2.0 IRE TEST METHOD PROTOCOL COMPONENTS

2.1 Overview of How the IRE Test Method is Conducted

The IRE test was developed by Burton and his colleagues at Unilever Research Laboratory, Colworth, United Kingdom as an *in vitro* alternative to the *in vivo* Draize rabbit eye test method for the assessment of eye irritation (Burton et al. 1981). In the IRE test method, liquid test substances are spread using a syringe and solids are pulverized and applied as a powder over the corneas of enucleated rabbit eyes. The principal advantages of this test method are that the animals are euthanized prior to ocular irritancy testing, eyes from animals used for other toxicological purposes or from the food chain can be used, and testing is performed on the cornea, the part of the eye that is generally given the highest weight for scoring ocular irritancy in the Draize test. The effects of the test substance on the cornea of the isolated eye are measured quantitatively as an increase in thickness (swelling), subjectively as scores for corneal opacity, the area of corneal involvement, and fluorescein penetration, and descriptively as morphological changes to the corneal epithelium. However, the number of ocular parameters and the number of time points measured varies from study to study. Two additional refinements of the IRE test method may be incorporated into the protocol or used *ad hoc* to supplement existing data. One is the use of histopathology to confirm or identify the extent of irritancy at the cellular level, especially when the degree of irritancy falls between moderate and severe. Another is the use of confocal microscopy to determine the extent and depth of ocular injury (Maurer et al. 2002). Many studies using the IRE test method evaluate single or multiple ocular endpoints at various times and then assign irritancy classifications to the substances tested (CEC 1991; Köeter and Prinsen 1995; Cooper et al. 2001; Jones et al. 2001), while others use mean data from one or more ocular endpoints assessed at various times after application of the test substance, typically 0.5 to 4 hours (Balls et al. 1995; Gettings et al. 1996). One protocol for the IRE test method was designed to specifically identify severe eye irritants (Guerriero et al. 2004). In this study, cut-off values for each ocular parameter tested were predetermined. If these cut-off values were achieved or exceeded in any single parameter over a period of 0.5 to 4 hours, including a significant change in the corneal epithelium, the test substance was classified as a severe eye irritant with potential to cause serious or irreversible damage to the human eye. Protocols developed and used at SafePharm and Unilever in the United Kingdom were provided (Jones P and Whittingham A, personal communications) and information on additional IRE protocols was obtained from reports in the literature (Gettings et al. 1966; Burton et al. 1981; Price and Andrews 1985; INVITTOX 1994; Balls et al. 1995; Chamberlain et al. 1997; Cooper et al. 2001; Jones et al. 2001; Guerriero et al. 2004). These protocols are compared in **Appendix A.**

2.2 Description and Rationale for the Test Method Components

Currently, there is no widely accepted, standardized IRE test method for detecting ocular corrosives and severe irritants. Evaluation of the IRE test method for its usefulness as a partial or full replacement for the Draize rabbit eye test has been confounded by the lack of a standardized protocol. Although initially developed by Burton et al. (1981) for the

assessment of severe eye irritants using a relatively small set of eleven test substances, the IRE test method has been modified for use in the assessment of either selective types of irritants (e.g., severe irritants) or for specific classes of chemical substances or products (e.g., surfactant-based chemicals, cosmetic and hair care products) (Gettings et al. 1966; Chamberlain et al. 1997; Cooper et al. 2001; Jones et al. 2001). In other studies, protocols were geared to evaluate a wider range of chemical classes over the entire range of irritancy for test method assessment or validation purposes (Price and Andrews 1985; Köeter and Prinsen 1985; CEC 1991; Balls et al. 1995; Gettings et al. 1996) or for interlaboratory trials (Whittle et al. 1992). Guerriero et al. (2004) modified the original IRE test method protocol to refine assessment of pharmaceutical worker safety by using decision criteria (Prediction Model) designed to identify severe eye irritants using a chemical database of 30 pharmaceutical ingredients, chemical intermediates, and raw materials and an additional 14 reference chemicals from ECETOC (1998).

The following sections describe in detail the essential components of the IRE test method for the identification of ocular corrosives or severe eye irritants. For each section, a summary is provided of the information obtained from reviewed reports and personal communications with expert scientists knowledgeable about the assay. Many of the components of these protocols have been included based on historical use, and the rationale for their selection is not known. For each test method component, a summary is presented of information obtained from:

- SafePharm Laboratories, a toxicology laboratory that has performed the enucleated rabbit eye assay in a GLP-compliant testing facility since 1999.
- INVITTOX Protocol No. IP-85 (1994). This protocol was used by the lead laboratory for the Balls et al. (1995) IRE validation study.
- A literature search and review of publicly available IRE protocols; which are based on the methodology first reported by Burton et al. (1981). These protocols are summarized in **Appendix A**.
- Discussion and personal communication with Ms. Penny Jones (Unilever) and Mr. Robert Guest (SafePharm Laboratories), scientific experts currently using the isolated rabbit eye test method, with additional information provided by Dr. Andrew Whittingham (SafePharm Laboratories) and Mr. Frederick Guerriero (GlaxoSmithKline).

