5.0 BCOP TEST METHOD DATA AND RESULTS

5.1 Description of the BCOP Test Method Protocols Used To Generate Data

As noted in **Section 3.1**, only a subset of the BCOP data obtained for this evaluation was useful for an accuracy analysis. These data were extracted from eight publications, data submissions, or study reports: Gautheron et al. (1994), Balls et al. (1995), Swanson et al. (1995), Gettings et al. (1996), Casterton et al. (1996), Southee (1998), Swanson and Harbell (2000), and Bailey (2004). The scientific methods described in these eight BCOP study reports provided various levels of detail. To the extent possible, information about the test method components discussed in **Section 2.0** was extracted from each publication and summarized in **Appendix A4**, so that any differences among the protocols are evident. Details about the following test method components are included in the appendix to the extent this information was available:

- collection of bovine eyes (e.g., transport conditions, temperature)
- cornea preparation
- pretreatment incubation/equilibration in corneal holder (e.g., duration and temperature)
- treatment groups used (i.e., number of corneas used per test substance)
- controls used
- treatment procedures for corneas
- endpoints assessed
- evaluation of test results
- calculation of *in vitro* score
- *in vitro* classification of ocular irritancy
- criteria for an acceptable test
- compliance with GLP

As is evident in **Appendix A4,** there are differences in various aspects of the test method protocols. These differences are summarized below:

- Four of the studies (Swanson et al. 1995; Gettings et al. 1996; Swanson and Harbell 2000; and Bailey et al. 2004) noted transporting the bovine eyes from the slaughterhouse to the testing facility over ice, as recommended in the proposed standardized protocol. Four other studies noted that isolated bovine eyes were transported at ambient temperature (Gautheron et al. 1994; Balls et al. 1995; Casterton et al. 1996; Southee 1998).
- Only four of the studies (Swanson et al. 1995; Southee 1998; Swanson and Harbell 2000; Bailey et al. 2004) noted transporting the bovine eyes in HBSS containing antibiotics.
- Although all studies reportedly used complete MEM for maintaining the isolated bovine corneas during incubations, only the more recent studies (Swanson et al. 1995; Casterton et al. 1996; Southee 1998; Swanson and Harbell 2000; Bailey et al. 2004) specified using MEM *without* phenol red for incubations.

- All studies used an opacitometer to measure opacity, except for Casterton et al. (1996), which used a UV/VIS spectrophotometer.
- The number of corneas used per test substance in each study varied from three to six per treatment group.
- All of the studies tested solid test substances as a 20% solution or suspension with an incubation period of four hours, with the exception of Casterton et al. (1996). In this study, solids were applied directly to the corneal surface and incubated for one hour.
- Casterton et al. (1996) independently evaluated both opacity and permeability for classifying the potential ocular irritancy of test substances. Gettings et al. (1996) used permeability values only to classify the *in vitro* ocular irritancy of the surfactant-based personal care formulation evaluated, because these materials are known to damage the epithelial layer of the cornea without producing significant opacity. In contrast, the remaining BCOP studies calculated an *in vitro* score equal to the mean opacity value plus 15 times the mean permeability value. This *in vitro* score was used to classify the ocular irritancy of test substances.
- The *in vitro* classification of severe ocular irritants was similar for Gautheron et al. (1994), Balls et al. (1995), and Southee (1998). Gautheron et al. (1994) defined a test substance as a severe irritant if it produced an *in vitro* score of 55.1 or greater. Balls et al. (1995) and Southee (1998) defined a severe irritant as one that produced an *in vitro* score between 55.1 and 80; an *in vitro* score greater than 80 was considered a very severe irritant. In contrast, Casterton et al. (1996) defined a severe irritant as a substance that produced either an opacity value greater than 1.300 or a permeability value greater than 0.600. For the surfactant-based personal care formulations evaluated by Gettings et al. (1996), it was recommended that a severe irritant be defined as a substance that produces a permeability value greater than 0.600 (Harbell J, personal communication), since these materials do not produce appreciable opacity in the isolated bovine cornea, but can damage the epithelium and increase permeability.
- Gautheron et al. (1994) evaluated the use of preserved corneas, in addition to using freshly isolated bovine corneas, in the BCOP assay.

