

4.0 ***IN VIVO* REFERENCE DATA USED FOR AN ASSESSMENT OF TEST METHOD ACCURACY**

4.1 **Description of Protocol Used to Generate *In Vivo* Data**

4.1.1 Draize Rabbit Eye Test

The test method protocol most widely accepted by regulatory agencies for the evaluation of ocular eye irritants is based on the Draize rabbit eye test method. The methodology, originally described by Draize et al. (1944), involves instillation of 0.1 mL of the test substance (e.g., liquids, solutions, and ointments) into the conjunctival sac of an albino rabbit eye. In this test method, one eye is treated while the other eye serves as the untreated control. The eye is examined at selected time intervals after exposure and any injuries to the cornea, conjunctiva, and the iris are scored. Scoring is subjective and based on a discrete, arbitrary scale (**Table 4-1**) for grading the severity of ocular lesions. The scores for the observed ocular injuries range from 1 to 2 for iris effects, from 1 to 3 for conjunctival redness and discharge, and from 1 to 4 for corneal effects and conjunctival chemosis. A score of zero is assigned when the eye is normal and no adverse effects are observed. In the original protocol, the eyes were observed up to four days after application of the test substance. However in current practice, these time points vary according to the degree of irritation, the clearing time, and testing requirements imposed by the various regulatory agencies.

The original Draize protocol describes a scoring system in which each ocular parameter is graded on a continuous numerical scale. The scores may be weighted (as shown in **Table 4-1**); however, most classification systems today do not use a weighting factor. The weighting of the score by Draize et al. (1944) is biased more heavily for corneal injury, since injury to the cornea has the greatest probability of producing irreparable eye damage. To illustrate, each ocular parameter shown in **Table 4-1** is evaluated for each rabbit. The product of the opacity and area scores is obtained, then multiplied by a weighting factor of 5; the maximum corneal score is 80. The iris score is multiplied by a weighting factor of 5; the maximum score is 10. The scores for the three conjunctival parameters are added together and then the total is multiplied by a weighting factor of 2; the maximum score is 20. The overall score for each rabbit is calculated by adding the values for each parameter; the maximum total score is 110.

While the current test method is widely used, it has limitations. For example, because of reflexive pawing at the eye or tearing after instillation of a test substance, the exact dose and/or concentration of the test substance is unknown. Additionally, if observations are made at 24-hour intervals, it may not always be clear whether observed effects are associated with the test substance or an unobserved reflexive behavior.

Table 4-1 Scale of Weighted Scores for Grading the Severity of Ocular Lesions¹

Lesion	Score ²
Cornea	
A. Opacity – Degree of density (area which is most dense is taken for reading)	
Scattered or diffuse area – details of iris clearly visible	1
Easily discernible translucent areas, details of iris slightly obscured	2
Opalescent areas, no details of iris visible, size of pupil barely discernible	3
Opaque, iris invisible	4
B. Area of cornea involved	
One quarter (or less), but not zero	1
Greater than one quarter, but less than one-half	2
Greater than one-half, but less than three quarters	3
Greater than three quarters up to whole area	4
Score equals A x B x 5 Total maximum = 80	
Iris	
A. Values	
Folds above normal, congestion, swelling, circumcorneal injection (any one or all of these or combination of any thereof), iris still reacting to light (sluggish reaction is positive)	1
No reaction to light, hemorrhage; gross destruction (any one or all of these)	2
Score equals A x 5 Total possible maximum = 10	
Conjunctiva	
A. Redness (refers to palpebral conjunctiva only)	
Vessels definitely injected above normal	1
More diffuse, deeper crimson red, individual vessels not easily discernible	2
Diffuse beefy red	3
B. Chemosis	
Any swelling above normal (includes nictitating membrane)	1
Obvious swelling with partial eversion of the lids	2
Swelling with lids about half closed	3
Swelling with lids about half closed to completely closed	4
C. Discharge	
Any amount different from normal (does not include small amount observed in inner canthus of normal rabbits)	1
Discharge with moistening of the lids and hairs just adjacent to the lids	2
Discharge with moistening of the lids and considerable area around the eye	3
Score equals (A + B + C) x 2 Total maximum = 20	

¹From Draize et al. (1944)²Scores of 0 are assigned for each parameter if the cornea, iris, or conjunctiva are normal.

4.1.2 Current *In Vivo* Ocular Irritation Test Method Protocols

Since the original description of the *in vivo* rabbit eye test method, regulatory agencies in the U.S., as well as in other countries, have modified the test method protocol to suit their specific needs and goals in protecting human health (**Table 4-2**). Regulatory agencies generally recommend using healthy adult albino rabbits (e.g., White New Zealand). The eyes of each test rabbit are examined within 24 hours prior to test initiation. A quantity of 0.1 mL (for liquids) or 0.1 g (for pulverized solid, granular, or particulate test substances) is placed into the conjunctival sac of one eye of each rabbit, after pulling the lower lid away from the eyeball. The other eye remains untreated. The lids are held together for about one second to decrease loss of test substance from the eye. Although the observation period

varies, the eyes are typically examined at 24-hour intervals for at least 72 hours after application of the test substance for adverse effects to the cornea, conjunctiva, and iris. The length of the observation period should be sufficient to evaluate reversibility of any of the observed effects, but generally does not exceed 21 days. The ocular effects observed were usually those described by Draize et al. (1944) in **Table 4-1**. For current uses, other lesions, such as pannus¹ and herniation of the cornea, also are noted. Corneal, iris, and conjunctival lesions are scored using the individual numerical grades described in **Table 4-1**, but weighted scores and an overall score for irritation are not typically calculated or used for U.S. or European regulatory purposes.