2.2.1 Materials, Equipment and Supplies Needed

2.2.1.1 *Source of Rabbit Eyes*

Typically, healthy New Zealand white rabbits weighing 2.5 to 4.0 Kg are used. Rabbits are usually purchased from laboratory animal suppliers. However, since a principal purpose of the IRE test method is to reduce animal use, eyes have been obtained from laboratory rabbits used for other purposes, such as skin testing, in which the eyes are not affected (e.g., mild or nonirritant substances or control eyes). However, where regulatory agencies do not permit animal reuse, it is possible to obtain eyes from an abattoir (e.g., PelFreeze, Rogers, AZ) where rabbits are routinely killed for food. Local abattoirs are available throughout the U.S. and Europe. There are 200,000 rabbit producers throughout the U.S. with turnover of 6 to 8 million rabbits per year [http://agalternatives.aers.psu.edu/other/rabbit/rabbit.pdf]. Eyes are typically shipped from a local laboratory or abattoir in a humidified container wetted with

saline or an appropriate buffer solution at room temperature for use within an hour or on ice for longer periods up to 24 hours. Eyes have been shipped overnight (i.e., PelFreeze, Rogers AR) on ice under conditions that do not have adverse effects on corneal transparency or physiological function when the abattoir is instructed on how to remove and package the eyes properly (Edelhauser H, personal communication). No ages have been reported for rabbits used in IRE test method in the literature. In terms of the weight range, there have been reports of differences in corneal thickness between rabbits that weigh less than 2000 grams and those that weigh more than 2000 grams (Burton et al., 1972). When two ranges of rabbit weights were compared with respect to corneal thickness, animals in the 1300 to 2000 gram range had corneal thickness measurements of 0.346 ± 0.02 mm (mean and standard deviation [SD], 156 eyes) versus 0.382 ± 0.017 mm (mean and SD, 18 eyes) for rabbits in the 2000-2700 gram range. However, there have been no reports regarding differences in the ability of the IRE test method to detect ocular corrosives and severe irritants depending on rabbit weight or corneal thickness. Published IRE studies typically report on the use of rabbits in the 2500 to 4000 gram weight range. Although corneal thickness in rabbits depends on animal weight, no studies have been conducted to evaluate whether differences in corneal thickness would alter the performance characteristics of the IRE test method. Furthermore, there are no reported studies comparing use of rabbit strains other than New Zealand White in the IRE test method and the consequence of the use of other strains (e.g., California, New Zealand Red) are unknown.

2.2.1.2 *Quality of Eyes*

Currently, there are no standardized criteria for the selection of rabbit eyes for the IRE assay. Most IRE studies reported that eyes were carefully examined visually for defects, including opacity, scratches or pitting, pannus or neovascularization, once they had arrived at the laboratory. A few studies also noted use of stereomicroscopes or loupes to assist in identifying damaged corneas. Some laboratories reported use of fluorescein to assist in the identification of corneal epithelial barrier defects.

The quality of the eyes is typically evaluated at later steps in the assay, as well. For example, an increase in corneal thickness measured just before and/or after equilibration of greater than 7 to 10% relative to that of the corneal thickness measurement taken during the initial eye examination following enucleation would result in rejection of the use of that eye in an assay.

2.2.1.3 *Preparation of the Eyes*

In general, rabbits are euthanized by an intravenous injection of a lethal dose of sodium pentobarbitone (approximately 200 mg/kg) into the marginal ear vein. The corneas are kept moist after sacrifice with drops of physiological saline (prewarmed from 31 to 32°C) applied throughout the dissection process. Although the dissection process is not typically described in the literature reports, scientists with expertise in performing the dissection have provided details of the procedure (Jones P, Guest R, personal communication). The nictitating membrane is deflected away using forceps and the conjunctivae are cut using angled forceps and curved scissors. The eyeball is proptosed by applying gentle pressure with fingers above and below the orbit. The remaining conjunctival tissue, the orbital muscles and the optic nerve (leaving approximately a 5-10 mm section to prevent loss of intraocular pressure) are

removed and the eyeball is lifted from the orbit. Any tissue adhering to the globe is then removed by careful dissection, and the eyeball is gently rinsed with a stream of physiological saline to remove any adherent debris. The eyes are prepared for immediate use or for shipment as described in **Section 2.2.1.1**.

2.2.1.4 *IRE Experimental Setup*

Burton et al. (1981) provided a description and drawing of the original IRE experimental eye incubation apparatus, which was termed a superfusion chamber, and most studies to date have used slight variations of this original instrumental setup, usually expanding the number of eyes that can be accommodated by the apparatus for a single experiment. The superfusion apparatus is a large Perspex chamber that has a water-jacketed surface that maintains the temperature of multiple (usually six to eleven) individual cells that house the isolated eyes during the experiment. The dissected eyes are mounted in specially designed Perspex, plastic or metal holders with rings or studs on both the upper and lower jaws of the holding clamp, which provide just enough pressure to keep the eyes from slipping while maintaining intraocular pressure. The holders are designed to fit into the individual cells of the superfusion apparatus with the eyes maintained in a vertical position. The holders can be readily removed and placed in a horizontal position for test substance application. The individual chambers are typically blackened to permit slit-lamp observations. A saline drip tube is mounted over the eye within the chamber and a steady drip of warm saline is used to maintain the eye in a hydrated condition. The temperature maintained in the cells of the superfusion apparatus is typically 31 to 32°C with a range of approximately 1.5°C. A water bath and two peristaltic pumps are used to heat and circulate the water and saline used for temperature control. The flow rates vary due to changes in ambient conditions in the laboratory, but are typically around 4 liters/minute for the water and range from 0.1-0.4 mL/min for the saline drip.

2.2.2 Dose-Selection Procedures, Including the Need for Any Dose Rangefinding Studies or Acute Toxicity Data Prior to Conducting a Study

As described below in **Section 2.2.4**, test substances are typically evaluated undiluted at a fixed volume of 0.1 mL liquid or a fixed weight of 100 mg powdered solid (prewetted if necessary). Because a fixed quantity of a substance is tested, dose selection is not a relevant issue.