The impact of how differences among test method protocols could impact the data and results is unknown.

5.2 Availability of Copies of Original Data Used to Evaluate the Accuracy and Reliability

NICEATM staff made several attempts to obtain original *in vitro* and *in vivo* data from BCOP test method studies. Two *Federal Register* (*FR)* Notices (Vol. 69, No. 57, pp. 13859-13861, March 24, 2004, and Vol. 70, No. 38, pp. 9661-9662, February 28, 2005; both available at http://iccvam.niehs.nih.gov/methods/eyeirrit.htm), were published requesting original BCOP test method data and *in vivo* reference data. In addition, authors of published BCOP studies were contacted to request original BCOP data and *in vivo* reference data from

the respective publications. As a result of these efforts, some original BCOP test method data (i.e., corrected opacity and OD_{490} values for individual corneas) were obtained. ECVAM provided corrected opacity and OD_{490} values in a written report for 16 substances evaluated in the European Community Prevalidation Study of the BCOP (Southee 1998). Dr. Joseph Sina also submitted corrected opacity and OD_{490} values electronically for 43 compounds; however, corresponding *in vivo* reference data was not obtained. ECVAM subsequently provided the mean opacity values, mean permeability values, and mean *in vitro* scores obtained for the 59 substances evaluated in the Balls et al. (1995) study. Dr. John Harbell submitted between-experiment (intralaboratory) permeability data for the Gettings et al. (1996) study. Dr. Freddy Van Goethem provided a summary table and individual cornea data for 52 compounds tested in the EEC validation study (Gautheron et al., 1994). S.C. Johnson & Son, Inc. provided transformed BCOP data (mean opacity, permeability, and *in vitro* scores) for the Swanson et al. (1995) and Swanson and Harbell (2000) studies, and ExxonMobil Biomedical Sciences, Inc. provided detailed study reports for the Bailey et al. (2004) study.

The majority of other published BCOP reports, which are discussed in **Section 9.0,** did not contain sufficient *in vitro* or *in vivo* data with which to conduct an accuracy analysis.

5.3 Description of the Statistical Approaches Used to Evaluate the Resulting Data

The BCOP studies included in the accuracy analysis in this document (**Section 6.0**) evaluated variability in the BCOP assay by calculating the mean (± SD) for the opacity values and the OD_{490} values for each treatment group and control group. The mean opacity and mean permeability (OD_{490}) values for each treatment group were then used to calculate an *in vitro* score for each treatment group:

In Vitro Irritancy Score = mean opacity value + (15 x mean OD_{490} value)

Sina et al. (1995) reported that this formula was derived empirically during in-house and interlaboratory studies. The data generated for a series of 36 compounds in a multilaboratory study were subjected to a multivariate analysis to determine the equation of best fit between *in vivo* and *in vitro* data. This analysis was performed by scientists at two separate companies, who generated nearly identical derived equations. The *In Vitro* Irritancy Score provides a numerical value that can be used to compare the relative irritancy of test substances.

The accuracy analysis in this document is focused on evaluating the ability of the BCOP test method to identify ocular corrosives and severe irritants as defined by the EPA (1996), EU (2001), and the GHS (UN 2003). A review of the BCOP test method protocols indicates that the decision criteria applied to *in vitro* data to classify a test substance as a severe ocular irritant or a nonsevere ocular irritant (i.e., mild irritant, moderate irritant) and/or nonirritant are similar for four BCOP protocols (Gautheron et al. 1994; Balls et al. 1995; Southee 1998; Bailey et al. 2004). The *in vitro* irritation classification scheme used in these studies is similar to the prediction model first proposed by Gautheron et al. (1994), for which *in vitro* irritancy categories were based on predetermined ranges of *in vitro* scores:

***In Vitro* Score Range**	***In Vitro* Classification**
0 - 25	Mild irritant
25.1 - 55	Moderate irritant
≥ 55.1	Severe irritant

This original classification system was based on studies with pharmaceutical intermediates in which bovine corneas were exposed for 10 minutes (liquids) or four hours (solids). The correlation of these categories to accepted regulatory categories for ocular irritation (i.e., GHS, EPA, EU) is unknown.