Depending on the regulatory agency, the number of rabbits required for a study of ocular irritation can vary. To minimize pain and suffering of rabbits exposed to potentially corrosive agents, the EPA and European regulatory agencies suggest that, if a test substance is anticipated to produce a severe effect (e.g., corrosive effect), a test in a single rabbit may be conducted. If a severe effect is observed in this rabbit, further testing does not need to be conducted and classification and labeling of a test substance can proceed on the effects observed in a single rabbit. In cases where more than one rabbit is tested, at least three should be examined to classify the ocular effects produced by the test substance (EU 2004; EPA 1998). In contrast, regulations for other U.S. agencies (e.g., CPSC, FDA) require at least six rabbits be examined to classify the effects produced by a test substance (CPSC 2003). The differences in current *in vivo* test protocols in the U.S. appear to reflect each agency's objectives for eye irritation testing; EPA regulates industrial chemicals, while the CPSC and FDA regulate household consumer products, pharmaceuticals, cosmetics, and toiletries.

Various data transformations have been developed to compare and rate irritants of varying severity. One is the MAS, in which the Draize scores obtained at each time point are averaged and the highest score obtained is the MAS. The MAS value was later modified to the MMAS (Modified Maximum Average Score), which is the highest average MAS value beginning with the 24-hour time point (ECETOC 1998).

4.1.3 Current *In Vivo* Ocular Irritancy Classification Systems

Although *in vivo* eye irritation test method protocols are similar across U.S. and international regulatory agencies, interpretation of the results from the *in vivo* test method varies considerably. Several classification systems are in use for regulatory ocular irritancy testing purposes (**Table 1-2**). In the United States, two major classification systems are currently used, the FHSA guideline (CPSC 1995), which is used by the FDA, OSHA, and CPSC, and the EPA guideline (EPA 1996).

¹ Pannus, also known as “chronic superficial keratitis”, describes a specific type of corneal inflammation. Pannus is caused by a local inflammatory response that begins within the conjunctiva, and with time spreads to the cornea. On a cellular level, the inflammation is composed of brown melanin pigment, red blood vessels, and pink scar tissue.

Table 4-2 Test Guidelines for *In Vivo* Ocular Irritation Test Methods

Test Method Component	Reference				
	Draize et al. (1944)	OECD TG 405 (April 2002)	FHSA Method 16CFR 1500.42 CPSC, FDA, OSHA (CPSC 2003)	FIFRA/TSCA Method EPA TG OPPTS 870.2400 (EPA 1998)	European Union Annex V B.5 (formerly EEC; EU 2004)
Evaluate existing animal and human eye data	NA	Yes	Yes ¹	NS	Yes
Results from dermal irritation study	NA	Yes	Yes ¹	Yes	Yes
Perform SAR for eye irritation	NA	Yes	Yes ¹	NS	Yes
Screen for pH	NA	Yes	Yes ¹	Yes	Yes
Results from validated alternative ocular methods	NA	Yes	Yes ¹	Yes	Yes
<i>Rabbit model/Number of rabbits</i>					
Rabbit species and strain	Albino rabbit	Healthy young adult albino rabbits.	New Zealand White rabbit	Healthy adult albino rabbits recommended. Other mammalian species may be substituted with justification.	Healthy young adult albino rabbits.
Sex and weight	NS	NS	Sex NS; 2.0-3.0 kg	NS	NS
Screen for severe effects	NS	1 rabbit – further testing not required if substance produces corrosive or severe effects.	NS	1 rabbit – further testing not required if substance produces corrosive or severe effects.	1 rabbit – further testing not required if substance produces corrosive or severe effects.
Main test/confirmatory test	NS	Up to 2 additional rabbits, tested sequentially. if irreversible effects are suspected. Test discontinued, if severe effects occur in 2 nd rabbit. Additional rabbits may be needed to confirm weak or moderate responses.	A minimum of 6 rabbits, and up to 18 rabbits for confirmatory tests.	≥ 3 rabbits	Up to 2 additional rabbits, tested sequentially, if irreversible effects are suspected. Test discontinued if severe effects occur in 2 nd rabbit.