2.2.3 Endpoints Measured

2.2.3.1 *Corneal Opacity*

The original developer of the IRE test method (Burton et al. 1981) noted changes in corneal opacity visually, by slit-lamp, and by fluorescein staining to assess the extent of corneal injury (i.e., effects on corneal stroma and/or epithelium). Corneal opacity was not formally scored in the original report, but was used in conjunction with corneal swelling measurements (see **Section 2.2.3.2)** to produce an arbitrary irritancy rating (e.g., negligible, slight, moderate, severe) for comparison of data on 10 test substances with *in vivo* rabbit eye data. The *in vivo* Draize et al. (1944) scoring system or a slightly modified version of it for assessment, although subjective, is now routinely used to score corneal opacity and area (Prinsen and Köeter 1985; Whittle et al. 1992; Balls et al. 1995; Jones et al. 2001). In addition, the McDonald-Shadduck ocular scoring system (Hacket and McDonald 1991) was

based on slit-lamp observations and proposed as an *in vivo* alternative to the Draize method to reduce subjectivity. Guerriero et al. (2004) used the McDonald-Shadduck scoring system for the identification of severe ocular irritants and it is the *in vitro* scoring method used routinely for isolated rabbit eye irritation testing at SafePharm Laboratories (Guest R, Whittingham A, personal communication).

2.2.3.2 *Corneal Swelling*

Burton et al. (1981) used a depth-measuring attachment on a slit-lamp microscope to measure corneal thickness in the isolated rabbit eye and then used relative changes in corneal thickness 4 hours after application of the test substance to provide a quantitative measurement of corneal swelling. Since then, corneal swelling is routinely used as an ocular endpoint at various times after application of a test substance (York et al. 1982; Price and Andrews 1985; Prinsen and Köeter 1985; CEC 1991; Balls et al. 1995; Gettings et al. 1996; Jacobs and Martens 1988; Cooper et al. 2001; Guerriero et al. 2004).

Corneal swelling may be calculated from corneal thickness measurements using the following equation:

$$\left(\frac{corneal\ thickness\ at\ time\ t \quad - \quad corneal\ thickness\ at\ time = 0}{corneal\ thickness\ at\ time = 0} \right) \quad \times \quad 100$$

In general, corneal thickness is measured as a quantitative endpoint and corneal swelling is typically calculated at 0.5, 1, 2, 3 and 4 hours.

2.2.3.3 *Fluorescein Penetration/Retention*

Although fluorescein staining is used routinely to assess the integrity of eyes used in the IRE test method, it is not used routinely as a scored endpoint. However, several investigators report the use of fluorescein retention or penetration as a scored ocular endpoint to supplement general observations regarding corneal opacity and to provide an indication of the area of the cornea affected and the type of lesion produced (e.g., diffuse, stippled, focal) or as an indicator of the depth of penetration of the injury to the cornea (i.e., intensity of fluorescein stain). Fluorescein staining has been applied by some investigators at 0.5 hours (CEC 1991), but is more often applied at 4 hours after the other endpoints are evaluated for any effects produced by the test substance (CEC 1991; Prinsen and Köeter 1985; Guerriero et al. 2004). Scoring systems for fluorescein area and intensity staining of isolated rabbit eyes are generally arbitrary, but well defined (e.g., scales of 0 to 3 or 0 to 4) in a manner similar to ocular opacity and area scores (Prinsen and Köeter 1985; Guerriero et al. 2004).

2.2.3.4 *Assessment of Epithelial Integrity*

Guerriero et al. (2004) reported the evaluation of epithelial integrity visually and by slit-lamp as an indicator of severe ocular corrosion or irritation. Any significant indication of corneal epithelial stippling, mottling, pitting, ulceration, pannus, or other significant or irreversible

corneal epithelial defects are considered sufficient to indicate that a substance is an ocular corrosive or severe irritant.

2.2.3.5 *Additional Endpoints*

Routine or selective use of histopathology may provide useful information regarding: 1) the depth of ocular injury; 2) characteristics of the injury at the cellular level; and 3) integrity of critical and irreplaceable cellular components such as endothelial or stem cells. Furthermore, histopathology can be combined with modern staining techniques (e.g., vital dyes, immunohistopathology, biochemical markers) to provide information on possible mechanisms of ocular toxicity. Unfortunately, not all laboratories are equipped to perform histopathology, or do not have access to or a collaboration with other laboratories that have this capability. Furthermore, the added cost of routine histopathology might impact consideration of the use of histopathology by an ocular toxicity-testing laboratory, especially if the burden of the increased cost falls on the client.

Confocal microscopy has been used with vital dyes to measure the depth of corneal injury as the level of penetration of the live/dead cell layer (Jester et al. 1996; Jester et al. 2001). The extent of this corneal injury has been suggested as the mechanistic basis for ocular irritation (Maurer et al. 2002). Again, the increased cost of a confocal microscope may be a significant burden to many laboratories.

2.2.4 Duration of Exposure

2.2.4.1 *Pre-exposure Preparations*

In most published studies, once the isolated rabbit eyes are assessed for their utility in the assay and deemed free of ocular defects, they are equilibrated in a superfusion apparatus based on the one originally described by Burton et al. (1981) for a period of 30 to 60 minutes at temperatures ranging from 31 to 32 °C. The eyes are mounted in special plastic or metal holders fitted with rings or pins used to gently clamp the eye in place and to prevent them from slipping and to minimize changes in intraocular pressure. Using these holders, the eyes are removed from the superfusion apparatus for application of the test substance, then returned to the original vertical position under the saline drip tube for incubation.

2.2.4.2 *Test Substance Exposure Duration*

Following equilibration, the isolated rabbit eye is typically exposed to the test substance for a total of 10 (± 2) seconds at which time it is gently rinsed off with a volume of 20 mL of physiological saline (prewarmed to 31 or 32°) using a syringe or other means of delivery. The 10-second exposure period is the standard time used by most investigators to identify and assess the ocular effects of severe eye irritants. A note is recorded if any particles of solid or precipitated material remain on the surface of the cornea after rinsing. Some authors have increased the time of exposure to 1 minute (Cooper et al. 2001; Jones et al. 2001) when evaluating select products such as hair shampoos that are generally in the mild to moderate range of eye irritation.