This same prediction model was used for the EC/HO validation study (Balls et al. 1995), with the exception that the investigators added a fourth classification of "very severe" for substances that produced an *in vitro* score greater than 80.1.

For the European Community prevalidation study, the investigators attempted to relate the prediction model to *in vivo* data (MMAS scores) (Southee 1998):

Draize Scale	**Draize Classification**	***In Vitro* Scale**	***In Vitro* Classification**
0 - 0.9	Minimal	0 - 3	Nonirritant
1 - 25	Minimal/slight	3.1 - 25	Mild irritant
26 - 56	Moderate	25.1 - 55	Moderate irritant
57 - 84	Marked	55.1 - 80	Severe irritant
85 - 110	Extreme	> 80.1	Very severe irritant

Gettings et al. (1996) did not report a classification scheme to assign irritancy potential to tested substances based on *in vitro* scores.

Casterton et al. (1996) assigned irritation classes based on the endpoint (opacity or permeability) with the highest score for its respective range:

***In Vitro* Opacity or Permeability Ranges**	***In Vitro* Classification**
Opacity < 0.400 or Permeability < 0.175	Mild irritant
0.400 ≤ Opacity < 1.300 or 0.175 ≤ Permeability < 0.600	Moderate irritant
Opacity > 1.300 or Permeability ≥ 0.600	Severe irritant

The rationale for the cutoffs used in this classification scheme was not provided and the correlation of these categories to accepted regulatory categories is unknown.
As described above, the surfactant-based personal care formulations evaluated by Gettings et al. (1996) do not produce appreciable opacity in the isolated bovine cornea, but can increase permeability by damaging the epithelium. Thus, it was recommended that a severe irritant be defined as a substance that produces a permeability value ≥ 0.600 (Harbell J, personal communication). Also, some companies, such as S.C. Johnson & Son, Inc., note that they do not use any of the classification systems described above to assign an ocular irritancy classification, but instead compare BCOP data for newly tested substances to benchmark materials, relying on a system of comparative toxicity instead of cutoff scores (Cuellar N and Swanson J, personal communication).

5.4 Summary of Results

BCOP data were collected for a total of 161 test substances among the eight studies evaluated. A summary of results used to evaluate test method accuracy is shown in **Appendix D**. **Appendix D1** provides a table, sorted first by reference then alphabetically by substance, with the name of the substance tested, the CASRN, the concentration tested, the available BCOP data (e.g., mean opacity value, mean OD_{490} value, standard deviation, number of replicates, mean *in vitro* score), the *in vitro* irritation classification of the test substance (based on the *in vitro* irritation classification scheme applied or noted by the study author), and the reference. **Appendix D2** provides the same information, but is sorted alphabetically by test substance to indicate which substances were tested in multiple studies. Other supporting information, such as the source, purity and physicochemical characteristics of the test substances, was included in the tables to the extent this information was available. No attempt was made to identify the source, purity, and physicochemical characteristics of a test substance, if the authors did not provide such information. If not provided, the CASRN was obtained from various sources, including the National Library of Medicine's ChemID database. Chemical and product classes were assigned based on the MeSH classification system (available at http://www.nlm.nih.gov/mesh). Each of the eight studies evaluated varied with respect to the level of detail of data that was provided, as described below.