Test Method Component	Reference				
	Draize et al. (1944)	OECD TG 405 (April 2002)	FHSA Method 16CFR 1500.42 CPSC, FDA, OSHA (CPSC 2003)	FIFRA/TSCA Method EPA TG OPPTS 870.2400 (EPA 1998)	European Union Annex V B.5 (formerly EEC; EU 2004)
Test substance (amount and method of application)					
Liquids	0.1 mL	0.1 mL	0.1 mL	0.1 mL	0.1 mL
Solids, pastes, particulates	NS	0.1 mL, or ≤ 100 mg	0.1 mL, or ≤ 100 mg	0.1 mL, or ≤ 100 mg	0.1 mL or 100 mg
Aerosols	NS	Single burst of about 1 second sprayed at 10 cm.	NS	Single burst of about 1 second sprayed at 10 cm.	Single burst of about 1 second sprayed at 10 cm.
Pump sprays	NS		NS	0.1 mL	Should not be used for instilling substances directly into eye.
Application of test substance	Test substance is placed in the conjunctival sac.	Test substance is placed in the conjunctival sac of one eye. Lids are gently held together for about 1 second.	Test substance is placed in the conjunctival sac of one eye.	Test substance is placed in the conjunctival sac of one eye. Lids are gently held together for about 1 second.	Test substance is placed in the conjunctival sac of one eye. Lids are gently held together for about 1 second.
Use of anesthetics prior to instillation of test substance	NS	Local anesthetic may be used, if the test substance is anticipated to cause pain.	Local anesthetic may be used prior to instillation of test substance.	Local anesthetic may be used, if the test substance is anticipated to cause pain.	Anesthetic may be used after 24 hours if it does not influence response of the eye to irritants.
Observation					
Observation Period	At least 48 hours. Extended if irritation persists.	At least 72 hours, except when rabbit shows severe pain or distress, or early severe/corrosive effects, upon which the rabbit is humanely killed. Otherwise, sufficient to evaluate reversibility or irreversibility within 21 days.	At least 72 hours. Extended if necessary.	At least 72 hours, but not more than 21 days. Should be sufficient enough to evaluate the reversibility or irreversibility of effects within a 21-day period.	At least 72 hours, except when rabbit shows severe pain or distress, or early severe/corrosive effects, upon which the rabbit is humanely killed. Can be extended up to 21 days if effects persist.

Test Method Component	Reference				
	Draize et al. (1944)	OECD TG 405 (April 2002)	FHSA Method 16CFR 1500.42 CPSC, FDA, OSHA (CPSC 2003)	FIFRA/TSCA Method EPA TG OPPTS 870.2400 (EPA 1998)	European Union Annex V B.5 (formerly EEC; EU 2004)
Examination times after treatment	1, 24, 48 hours, and 4, 7 days.	1, 24, 48, 72 hours, 7, 14, 21 days.	24, 48, 72 hours, and 7 days.	1, 24, 48, and 72 hours. Extended up to 21 days to assess reversibility.	1, 24, 48, and 72 hours. Can be extended up to 21 days. Observations of mild to moderate lesions until they clear or for 21 days. Observations at 7,14 and 21 days to determine reversibility.
Observation aids	NS	Binocular loupe, hand slit-lamp, biomicroscope or other suitable devices can be used. Fluorescein may be used after 24 hours.	Binocular loupe, hand slit-lamp, biomicroscope or other suitable devices can be used. Fluorescein may be used after 24 hours.	Binocular loupe, hand slit-lamp, biomicroscope or other suitable devices can be used. Fluorescein may be used after 24 hours.	Binocular loupe, hand slit-lamp, biomicroscope or other suitable devices can be used. Fluorescein may be used after 24 hours.
Irrigation					
Washout	NS	Generally, eyes may not be washed until after 24 hours post-treatment, except for solids, which may be removed with saline or water after 1 hour.	After 24 hours post-treatment, eyes may be washed with a sodium chloride solution.	After 24 hours post-treatment, eyes may be washed with water to show whether washing palliates or exacerbates irritation.	Generally, eyes may not be washed until after 24 hours post-treatment, except for solids, which may be removed with saline or water after 1 hour.

Test Method Component	Reference				
	Draize et al. (1944)	OECD TG 405 (April 2002)	FHSA Method 16CFR 1500.42 CPSC, FDA, OSHA (CPSC 2003)	FIFRA/TSCA Method EPA TG OPPTS 870.2400 (EPA 1998)	European Union Annex V B.5 (formerly EEC; EU 2004)
Additional testing to determine effects of timely irrigation	NS	Not recommended unless scientifically justified.	NS	Indicated when substances are shown to be irritating. At 30 seconds after exposure, the eyes are washed with water for 30 seconds	Possibility of washing out in case of immediate corrosive or irritating effects. Use of satellite group to investigate influence of washing is not recommended unless scientifically justified.

Abbreviations: CPSC = U.S. Consumer Product Safety Commission; EEC = European Economic Commission; EPA = U.S. Environmental Protection Agency; FDA = U.S. Food and Drug Administration; FIFRA = Federal Insecticide, Fungicide, and Rodenticide Act; NA = Not applicable; NS = Not specified; OECD = Organization for Economic Cooperation and Development; OPPTS = Office of Prevention, Pesticide, and Toxic Substances; OSHA = U.S. Occupational Safety and Health Administration; SAR = Structure activity relationships; TG = Test guideline; TSCA = Toxic Substances Control Act.

¹ Use of this information is not provided in the regulations cited, but in the CPSC Animal Testing Policy guideline (CPSC 1984) states that prior human experience, literature sources which record prior animal testing or limited human tests, and expert opinion may be used in making appropriate hazard determinations.

