
| **Physical Form: IS(B)-100** | | | | | |
| **Liquids** | 60 | 65 | 33/51 | 0 | 0/9 |
| **Solids** | 41 | 67 | 16/24 | 24 | 4/17 |
| **Unknown** | 37 | 38 | 9/24 | 8 | 1/13 |
| **Surfactant – Total IS(B)-100** | 2 | 50 | 1/2 | - | 0/0 |
| **-nonionic** | 2 | 50 | 1/2 | - | 0/0 |
| **-anionic** | 0 | - | - | - | - |
| **-cationic** | 0 | - | - | - | - |
| **Surfactant-Based Formulations – IS(B)-10** | 24 | 0 | 0/8 | 44 | 7/16 |
| **pH – IS(B)-10** | 35[4] | 58 | 11/19 | 13 | 2/16 |
| **- acidic (pH < 7.0)** | 24 | 50 | 7/14 | 20 | 2/10 |
| **- basic (pH > 7.0)** | 11 | 80 | 4/5 | 0 | 0/6 |

| Category | n[1] | False Positive Rate[2] | | False Negative Rate[2] | |
|---|---|---|---|---|---|
| | | % | No. | % | No. |
| **pH – IS(B)-100** | 35[4] | 68 | 13/19 | 13 | 2/16 |
| **- acidic (pH < 7.0)** | 23 | 69 | 9/13 | 10 | 1/10 |
| **- basic (pH > 7.0)** | 12 | 67 | 4/6 | 17 | 1/6 |
| **Category 1 Subgroup-IS(B)-10[5]** | | | | | |
| **- Total** | 40 | - | - | 30 | 12/40 |
| **- 4 (CO=4 at any time)** | 13 | - | - | 15 | 2/13 |
| **- 3 (severity/persistence)** | 0 | - | - | - | - |
| **- 2 (severity)** | 0 | - | - | - | - |
| **- 2-4 combined[6]** | 13 | - | - | 15 | 2/13 |
| **- 1 (persistence)** | 27 | - | - | 37 | 10/27 |
| **Category 1 Subgroup-IS(B)-100[5]** | | | | | |
| **- Total** | 38[7] | - | - | 11 | 4/38 |
| **- 4 (CO=4 at any time)** | 19 | - | - | 11 | 2/19 |
| **- 3 (severity/persistence)** | 1 | - | - | 100 | 1/1 |
| **- 2 (severity)** | 2 | - | - | 0 | 0/2 |
| **- 2-4 combined[6]** | 22 | - | - | 14 | 3/22 |
| **- 1 (persistence)** | 16 | - | - | 6 | 1/16 |

Abbreviations: CO = corneal opacity, GHS = Globally Harmonized System (UN 2003).
[1]n = number of substances
[2]False Positive Rate = the proportion of all negative substances that are falsely identified as negative *in vitro*; False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*.
[3] Chemical classes included in this table are represented by at least five substances evaluated by the analysis method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh).
[4] Total number of GHS Category 1 substances for which pH information was obtained.
[5] NICEATM-defined subgroups assigned based on the lesions that drove classification of a GHS Category 1 substance. 4: corneal opacity (CO) = 4 at any time; 3: based on lesions that are both severe (not including CO=4) and persistent; 2: based on lesions that are severe (not including CO=4); 1: based on lesions that are persistent
[6] Subcategories 2 to 4 combined to allow for a direct comparison of GHS Category 1 substances classified *in vivo* based on some lesion severity component and those classified based on persistent lesions alone.
[7]The number of substances evaluated in the Category 1 subgroup analysis may be less than the total number of *in vivo* Category 1 substances evaluated since some substances could not be classified into the subgroups used in the evaluation.

*Q-Score Analysis Method*

As is shown in **Table 6-5**, the false positive and negative rates of the Q-Score analysis method are 57% (16/28) and 0% (0/15), respectively.

As indicated in **Table 6-5**, the chemical classes that were overpredicted by the Q-Score analysis method were alcohols and esters. The false negative rate was 0% for all chemical classes shown in the table.

With regard to physical form of the substances, the false positive and false negative rates were 59% (16/27) and 0% (0/11), respectively for liquids and 0% (0/1) and 0% (0/4) for solids.

**Table 6-5 False Positive and False Negative Rates of the Q-Score Analysis Method, by Chemical Class and Properties of Interest, for the GHS Classification System**

| Category | n[1] | False Positive Rate[2] | | False Negative Rate[2] | |
|---|---|---|---|---|---|
| | | % | No. | % | No. |
| **Overall Q-Score** | 43 | 57 | 16/28 | 0 | 0/15 |
| ***Chemical Class***[3] | | | | | |
| **Alcohols** | 10 | 50 | 4/8 | 0 | 0/2 |
| **Carboxylic Acid/Carboxylic Acid Salt** | 5 | 100 | 2/2 | 0 | 0/3 |
| **Ester** | 7 | 43 | 3/7 | - | 0/0 |
| **Heterocyclic compound** | 7 | 50 | 1/2 | 0 | 0/5 |
| **Onium** | 7 | 0 | 0/2 | 0 | 0/5 |
| ***Properties of Interest*** | | | | | |
| **Physical Form:** | | | | | |
| **Liquids** | 38 | 59 | 16/27 | 0 | 0/11 |
| **Solids** | 5 | 0 | 0/1 | 0 | 0/4 |
| **Category 1 Subgroup**[4] | | | | | |
| **- Total** | 14[6] | - | - | 0 | 0/14 |
| **- 4 (CO=4 at any time)** | 7 | - | - | 0 | 0/7 |
| **- 3 (severity/persistence)** | 0 | - | - | - | 0/0 |
| **- 2 (severity)** | 3 | - | - | 0 | 0/3 |
| **- 2-4 combined**[5] | 10 | - | - | 0 | 0/10 |
| **- 1 (persistence)** | 4 | - | - | 0 | 0/4 |

Abbreviations: CO = corneal opacity, GHS = Globally Harmonized System (UN 2003).
[1]n = number of substances
[2]False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro*; False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*.
[3]Chemical classes included in this table are represented by at least five substances evaluated by the analysis method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh).
[4] NICEATM-defined subgroups assigned based on the lesions that drove classification of a GHS Category 1 substance. 4: corneal opacity (CO) = 4 at any time; 3: based on lesions that are both severe (not including CO=4) and persistent; 2: based on lesions that are severe (not including CO=4); 1: based on lesions that are persistent
[5] Subcategories 2 to 4 combined to allow for a direct comparison of GHS Category 1 substances classified *in vivo* based on some lesion severity component and those classified based on persistent lesions alone.
[6] The number of substances evaluated in the Category 1 subgroup analysis may be less than the total number of *in vivo* Category 1 substances evaluated since some substances could not be classified into the subgroups used in the evaluation.

Since there was an overall 0% (0/15) false negative rate, there was a 0% false negative rate when the accuracy of the analysis method was evaluated when compared to different *in vivo* lesion types.

*S-Score Analysis Method*

As is shown in **Table 6-6**, the false positive and false negative rates of the S-Score analysis method are 40% (2/5) and 64% (7/11).

**Table 6-6 False Positive and False Negative Rates of the S-Score Analysis Method, by Chemical Class and Properties of Interest, for the GHS Classification System**

| Category | n[1] | False Positive Rate[2] | | False Negative Rate[2] | |
|---|---|---|---|---|---|
| | | % | No. | % | No. |
| **Overall S-Score** | 16 | 40 | 2/5 | 64 | 7/11 |
| ***Chemical Class[3]*** | | | | | |
| **Carboxylic Acid/Carboxylic Acid Salt** | 5 | 0 | 0/1 | 75 | 3/4 |
| ***Properties of Interest*** | | | | | |
| **Physical Form:** | | | | | |
| **Liquids** | 0 | - | - | - | - |
| **Solids** | 16 | 40 | 2/5 | 64 | 7/11 |
| **Category 1 Subgroup[4]** | | | | | |
| **- Total** | 10[6] | - | - | 60 | 6/10 |
| **- 4 (CO=4 at any time)** | 5 | - | - | 80 | 4/5 |
| **- 3 (severity/persistence)** | 1 | - | - | 100 | 1/1 |
| **- 2 (severity)** | 1 | - | - | 100 | 1/1 |
| **- 2-4 combined[5]** | 7 | - | - | 86 | 6/7 |
| **- 1 (persistence)** | 3 | - | - | 0 | 0/3 |

Abbreviations: CO = corneal opacity, GHS = Globally Harmonized System (UN 2003).
[1]n = number of substances
[2]False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*; False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro*.
[3] Chemical classes included in this table are represented by at least five substances evaluated by the analysis method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh).
[4] NICEATM-defined subgroups assigned based on the lesions that drove classification of a GHS Category 1 substance. 4: corneal opacity (CO) = 4 at any time; 3: based on lesions that are both severe (not including CO=4) and persistent; 2: based on lesions that are severe (not including CO=4); 1: based on lesions that are persistent
[5] Subcategories 2 to 4 combined to allow for a direct comparison of GHS Category 1 substances classified *in vivo* based on some lesion severity component and those classified based on persistent lesions alone.
[6] The number of substances evaluated in the Category 1 subgroup analysis may be less than the total number of *in vivo* Category 1 substances evaluated since some substances could not be classified into the subgroups used in the evaluation.