5.4.1 Gautheron et al. (1994)

In this interlaboratory evaluation of the BCOP test method, BCOP data were extracted for 52 test substances, which were evaluated in 11 or 12 laboratories. Four of these laboratories (numbers 2, 3, 8, and 10) used a modified protocol, in which preserved corneas were used in place of freshly collected corneas. Laboratory 10 completed studies on only 23 of the test substances. NICEATM classified each test substance based on the *in vitro* classification system described by the authors of the study.

The *in vivo* reference data were provided by Dr. Philippe Vanparys, allowing for an accuracy analysis of up to 50 substances in relation to the *in vivo* ocular irritancy classifications assigned by NICEATM for the substances according to the EPA (EPA 1996), EU (EU 2001), and GHS (UN 2003) classification systems. Not all of the 52 substances tested could be evaluated because some ($n \leq 3$) of the *in vivo* studies did not provide sufficient data to assign an ocular irritancy classification for each classification scheme. However, because the 52

test substances were tested *in vitro* using a standardized protocol in eight laboratories, an interlaboratory reliability analysis also could be conducted for this study.

5.4.2 Balls et al. (1995)

In this evaluation of the BCOP test method, 51 chemicals were evaluated in five laboratories. Four of these chemicals were tested at two different concentrations and two were tested at three concentrations, for a total of 59 different test substances. BCOP test method data on the 59 tests were not included in the published report. Rather, the study report included scatter plots showing the relationship between the BCOP data (i.e., mean opacity value, mean permeability value, and mean *in vitro* score) obtained in the lead laboratory and the MMAS for the entire set of test substances. However, the mean opacity value, the mean permeability (OD_{490}) value, and the mean *in vitro* score obtained for each substance in each laboratory were provided by ECVAM for all 59 test substances. Detailed *in vivo* data are available for the 59 test substances (including the different concentrations tested) in ECETOC (1998), allowing for an accuracy analysis of the 59 substances in relation to the *in vivo* ocular irritancy classifications assigned by NICEATM for the substances according to the EPA (1996), GHS (UN 2003) and EU (2001) classification systems. Because the 59 test substances were tested using a standardized protocol in five laboratories, an interlaboratory reliability analysis could be conducted for this study. Although the *in vitro* classification for each test substance was not provided in the study report, NICEATM used the *in vitro* classification system noted by the authors of the study to classify each test substance.

5.4.3 Swanson et al. (1995)

In this study of 20 full-strength cleaners and floor strippers, *in vitro* data were extracted for 13 formulations with sufficient *in vivo* reference data to allow for an accuracy analysis. The mean opacity value, the mean permeability (OD_{490}) value, and the mean *in vitro* score obtained for each substance (in one laboratory) were provided by S.C. Johnson & Son, Inc. Although the *in vitro* classification for each test substance was not provided in the study report, NICEATM used the *in vitro* classification system noted by Gautheron et al. (1994) to classify each test substance.

5.4.4 Gettings et al. (1996)

In the CTFA Evaluation of Alternatives Program – Phase III, 25 surfactant-based personal care cleansing formulations were evaluated in one laboratory. The mean permeability (OD_{490}) and mean *in vitro* score were provided for each formulation in the study report. Although the *in vitro* classification for each test substance was not provided in the study report, NICEATM assigned a classification to each test substance based on the mean permeability value obtained for each substance. A substance that produced a permeability value ≥ 0.600 was classified as a severe ocular irritant.

5.4.5 Casterton et al. (1996)

For this study, *in vitro* data were extracted for 15 personal care product formulations, 13 household cleaning product formulations, and 32 chemicals with available *in vivo* reference data. The mean opacity value and the mean permeability value were provided in the study report, as well as the laboratory specific *in vitro* classification for each test substance. These data were obtained from one laboratory.

Although BCOP data were reported in this publication for an additional 37 chemicals and consumer product formulations, detailed *in vivo* reference data were not available for these substances, precluding an accuracy analysis for this set of substances. Therefore, the BCOP data for these 37 substances are not included in this document.