The FHSA guideline states that a test substance is considered an eye irritant if four or more of six rabbits have positive ocular scores in nonirrigated eyes within 72 hours after instillation of the test substance (CPSC 2003). A positive score is defined by corneal opacity or iritis scores of ≥ 1 , or conjunctival redness or chemosis scores of ≥ 2 . In addition, if only one of the six rabbits shows ocular effects within 72 hours, the test substance is considered nonirritating to the eye. If two or three rabbits have positive ocular scores, the test is repeated in a second group of six rabbits. Then, if the criteria for an ocular irritant for the second test (three or more positive rabbits) or a nonirritant (0 positive rabbits) are met, a classification is made. However, if only one or two rabbits have positive scores in the second test, the test is repeated a third and final time. If one or more rabbits have positive ocular scores in the third test, the test substance is classified as an ocular irritant. If none of the rabbits have positive ocular scores in the third test, the test substance is classified as a nonirritant (CPSC 2003).

The EPA classification guideline considers the kinds of ocular effects produced in the *in vivo* rabbit eye test, as well as the reversibility and the severity of the effects (EPA 1996). However, unlike the FSHA system, incidence is not considered, as classification is based on the rabbit that exhibits the most severe response in a group of three or more rabbits. Data from all observation times are used for EPA classification. Corneal opacity or iritis scores of

≥ 1 , or conjunctival redness or chemosis scores of ≥ 2 define a positive score. EPA labeling regulations also require an assessment of the reversibility of positive scores. If a positive score persists for > 21 days, the substance is classified as a Category I eye irritant, which is defined as “corrosive (irreversible destruction of ocular tissue) or corneal involvement or irritation persisting for > 21 days.” Substances that cause positive corneal opacity, iritis, or conjunctival scores that clear in 8 to 21 days are designated as Category II eye irritants. If positive scores induced by a substance clear within 7 days, the substance is labeled Category III. A minimal effect (i.e., inconsequential or complete lack of irritation) or an effect that clears within 24 hours of application is designated as Category IV.

In the current EU classification system for eye irritation, risk phrases are assigned based on whether (a) two or more of three rabbits exhibit a positive score, averaged across the 24-, 48- and 72-hour observation times, or (b) the score of four or more rabbits, averaged across the 24-, 48-, and 72-hour observation times, for each ocular lesion that falls within or above certain ranges of scores (**Table 1-2**) (EU 2001). Hazard classification in the EU system corresponds to the following risk phrases: (1) R36 denotes “Irritating to eyes”; (2) R41 denotes “Risk of serious damage to the eyes.” An *in vivo* rabbit eye study that results in (1) a mean corneal opacity score ≥ 3 ; (2) a mean iris score of 2 in two or more of three rabbits; (3) an overall mean corneal opacity ≥ 3 ; or (4) a mean iris score ≥ 1.5 in four or more rabbits, would be assigned the R41 risk phrase. Additionally, if a positive score persists to ≥ 21 days, the substance is assigned the R41 risk phrase. Criteria for assigning the risk phrase R36 are provided in detail in **Table 1-2**.

The GHS for the classification and labeling of hazardous chemicals (UN 2003) is an initiative developed through the cooperative efforts of the International Labour Office, the OECD, and the UN to promote an internationally-harmonized approach for classifying chemicals according to their health hazards. For the purpose of harmonizing classification of ocular irritants, the UN adopted an approach put forth by the OECD in its *Final Report of the OECD Workshop on Harmonisation of Validation and Acceptance Criteria for Alternative Toxicological Test Methods* (OECD 1996). A tiered testing and evaluation strategy using available data from dermal irritation studies, data from validated alternative toxicological methods, knowledge of structure activity relationships, and screening for pH extremes (≤ 2 or ≥ 11.5 ; considering acid or alkaline reserve) has been proposed (UN 2003). In addition, a single harmonized hazard category is proposed for irreversible effects on the eye/serious damage to eye (Category 1). Irreversible effects according to the GHS system include grade 4 corneal lesions at any time during the *in vivo* test, positive responses on day 21 (e.g., score > 0 for any endpoint evaluated), and cases where two or more of three rabbits exhibit a mean score (24, 48, 72 hours) for corneal opacity ≥ 3 and/or iritis > 1.5 . A single harmonized hazard category, Category 2, is proposed for reversible effects on the eye; however, for regulatory authorities that prefer to distinguish irritants in this group, subcategories have been developed based on whether effects reverse within 7 or 21 days. Category 2A is defined as an eye irritant with effects that fully reverse within 21 days. Category 2B is considered mildly irritating to the eyes, and is designated for substances whose effects reverse fully within 7 days. Reversible effects include positive responses in two or more of three rabbits, where the mean score (24, 48, 72 hours) for corneal opacity or iritis ≥ 1 (but < 3

or < 1.5 , respectively), or conjunctival redness or chemosis ≥ 2 . Additional details on the GHS classification system are provided in **Section 4.3**.

4.2 Detailed Reference Data Used to Assess *In Vitro* Test Method Accuracy

The IRE studies evaluated in this document include *in vivo* reference data generated using the basic procedures described above for the *in vivo* rabbit eye test method.

For the EC/HO validation study (Balls et al. (1995), MMAS were calculated for the 59 studies from existing and concurrently run *in vivo* studies, all of which were performed according to OECD TG 405 (OECD 2002) and following GLP guidelines. The data were generated since 1981 and met the following criteria.

- Normally used at least three New Zealand White rabbits tested at the same time.
- 0.1 mL or the equivalent weight of substance was instilled into the conjunctival sac.
- Anesthesia was not used.
- Observations were made at least at 1, 2, and 3 days after instillation.