The only chemical class with sufficient substances to conduct an analysis was carboxylic acids/carboxylic acid salts. In this chemical class, the false negative rate was 75% (3/4) while the false positive rate was 0% (0/1). With regard to physical form of the substances, all substances tested using this analysis method were solids; thus, the false negative rate was 64% (7/11). Finally, substances were more likely to be underpredicted if (a) the *in vivo* effect was based on a corneal opacity of 4 at any time.

### 6.1.2 EPA Classification System: HET-CAM Test Method Accuracy

#### 6.1.2.1 *Overall Test Method Accuracy*

Accuracy analyses for ocular corrosives and severe irritants, as defined by the EPA classification system (EPA 1996), were evaluated for the following reports: Gettings et al. (1991, 1994, 1996), Bagley et al. (1992), Vinardell and Macián (1994), Balls et al. (1995), Kojima et al. (1995), Spielmann et al. (1996), Gilleron et al. (1997), and Hagino et al. (1999). Of these reports, Balls et al. (1995), Spielmann et al. (1996), and Hagino et al. (1999) provided HET-CAM test data for substances tested in multiple laboratories.

In these studies, HET-CAM test data was provided for a total of 376 substances, 256 of which had sufficient *in vivo* data to be assigned an ocular irritancy classification as defined by the EPA classification system (EPA 1996)[5]. Based on results from the *in vivo* rabbit eye test, 76 of these 256 substances were classified as severe irritants (i.e., Category I). *In vivo* and *in vitro* irritancy classifications of test substances are provided in **Appendix C** and **Appendix D**.

As described in the previous section (see **Section 6.1.1**), a large number of substances were available to compare the accuracy of the test method when substances were evaluated at a 10% and 100% concentration *in vitro* and 100% *in vivo*. As conducted previously, to include the additional HET-CAM test data, which were tested at 10% and 100% concentrations, appropriate data were combined with each of the Spielmann et al. data sets. These combined data sets were used to evaluate the accuracy of the IS(B) test method, when using a 10% (IS(B)-10) or 100% (IS(B)-100) concentration *in vitro*, to predict the effects produced *in vivo* at 100%. As a corollary to this evaluation, the accuracy of the IS(A) method, when substances were tested at 10% or 100% concentration *in vitro*, to predict the effects produced *in vivo* at 100% concentration was evaluated.

Based on the data provided in the ten reports and when results across multiply tested substances were combined to generate a single consensus call per test substance, the HET-CAM test method has an accuracy in predicting substances classified as corrosives or severe irritants, according to the EPA classification system (EPA 1996), of 38% to 83%, a sensitivity of 24% to 100%, a specificity of 12% to 100%, a false positive rate of 0% to 88%, and a false negative rate of 0% to 76%[6]. The performance characteristics for each report are provided in **Table 6-7**.

The performance statistic ranges for Balls et al. (1995), Spielmann et al. (1996) and Hagino et al. (1999), when results from different testing laboratories are considered separately rather than combined, are: of 53% to 72% for accuracy, 32% to 94% for sensitivity, 35% to 83% for specificity, 17% to 65% for the false positive rate, and 6% to 68% for the false negative rate. These performance characteristics are provided in **Table 6-7**.

The overall performance statistics, arranged by HET-CAM data analysis method, are provided in **Table 6-8**. Based on the combined test result approach, the HET-CAM test method has an accuracy in predicting substances classified as corrosives or severe irritants, according to the EPA classification system (EPA 1996), of 48% to 85%, a sensitivity of 24% to 100%, a specificity of 41% to 100%, a false positive rate of 0% to 59%, and a false negative rate of 0% to 76%.

[5] For the purpose of this accuracy analysis, *in vivo* rabbit study results were used to identify GHS Category I irritants (i.e., severe irritants); substances classified as EPA Category II, III, and IV were identified as nonsevere irritants.
[6] For substances where there were two *in vivo* studies with discordant results (e.g., one study classified the substance as a Category I and a second study classified the substance as a Category II), the more severe irritancy classification was used for the accuracy analysis.

**Table 6-7 Evaluation of the Performance of the HET-CAM Test Method In Predicting Ocular Corrosives and Severe Irritants Compared to the *In Vivo* Rabbit Eye Test Method, as Defined by the EPA Classification System, by Study**

| Data Source | Anal.[1] | n[2] | Accuracy | | Sensitivity | | Specificity | | Positive Predictivity | | Negative Predictivity | | False Positive Rate | | False Negative Rate | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | % | No.[3] | % | No. | % | No. | % | No. | % | No. | % | No. | % | No. |
| **Gettings et al. (1991)** | IS(B) | 9/10 | 78 | 7/9 | 100 | 3/3 | 67 | 4/6 | 60 | 3/5 | 100 | 4/4 | 33 | 2/6 | 0 | 0/3 |
| **Gettings et al. (1994)** | IS(A) | 18/18 | 83 | 15/18 | 100 | 1/1 | 82 | 14/17 | 25 | 1/4 | 100 | 14/14 | 0 | 3/17 | 0 | 0/1 |
| **Gettings et al. (1994)** | IS(B) | 18/18 | 78 | 14/18 | 100 | 1/1 | 76 | 13/17 | 20 | 1/5 | 100 | 13/13 | 24 | 4/17 | 0 | 0/1 |
| **Gettings et al. (1996)** | IS(A) | 25/25 | 48 | 12/25 | 24 | 4/17 | 100 | 8/8 | 100 | 4/4 | 38 | 8/21 | 0 | 0/8 | 76 | 13/17 |
| **Gettings et al. (1996)** | IS(B) | 25/25 | 72 | 18/25 | 59 | 10/17 | 100 | 8/8 | 100 | 10/10 | 53 | 8/15 | 0 | 0/8 | 41 | 7/17 |
| **Bagley et al. (1992)** | IS(A) | 2/32 | 0 | 0/2 | - | - | 0 | 0/2 | 0 | 0/2 | - | - | 100 | 2/2 | - | - |
| **Vinardell and Macián (1994)** | IS(B) | 2/13 | 50 | 1/2 | - | - | 50 | 1/2 | 0 | 0/1 | 100 | 1/1 | 50 | 1/2 | - | - |
| **Balls et al. (1995)** | Q | 44/59 | 61 | 27/44 | 100 | 14/14 | 43 | 13/30 | 45 | 14/31 | 100 | 13/13 | 57 | 17/30 | 0 | 0/14 |
| | Q* | 163/177 | 63 | 103/163 | 92 | 44/48 | 51 | 59/115 | 44 | 44/100 | 94 | 59/63 | 49 | 56/115 | 8 | 4/48 |
| **Balls et al. (1995)** | S | 15/20 | 53 | 8/15 | 44 | 4/9 | 67 | 4/6 | 67 | 4/6 | 44 | 4/9 | 33 | 2/6 | 56 | 5/9 |
| | S* | 43/54 | 53 | 23/43 | 32 | 8/25 | 83 | 15/18 | 73 | 8/11 | 47 | 15/32 | 17 | 3/18 | 68 | 17/25 |
| **Kojima et al. (1995)** | IS(A) | 5/24 | 80 | 4/5 | 67 | 2/3 | 100 | 2/2 | 100 | 2/2 | 67 | 2/3 | 0 | 0/2 | 33 | 1/3 |
| **Spielmann et al. (1996)** | IS(B)-10 | 74/120 | 64 | 47/74 | 80 | 12/15 | 59 | 35/59 | 33 | 12/36 | 92 | 35/38 | 41 | 24/59 | 20 | 3/15 |
| | IS(B)-10* | 148/236 | 72 | 107/148 | 72 | 21/29 | 72 | 86/119 | 39 | 21/54 | 91 | 86/94 | 28 | 33/119 | 28 | 8/29 |
| **Spielmann et al. (1996)** | IS(B)-100 | 71/120 | 51 | 36/71 | 93 | 14/15 | 39 | 22/56 | 29 | 14/48 | 96 | 22/23 | 61 | 34/56 | 7 | 1/15 |
| | IS(B)-100* | 141/236 | 55 | 77/141 | 89 | 25/28 | 46 | 52/113 | 29 | 25/86 | 95 | 52/55 | 54 | 61/113 | 11 | 3/28 |
| **Gilleron et al. (1997)** | IS(B) | 53/60 | 38 | 20/53 | 84 | 16/19 | 12 | 4/34 | 35 | 16/46 | 57 | 4/7 | 88 | 30/34 | 16 | 3/19 |
| **Hagino et al. (1999)** | IS(A) | 15/17 | 73 | 11/15 | 100 | 7/7 | 50 | 4/8 | 64 | 7/11 | 100 | 4/4 | 50 | 4/8 | 0 | 0/7 |
| | IS(A)* | 75/135 | 63 | 47/75 | 94 | 33/35 | 35 | 14/40 | 56 | 33/59 | 88 | 14/16 | 65 | 26/40 | 6 | 2/35 |

Abbreviation: EPA = U.S. Environmental Protection Agency.