2.2.4.3 *Application and Amount of the Test Substance*

A volume of 0.1 mL of a neat liquid test substance or 0.1 gram of a solid ground to a fine powder is typically applied to the isolated rabbit eye using a syringe. The isolated eye is

removed from the equilibration chamber and placed in a horizontal position with the cornea facing upward for application of the test substance. This is the standard scientific practice for volume or weight used by most investigators in the *in vivo* rabbit eye test in the relevant literature and it is the application volume and weight currently accepted by the U.S. (EPA 1998) and EU (EU 2001, UN 2003 [GHS]) regulatory agencies for *in vivo* studies.

2.2.4.4 *Number of Eyes Required per Test Substance and Controls*

Historically, in IRE studies, one to three isolated rabbit eyes have been used to assess the ability of a test substance to induce corrosion or irritation; one isolated rabbit eye has been used as the negative control and, if included, one isolated rabbit eye has been used to assess the ability of a positive control substance to induce an appropriate response (Whittle et al. 1992; Balls et al. 1995; Gettings et al 1996; Jones et al. 2001). Some authors include benchmark controls (Jones et al. 2001).

2.2.4.5 *Concentration of Test Substance*

For regulatory purposes, substances are generally tested neat in the IRE. However, there have been reports comparing neat liquid test substances with 10% dilutions (Cooper et al. (2001; Jones et al. 2001) at 10 and 60 second exposure times to differentiate mild and moderate eye irritants for select product types such as hair shampoos.

2.2.5 Known Limits of Use

The IRE was designed as an *in vitro* test method to measure effects on the isolated eye; namely effects on the cornea. However, scoring of irritancy of the iris similar to that reported in the Draize assay such as swelling, injection, reaction to light, and hemorrhage is not possible, because the physiological mechanisms for their production (principally blood flow and muscular activity) are not present in the isolated eye. In addition, the dissection and removal of the conjunctiva during removal of the eye precludes testing irritant effects on this tissue. Although a severe ocular irritant may produce significant corneal damage, the EPA has documented cases in which severe conjunctival irritancy persisted for 21 days in the absence of significant corneal opacity, and the test substance was therefore labeled a severe ocular irritant according to the EPA (EPA 1996) classification system (Lewis M, personal communication). Thus, severe irritancy of a test substance resulting from severe effects on the conjunctiva cannot be identified in the IRE assay. Finally, reversible ocular effects cannot be evaluated in the IRE test method.

2.2.6 Nature of the Response Assessed

As noted in **Section 2.2.3**, the corneal endpoints observed in the IRE are opacity, swelling, fluorescein retention, and morphological effects on the epithelium. The severity of each response is graded at each time point (with the exception of fluorescein retention which is generally assessed only at four hours so as not to interfere with the other endpoint evaluations). The data to be collected includes both numerical and descriptive data. The numerical data includes scores for corneal opacity, corneal thickness, and fluorescein retention, while the descriptive data represents morphological and/or histopathological findings. Alternative endpoints such as histopathology and confocal microscopy for evaluation of depth of corneal injury are available if it becomes necessary to differentiate a moderate response from a severe response, when the existing endpoints do not permit this

level of differentiation. Use of vital dyes alone or with confocal microscopy, immunohistopathology, or biochemical markers may be useful to assess mechanistic aspects of a severe irritant.

2.2.7 Appropriate Controls and the Basis for their Selection

2.2.7.1 Negative Controls

The negative control provides a baseline for the assay endpoints, ensures that the experimental conditions do not inappropriately result in an irritant response, and permits detection of nonspecific changes in the test system. The most frequently reported negative control in published IRE studies is isotonic saline. This would appear to be the most suitable control since the test method is conducted using isotonic saline to bathe the rabbit eyes as well as for the requisite rinsing steps. Treating the negative control eyes with isotonic saline ensures that any mechanical alterations (i.e., those not related to the test substance) are properly controlled. There have been no formal studies to assess whether results obtained using buffered salt solutions (e.g., Ringer's) would be similar to those using isotonic saline. For example, it is not known whether buffered salt solutions would be an impediment to proper evaluation of an acidic or basic test substance due to pH control, or if they would increase or reduce the false positive or false negative rates and impact accuracy.

2.2.7.2 Solvent/Vehicle Controls

Based on a review of published IRE studies, it appears that concurrent solvent controls have not been used. However, it is scientifically critical to know that the vehicle for the test material, if different from isotonic saline, has an impact on the outcome of the study by producing irritancy on its own and possibly impeding the detection of irritancy of the test substance.

2.2.7.3 Positive Controls

As discussed by Harbell and Curren (2002), the function of the positive control is to ensure that the test method is operating within normal limits and that each experiment is properly executed, such that the toxic effects of interest can be properly detected. A concurrent positive control substance is included in each experiment to develop a historical database. Results from the concurrent positive control are compared to the historical control range, which is used to determine whether a particular experiment is acceptable. Because the positive control should allow for detection of an over- or under-response in the assay, the selected positive control should not produce responses at either the extreme low or the extreme high end of assay response. In the literature, positive controls have not historically been used in the IRE test method, because positive controls would typically be severe irritants such as 10% sodium hydroxide, by nature of the design of the original IRE test method (Burton et al. 1981). As discussed by Harbell and Curren (2002), the importance of a positive control cannot be overemphasized, and perhaps severe irritants producing less than a maximal level of ocular damage could be considered as positive controls to permit assessment of variability over time and to insure the integrity of the test system and its proper execution.