The MMAS were calculated for each test substance. Detailed *in vivo* data, consisting of cornea, iris and conjunctiva scores for each rabbit, for each of these substances are available in the ECETOC Reference Chemicals data bank (ECETOC 1998). These substances have been classified by NICEATM according to the EPA (1996), the EU (2001), and the GHS (UN 2003) ocular irritancy classification systems (**Appendix D**).

For the CEC (1991) study, *in vivo* irritancy data was obtained from historical data on 21 chemically-diverse test substances and an irritancy classification assigned, to the extent possible, according to Directive 83/467/EEC, Part II (B) of Appendix 6, Dangerous Substances Directive, 5th Adaptation, using NI (nonirritating), R36 (irritating), or R41 (severely irritating, serious risk to eyes). A total of three substances were classified as either R36 and R41 (1), or NI and R36 (2), based on different results in different laboratories. However, the greater level of irritant classification was assigned for the accuracy analysis. The rationale for use of the 21 test substances was based on inclusion of as much chemical (e.g., acid, base, inorganic and organic salt, substituted benzene, heterocycle, surfactant) and product (e.g., pesticide, detergent, antimicrobial, solvent) diversity. It was also important to have test substances covering the complete range of irritancy from nonirritant (10), to irritant (4), and severe irritant (7).

For the CTFA study, data were obtained from a modified Draize eye test. Details of the protocol are provided in Gettings et al. (1991, 1994, 1996). Six rabbits (three male, three female) were used for each test substance. The right eye of each rabbit was anesthetized prior to instillation of 0.1 mL of test substance into the conjunctival sac. Ocular irritation was evaluated at 1 hour, and at 1, 2, 3, 4 and 7 days. If irritation persisted, ocular responses were observed at 7-day intervals up to a maximum of 21 days. MAS were determined according to Williams et al. (1982). Data were classified according to the scheme proposed by the FHSA (1988). MAS, maximum average total scores for each endpoint (i.e., cornea,

iris, conjunctiva), number of positive responses, maximum day to clear, and FHSA categories are reported in the papers for all the tested substances. Detailed *in vivo* data, consisting of cornea, iris and conjunctiva scores for each rabbit, for each of these substances were provided by the CTFA. The substances have been classified by NICEATM according to the EPA (1996), the EU (2001), and the GHS (UN 2003) ocular irritancy classification systems (**Appendix D**).

For Guerriero et al. (2002, 2004) studies, data were obtained on 30 test substances (pharmaceutical process materials), respectively, from an *in vivo* Draize rabbit eye test performed approximately at the same time as the *in vitro* IRE test method. The *in vivo* data on the remaining 14 substances was obtained from historical data (ECETOC 1998). Since individual rabbit eye test data was kindly provided by Frederick Guerriero and GlaxoSmithKline, the *in vivo* data was classified according to the GHS (UN 2003), EPA (1996) and EU regulatory classification system (EU 2001) for comparison with *in vitro* data obtained using the IRE test method. The individual animal *in vivo* rabbit eye data that could be obtained for substances tested using IRE are provided in **Appendix D**.

4.3 *In Vivo* Classification Criteria Used for BRD Analysis

The *in vivo* rabbit eye database used to conduct a retrospective analysis of the accuracy of the IRE test method includes studies that were conducted using from one to six rabbits. However, some of the *in vivo* classification systems considered for the accuracy analyses are currently devised to be applied to studies using no more than three rabbits. Thus, to maximize the amount of data used for the evaluation of IRE, as well as for the three other *in vitro* test methods (ICE, BCOP, HET-CAM) being evaluated, the decision criteria for each classification system were expanded to include studies that used more than three rabbits in their evaluation.

All classification systems require the scoring of rabbits using the Draize scoring system (see **Table 4-1**). Scoring of rabbits occurs until the effect is cleared, but usually not beyond 21 days after the substance is applied to the eye of the rabbit. In order for a substance to be included in the accuracy evaluations in this BRD, four criteria must apply. These criteria were:

- At least three rabbits were tested in the study, unless a severe effect (e.g., corrosion of the cornea) was noted in a single rabbit. In such cases, substance classification could proceed based on the effects observed in less than three rabbits.
- A volume of 0.1 mL or 0.1 g was tested in each rabbit. A study in which a lower quantity was applied to the eye was accepted for substance classification, provided that a severe effect (e.g., corrosion of the cornea, lesion persistence) was observed in a rabbit.
- Observations of the eye must have been made, at minimum, at 24, 48, and 72 hours following test substance application if no severe effect was observed.
- Observations of the eye must have been made until reversibility was assessed, typically meaning that all endpoint scores were cleared. Results from a study

terminated early were not used, unless the reason for the early termination was documented.

If any of the above criteria were not fulfilled, then the data for that substance were not used for the accuracy analyses.

4.3.1 GHS Classification Rules Used for BRD Analysis

The classification of substances using the GHS classification system (UN 2003) was conducted sequentially. Initially, each rabbit tested was classified into one of four categories (Category 1, Category 2A, Category 2B, and nonirritant) based on the criteria outlined in **Table 4-3**. The criteria provided in this table are identical to those described in the GHS classification and labeling manual (UN 2003). Once all rabbits were categorized, the substance classification was determined based on the proportion of rabbits with a single irritancy category.