5.4.6 Southee (1998)

In this study, 16 test substances were evaluated in three laboratories. The mean opacity value, mean permeability value (OD_{490}), number of replicates, mean *in vitro* score, SD for all mean values, and *in vitro* classification were provided for each test substance. Each laboratory tested each substance on at least two separate occasions. Imidazole, ethanol, and benzalkonium chloride were each tested in at least seven different experiments by each laboratory.

5.4.7 Swanson and Harbell (2000)

In this study of 13 ethanol containing insect repellent formulations, *in vitro* data were extracted for ethanol and eight formulations with sufficient *in vivo* reference data to allow for an accuracy analysis. The mean opacity value, the mean permeability (OD_{490}) value, and the mean *in vitro* score obtained for each substance (in one laboratory) were provided by S.C. Johnson & Son, Inc. Although the *in vitro* classification for each test substance was not provided in the study report, NICEATM used the *in vitro* classification system noted by Gautheron et al. (1994) to classify each test substance.

5.4.8 Bailey et al. (2004)

In this study of 16 petrochemical products, *in vitro* data were extracted for all of the test substances, which had sufficient *in vivo* reference data to allow for an accuracy analysis. The mean opacity value, the mean permeability (OD_{490}) value, and the mean *in vitro* score obtained for each substance (in one laboratory) were provided by ExxonMobil Biomedical Sciences, Inc. Although the *in vitro* classification for each test substance was not provided in the study report, NICEATM used the *in vitro* classification system noted by Gautheron et al. (1994) to classify each test substance.

5.5 Use of Coded Chemicals and Compliance with GLP Guidelines

Ideally, all data supporting the validity of a test method should be obtained and reported in accordance with GLP guidelines and with the use of coded chemicals (OECD 1998; EPA 2003a, 2003b; FDA 2003). The data quality was evaluated by a review of the methods section in literature references and the submitted reports. The data quality presented in the reviewed literature references can be evaluated to the extent this information was provided in the published reports. Based on the available information, the reports that identified following GLP guidelines or used data obtained according to GLP guidelines were Balls et al. (1995), Swanson et al. (1995), Gettings et al. (1996), Swanson and Harbell (2000), and Bailey et al. (2004). Likewise, the reports that identified using coded chemicals were Gautheron et al. (1994), Balls et al. (1995), Swanson et al. (1995), Gettings et al. (1996), Southee (1998), Swanson and Harbell (2000), and Bailey et al. (2004).

5.6 Lot-to-lot Consistency of Test Substances

Ideally, the lot-to-lot consistency of test substances is evaluated to ensure that the same substance, with the same physicochemical properties, is being evaluated over the duration of the study. A description of the procedures used to evaluate the lot-to-lot consistency was provided in the published reports. No attempt was made to review original records to assess the procedures used to evaluate different batches of tested substances.

Gettings et al. (1996) noted that all samples were dispensed from a single source to ensure test substance consistency. The samples were placed into a secondary container, labeled with appropriate chemical code information and then forwarded to the participating testing laboratories. There is no information about the time frame in which the studies were conducted or whether additional aliquots of the samples were forwarded to specific testing laboratories.

Balls et al. (1995) noted that substances with the same source and specification as those tested *in vivo* were obtained, whenever possible, to test *in vitro*. When such a situation was not possible, a specification as close as possible to what was evaluated *in vivo* was selected. Aliquots of each test substance were prepared at one time and forwarded to the participating testing laboratories. There is no information about the time frame in which the studies were conducted or whether additional aliquots of the samples were forwarded to specific testing laboratories.

There was no information about maintaining lot-to-lot consistency in any of the other reports reviewed.

5.7 Availability of Data for External Audit

Study notebooks and other supporting records are known to be available, upon request, for an external audit, for the following studies: Swanson et al. (1995), Gettings et al. (1996), Swanson and Harbell (2000), and Bailey et al. (2004). The availability of data for an external audit for the other reports described in this section has not been determined.