2.2.7.4 Benchmark Controls

Benchmark substances are often used during the testing of substances of unknown toxicity potential. The toxicity of the benchmark substance is generally well characterized (i.e., adequate human or animal toxicity data are available). A benchmark is selected to match the chemical or product type of the unknown substance, and is used to set an upper or a lower limit of response against which the unknown is compared (Harbell and Curren 2002). Benchmark substances are often selected from a list of reference chemicals for the assay and have the following properties:

- consistent and reliable source(s)
- structural and functional similarity to the class of the substance being tested
- known physical/chemical characteristics
- supporting data on known effects in the *in vivo* rabbit eye test
- known potency in the range of the desired response

They are useful for evaluating the ocular irritancy potential of unknown chemicals of a specific chemical or product class, or for evaluating the relative irritancy potential of an ocular irritant within a specific range of irritant responses.

2.2.8 Acceptable Range of Control Responses and the Basis for the Acceptable Ranges

2.2.8.1 *Negative/Solvent Controls*

Negative controls are generally considered acceptable if they produce a nonirritant response. Solvent controls are not typically used in the IRE test method, but would also be expected to produce a nonirritant response. A positive solvent response (mild or moderate irritation) would generally require replacement of the irritating solvent with a nonirritating solvent, unless it was part of a formulation that could not be changed. The basis for the acceptable range of negative controls were derived from observations made by laboratories experienced in the performance of the IRE assays (Jones P, Prinsen M, Harbell J, personal communications) and from information in articles that describe the IRE test method in the literature.

2.2.8.2 *Positive Controls*

Because positive controls have not been traditionally employed in this test method, a defined range of response has not been described previously. However, the positive control substance should produce a response that is appropriate based on its historical classification as a severe irritant in the *in vivo* rabbit eye test. If adequate historical IRE test method data are not available for a particular positive control, pilot studies may have to be conducted to provide this information. Positive controls should produce the anticipated response in order to ensure that the test method is performing correctly.

2.2.8.3 *Benchmark Controls*

Benchmark substances may be useful in demonstrating that the test method is functioning properly for detecting the ocular irritancy potential of chemicals of a specific chemical class or a specific range of response, or for evaluating the relative irritancy potential of an ocular irritant. Therefore, benchmark substances should produce an irritation response that is within acceptable limits of historical data.

2.2.9 Nature of the Data to be Collected and the Methods Used for Data Collection

2.2.9.1 *Corneal Opacity and Area of Involvement*

Corneal opacity and area are typically evaluated visually and by use of a slit-lamp microscope. The most common scheme used to quantify corneal opacity and area of involvement uses the Draize scoring system (Draize et al. 1944). In this method, the severity of corneal cloudiness and the area of the cornea involved are graded and a score is assigned for each parameter using various scoring schemes. However, not all authors include an assessment of the area of corneal involvement in the IRE test method. The reason for this is unclear. The method described by Hackett and McDonald (1991) is an updated version of the original McDonald-Shadduck scoring system (McDonald and Shadduck (1977). The updated version of this scoring system is presented in **Table 2-1**. This method is similar in scoring to the Draize method, but is more specifically targeted to slit-lamp observations and describes corneal effects in terms of what is actually observed with the slit-lamp at each increasing level of corneal damage and score. Like the Draize assay, this method also allows for separate examination and scoring of the area of corneal involvement. Raw data are typically recorded in notebooks and electronically.

Table 2-1 Evaluation of Corneal Irritation[1]

Appearance	**Score**
Normal cornea. Appears with the slit-lamp as having a bright grey line on the epithelial surface and a bright grey appearance on the stroma.	0
Some loss of transparency. Only the anterior half of the stroma is involved as observed with an optical section of the slit-lamp. The underlying structures are clearly visible with diffuse illumination, although some cloudiness can be readily apparent with diffuse illumination.	1
Moderate loss of transparency. In addition to involving the anterior stroma, the cloudiness extends all the way to the endothelium. The stroma has lost its marble-like appearance and is homogenously white. With diffuse illumination, underlying structures are clearly visible.	2
Involvement of the entire thickness of the stroma with endothelium intact. With the optical section, the endothelial surface is still visible. However, with diffuse illumination the underlying structures are just visible.	3
Involvement of the entire thickness of the stroma with endothelium damaged. With the optical section cannot clearly visualize the endothelium. With diffuse illumination, the underlying structures cannot be seen.	4
Area	**Score**
Normal cornea with no area of cloudiness	0
1 to 25% area of stromal cloudiness	1
26 to 50% area of stromal cloudiness	2
51 to 75% area of stromal cloudiness	3
76 to 100% area of stromal cloudiness	4
Overall Corneal Opacity/Area	**Product Score**
Corneal Opacity x Area[2]	Maximum of 16

[1] From: Hackett and McDonald (1991)

[2]The overall corneal opacity score is the product of the corneal opacity score and the corneal area score. The product of individual scores of 1 and 4 (Product Score of 4) or 2 and 2 (Product Score of 4), for example, would each qualify for a severe irritant rating based on the overall corneal opacity/area score.