Table 4-3 Criteria for Classification of Rabbits According to the GHS Classification System

GHS Category	Rabbit Criteria Necessary for Classification
Category 1	<p><u>Group A:</u></p> <ul style="list-style-type: none"> - Effects in the cornea, iris, or conjunctiva that were not expected to reverse or did not fully reverse¹ within the observation period of 21 days, or - A corneal opacity score of 4 at any time during the test <p><u>Group B:</u></p> <ul style="list-style-type: none"> - Rabbit with mean scores (average of the scores on day 1, 2, and 3) for opacity ≥ 3 and/or iritis ≥ 1.5
Category 2A	<ul style="list-style-type: none"> - Rabbit with mean scores (rabbit values are averaged across observation days 1, 2, and 3) for one of more of the following: <ul style="list-style-type: none"> Iritis ≥ 1 but < 1.5 Corneal opacity ≥ 1 but < 3 Redness ≥ 2 Chemosis ≥ 2 <p>and the effects fully reverse within 21 days</p>
Category 2B	<ul style="list-style-type: none"> - Rabbit with mean scores (rabbit values are averaged across observation days 1, 2, and 3) for one of more of the following: <ul style="list-style-type: none"> Iritis ≥ 1 but < 1.5 Corneal opacity ≥ 1 but < 3 Redness ≥ 2 Chemosis ≥ 2 <p>and the effect fully reversed within 7 days</p>
Nonirritant	Rabbit mean scores fall below threshold values for Category 1, 2A, and 2B

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

¹Full reversal of the effects was defined as corneal opacity, iritis, redness, and chemosis = 0.

After each rabbit was categorized, the ocular irritancy potential of the substance was determined. As shown in **Table 4-4**, substance classification depended on the proportion of

rabbits that produced the same response. As noted above, if a substance was tested in more than three rabbits, decision criteria were expanded. Generally, the proportionality needed for classification was maintained (e.g., one out of three or two out of six rabbits were required for classification for most categories). However, in some cases, additional classification rules were necessary to include the available data. These additional rules are distinguished by italicized text in **Table 4-4**.

Table 4-4 Criteria for Classification of Substances According to the GHS Classification System (Modified from UN 2003)

GHS Category	Criteria Necessary for Substance Classification
Category 1	<ol style="list-style-type: none"> 1. At least 1 of 3 rabbits or 2 of 6 rabbits classified as Category 1, Group A 2. <i>One of 6 rabbits classified as Category 1, Group A and at least 1 of 6 rabbits classified as Category 1, Group B</i> 3. At least 2 of 3 rabbits or 4 of 6 rabbits classified as Category 1, Group B
Category 2A	<ol style="list-style-type: none"> 1. At least 2 of 3 rabbits or 4 of 6 rabbits classified as Category 2A 2. <i>One of 3 (2 of 6) rabbits classified as Category 2A and 1 of 3 (2 of 6) rabbits classified as Category 2B</i>
Category 2B	At least 2 of 3 rabbits or 4 of 6 rabbits classified as Category 2B
Nonirritant	At least 2 of 3 rabbits or 4 of 6 rabbits classified as nonirritant

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

Italicized text indicates rules that were developed to include additional data.

If an unequivocal substance classification could not be made due to the response pattern of the tested rabbits for a substance (e.g., one rabbit classified as Category 1, Group B; two rabbits classified as Category 2B; three rabbits classified as nonirritant), the data were not used in the analysis.

4.3.2 EPA Classification Rules Used for BRD Analysis

The classification of substances using the EPA classification system (EPA 1996) was conducted sequentially. Initially, each rabbit was classified into one of four categories (Category I to Category IV) (**Table 4-5**.)

Substance classification was dependent upon the most severe category observed among the tested rabbits. Thus, a single rabbit in a more severe category than the remaining animals would lead to classification of the substance into that category (i.e., classification of a substance was not based on the majority classification among rabbits tested).

Table 4-5 Criteria for Classification of Rabbits According to the EPA Classification System (EPA 1996)

EPA Category	Criteria for Rabbit Classification
Category I	- Corrosive, corneal involvement or irritation (iris or cornea score ≥ 1 or redness or chemosis ≥ 2) persisting more than 21 days or - Corneal effects that are not expected to reverse by 21 days
Category II	- Corneal involvement of irritation clearing ¹ in 8 to 21 days
Category III	- Corneal involvement of irritation clearing in 7 days or less
Category IV	- Minimal or no effects clearing in less than 24 hours

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

¹For the purposes of this analysis, clearing was defined as iritis or corneal opacity score < 1 and redness or chemosis score < 2 .

4.3.3 EU Classification Rules Used for BRD Analysis

Substance classification using the EU classification system was conducted sequentially (EU 2001). While average Draize scores are used for classification, the calculation of average scores for the EU system depends on the number of rabbits tested in a study (see **Section 4.1.3** for additional details). Depending on the number of rabbits tested, the appropriate average scores were calculated, then the substance was classified based on the number of rabbits with a minimal positive average (for studies that used three rabbits) or the overall average (for studies that used more than three rabbits). The criteria used for substance classification are in **Table 4-6**.