[1]Anal. = Data collection/analysis method used to transform the sample data into HET-CAM scores. IS(A) = Method described in Luepke (1985); IS(B), IS(B)-10, and IS(B)-100 = Method described in Kalweit et al. (1987); Q = Q-Score, Method described in Balls et al. (1995); S = S-Score, Method described in Balls et al. (1995). For those analysis methods marked with an "*", *in vitro* results across multiple testing laboratories were not pooled to develop an overall HET-CAM classification for test substances. In these analyses, the accuracy evaluation was based on individual study results for substances evaluated in multiple laboratories. Additional information on this approach is provided in **Section 6.1**.
[2]n = Number of substances included in this analysis/the total number of substances in the study.
[3]The data on which the percentage calculation is based.

**Table 6-8 Evaluation of the Performance of the HET-CAM Test Method In Predicting Ocular Corrosives and Severe Irritants Compared to the *In Vivo* Rabbit Eye Test Method, as Defined by the EPA Classification System, by HET-CAM Analysis Method**

| Analysis Method[1] | n[2] | Accuracy | | Sensitivity | | Specificity | | Positive Predictivity | | Negative Predictivity | | False Positive Rate | | False Negative Rate | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | % | No.[3] | % | No. | % | No. | % | No. | % | No. | % | No. | % | No. |
| IS(A)-10[4] | 25 | 48 | 12/25 | 24 | 4/17 | 100 | 8/8 | 100 | 4/4 | 38 | 8/21 | 0 | 0/8 | 76 | 13/17 |
| IS(A)-100[4] | 20 | 85 | 17/20 | 100 | 2/2 | 83 | 15/18 | 40 | 2/5 | 100 | 15/15 | 17 | 3/18 | 0 | 0/2 |
| IS(A) | 61 | 66 | 40/61 | 46 | 12/26 | 74 | 26/35 | 57 | 12/21 | 65 | 26/40 | 26 | 9/35 | 54 | 14/26 |
| IS(B)-10[4] | 98 | 65 | 64/98 | 68 | 21/31 | 64 | 43/67 | 47 | 21/45 | 81 | 43/53 | 36 | 24/67 | 32 | 10/31 |
| IS(B)-100[4] | 133 | 52 | 69/133 | 89 | 25/28 | 42 | 44/105 | 29 | 25/86 | 94 | 44/47 | 58 | 61/105 | 11 | 3/28 |
| IS(B)[5] | 106 | 57 | 60/106 | 78 | 31/40 | 49 | 29/66 | 46 | 31/68 | 76 | 29/38 | 56 | 37/66 | 22 | 9/40 |
| Q-Score | 44 | 61 | 27/44 | 100 | 14/14 | 43 | 13/30 | 45 | 14/31 | 100 | 13/13 | 57 | 17/30 | 0 | 0/14 |
| S-Score | 15 | 53 | 8/15 | 44 | 4/9 | 67 | 4/6 | 67 | 4/6 | 44 | 4/9 | 33 | 2/6 | 56 | 5/9 |

Abbreviation: EPA = U.S. Environmental Protection Agency.
[1]IS(A), IS(A)-10, IS(A)-100 = Method described in Luepke (1985); IS(B), IS(B)-10, IS(B)-100 = Method described in Kalweit et al. (1987); Q = Q-Score, Method described in Balls et al. (1995); S = S-Score, Method described in Balls et al. (1995).
[2]n = Number of substances evaluated in each study.
[3]The data on which the percentage calculation is based.
[4]The analysis compares the ability of the specified concentration tested *in vitro* (IS(A)-10 represents the 10% concentration tested *in vitro*) to predict the effect produced by the undiluted test substance tested *in vivo*.
[5]This analysis excluded substances evaluated in Spielmann et al. (1996).

The IS(A)-100 analysis approach, when substances were tested *in vitro* at a concentration of 100% and compared to substances tested *in vivo* at 100%, had the highest accuracy for predicting ocular corrosives and severe irritants (85% [17/20]), as classified by the EPA (EPA 1996). It is noted that the database used for this evaluation represents 20 substances that are mostly formulations. Comparatively, the IS(B) approach (which has a larger database and contains many individual chemicals) had the highest accuracy when 10% concentration tested *in vitro* was compared to 100% concentration tested *in vivo*. The false positive and false negative rates for this analysis method were 36% (24/67) and 32% (10/31), respectively.

6.1.2.2 *Discordant Results According to the EPA Classification System*

To evaluate discordant responses of the HET-CAM test method relative to the *in vivo* hazard classification, several accuracy sub-analyses were performed for each analysis method evaluated. These included specific classes of chemicals with sufficiently robust numbers of substances ($n \geq 5$), as well as certain properties of interest considered relevant to ocular toxicity testing (e.g., surfactants and surfactant-based formulations, pH, physical form).

*IS(A)-10 and IS(A)-100 Analysis Methods*

The overall false positive and false negative rates for the test substances evaluated are provided for two different groups: (a) substances tested at a 10% concentration, and (b) substances tested at a 100% concentration. As is shown in **Table 6-9**, the false negative rate of the IS(A) analysis method is higher when test substances are tested at a 10% concentration (76% [13/17]) when compared to 100% (0% [0/2]). However, the false positive rate of the IS(A) analysis method is lower for the 10% concentration (0% [0/8]) compared to the 100% concentration (17% [3/18]).

As indicated in **Table 6-9**, formulations were the only chemical class with a sufficient number of substances that allowed for an evaluation. Thirteen out of seventeen (76%) of formulations (all of which were surfactant-based formulations and all of which were tested as liquids *in vivo*) were underpredicted by the IS(A)-10 analysis method. Comparatively, 18% (3/17) formulations (oil water formulations) evaluated by the IS(A)-100 analysis method were overpredicted. With regard to physical form for the IS(A)-100 analysis method, the false positive and false negative rates were 17% (3/18) and 0% (0/1), respectively for liquids.

*IS(B)-10 and IS(B)-100 Analysis Methods*

Due to the various concentrations of test substances evaluated in this test method, different permutations of these sub-analyses are provided for comparative purposes. The overall false positive and false negative rates for the test substances evaluated are provided for two different groups: (a) substances tested at a 10% concentration in the entire database and (b) substances tested at a 100% concentration in the entire database. As is shown in **Table 6-10**, the false negative rate of the IS(B) analysis method is higher when test substances are tested at a 10% concentration (32%, 10/31) when compared to 100% (11%, 3/28). However, the false positive rate of the IS(B) analysis method is lower for the 10% concentration (36%, 26/67) compared to the 100% concentration (58%, 61/105).

**Table 6-9 False Positive and False Negative Rates of the IS(A)-10 and IS(A)-100 Analysis Methods, by Chemical Class and Properties of Interest, for the EPA Classification System**

| Category | n[1] | False Positive Rate[2] | | False Negative Rate[2] | |
|---|---|---|---|---|---|
| | | % | No. | % | No. |
| **Overall IS(A)-10** | 25 | 0 | 0/8 | 76 | 13/17 |
| **Overall IS(A)-100** | 20 | 17 | 3/18 | 0 | 0/2 |
| ***Chemical Class[3]-IS(A)-10*** | | | | | |
| **Formulation** | 25 | 0 | 0/8 | 76 | 13/17 |
| ***Chemical Class[3]-IS(A)-100*** | | | | | |
| **Formulation** | 18 | 18 | 3/17 | 0 | 0/1 |
| ***Properties of Interest*** | | | | | |
| **Physical Form: IS(A)-10 Liquids** | 25 | 0 | 0/8 | 76 | 13/17 |
| **Physical Form: IS(A)-100 Liquids** | 20 | 17 | 3/18 | 0 | 0/2 |
| **Solids** | 0 | - | - | - | - |
| **Surfactant-Based Formulations – IS(A)-10** | 25 | 0 | 0/8 | 76 | 13/17 |

Abbreviations: EPA = U.S. Environmental Protection Agency (EPA 1996).
[1]n = number of substances
[2]False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro*; False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*.
[3]Chemical classes included in this table are represented by at least five substances evaluated by the analysis method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh).

As indicated in **Table 6-10**, there were some trends in the performance of the HET-CAM test method among subgroups of the tested substances. The chemical class of substances that was consistently overpredicted according the EPA classification system (i.e., were false positives) by both analysis methods is alcohols. Eight out of a total of 15 (89% [8/9]) and 14 out of a total of 23 alcohols (88% [14/16]) were overpredicted by the IS(B)-10 and IS(B)-100 analysis methods, respectively. Additional chemical classes that were overpredicted by both analysis methods were ethers, amines, organic salts, and heterocyclic compounds. Formulations appeared to have the lowest false positive rates for both analysis methods (0% [0/8]) and 26% [6/23]). Generally, the false negative and false positive rates for the same chemical class were higher for the IS(B)-100 analysis method when compared to the IS(B)-10 analysis method.

