5.0 ICE TEST METHOD DATA AND RESULTS

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

A total of five reports, three published (Prinsen and Koëter 1993; Balls et al. 1995; Prinsen 1996) and two unpublished (Prinsen 2000; Prinsen 2005) contained sufficient data to do an accuracy analysis of the ICE test method. The test method components of the ICE protocols used in these studies (discussed in **Section 2.0**) are summarized in **Appendix A**. As discussed in **Sections 1.0** and **2.0**, only one modification to the original ICE test method protocol (Prinsen and Koëter, 1993) was made; i.e., the number of chicken eyes evaluated was reduced from five to three per test substance. Reportedly, the reduction has no effect on the overall accuracy of the ICE test method (Prinsen M, personal communication). However, a formal evaluation of the effect of the number of eyes per test substance on the accuracy or the reliability of the ICE test method has not been done. Historically, positive controls have not been used in the ICE test method, and therefore do not appear in any of the previously published protocols. The only negative control used to date has been isotonic saline – an untreated negative control. **Section 2.2.7.2** describes the need for a solvent control when a test substance is dissolved in a solvent other than water or isotonic saline.

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

The NICEATM staff made several attempts to obtain original ICE data for substances that had also been tested *in vivo* using the standard rabbit eye test. *Federal Register* (*FR*) notices were published on March 24, 2004 (Vol. 69. No. 57, pp. 13589-12861; available at http://iccvam.niehs.nih.gov/methods/ eyeirrit.htm) and February 28, 2005 (Vol. 70, No. 38, pp. 9661-9662; http://iccvam.niehs.nih.gov/methods/eyeirrit.htm), requesting original ICE data, comparative *in vivo* rabbit data, as well as any human exposure data (either via ethical human studies or accidental exposure). In addition, the NICEATM staff contacted authors of published ICE studies to request original ICE data used to support the authors' conclusions. In response to these efforts, summaries of ICE results (i.e., total scores) but not original data were obtained for the 60 substances evaluated by Balls et al. (1995). NICEATM also received original study records, containing data for the substances screened with the ICE test method in Prinsen (1996), Prinsen (2000), and Prinsen (2005), kindly provided by Mr. Menk Prinsen of TNO (TNO Nutrition and Food Research, Toxicology and Applied Pharmacology, Zeist, The Netherlands).

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

As noted in **Section 2.2.12**, statistical analyses to compare ICE test method results to those from the *in vivo* reference test method have been done predominantly by comparing the ICE Irritation Index and the maximum mean scores of its individual components (i.e., corneal swelling, corneal opacity, fluorescein retention) to a numerical *in vivo* rabbit eye score (e.g., MMAS). However, because this BRD is concerned with the regulatory applicability of the ICE test method and MMAS scores are not used for regulatory classification, this approach was not taken in the analyses done for this BRD. Rather, the *in vitro* classification system

described in **Section 2.2.13** was used to assign an *in vitro* ocular irritation classification for each test substance. This approach entails calculation of mean corneal opacity, corneal swelling, and fluorescein retention scores at each time point for each test substance (see **Section 2.2.9**) and relating the maximum scores for each endpoint to an *in vitro* irritancy category. Interpretation of corneal thickness, corneal opacity, and fluorescein retention using four irritancy categories is done according to the scales shown below, provided by endpoint.

Corneal Thickness

Mean Corneal Swelling (%)	Category
0 to 5	I
> 5 to 12	II
> 12 to 18 (>75 min after treatment)	II
> 12 to 18 (≤75 min after treatment)	III
> 18 to 26	III
> 26 to 32 (>75 min after treatment)	III
> 26 to 32 (≤75 min after treatment)	IV
> 32	IV

Corneal Opacity

Mean Maximum Opacity Score	Category
0.0-0.5	I
0.6-1.5	II
1.6-2.5	III
2.6-4.0	IV

Fluorescein Retention

Mean Fluorescein Retention Score at 30 Minutes Post-treatment	Category
0.0-0.5	I
0.6-1.5	II
1.6-2.5	III
2.6-3.0	IV

The categories for each individual ICE test method endpoint can then be combined into an overall *in vitro* ocular irritancy classification for comparison to the *in vivo* ocular irritancy classification. For assigning the classification of severe irritant to a test substance, the combinations of results shown below was used (Prinsen and Koëter 1993).

Classification	**Combinations of the 3 Endpoints**
Severely Irritating	3 x IV
	2 x IV, 1 x III
	2 x IV, 1 x II[1]
	2 x IV, 1 x I[1]
	Corneal opacity ≥ 3 at 30 min (≥ 2 eyes)
	Corneal opacity = 4 at any time point (≥ 2 eyes)
	Severe loosening of the epithelium (≥ 1 eye)

[1]Combinations less likely to occur.

To date, this method has been published only as an application to the EU classification system. However, using the same classification system, ICE results have also reportedly been used to predict the *in vivo* classification of substances according to the GHS classification system (Prinsen M, personal communication). For this BRD, the *in vitro* classification was compared to the *in vivo* classification based on the EU, GHS, and EPA classification systems (EPA 1996; EU 2001; UN 2003), when feasible; i.e., when adequate *in vivo* data were available to assign a classification. To conduct this analysis, no modifications to the *in vitro* classification system were made.