Table 4-6 Criteria for Classification of Substances According to the EU Classification System (EU 2004)

EU Category	Three Rabbits Tested	Greater than Three Rabbits Tested
R41	Two or more rabbits where the average rabbit Draize scores over Days 1, 2, and 3 were: Opacity ≥ 3 Iritis = 2 Or At least one rabbit (at end of observation period), where the effect has not reversed ¹	Overall mean rabbit Draize scores over Days 1, 2, and 3 were: Opacity ≥ 3 or Iritis > 1.5 Or At least one rabbit (at end of observation period), where the effect has not reversed
R36	Two or more rabbits where the average rabbit Draize scores over Days 1, 2, and 3 were: $2 \leq \text{Opacity} < 3$ $1 \leq \text{Iritis} < 2$ Redness ≥ 2.5 Chemosis ≥ 2	Overall mean rabbit Draize scores over Days 1, 2, and 3 were: $2 \leq \text{Opacity} < 3$ $1 \leq \text{Iritis} < 1.5$ Redness ≥ 2.5 Chemosis ≥ 2

Abbreviation: EU = European Union.

¹Full reversal of the effects was defined as corneal opacity, chemosis, redness, or iritis = 0.

4.4 Availability of Original Records for the *In Vivo* Reference Data

Much of the published data on the prediction of ocular irritancy potential for test chemicals using the *in vivo* test method was limited to average score data, average animal data, or irritancy classification. An attempt to obtain the original records and/or compiled reports for the *in vivo* reference data was made. However, much of the information and data was either not readily available or not provided in the requested format.

4.5 *In Vivo* Data Quality

Ideally, all data supporting the validity of a test method should be obtained and reported from studies conducted in accordance with GLP guidelines, which are nationally and internationally recognized rules designed to produce high-quality laboratory records (OECD 1998; EPA 2003a, 2003b; FDA 2003). GLP guidelines provide an internationally standardized approach for the conduct of studies, reporting requirements, archival of study data and records, and information about the test protocol, in order to ensure the integrity, reliability, and accountability of a study.

The extent to which the *in vivo* rabbit eye studies, which were used to provide the comparative data in the published IRE validation studies, were compliant with GLP guidelines is based on the information provided in the published reports. Although an attempt was made to obtain the original study records, such records could not be obtained. Based on the available information, the *in vivo* rabbit data used in three of these four reports (Balls et al. 1995; Gettings et al. 1996; Guerriero et al. 2004) were obtained in compliance with GLP guidelines.

4.6 Availability and Use of Toxicity Information from the Species of Interest

Due to the possibility of irreversible eye injury that could impair vision or cause blindness, human ocular irritancy studies are not routinely conducted. The only exceptions are for products intended for actual human eye use (e.g., contact lens solutions, ophthalmic pharmaceuticals) or cosmetic/personal care products that are known not to cause more than minimal to mild responses in rabbits. Bruner et al. (1998) and Cater et al. (2004) reported on studies conducted in humans of cosmetic and surfactant-based personal care formulations. However, all of the substances tested were classified as mild irritants or nonirritants and corresponding IRE tests were not conducted. Procter & Gamble provided information from human exposures to three consumer-product formulations as a comparison to the EU ocular toxicity classifications (EU 2001), assigned based on results from the low volume eye test (LVET). However, because all three of these formulations were classified as nonirritants or mild irritants, based on results obtained in LVET, evaluation of the accuracy of the IRE test method for identifying ocular corrosives and severe irritants in humans is not possible.

It may be possible to consider accidental human exposure injury data to identify substances or products capable of producing severe or irreversible eye injuries in humans. These data could then be compared with available rabbit data and hazard classifications to determine if the potential for severe human effects was not predicted by the rabbit test. A query to all ICCVAM regulatory agencies did not yield any substances or products known to produce

severe or irreversible human eye injury not predicted by the rabbit test. However, this lack of such substances or products must be considered in light of the surveillance and reporting systems for such injuries.

Several U.S. Federal agencies (OSHA, CPSC, and the National Institute for Occupational Safety and Health [NIOSH]) were contacted for data resulting from accidental human exposures. Based on emergency department reports for work related eye-injuries, NIOSH estimated that approximately 39,200 chemical-related eye injuries occurred in 1998, (NIOSH 2004). Approximately 10,000 of these cases were attributed to an unidentified or unspecified chemical. Additional cases (< 2500 each) were reported for injuries related to specific chemicals or chemical/product classes, which included²:

- acids (unspecified)
- adhesives/glues
- cement/mortar mix
- chlorine/chlorine bleach
- cleaning/polishing agents
- detergents/shampoos
- disinfectants
- drain/oven cleaners
- gasoline/jet fuels/diesel fuel
- hydrochloric acid
- nonchlorine bleach
- paint removers/thinners
- paints
- soaps
- sodium hydroxide, potassium hydroxide, and potassium carbonate
- solvents/degreasers
- sulfuric acid

However, for the product classes listed above, specific information on which products were involved are not available. No human data were provided for any of these substances, nor were details of the types of ocular injuries sustained described.

In addition, according to U.S. Bureau of Labor Statistics (BLS), 6303 lost workdays attributable to occupational eye injuries from chemical exposures were reported in 2002 (BLS 2004). These numbers may be underestimates of the actual incidence, since not all employers are required to report such injuries. The specifics of the exposures are not provided.

Without more detail about the specific nature of the substances and exposure conditions, these types of accidental human exposure injury data are not useful for evaluating the accuracy of the IRE test method for predicting human ocular hazard.