With regard to physical form of the substances overpredicted by the IS(B)-10 analysis method, the false positive and false negative rates were 19% (3/16) and 37% (7/19), respectively for liquids and 65% (15/23) and 0% (0/1) for solids. For the IS(B)-100 analysis method, the false positive and false negative rates were 65% (33/51) and 0% (0/9), respectively for liquids and 66% (19/29) and 25% (2/8) for solids.

**Table 6-10 False Positive and False Negative Rates of the IS(B)-10 and IS(B)-100 Analysis Methods, by Chemical Class and Properties of Interest, for the EPA Classification System**

| Category | n[1] | False Positive Rate[2] | | False Negative Rate[2] | |
|---|---|---|---|---|---|
| | | % | No. | % | No. |
| **Overall IS(B)-10** | 98 | 36 | 24/67 | 32 | 10/31 |
| **Overall IS(B)-100** | 133 | 58 | 61/105 | 11 | 3/28 |
| ***Chemical Class[3]-IS(B)-10*** | | | | | |
| **Alcohols** | 15 | 89 | 8/9 | 33 | 2/6 |
| **Aldehyde** | 6 | 80 | 4/5 | 0 | 0/1 |
| **Amine** | 5 | 60 | 3/5 | - | 0/0 |
| **Ether** | 11 | 50 | 5/10 | 0 | 0/1 |
| **Formulation** | 25 | 0 | 0/8 | 41 | 7/17 |
| **Heterocyclic compound** | 8 | 88 | 7/8 | - | 0/0 |
| **Organic salt** | 7 | 57 | 4/7 | - | 0/0 |
| ***Chemical Class[3]-IS(B)-100*** | | | | | |
| **Alcohols** | 23 | 88 | 14/16 | 14 | 1/7 |
| **Aldehyde** | 6 | 80 | 4/5 | 0 | 0/1 |
| **Amine** | 8 | 71 | 5/7 | 100 | 1/1 |
| **Carboxylic acid/Carboxylic acid salt** | 9 | 60 | 3/5 | 25 | 1/4 |
| **Ester** | 11 | 90 | 9/10 | 0 | 0/1 |
| **Ether** | 13 | 50 | 6/12 | 0 | 0/1 |
| **Formulation** | 27 | 26 | 6/23 | 0 | 0/4 |
| **Heterocyclic compound** | 12 | 78 | 7/9 | 33 | 1/3 |
| **Inorganic salt** | 5 | 100 | 4/4 | 0 | 0/1 |
| **Ketone** | 6 | 67 | 4/6 | - | 0/0 |
| **Organic salt** | 10 | 75 | 6/8 | 0 | 0/2 |
| ***Properties of Interest*** | | | | | |
| **Physical Form: IS(B)-10** | | | | | |
| **Liquids/Solutions** | 35 | 19 | 3/16 | 37 | 7/19 |
| **Solids** | 24 | 65 | 15/23 | 0 | 0/1 |
| **Unknown** | 39 | 21 | 6/28 | 27 | 3/11 |
| **Physical Form: IS(B)-100** | | | | | |
| **Liquids** | 60 | 65 | 33/51 | 0 | 0/9 |
| **Solids** | 37 | 66 | 19/29 | 25 | 2/8 |
| **Unknown** | 36 | 36 | 9/25 | 9 | 1/11 |
| **Surfactant – Total IS(B)-100** | 2 | 50 | 1/2 | - | 0/0 |
| **-nonionic** | 2 | 50 | 1/2 | - | 0/0 |
| **-anionic** | 0 | - | - | - | - |
| **-cationic** | 0 | - | - | - | - |
| **Surfactant-Based Formulations – IS(B)-10** | 25 | 0 | 0/8 | 41 | 7/17 |
| **pH – IS(B)-10** | 32[4] | 58 | 14/24 | 0 | 0/8 |
| **- acidic (pH < 7.0)** | 19 | 53 | 8/15 | 0 | 0/4 |
| **- basic (pH > 7.0)** | 13 | 78 | 7/9 | 0 | 0/4 |

| Category | n[1] | False Positive Rate[2] | | False Negative Rate[2] | |
|---|---|---|---|---|---|
| | | % | No. | % | No. |
| **pH – IS(B)-100** | 30[4] | 73 | 16/22 | 0 | 0/8 |
| **- acidic (pH < 7.0)** | 18 | 64 | 9/14 | 0 | 0/4 |
| **- basic (pH > 7.0)** | 12 | 88 | 7/8 | 0 | 0/4 |

Abbreviation: EPA = U.S. Environmental Protection Agency.
[1]n = number of substances
[2]False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro*; False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*.
[3]Chemical classes included in this table are represented by at least five substances evaluated by the analysis method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh).
[4] Total number of EPA Category I substances for which pH information was obtained.

Information regarding the pH of test substances was available for a subset of the substances tested. Among the substances that were tested at a 10% concentration, zero out of a total of 32 test substances were underpredicted (false negative rate: 0% [0/8]). The false positive rate for substances (a) tested at 10% concentration and (b) pH information was available was 58% (14/24). When these substances were separated based on pH, the false positive rate for acidic substances was 53% (8/15) and for basic substances was 78% (7/9). For substances tested at a 100% concentration, the false negative rate for substances for which pH information was available was 0% (0/8). Basic test substances had a higher false positive rate than acidic substances when tested at a 100% concentration *in vitro* (false positive rate of basic substances = 88% [7/8] vs. false positive rate of acidic substances: 64% [9/14]) (see **Table 6-10**).

*Q-Score Analysis Method*
As is shown in **Table 6-11**, the false positive and negative rates of the Q-Score analysis method are 57% (17/30) and 0% (0/14), respectively.

As indicated in **Table 6-11**, there were some trends in the performance of the Q-Score analysis method among subgroups of the tested substances. The chemical classes that were overpredicted according the EPA classification system (i.e., were false positives) were alcohols, carboxylic acids/carboxylic acid salts, esters, and heterocyclic compounds. The false negative rate was 0% for all chemical classes shown in the table.

With regard to physical form of the substances overpredicted by the Q-Score analysis method, the false positive and false negative rates were 61% (17/28) and 0% (0/10), respectively for liquids, and 0% (0/1) and 0% (0/4) for solids.

*S-Score Analysis Method*
As is shown in **Table 6-12**, the false positive and false negative rates of the S-Score analysis method are 33% (2/6) and 56% (5/9).

There were insufficient substances in any single chemical class evaluated ($n \geq 5$) to assess the ability of the S-Score analysis method to predict specific classes. With regard to physical form of the substances, most of the substances evaluated with this method were solids. The false positive rate and false negative rate of solids was 33% (2/6) and 56% (5/9), respectively (**Table 6-12**).

**Table 6-11 False Positive and False Negative Rates of the Q-Score Analysis Method, by Chemical Class and Properties of Interest, for the EPA Classification System**

| Category | n[1] | False Positive Rate[2] | | False Negative Rate[2] | |
|---|---|---|---|---|---|
| | | % | No. | % | No. |
| **Overall Q-Score** | 44 | 57 | 17/30 | 0 | 0/14 |
| ***Chemical Class***[3] | | | | | |
| **Alcohols** | 10 | 50 | 4/8 | 0 | 0/2 |
| **Carboxylic Acid/Carboxylic Acid Salt** | 6 | 100 | 2/2 | 0 | 0/4 |
| **Ester** | 7 | 43 | 3/7 | - | 0/0 |
| **Heterocyclic compound** | 6 | 50 | 1/2 | 0 | 0/4 |
| **Onium** | 6 | 0 | 0/2 | 0 | 0/4 |
| ***Property of Interest*** | | | | | |
| **Physical Form:** | | | | | |
| **Liquids/Solutions** | 38 | 61 | 17/28 | 0 | 0/10 |
| **Solids** | 6 | 0 | 0/2 | 0 | 0/4 |

Abbreviation: EPA = U.S. Environmental Protection Agency.
[1]n = number of substances
[2]False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro*; False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*.
[3]Chemical classes included in this table are represented by at least five substances evaluated by the analysis method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh).

**Table 6-12 False Positive and False Negative Rates of the S-Score Analysis Method, by Chemical Class and Properties of Interest, for the EPA Classification System**

| Category | n[1] | False Positive Rate[2] | | False Negative Rate[2] | |
|---|---|---|---|---|---|
| | | % | No. | % | No. |
| **Overall S-Score** | 15 | 33 | 2/6 | 56 | 5/9 |
| ***Property of Interest*** | | | | | |
| **Physical Form** | | | | | |
| **Liquids** | 0 | - | - | - | - |
| **Solids** | 15 | 33 | 2/6 | 56 | 5/9 |

Abbreviation: EPA = U.S. Environmental Protection Agency.
[1]n = number of substances
[2]False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro*; False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*.

### 6.1.3 EU Classification System: HET-CAM Test Method Accuracy

#### 6.1.3.1 *Overall Test Method Accuracy*

Accuracy analyses for ocular corrosives and severe irritants, as defined by the EU classification system (EU 2001), were evaluated for the following reports: CEC (1991), Gettings et al. (1991, 1994, 1996), Bagley et al. (1992), Vinardell and Macián (1994), Balls et al. (1995), Kojima et al. (1995), Spielmann et al. (1996), Gilleron et al. (1996, 1997), and Hagino et al. (1999). Of these reports, CEC (1991), Balls et al. (1996), Spielmann et al. (1996), and Hagino et al. (1999) provided HET-CAM data for substances tested in multiple laboratories.