2.2.9.2 *Corneal Thickness and Calculation of Corneal Swelling*

Corneal thickness is measured quantitatively using an optical pachymeter (Attachment No. 1) for the Haag-Streit slit-lamp biomicroscope (e.g., Haag-Streit AG or equivalent, Liebefeld-

Bern, Switzerland), or by an ultrasonic pachymeter (e.g., DGH Technology Inc., Solana Beach, California, USA) (Jones P, Guest R, personal communication). The optical pachymeter measures degree of reflection or refraction from a normal light slit imposed into the corneal surface from the slit-lamp. The ultrasonic pachymeter measures the transit time of high -frequency sound pulses beamed into the eye and reflected off tissue interfaces where high impedance gradients are encountered (Jacobs and Martens 1988). From the known acoustic velocity of tissue, transit times may be converted to distance. Corneal swelling results induced by a variety of test substances using both measuring systems were found to be comparable (Jacobs and Martens 1988). One advantage of the ultrasonic pachymeter is that measurement of corneal thickness is usually possible even when corneal opacity has been induced, while this may not always be possible when using the optical pachymeter. Another advantage is that measurement of corneal thickness can be conducted at any position on the cornea, which is not possible with the optical pachymeter. Corneal thickness can therefore be measured at various positions on the cornea and a mean value obtained for each eye. However, if injury to the cornea is restricted to a small area, it may be more appropriate to measure corneal thickness at this position.

Corneal swelling is measured as the percent increase in thickness at each time point relative to the measurement at T0 (after equilibration, before treatment) as follows:

$$\left(\frac{\textit{corneal thickness at time } t \;-\; \textit{corneal thickness at time} = 0}{\textit{corneal thickness at time} = 0} \right) \times 100$$

The level of corneal swelling needed for a test substance to be considered an irritant varies in the literature depending on what type of instrument is used and the experience of the laboratory performing the study. Levels of irritancy may be assigned to a test substance based on 20 or 25% swelling over an entire range of time (e.g., 0 to 4 hours) (Balls et al. (1995; Gettings et al. 1996; Guerriero et al. 2004) or based on differences in swelling over time (CEC 1991; Prinsen and Köeter 1995). For example, 20-25% swelling in one hour may produce a higher irritancy rating than 20-25% in 5 hours. For the purposes of the analyses used in this BRD, a decision criterion for identification of a severe irritant is a corneal swelling value equal to or exceeding 25%.

2.2.9.3 Fluorescein Penetration

Fluorescein is used as an aid to further define epithelial damage (Norn 1971). In the IRE literature, a drop or two of a 1 to 2% solution of fluorescein sodium (sufficient to cover the cornea) is generally applied and left for several seconds followed by rinsing with isotonic saline. Fluorescein penetration is typically measured at 4 hours after application of the test substance. Most authors use a scoring system similar to that shown in **Table 2-2**. However, some protocols use a broader range of fluorescent terms (e.g., moderately bright, extremely bright) and scores from 0 to 5. Although the use of 1 to 2% fluorescein is recommended, variations in concentration of fluorescein from batch to batch may require some adjustment to achieve the desired corneal effect (Chambers W, personal communication).

Table 2-2 Fluorescein Penetration Scoring System[1]

Description	Scores
	Area/Intensity
Negligible – No staining.	0
Slight staining confined to small focal area. Some loss of detail in underlying structures with diffuse illumination.	1
Moderate staining confined to a small focal area. Some loss of detail in underlying structures on diffuse illumination.	2
Marked staining involving a larger portion of the cornea. Underlying structures are barely visible, but not completely obliterated with diffuse illumination	3
Extreme staining with no visibility of underlying structures.	4
Overall Fluorescent Area/Intensity	**Product Score**
Fluorescent Area x Intensity[2]	Maximum score of 16

[1]From: Hackett and McDonald (1991)

[2] Fluorescent area and intensity scores are determined individually and the overall Fluorescent Penetration Score is the product of both measurements. For example, an area score of 1 and intensity score of 4 would produce an overall score of 4, which meets the criteria for a severe irritant. An area score of 2 and intensity score of 2 would produce a product score of 4, which also exceeds the cut-off for a severe irritant, although each individual score did not.

2.2.9.4 Evaluation of Corneal Epithelial Integrity

As described by Guerriero et al. (2004), the cornea may also be observed macroscopically or microscopically using a slit-lamp to evaluate any injury to the epithelium. Stippling, pitting, mottling, sloughing, ulceration, or any other unusual effects on the epithelium are noted and reported.

Observations of the integrity of the corneal epithelium may be done visually or with a slit-lamp when laboratories are not equipped to perform routine histology. However, histology may also be used as an additional method for more precise evaluation of the integrity of the corneal epithelium.

2.2.9.5 *Overall Scoring System for Identification of a Severe Irritant*

Assignment of irritant classification or categories to test substances evaluated in the IRE test method varies from study to study. For example, in the CEC (1991) study, irritancy ratings of A, B, C, or D were assigned ranging from least to most severe that were arbitrarily based on the results from a combination of endpoints (corneal opacity, corneal swelling, and fluorescein retention). In the Balls et al. (1995) study, irritancy of a test substance was based on the mean corneal swelling measurement or corneal opacity score and then ranked accordingly. In the Gettings et al. (1996) study, an irritancy rating was assigned based on a predetermined cutoff using the percentages of corneal swelling. In the Guerriero et al. (2004) study, an overall scoring system for the identification of severe irritants was based on a test substance meeting or exceeding predetermined cutoff values in any of four ocular endpoints evaluated (corneal opacity and area, corneal swelling, fluorescein area and intensity, and integrity of the epithelium) (**Table 2-3**).