² These specific chemicals or chemical/product classes are listed in alphabetic order; actual numbers of cases for each specific chemical or chemical/product class are not provided.

4.7 Information About Accuracy and Reliability of the *In Vivo* Test Method

4.7.1 Information About the Accuracy of the *In Vivo* Test Method

Accuracy of the *in vivo* test method would ideally be assessed by comparison of ocular effects observed in the rabbit to those effects produced in humans. A review of the literature indicates that there are few studies in which rabbit and human responses have been carefully compared under controlled conditions to assess the accuracy of the *in vivo* test method. Therefore, most studies conduct retrospective evaluations and comparisons of responses between humans and rabbits. A review indicates that a number of studies show that responses to mild to moderate irritants were generally similar between rabbits and humans (Lewin and Guillery 1913; Suler 1913; Leopold 1945; Carpenter and Smyth 1946; McLaughlin 1946; Nakano 1958; Barkman 1969; Grant 1974). A review of these studies can be found in McDonald et al. (1987). For a severe irritant, Grant (1974) and Butscher (1953) showed that accidental exposure to neat thioglycolic acid produced similar responses in humans and rabbits.

In comparison, there have been studies where the responses to ocular irritants differ between humans and rabbits. In some cases, test substances produced more severe responses in humans than in rabbits (Lewin and Guillery 1913; Gartner 1944; Estable 1948; Marsh and Maurice 1971; Grant 1974). For example, Marsh and Maurice (1971) evaluated the effects of a 1% concentration of nonionic detergents in humans. The most severe symptoms (e.g., blurred vision and halos with corneal epithelial bedewing; most effects disappearing with 24 hours) were associated with 1% Brij 58. Comparatively, Grant (1974) showed that, in general, nonionic detergents did not damage the rabbit eye, even when tested at higher concentrations. Additional examples of disparate effects between humans and rabbits are summarized in McDonald et al. (1987). Studies with some soaps and surfactants indicated that more severe responses were produced in rabbits than in humans (Calabrese 1983). Differences between humans and rabbits with respect to anatomy and physiology, pain thresholds, exposure parameters (e.g., volume administered, length of exposure period), and potential differences in mechanism of action of test substances have been proposed as reasons for the discordant responses.

4.7.2 Information About the Reliability of the *In Vivo* Test Method

Based largely on the protocol of Draize et al. (1944), the original regulatory requirements for eye irritation testing mandated the use of at least six rabbits. In recognition of animal welfare concerns, several evaluations were conducted to assess the reliability of the test method and the consequences of reducing the number of rabbits per test from six to as few as two (DeSousa et al. 1984; Solti and Freeman 1988; Talsma et al. 1988; Springer et al. 1993; Dalbey et al. 1993; Berdasco et al. 1996). With the exception of Dalbey et al. (1993), each study concluded that reducing the number of rabbits from six to three would not have an unacceptable reduction on the predictivity of ocular irritancy classification/categorization. Analyses were performed using MAS, internal irritancy classification schemes, and/or regulatory classification schemes as endpoints for comparison. Several of these studies (DeSousa et al. 1984; Talsma et al. 1988; Dalbey et al. 1993) revealed that correlations between three-rabbit and six-rabbit classifications were the highest among substances classified on the extreme ends of the irritancy range (i.e., nonirritants and severe irritants).

These studies noted that the majority of variability among rabbit responses was observed among substances classified in the middle range of irritation (i.e., mild and moderate irritants). Accordingly, Dalbey et al. (1993) concluded that the observed variability in the middle range of irritation justified the continued routine use of six rabbits. However, based primarily on the results of these evaluations, the EPA (EPA 1998), EU (EU 2001), and the OECD (in revised TG 405), recommended the use of a maximum of three rabbits, although additional rabbits could be tested under certain circumstances (e.g., to confirm weak or moderate responses).

To further address the reliability of the rabbit eye test, ICCVAM and NICEATM used the available *in vivo* data to estimate the likelihood of underclassifying a positive substance or overclassifying a negative substance in the current one to three rabbit sequential test. Data from Draize eye testing using three to six rabbits was obtained for approximately 900 substances from U.S. Federal regulatory agencies, published studies, and scientists and organizations. Ocular irritation categories were assigned for each substance based on the GHS classification system (UN 2003). Using the available *in vivo* rabbit eye test database of 181 severe irritant studies, the distribution of individual rabbit responses within each severity class was used to estimate the likelihood of under- and over-classification rates for a sequential one to three rabbits testing strategy. Based on three different assumptions about the variability in response among substances within each classification category, the estimated underclassification rate for corrosives/severe irritants (GHS Category 1) as nonsevere irritants (GHS Category 2) or nonirritants ranged from 4% to 13%. Analyses based on physical form of the test substance suggested that underclassification rates for solids were lower than liquids (2.9% to 8.3% vs. 5.4% to 15.8%, respectively), although these differences are not statistically significant. Estimated underclassification rates were higher when a corrosive/severe irritant classification was based solely on persistent lesions present at observation day 21. By chemical class, carboxylic acids had the highest underclassification rate (16.64%). Overclassification rates of substances as corrosive/severe irritants, based on 596 studies, were estimated to be 7% to 8% for Category 2A substances, 1% for Category 2B substances, and 0% for nonirritants.

[This Page Intentionally Left Blank]