In these studies, HET-CAM test data was provided for a total of 381 substances, 312 of which had sufficient *in vivo* data to be assigned an ocular irritancy classification as defined by the EU

classification system (EU 2001)[7]. Based on results from the *in vivo* rabbit eye test, 85 of these 312 substances were classified as severe irritants (i.e., R41). *In vitro* and *in vivo* classifications of these substances are provided in **Appendix C** and **Appendix D**.

As described in **Section 6.1.1.1**, a large number of substances were available to compare the accuracy of the test method when substances were evaluated at a 10% and 100% concentration *in vitro* and 100% *in vivo*. As conducted previously, to include the additional HET-CAM test data, which were tested at 10% and 100% concentrations, appropriate data were combined with each of the Spielmann et al. data sets. These combined data sets were used to evaluate the overall accuracy of the IS(B) test method, when using a 10% (IS(B)-10) or 100% (IS(B)-100) concentration *in vitro*, to predict the effects produced *in vivo* at 100% concentration. As a corollary to this evaluation, the accuracy of the IS(A) method, when substances were tested at 10% or 100% concentration *in vitro*, to predict the effects produced *in vivo* at 100% concentration was evaluated.

In addition to the analysis methods described previously, two additional analysis methods were evaluated for performance when compared to the EU classification system. These methods, the IS and ITC and the mtc, were evaluated and the results presented in Spielmann et al. (1996). The results of the analysis discussed in the report were included in this section for comparison.

Based on the data provided in these reports and when results across multiply tested substances were combined to generate a single consensus call per test substance, the HET-CAM test method has an accuracy in predicting substances classified as corrosives or severe irritants, according to the EU classification system (EU 2001), of 40% to 88%, a sensitivity of 25% to 100%, a specificity of 10% to 100%, a false positive rate of 0% to 90%, and a false negative rate of 0% to 75%[8]. The performance characteristics for each report are provided in **Table 6-13**.

The performance statistic ranges for CEC (1991), Balls et al. (1995), Spielmann et al. (1996) and Hagino et al. (1999), when results from different testing laboratories are considered separately rather than combined, are: of 55% to 76% for accuracy, 35% to 91% for sensitivity, 38% to 82% for specificity, 18% to 62% for the false positive rate, and 9% to 65% for the false negative rate. These performance characteristics are provided in **Table 6-13**.

The overall performance statistics, arranged by HET-CAM data analysis method, are provided in **Table 6-14**. Based on the combined test result approach, the HET-CAM test method has an accuracy in predicting substances classified as corrosives or severe irritants, according to the EU classification system (EU 2001), of 50% to 85%, a sensitivity of 25% to 100%, a specificity of 46% to 100%, a false positive rate of 0% to 54%, and a false negative rate of 0% to 75%.

[7] For the purpose of this accuracy analysis, *in vivo* rabbit study results were used to identify EU R41 irritants (i.e., severe irritants); substances classified R36 and nonirritants were identified as nonsevere irritants.

[8] For substances where there were two *in vivo* studies with discordant results (e.g., one study classified the substance as a Category I and a second study classified the substance as a Category II), the more severe irritancy classification was used for the accuracy analysis.

**Table 6-13 Evaluation of the Performance of the HET-CAM Test Method In Predicting Ocular Corrosives and Severe Irritants Compared to the *In Vivo* Rabbit Eye Test Method, as Defined by the EU Classification System, by Study**

| Data Source | Anal.[1] | n[2] | Accuracy | | Sensitivity | | Specificity | | Positive Predictivity | | Negative Predictivity | | False Positive Rate | | False Negative Rate | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | % | No.[3] | % | No. | % | No. | % | No. | % | No. | % | No. | % | No. |
| **CEC (1991)** | IS(B) | 26/32 | 62 | 16/26 | 86 | 6/7 | 53 | 10/19 | 40 | 6/15 | 91 | 10/11 | 47 | 9/19 | 14 | 1/7 |
| | IS(B)* | 130/130 | 68 | 88/130 | 67 | 22/33 | 68 | 66/97 | 42 | 22/53 | 86 | 66/77 | 32 | 31/97 | 33 | 11/33 |
| **Gettings et al. (1991)** | IS(B) | 8/10 | 88 | 7/8 | 100 | 3/3 | 80 | 4/5 | 75 | 3/4 | 100 | 4/4 | 20 | 1/5 | 0 | 0/3 |
| **Gettings et al. (1994)** | IS(A) | 18/18 | 83 | 15/18 | 100 | 1/1 | 82 | 14/17 | 25 | 1/4 | 100 | 14/14 | 18 | 13/17 | 0 | 0/1 |
| **Gettings et al. (1994)** | IS(B) | 18/18 | 78 | 14/18 | 100 | 1/1 | 76 | 13/17 | 20 | 1/5 | 100 | 13/13 | 24 | 4/17 | 0 | 0/1 |
| **Gettings et al. (1996)** | IS(A) | 24/25 | 50 | 12/24 | 25 | 4/16 | 100 | 8/8 | 100 | 4/4 | 40 | 8/20 | 0 | 0/8 | 75 | 12/16 |
| **Gettings et al. (1996)** | IS(B) | 24/25 | 71 | 17/24 | 56 | 9/16 | 100 | 8/8 | 100 | 9/9 | 53 | 8/15 | 0 | 0/8 | 44 | 7/16 |
| **Bagley et al. (1992)** | IS(A) | 2/32 | 0 | 0/2 | - | - | 0 | 0/2 | 0 | 0/2 | - | - | 100 | 2/2 | - | - |
| **Vinardell and Macián (1994)** | IS(B) | 2/13 | 50 | 1/2 | - | - | 50 | 1/2 | 0 | 0/1 | 100 | 1/1 | 50 | 1/2 | - | - |
| **Balls et al. (1995)** | Q | 39/49 | 64 | 25/39 | 100 | 13/13 | 46 | 12/26 | 48 | 13/27 | 100 | 12/12 | 54 | 14/26 | 0 | 0/13 |
| | Q* | 146/177 | 64 | 94/146 | 91 | 40/44 | 53 | 54/102 | 45 | 40/88 | 93 | 54/58 | 47 | 48/102 | 9 | 4/44 |
| **Balls et al. (1995)** | S | 14/59 | 50 | 7/14 | 44 | 4/9 | 60 | 3/5 | 67 | 4/6 | 38 | 3/8 | 40 | 2/5 | 56 | 5/9 |
| | S* | 40/54 | 55 | 22/40 | 35 | 8/23 | 82 | 14/17 | 73 | 8/11 | 48 | 14/29 | 18 | 3/17 | 65 | 15/23 |
| **Kojima et al. (1995)** | IS(A) | 4/24 | 75 | 3/4 | 67 | 2/3 | 100 | 1/1 | 100 | 2/2 | 50 | 1/2 | 0 | 0/1 | 33 | 1/3 |
| **Spielmann et al. (1996)[4]** | IS and ITC | 118/118 | 71 | 84/118 | 42 | 19/45 | 89 | 65/73 | 70 | 19/27 | 71 | 65/91 | 11 | 8/73 | 58 | 26/45 |
| **Spielmann et al. (1996)[4]** | mtc10 | 142 | 76 | 108/142 | 52 | 25/48 | 88 | 83/94 | 70 | 25/36 | 78 | 83/106 | 12 | 11/94 | 48 | 23/48 |
| **Spielmann et al. (1996)[4]** | mtc10 | 189 | 77 | 145/189 | 53 | 30/57 | 87 | 115/132 | 64 | 30/47 | 81 | 115/142 | 13 | 17/132 | 47 | 27/57 |
| **Spielmann et al. (1996)** | IS(B)-10 | 71/120 | 66 | 47/71 | 82 | 14/17 | 61 | 33/54 | 40 | 14/35 | 92 | 33/36 | 39 | 21/54 | 18 | 3/17 |
| | IS(B)-10* | 144/236 | 76 | 109/144 | 77 | 27/35 | 75 | 82/109 | 50 | 27/50 | 91 | 82/90 | 25 | 27/109 | 23 | 8/35 |

| Data Source | Anal.[1] | n[2] | Accuracy | | Sensitivity | | Specificity | | Positive Predictivity | | Negative Predictivity | | False Positive Rate | | False Negative Rate | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | % | No.[3] | % | No. | % | No. | % | No. | % | No. | % | No. | % | No. |
| **Spielmann et al. (1996)** | IS(B)-100 | 69/120 | 52 | 36/69 | 94 | 16/17 | 38 | 20/52 | 33 | 16/48 | 95 | 20/21 | 62 | 32/52 | 6 | 1/17 |
| | IS(B)-100* | 138/236 | 70 | 97/138 | 91 | 30/33 | 45 | 47/105 | 34 | 30/88 | 94 | 47/50 | 55 | 58/105 | 9 | 3/33 |
| **Gilleron et al. (1996)** | IS(B) | 46/46 | 57 | 26/46 | 67 | 2/3 | 56 | 24/43 | 10 | 2/21 | 96 | 24/25 | 44 | 19/43 | 33 | 1/3 |
| **Gilleron et al. (1997)** | IS(B) | 48/60 | 40 | 19/48 | 89 | 16/18 | 10 | 3/30 | 37 | 16/43 | 60 | 3/5 | 90 | 27/30 | 11 | 2/18 |
| **Hagino et al. (1999)** | IS(A) | 15/17 | 73 | 11/15 | 100 | 7/7 | 50 | 4/8 | 64 | 7/11 | 100 | 4/4 | 50 | 4/8 | 0 | 0/7 |
| | IS(A)* | 75/85 | 63 | 47/75 | 91 | 32/35 | 38 | 15/40 | 56 | 32/57 | 83 | 15/18 | 62 | 25/40 | 9 | 3/35 |

Abbreviation: EU = European Union.