Four of the five studies considered for this BRD (Prinsen and Koëter 1993; Prinsen 1996; Prinsen 2000; Prinsen 2005) assigned the *in vitro* classification of test substances based on this system. However, because one study (Balls et al. 1995) did not use this approach, the data generated in this study was used to assign an *in vitro* classification (as directed in **Section 2.2.13.1**). Once the *in vitro* classification was established for the substances tested in all relevant studies, an accuracy assessment was done for each parameter investigated (i.e., ICE classification versus the *in vivo* classification according to the rules applied by each regulatory agency, if adequate *in vivo* data were available to assign each classification).

5.4 Summary of Results

When provided, the specific information extracted for each substance included its name, CASRN (if available), chemical class, product class, concentration tested, form tested, ICE test method endpoint values (maximum mean), *in vitro* classification, and reference. No attempt was made to identify the source and purity 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 (available at http://chem2.sis.nlm.nih.gov/chemidplus). All substances with the same CASRN were listed under the same name, regardless of the synonym used in the original report. Chemical and product classes were assigned based on the classification of the National Library of Medicine's Medical Subject Heading (MeSH; available at http://www.nlm.nih.gov/mesh). **Appendix B** provides information on the names, synonyms, CASRN, and chemical/product

class, where available, for each substance while **Appendix C** contains the *in vitro* ICE test method data sorted by reference and alphabetically by substance name. The type of data contained in each study evaluated for this BRD varied, as discussed in Sections 5.4.1 to 5.4.5.

5.4.1 Prinsen and Koëter (1993)

The mean percentage corneal swelling at each time point, mean corneal opacity at each time point, and mean fluorescein retention at 30 minutes were generated for all 21 test substances. However, individual scores for each eye were not provided. No i*n vivo* scores were provided, but an irritation classification (according to the EU classification system [EU 2001]) was provided for all test substances.

5.4.2 Balls et al. (1995)

Neither the scores for each ICE test method endpoint nor the Irritation Index are included in the published report. Rather, the study report includes scatter plots showing the relationship between mean corneal swelling, mean opacity score, mean fluorescein retention score, and ICE Index score, as obtained in the lead laboratory, to the MMAS for the entire set of test substances. However, the maximum mean percentage corneal swelling and corneal opacity and the mean fluorescein retention at 30 minutes, along with the Irritation Index, was provided for all 59 test substances following a request to the European Centre for the Validation of Alternative Methods (ECVAM) by NICEATM.

5.4.3 Prinsen (1996)

Forty-four test substances were assayed in the ICE test method. Thirty-nine of the 44 test substances were evaluated in both the ICE test method and the *in vivo* rabbit eye test method. Five of the test substance were labeled as corrosive to skin and thus were not evaluated in the rabbit eye test, but rather presumed to be severely irritating to the eye (i.e., EU classification of R41 [EU 2001]). Seven substances were evaluated that had an *in vivo* classification of R41. For the *in vitro* test method, the mean percentage corneal swelling at each time point, mean corneal opacity at each time point, and mean fluorescein retention at 30 minutes were provided on all test substances, although individual eye scores were not. However, Mr. Menk Prinsen (TNO) subsequently provided this information.

5.4.4 Prinsen (2000)

This report contained ICE test method data for four substances. For the *in vitro* test method, individual eye scores for corneal thickness and corneal opacity were provided for each time point, and mean fluorescein retention at 30 minutes was provided for all test substances. The EU classification for each substance was provided, but the corresponding *in vivo* rabbit eye test data were not.

5.4.5 Prinsen (2005)

This report contained ICE test method data for 50 substances. For the *in vitro* test method, individual eye scores for corneal thickness and corneal opacity were provided for each time point, and mean fluorescein retention at 30 minutes was provided for all test substances. Corresponding *in vivo* data were also provided for each test substance, although, in some

cases, this data was inadequate to assign an irritancy classification in a particular classification system.

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 only be evaluated to the extent such information was provided in the published reports. Based on the available information, all ICE test method studies evaluated were conducted according to GLP guidelines.

Based on the information in the five studies evaluated, Balls et al. (1995) was the only study that employed specific mechanisms to code the chemicals that were tested (See **Section 3.4.2**).

5.6 Lot-to-lot Consistency of Test Substances

Ideally, a single lot of each substance is used during the validation of a test method. In situations where multiple lots of a chemical must be used, the lot-to-lot consistency of a test substance must be evaluated to ensure the consistency of the substance evaluated over the course of the study. A description of the procedures used to evaluate lot-to-lot consistency was provided in the reports. No attempt was made to review original records to assess the procedures used to evaluate different batches of substances.

One selection criterion for reference chemicals selected for the ECETOC evaluation was known high consistency and purity. Test substances for the Balls et al. (1995) evaluation were selected from the ECETOC database, and where feasible, the same source and specification was used. If obtaining the test substance from the same source and/or specification was not feasible, a test substance with a specification as close to that included in the *in vivo* testing was selected.

Based on the limited chemical information provided in the remaining reports (Prinsen and Koëter 1993; Prinsen 1996; Prinsen 2000; Prinsen 2005), and the absence of specifically cited selection criteria in these studies, an accurate assessment of lot-to-lot consistency of the test substances evaluated was not feasible. Prinsen (1996) and Prinsen (2005) appear to have used the same batch of test substances in both the ICE and *in vivo* test methods, thus ensuring an optimum level of consistency for both test methods used in these studies.

5.7 Availability of Data for External Audit

The availability of the original study records, for the reports included in the accuracy and reliability evaluation of the ICE test method, for external audit was not determined.

[This Page Intentionally Left Blank]