Table 2-3 Overall Scoring System for Corneal Damage and Irritation[1]

Ocular Parameter	Cutoff Value to Detect Severe Eye Irritants
Maximum Corneal Opacity[2] (Cloudiness x Area)	Greater than or equal to a score of 3
Maximum Fluorescein Uptake[3] (Intensity x Area)	Greater than or equal to a score of 4
Mean Corneal Swelling[4] 0.5 hours 1 hour 2 hours 3 hours 4 hours	Greater than or equal to 25%
Corneal Epithelial Observations[5]	Any pitting, mottling, stippling, sloughing, or ulceration of epithelium

[1] From: Guerriero et al. (2002)
[2] Represents maximum score obtained in three eyes
[3] Represents maximum score obtained in three eyes
[4] Represents mean swelling calculated for three eyes
[5] Represents information obtained for any single animal

2.2.10 Types of Media in Which Data are Stored

Although not specifically mentioned in published IRE protocols, it is reasonable to assume that data from studies performed in compliance with GLP guidelines (Balls et al. 1995; Gettings et al. 1996; Guerriero et al. 2004) were stored in a manner suitable for GLP compliant studies. It would seem appropriate that data from the IRE be stored and archived in a manner consistent with international GLP guidelines (OECD 1998; EPA 2003a, 2003b; FDA 2003). GLP guidelines are nationally and internationally recognized rules designed to produce high-quality laboratory records. These guidelines provide a standardized approach to report and archive laboratory data and records, and information about the test protocol, to ensure the integrity, reliability, and accountability of a study (EPA 2003a,b; FDA 2003). Materials that should be retained include, but are not limited to, raw data, documentation, protocols, final reports, records and reports of the maintenance and calibration of apparatus, validation documentation for computerized systems, the historical file of all Standard Operating Procedures, and environmental monitoring records. The archives should be organized and indexed so that retrieval of all information can be done expediently and conditions of storage should minimize deterioration of the documents. An individual should be identified as responsible for these data archives. All raw data from the experiment should be recorded using a system that meets institutional and GLP requirements.

2.2.11 Measures of Variability

Both numerical and descriptive data are generated using IRE. Variability of numerical data is typically assessed through calculation of the mean along with the standard deviation for each numerical endpoint. Other descriptive statistics (e.g., coefficient of variation or CV) may be used in the analysis of variability. These values allow for an assessment of the performance of the test conducted and whether the observed variability between replicates or groups of replicates is greater than would be considered acceptable. Descriptive data may also provide an additional subjective measurement of variability.

2.2.12 Statistical or Nonstatistical Methods Used to Analyze the Resulting Data

For statistical analysis, most studies rely on calculation of the mean and standard deviation of individual endpoint data produced by a test substance (CEC 1991; Balls et al. 1995; Gettings et al. 1996; Guerriero et al. 2004). Other descriptive statistical methods may also be used in analysis of *in vitro* data, or for comparison of the *in vitro* data with *in vivo* rabbit eye test data (e.g., coefficient of variation or CV, ANOVA, regression, rank correlation). As an example, in the EC/HO validation study, Balls et al. (1995) used regression analysis of MMAS scores with mean results from the IRE test data (i.e., corneal opacity or corneal swelling at two time points) and obtained Spearman's rank correlation coefficients to compare *in vivo* and *in vitro* test results for a set of 59 test substances.

2.2.13 Decision Criteria and the Basis for the Prediction Model Used to Classify a Test Chemical as a Severe Eye Irritant

Once the individual mean endpoint data are obtained, studies vary in the methods used to assign an irritation classification based on the degree of severity of the ocular response using composite endpoint data. The irritant classifications assigned may be either descriptive (e.g., nonirritant, mild, moderate or severe) (Cooper et al. 2001; Jones et al. 2001) or scaled rankings of increasing or decreasing irritancy (e.g., numerical [0 to 4] or alphabetical [A to D]), based on predetermined, arbitrary endpoint values and, are occasionally, time dependent (i.e., corneal swelling) (CEC 1991; Köeter and Prinsen 1995). In addition, identification of severe irritants may be based on meeting or exceeding predetermined cutoff values (Guerriero et al. 2004). These predetermined endpoint values may be selected on the basis of statistically derived decision criteria (Prediction Model) using biostatistical approaches such as discriminant analysis. These decision criteria (Prediction Model) may be targeted to either a general population or to a select population of test substances (e.g., surfactant-based products) that may vary with respect to prevalence of a particular level or range of severity of ocular irritation. The statistical methods used for the determination of these decision criteria are not usually provided in publications. An example of the decision criteria used by Guerriero et al. (2004) for the identification of severe ocular irritants is shown in **Table 2-3**.

2.2.14 Information and Data That Will be Included in the Study Report and Availability of Standard Forms for Data Collection and Submission

It would seem appropriate that the test report include the following information, if relevant to the conduct of the study:

Test and Control Substances

- Chemical name(s) such as the structural name used by the Chemical Abstracts Service (CAS), followed by other names, if known
- The CAS Registry Number (RN), if known
- Purity and composition of the substance or preparation (in percentage[s] by weight)
- Physicochemical properties such as physical state, volatility, pH, stability, chemical class, water solubility relevant to the conduct of the study
- Treatment of the test/control substances prior to testing, if applicable (e.g., warming, grinding)
- Stability, if known

Information Concerning the Sponsor and the Test Facility

- Name and address of the Sponsor
- Name and address of the test facility
- Name and address of the Study Director

Justification of the Test Method and Protocol Used

Test Method Integrity

- The procedure used to ensure the integrity (i.e., accuracy and reliability) of the test method over time (e.g., periodic testing of proficiency substances, use of historical negative and positive control data)

Criteria for an Acceptable Test

- Acceptable concurrent negative control ranges based on historical data
- Acceptable concurrent positive control ranges based on historical data
- If applicable, acceptable concurrent benchmark control ranges based on historical data

Test Conditions

- Experimental starting and completion dates
- Details of test procedure used
- Test concentration(s) used
- Description of any modifications of the test procedure
- Reference to historical data of the model (e.g., negative and positive controls, proficiency substances, benchmark substances)
- Description of evaluation criteria used

Results

- Tabulation of data from individual test samples (e.g., irritancy scores for the test substance and the positive, negative, and benchmark controls, including data from replicate repeat experiments as appropriate, and means ± the SDs for each experiment)

Description of Other Effects Observed

Discussion of the Results

Conclusion

A Quality Assurance Statement for Good Laboratory Practice (GLP)-Compliant Studies

- This statement indicates all inspections made during the study, and the dates any results were reported to the Study Director. This statement also serves to confirm that the final report reflects the raw data.