[1]Anal. = Data collection/analysis method used to transform the sample data into HET-CAM scores. IS(A) = Method described in Luepke (1985); IS(B), IS(B)-10, and IS(B)-100 = Method described in Kalweit et al. (1987); Q = Q-Score, Method described in Balls et al. (1995); S = S-Score, Method described in Balls et al. (1995). For those analysis methods marked with an "*", *in vitro* results across multiple testing laboratories were not pooled to develop an overall HET-CAM classification for the test substances. In these analyses, the accuracy evaluation was based on individual study results for substances evaluated in multiple laboratories. Additional information on this approach is provided in **Section 6.1**.

[2]n = Number of substances included in this analysis/the total number of substances in the study.

[3]The data on which the percentage calculation is based

[4]Results were calculated based on the results presented in Spielmann et al. (1996). Classification of *in vivo* results is described in Spielmann et al. (1996).

**Table 6-14 Evaluation of the Performance of the HET-CAM Test Method In Predicting Ocular Corrosives and Severe Irritants Compared to the *In Vivo* Rabbit Eye Test Method, as Defined by the EU Classification System, by HET-CAM Analysis Method**

| Analysis Method[1] | N[2] | Accuracy | | Sensitivity | | Specificity | | Positive Predictivity | | Negative Predictivity | | False Positive Rate | | False Negative Rate | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | % | No.[3] | % | No. | % | No. | % | No. | % | No. | % | No. | % | No. |
| IS(A)-10[4] | 24 | 50 | 12/24 | 25 | 4/16 | 100 | 8/8 | 100 | 4/4 | 40 | 8/16 | 0 | 0/8 | 75 | 12/16 |
| IS(A)-100[4] | 20 | 85 | 17/20 | 100 | 2/2 | 83 | 15/18 | 40 | 2/5 | 100 | 15/15 | 17 | 3/18 | 0 | 0/2 |
| IS(A) | 59 | 66 | 39/59 | 48 | 12/25 | 74 | 25/34 | 57 | 12/21 | 66 | 25/38 | 26 | 9/34 | 52 | 13/25 |
| IS(B)-10[4] | 95 | 67 | 64/95 | 70 | 23/31 | 66 | 41/62 | 52 | 23/44 | 80 | 41/51 | 34 | 21/62 | 30 | 10/33 |
| IS(B)-100[4] | 164 | 57 | 94/164 | 93 | 31/33 | 48 | 63/131 | 31 | 31/99 | 97 | 63/65 | 52 | 68/131 | 6 | 2/33 |
| IS(B)[5] | 161 | 60 | 97/161 | 80 | 37/46 | 52 | 60/115 | 40 | 37/92 | 87 | 60/69 | 48 | 55/115 | 20 | 9/46 |
| Q-Score | 39 | 64 | 25/39 | 100 | 13/13 | 46 | 12/26 | 48 | 13/27 | 100 | 12/12 | 54 | 14/26 | 0 | 0/13 |
| S-Score | 14 | 50 | 7/14 | 44 | 4/9 | 60 | 3/5 | 67 | 4/6 | 38 | 3/8 | 40 | 2/5 | 56 | 5/9 |
| mtc10[6] | 142 | 76 | 108/142 | 52 | 25/48 | 88 | 83/94 | 70 | 25/36 | 78 | 83/106 | 12 | 11/94 | 48 | 23/48 |
| mtc10[6] | 189 | 77 | 145/189 | 53 | 30/57 | 87 | 115/132 | 64 | 30/47 | 81 | 115/142 | 13 | 17/132 | 47 | 27/57 |
| IS and ITC[6] | 118 | 71 | 84/118 | 42 | 19/45 | 89 | 65/73 | 70 | 19/27 | 71 | 65/91 | 11 | 8/73 | 58 | 26/45 |

Abbreviation: EU=European Union (EU [2001]).

[1]IS(A), IS(A)-10, IS(A)-100 = method described in Luepke (1985); IS(B), IS(B)-10, IS(B)-100 = method described in Kalweit et al. (1987); Q = Q-Score, method described in Balls et al. (1995); S = S-Score, method described in Balls et al. (1995).

[2]N = number of substances evaluated in each study.

[3]Data used to calculate the percentage.

[4]The analysis compares the ability of the specified concentration tested *in vitro* (IS(A)-10 represents the 10% concentration tested *in vitro*) to predict the effect produced by the undiluted test substance tested *in vivo*.

[5]This analysis excluded substances evaluated in Spielmann et al. (1996).

[6]Results were calculated based on the results presented in Spielmann et al. (1996). Classification of *in vivo* results is described in Spielmann et al. (1996).

The IS(A)-100 analysis approach, when substances were tested *in vitro* at a concentration of 100% and compared to substances tested *in vivo* at 100%, had the highest accuracy for predicting ocular corrosives and severe irritants (85% [17/20]), as classified by the EU (EU 2001). It is noted that these results that the database used for this evaluation represents 20 substances that are mostly formulations. Comparatively, the IS(B) approach (which has a larger database and contains many individual chemicals) had the highest accuracy when 10% concentration tested *in vitro* was compared to 100% concentration tested *in vivo*. The false positive and false negative rates for this analysis method were 34% (21/62) and 30% (10/33), respectively.

6.1.3.2 *Discordant Results According to the EU Classification System*

To evaluate discordant responses of the HET-CAM test method relative to the *in vivo* hazard classification, several accuracy sub-analyses were performed for each analysis method evaluated. These included specific classes of chemicals with sufficiently robust numbers of substances ($n \geq 5$), as well as certain properties of interest considered relevant to ocular toxicity testing (e.g., surfactants and surfactant-based formulations, pH, physical form).

*IS(A)-10 and IS(A)-100 Analysis Method*

The overall false positive and false negative rates for the test substances evaluated are provided for two different groups: (a) substances tested at a 10% concentration, and (b) substances tested at a 100% concentration. As is shown in **Table 6-15**, the false negative rate of the IS(A) analysis method is higher when test substances are tested at a 10% concentration (75% [12/16]) when compared to 100% (0% [0/2]). However, the false positive rate of the IS(A) analysis method is lower for the 10% concentration (0% [0/8]) compared to the 100% concentration (17% [3/18]).

As indicated in **Table 6-15**, formulations were the only chemical class with a sufficient number of substances that allowed for an evaluation. Twelve out of sixteen (75%) of formulations (all of which were surfactant-based formulations and all of which were tested as liquids *in vivo*) were underpredicted by the IS(A)-10 analysis method. Comparatively, 18% (3/17) formulations (oil-water formulations) evaluated by the IS(A)-100 analysis method were overpredicted. With regard to physical form for the IS(A)-100 analysis method, the false positive and false negative rates were 17% (3/18) and 0% (0/2), respectively for liquids.

*IS(B)-10 and IS(B)-100*

Due to the various concentrations of test substances evaluated in this test method, different permutations of these sub-analyses are provided for comparative purposes. The overall false positive and false negative rates for the test substances evaluated are provided for two different groups: (a) substances tested at a 10% concentration in the entire database, and (b) substances tested at a 100% concentration in the entire database. As is shown in **Table 6-16**, the false negative rate of the IS(B) analysis method is higher when test substances are tested at a 10% concentration (30%, 10/33) when compared to 100% (6%, 2/33). However, the false positive rate of the IS(B) analysis method is lower for the 10% concentration (34%, 21/62) compared to the 100% concentration (52%, 68/131).

**Table 6-15 False Positive and False Negative Rates of the IS(A)-10 and IS(A)-100 Analysis Methods, by Chemical Class and Properties of Interest, for the EU Classification System**

| Category | n[1] | False Positive Rate[2] | | False Negative Rate[2] | |
|---|---|---|---|---|---|
| | | % | No. | % | No. |
| **Overall IS(A)-10** | 24 | 0 | 0/8 | 75 | 12/16 |
| **Overall IS(A)-100** | 20 | 17 | 3/18 | 0 | 0/2 |
| ***Chemical Class[3]-IS(A)-10*** | | | | | |
| **Formulation** | 24 | 0 | 0/8 | 75 | 12/16 |
| ***Chemical Class[3]-IS(A)-100*** | | | | | |
| **Formulation** | 18 | 18 | 3/17 | 0 | 0/1 |
| ***Properties of Interest*** | | | | | |
| **Physical Form: IS(A)-10 Liquids** | 24 | 0 | 0/8 | 75 | 12/16 |
| **Physical Form: IS(A)-100 Liquids** | 20 | 17 | 3/18 | 0 | 0/2 |
| **Solids** | 0 | - | - | - | - |
| **Surfactant-Based Formulations – IS(A)-10** | 24 | 0 | 0/8 | 75 | 12/16 |

Abbreviation: EU = European Union.
[1]n = number of substances
[2]False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro;* False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*.
[3]Chemical classes included in this table are represented by at least five substances evaluated by the analysis method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh).

As indicated in **Table 6-16**, there were some trends in the performance of the HET-CAM test method among subgroups of the tested substances. The chemical class of substances that was consistently overpredicted according the GHS classification system (i.e., were false positives) by both analysis methods is alcohols. Nine out of a total of 15 (90% [9/10]) and 19 out of at total of 31 alcohols (79% [19/24]) were overpredicted by the IS(B)-10 and IS(B)-100 analysis methods, respectively. Additional chemical classes that were overpredicted by both analysis methods were ethers, organic salts, and heterocyclic compounds. Formulations appeared to have the lowest false positive rates for both analysis methods (0% [0/8] and 23% [5/22]). The chemical classes that were underpredicted by both the IS(B)-10 and IS(B)-100 analysis methods were amines. Generally, the false negative and false positive rates for the same chemical class were higher for the IS(B)-100 analysis method when compared to the IS(B)-10 analysis method.

With regard to physical form of the substances overpredicted by the IS(B)-10 analysis method, the false positive and false negative rates were 19% (3/16) and 38% (7/18), respectively for liquids and 60% (12/20) and 0% (0/3) for solids. For the IS(B)-100 analysis method, the false positive and false negative rates were 61% (40/66) and 0% (0/8), respectively for liquids and 48% (19/40) and 8% (1/13) for solids.

**Table 6-16 False Positive and False Negative Rates of the IS(B)-10 and IS(B)-100 Analysis Methods, by Chemical Class and Properties of Interest, for the EU Classification System**

| Category | n[1] | False Positive Rate[2] | | False Negative Rate[2] | |
|---|---|---|---|---|---|
| | | % | No. | % | No. |
| **Overall IS(B)-10** | 95 | 34 | 21/62 | 30 | 10/33 |
| **Overall IS(B)-100** | 164 | 52 | 68/131 | 6 | 2/33 |
| ***Chemical Class[3]-IS(B)-10*** | | | | | |
| **Alcohols** | 15 | 90 | 9/10 | 40 | 2/5 |
| **Aldehyde** | 5 | 0 | 0/4 | 100 | 1/1 |
| **Amine** | 14 | 30 | 3/10 | 50 | 2/4 |
| **Ether** | 12 | 50 | 5/10 | 0 | 0/2 |
| **Formulation** | 24 | 0 | 0/8 | 44 | 7/16 |
| **Heterocyclic compound** | 7 | 86 | 6/7 | - | 0/0 |
| **Organic salt** | 7 | 57 | 4/7 | - | 0/0 |
| ***Chemical Class[3]-IS(B)-100*** | | | | | |
| **Alcohols** | 31 | 79 | 19/24 | 14 | 1/7 |
| **Aldehyde** | 6 | 80 | 4/5 | 0 | 0/1 |
| **Amine** | 15 | 64 | 7/11 | 25 | 1/4 |
| **Carboxylic acid/Carboxylic acid salt** | 15 | 50 | 5/10 | 0 | 0/5 |
| **Ester** | 12 | 90 | 9/10 | 0 | 0/2 |
| **Ether** | 17 | 47 | 7/15 | 0 | 0/2 |
| **Formulation** | 27 | 23 | 5/22 | 0 | 0/4 |
| **Heterocyclic compound** | 16 | 58 | 7/12 | 25 | 1/4 |
| **Ketone** | 10 | 70 | 7/10 | - | 0/0 |
| **Organic salt** | 12 | 80 | 8/10 | 0 | 0/2 |
| **Organic sulfur containing compound** | 7 | 50 | 2/4 | 0 | 0/3 |
| ***Properties of Interest*** | | | | | |
| **Physical Form: IS(B)-10** | | | | | |
| **Liquids/Solutions** | 34 | 19 | 3/16 | 38 | 7/18 |
| **Solids** | 23 | 60 | 12/20 | 0 | 0/3 |
| **Unknown** | 38 | 23 | 6/26 | 25 | 3/12 |
| **Physical Form: IS(B)-100** | | | | | |
| **Liquids** | 74 | 61 | 40/66 | 0 | 0/8 |
| **Solids** | 53 | 48 | 19/40 | 8 | 1/13 |
| **Unknown** | 37 | 36 | 9/25 | 83 | 1/12 |
| **Surfactant – Total IS(B)-100** | 10 | 44 | 4/9 | 0 | 0/1 |
| **-nonionic** | 6 | 33 | 2/6 | - | 0/0 |
| **-anionic** | 1 | 100 | 1/1 | - | 0/0 |
| **-cationic** | 2 | 100 | 1/1 | 0 | 0/1 |
| **-zwitterionic** | 1 | 0 | 0/1 | - | 0/0 |
| **Surfactant-Based Formulations – IS(B)-10** | 24 | 0 | 0/8 | 44 | 7/16 |
| **pH – IS(B)-10** | 30[4] | 58 | 11/19 | 0 | 0/11 |
| **- acidic (pH < 7.0)** | 21 | 50 | 7/14 | 0 | 0/7 |

| Category | n[1] | False Positive Rate[2] | | False Negative Rate[2] | |
|---|---|---|---|---|---|
| | | % | No. | % | No. |
| **- basic (pH > 7.0)** | 9 | 80 | 4/5 | 0 | 0/4 |
| **pH – IS(B)-100** | 29[4] | 72 | 13/18 | 0 | 0/11 |
| **- acidic (pH < 7.0)** | 20 | 69 | 9/13 | 0 | 0/7 |
| **- basic (pH > 7.0)** | 9 | 80 | 4/5 | 0 | 0/4 |

Abbreviation: EU = European Union.
[1]n = number of substances
[2]False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro;* False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*.
[3]Chemical classes included in this table are represented by at least five substances evaluated by the analysis method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh).
[4]Total number of EU R41 substances for which pH information was obtained.

Information regarding the pH of test substances was available for a subset of the substances tested. Among the substances that were tested at a 10% concentration, 11 out of 30 test substances were overpredicted (false positive rate: 58% [11/19]). Among these, seven were acidic (pH < 7.0), and four were basic. For substances tested at a 100% concentration, 13 out of 29 test substances were overpredicted. Of these substances, nine were acidic (pH < 7.0), and four were basic (pH > 7.0). For substances that were underpredicted, there was a 0% false negative rate for both analysis methods (see **Table 6-16**).

*Q-Score Analysis Method*
As is shown in **Table 6-17**, the false positive and negative rates of the Q-Score analysis method are 56% (14/26) and 0% (0/13), respectively.

**Table 6-17 False Positive and False Negative Rates of the Q-Score Analysis Method, by Chemical Class and Properties of Interest, for the EU Classification System**

| Category | n[1] | False Positive Rate[2] | | False Negative Rate[2] | |
|---|---|---|---|---|---|
| | | % | No. | % | No. |
| **Overall Q-Score** | 39 | 56 | 14/26 | 0 | 0/13 |
| ***Chemical Class[3]*** | | | | | |
| **Alcohols** | 9 | 43 | 3/7 | 0 | 0/2 |
| **Ester** | 7 | 43 | 3/7 | - | 0/0 |
| **Heterocyclic compound** | 7 | 50 | 1/2 | 0 | 0/5 |
| **Onium** | 7 | 0 | 0/2 | 0 | 0/5 |
| ***Properties of Interest*** | | | | | |
| **Physical Form:** | | | | | |
| **Liquids** | 34 | 56 | 14/25 | 0 | 0/9 |
| **Solids** | 5 | 0 | 0/1 | 0 | 0/4 |

Abbreviation: EU = European Union.
[1]n = number of substances
[2]False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro*; False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*.
[3]Chemical classes included in this table are represented by at least five substances evaluated by the analysis method and assignments are based on the MeSH categories (www.nlm.nih.gov/mesh).

There were some trends in the performance of the Q-Score analysis method among subgroups of the tested substances. The chemical class of substances that were consistently overpredicted was alcohols, ester, and heterocyclic compounds. The false negative rate was 0% for all chemical classes shown in the table.

With regard to physical form of the substances overpredicted by the Q-Score analysis method, the false positive and false negative rates were 56% (14/25) and 0% (0/9) for liquids and 0% (0/1) and 0% (0/4) for solids, respectively.

*S-Score Analysis Method*
As is shown in **Table 6-18**, the false positive and false negative rates of the S-Score analysis method are 40% (2/5) and 56% (5/9).