Additional reporting requirements for GLP-compliant studies are provided in the relevant guidelines (e.g., OECD 1998; EPA 2003a, 2003b; FDA 2003).

The use of a standardized scoring form may be appropriate (**Figure 2-1**).

2.3 Basis for Selection of the Test Method System

The IRE test method was designed as an *in vitro* assay to measure the direct effects of severe irritant test substances on the corneal tissue of a rabbit eye (Burton et al. 1972, 1981). The purpose of the IRE was to use rabbits that had been previously euthanized, thus preventing the pain and suffering associated with application of severe eye irritants to live animals. Furthermore, rabbits used for this assay could be obtained from other laboratories (e.g., rabbits used for skin irritancy testing or physiology studies in which the eyes were unaffected) minimizing the need for additional animals. For larger numbers of animals, the rabbits could be obtained from a local abattoir where the animals are bred and used as a food source. Therefore, the use of the IRE as a prescreen or as a replacement assay could reduce the number of animals used in the *in vivo* Draize rabbit eye test. Furthermore, the assay is refined by elimination of pain and suffering. As discussed in **Section 1.1.1**, the cornea is one of the main tissues targeted during accidental eye exposures. In addition, corneal effects are weighed heavily in the original *in vivo* ocular irritancy scoring systems (e.g., 80 out of a possible 110 points in the Draize eye test scoring system). Therefore, although conjunctival and iridal tissue are not available for use as endpoints in the IRE, most of the scoring capacity of the *in vivo* rabbit test method is maintained and other endpoints such as corneal swelling and fluorescein penetration may be incorporated. Furthermore, use of histopathology and/or confocal microscopy can be used to qualitatively or quantitatively assess the depth of penetration of ocular injury in the IRE.

2.4 Proprietary Components

The IRE assay does not employ any proprietary components.

2.5 Basis for the Number of Replicate and Repeat Experiments

The irritancy of a test substance is normally determined using three rabbit eyes if the distribution of the data is within acceptable limits (no single value exceeds a statistically acceptable deviation from the mean group response). The experiment is typically repeated when an individual data point is outside the range of random distribution as determined by appropriate statistical methods or failure to meet predetermined acceptance criteria, such as the various controls are outside the range of historically acceptable data, corneal swelling in a negative control eye exceeds 7 to 10%, and/or when equivocal results are obtained for the test substance (Jones P, Guest R, personal communication).

Figure 2-1 Sample Scoring Form for the IRE Test Method

Substance Name:	**Test Facility:**	**Study No.:**
CASRN:	**Chem Class:**	**Prod Class:**
Date:	**pH:**	
Color of Material:	**Liquid or Solid:**	**Viscosity of Material**

	Corneal Opacity/Area Score							Fluorescein Penetration Score	
	Hours after Application of Test Substance								
Eye #	**-1**	**0**	**0.5**	**1**	**2**	**3**	**4**	**0.5**	**2**
1									
2									
3									
4									
5									
6									

	Corneal Thickness (Instrument Units)							Additional Comments
Eye #	**-1**	**0**	**0.5**	**1**	**2**	**3**	**4**	
1								
2								
3								
4								
5								
6								

	Corneal Swelling (Percent of Time 0)							
Eye #	**-1**	**0**	**0.5**	**1**	**2**	**3**	**4**	
1								
2								
3								
4								
5								
6								

	Corneal Epithelium Damage Assessment (See Legend Below)							Epithelium Notations
Eye #	**-1**	**0**	**0.5**	**1**	**2**	**3**	**4**	
1								
2								
3								
4								
5								
6								

Corneal Epithelium Damage Assessment: Pitting = PT; Stippling = ST; Mottling = MT; Sloughing; SL; Ulceration = UL; Other = OT (Describe Other Effects in Epithelium Notations)

2.6 Compliance with Good Laboratory Practice

GLP compliant studies are performed in compliance with regulatory GLP Guidelines (OECD 1998; EPA 2003a, 2003b; FDA 2003) to increase confidence in the quality and reliability of the test data. For potential submission of data using these test methods, compliance with appropriate GLP guidelines would be required.

2.7 Study Acceptance Criteria

A test is acceptable if the positive control gives a score for each ocular test parameter that falls within two SDs of the current historical mean, which should be updated on a regular basis. The negative/solvent control responses should be nonirritating and corneal swelling in each of negative control eyes should not exceed 7 to 10% (Jones P, Guest R, personal communication). As described in previous sections in detail, the McDonald-Shadduck scoring methodology from Hackett and McDonald (1991) is used to assess corneal opacity and is based on a description of slit-lamp observations of corneal damage from the epithelium to the endothelium. The decision criteria were designed to identify severe versus nonsevere irritants and are based on exceeding maximal cut-off values in any of four ocular test parameters. The cut-off values are based on a maximum corneal opacity score (opacity x area), maximum fluorescein penetration score (area x intensity), maximal corneal swelling, and observation of the corneal epithelium in which any indication of epithelial damage (e.g., pitting, mottling, stippling, sloughing, or ulceration) constitutes a severe irritant classification. A positive control, in addition to the negative control, is needed to ensure that operation of the test system is within normal limits. Benchmark controls should also be used to demonstrate test method function within an applicability domain (e.g., surfactant formulations). Ideally, a set of quality reference substances should be used for validation efforts such as that used by Balls et al. (1995).