**Table 6-18 False Positive and False Negative Rates of the S-Score Analysis Method, by Chemical Class and Properties of Interest, for the EU Classification System**

| Category | n[1] | False Positive Rate[2] | | False Negative Rate[2] | |
|---|---|---|---|---|---|
| | | % | No. | % | No. |
| **Overall S-Score** | 14 | 40 | 2/5 | 56 | 5/9 |
| ***Properties of Interest*** | | | | | |
| **Physical Form:** | | | | | |
| **Liquids** | 0 | - | - | - | - |
| **Solids** | 14 | 40 | 2/5 | 56 | 5/9 |

Abbreviation: EU = European Union.
[1]n = number of substances
[2]False Positive Rate = the proportion of all negative substances that are falsely identified as positive *in vitro;* False Negative Rate = the proportion of all positive substances that are falsely identified as negative *in vitro*.

There were insufficient substances in any single chemical class evaluated ($n \geq 5$) to assess the ability of the S-Score analysis method to predict specific classes. With regard to physical form of the substances, all substances tested using this analysis method were solids. As shown in **Table 6-18**, the false positive and false negative rates for solids ranged from 40% (2/5) to 56% (5/9).

*IS and ITC Analysis Method*
Based on the information provided in Spielmann et al. (1996), there were eight substances that were identified as false positives. These substances were:

- (-)-phenylephrine
- theophylline sodium acetate
- (+)-phenylephrine
- sodium cyanate
- sodium lauryl ether sulfate
- hyton
- *p*-anisidine
- rubinrot Y

## 6.2 Accuracy of the HET-CAM Test Method for Identifying Ocular Corrosives and Severe Irritants – Summary of Results

While there were some differences in results among the three hazard classification systems evaluated (i.e., EPA [EPA 1996], EU [EU 2001], and GHS [UN 2003]), the accuracy analysis revealed that HET-CAM test method performance was comparable among the three systems (see **Table 6-19**).

**Table 6-19 Ranges of Performance Statistics for Evaluated Analysis Methods for GHS, EPA, and EU Classification Systems**

| Analysis Methods | Accuracy | Sensitivity | Specificity | False Positive Rates | False Negative Rates |
|---|---|---|---|---|---|
| **IS(A)-10** | 48-50% | 24-25% | 100% | 0% | 75-76% |
| **IS(A)-100** | 85% | 100% | 83% | 17% | 0% |
| **IS(B)-10** | 65-68% | 68-70% | 64-67% | 33-36% | 30-32% |
| **IS(B)-100** | 51-57% | 87-93% | 40-47% | 52-59% | 6-13% |
| **Q-Score** | 61-64% | 100% | 43-46% | 54-57% | 0% |
| **S-Score** | 44-50% | 36-44% | 60-67% | 33-40% | 56-64% |

Abbreviations: EPA = U.S. Environmental Protection Agency, EU = European Union, GHS = Globally Harmonized System.
Based on data presented in **Tables 6-2, 6-8,** and **6-14**. A single value indicates the same percentage results for all three hazard classification systems.

Given the relatively homogeneous performance of the HET-CAM test method among the three classification systems, the discussion for the IS(A)-10 and IS(A)-100, IS(B)-10 and IS(B)-100, Q-Score, and S-Score analysis methods encompasses all three hazard classification systems, unless otherwise indicated. Additional information on the mtc and/or the IS and ITC analysis method can be obtained from Spielmann et al. (1996) (**Section 9.0** provides a summary of the report).

### 6.2.1 Discordance Among Chemical Classes

#### 6.2.1.1 *IS(A)-10 and IS(A)-100 Analysis Methods*

Most of the substances evaluated by these analysis methods were formulations. For the IS(A)-10 analysis method, which evaluated mostly surfactant-based formulations, the false negative rates ranged from 75% to 76%, while the false positive rate was 0% for all classification systems. Comparatively, the IS(A)-100 analysis method, which evaluated primarily oil-water formulations, had a higher false positive rate than false negative rate.

#### 6.2.1.2 *IS(B)-10 and IS(B)-100 Analysis Methods*

The chemical class of substances that was consistently overpredicted according the GHS classification system (i.e., were false positives) by both analysis methods is alcohols (89% to 90% for the IS(B)-10 analysis method and 79% to 88% for the IS(B)-100 analysis method).

Additional chemical classes that were overpredicted by both analysis methods were ethers (50% for IS(B)-10 and 47 to 50% for IS(B)-100), organic salts (57% for IS(B)-10 and 75% to 86% for IS(B)-100), and heterocyclic compounds (86% to 88% for IS(B)-10 and 58 to 78% for IS(B)-100). Formulations appeared to have the lowest false positive rates for both analysis methods (0% for IS(B)-10 and 23% to 26% for IS(B)-100). The chemical classes that were underpredicted by both the IS(B)-10 and IS(B)-100 analysis methods were amines. Generally, the false negative and false positive rates for the same chemical class were higher for the IS(B)-100 analysis method when compared to the IS(B)-10 analysis method.

The broad range in the accuracy results from some of the chemical classes evaluated (e.g., heterocyclic compounds evaluated with the IS(B)-100 analysis method) appears to be due to the greater number of substances within this chemical class that were evaluated by the EU classification system and not the GHS or EPA classification systems. As mentioned earlier in this section (see **Section 6.1**), insufficient *in vivo* data was available for some of the substances evaluated, which did not allow for classification according to all three classification systems.

#### 6.2.1.3 *Q-Score Analysis Method*

The accuracy analysis indicated that alcohols and esters are often overpredicted (43 to 50% and 43%) false positive rate, depending on the classification system used) in the Q-score analysis method. The numbers of substances among the remaining chemical classes were too few to resolve any definitive trends in overprediction by the Q-Score analysis method. The false negative rate for all chemical classes with a sufficient number of substances ($n \geq 5$) was 0%.

#### 6.2.1.4 *S-Score Analysis Method*

Due to the limited database for this analysis method, a chemical class evaluation could only be conducted for carboxylic acids/carboxylic acid salts for the GHS classification system. For this chemical class and classification system, the false negative rate was 75% (3/4) and the false positive rate was 0% (0/1).

### 6.2.2 Discordance Among Physical or Chemical Properties of Interest

#### 6.2.2.1 *IS(A)-10 and IS(A)-100 Analysis Methods*

With regard to physical form of the substances tested by these analysis methods, a majority of the substances were tested as liquids/solutions *in vitro* and *in vivo*. Therefore, the false negative and false positive rates for these analysis methods were similar or the same as to the overall false positive and false negative rates. That is the false negative and false positive rates for liquids were 75% to 76% and 0% for the IS(A)-10 analysis method and 0% and 18% for the IS(A)-100 analysis method. No solids were evaluated using the IS(A)-10 analysis method, while the false negative and false positive rates were 0% for the IS(A)-100 analysis method.

For the GHS classification scheme, the evaluation indicated that substances were more likely to be underpredicted if (a) the *in vivo* lesion was based on persistence of effect and (b) if the *in vitro* test concentration was 100%.

6.2.2.2 *IS(B)-10 and IS(B)-100 Analysis Methods*

With regard to physical form of the substances overpredicted by the IS(B)-10 analysis method, the false positive and false negative rates were 19% and 37% to 38%, respectively for liquids and 58% to 65% and 0% to 13% for solids. For the IS(B)-100 analysis method, the false positive and false negative rates were 61% to 65% and 0%, respectively for liquids and 48% to 67% and 8% to 24% for solids. The physical form of many of the tested substances was unknown based on the available information. Therefore, there were numerous tested compounds (36 to 39 substances) for each hazard classification system that were not included in this evaluation.

The broad range in the accuracy results from some of the physical properties (e.g. IS(B)-100 solids) evaluated appears to be due to the greater number of substances within this class that were evaluated by the EU classification system and not the GHS or EPA classification systems. As mentioned earlier in this section (see **Section 6.1**), insufficient *in vivo* data was available for some of the substances evaluated, which did not allow for classification according to all three classification systems.

Information regarding the pH of test substances was available for a subset of the substances tested (29 to 35 substances). Overall, substances were observed to have a higher false positive rate when (a) tested at a 100% concentration and (b) had a pH greater than 7.0.

For the GHS classification scheme, the evaluation indicated that substances were more likely to be underpredicted if (a) the *in vivo* lesion was based on persistence of effect and (b) if the *in vitro* test concentration was 10%.

6.2.2.3 *Q-Score Analysis Method*

With regard to physical form of the substances overpredicted by the Q-Score analysis method, 14 to 17 were liquids and none were solids. The ranges of false positive and false negative rates for liquids were 56% to 61% and 0%, respectively. The false positive and false negative rates for solids were 0% for both parameters. There was insufficient information for the other evaluated categories (e.g., surfactant-based formulations) to conduct an analysis.

6.2.2.4 *S-Score Analysis Method*

With regard to physical form of the substances overpredicted by the S-Score analysis method, 14 to 16 were solids. There were no liquids evaluated with analysis method. The false negative rates for solids ranged from 56% to 64% (5/9 to 7/11) and the false positive rates ranged from 33% to 40% (2/6 to 2/5). There was insufficient information for the other evaluated categories (e.g., surfactant-based formulations) to conduct an analysis.