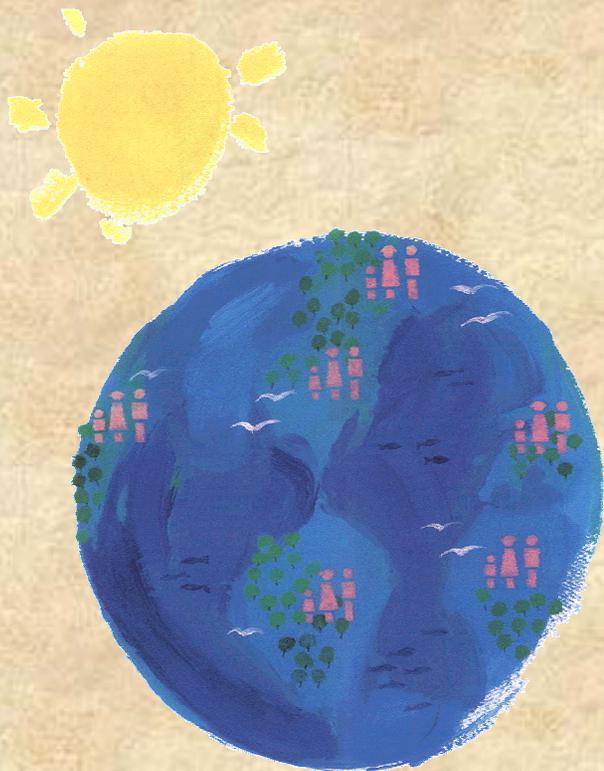


# NICEATM

National Toxicology Program Interagency  
Center for the Evaluation Of Alternative  
Toxicological Methods

# ICCVAM

Interagency Coordinating Committee  
on the Validation of Alternative  
Methods



## Preliminary Evaluation of the Underprediction Rate of the *In Vivo* Ocular Irritation Test Method Part I: Introduction

**William S. Stokes, D.V.M., D.A.C.L.A.M.**  
Director, NICEATM

**Scientific Advisory Committee on Alternative  
Toxicological Methods**

**October 20, 2004**

**Research Triangle Park, NC**



ICCVAM  
NICEATM



# Acknowledgements

---

## ICCVAM Ocular Toxicity Working Group (OTWG)

Kailash Gupta, CPSC

Harry Salem, DOD

Steve Hwang, DOT

Angela Auletta, EPA

Meta Bonner, EPA

Andrew Geller, EPA

Karen Hammernik, EPA (Co-Chair)

Karen Hicks, EPA

Marianne Lewis, EPA

Debbie McCall, EPA

Mark Perry, EPA

John Redden, EPA

Amy Rispin, EPA

William Stokes, NIEHS

Raymond Grissom, NIEHS

Surrender Ahir, NIOSH

Leonard Schechtman, FDA

Robert Bronaugh, FDA

Paul Brown, FDA

Wiley Chambers, FDA

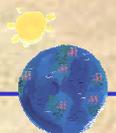
Zhou Chen, FDA

Abby Jacobs, FDA

Donnie Lowther, FDA

Jill Merrill, FDA (Co-Chair)

Josie Yang, FDA



# Acknowledgements

---

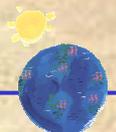
- **Dr. Joe Haseman, NIEHS/NTP (retired)**
  - Biostatistician
- **NICEATM Staff**
  - Dr. Neepa Choksi, ILS Inc.
  - Dr. Dave Allen, ILS Inc.
  - Dr. Ray Tice, ILS Inc.
- **Data Sources**
  - ECETOC
  - Cosmetic, Toiletry, and Fragrance Association
  - Food and Drug Administration
  - Toxic Substances Control Act Database
  - Yasuo Ohno, National Institute of Health Sciences, Japan
  - European Isocyanate & Polyol Producer's Association (ISOPA)
  - Access Business Group



# Outline

---

- **Introduction**
  - Background
  - Current Testing Procedures
  - Prior Analyses
  - Study Objectives
  - Database
  - Classification rules
  - Future Plans
  
- **Data Analysis**



# Background

---

- **Draize eye irritation test method**
  - Used since the 1940's to identify ocular corrosives and irritants
- **Serious eye damage**: *the production of tissue damage in the eye following application of a test substance to the anterior surface of the eye, which is **not fully reversible** within 21 days of application or results in serious physical decay of vision (UN 2003)*
  - Includes **irreversible effects (permanent damage)** from corrosive substances
  - Includes reversible severe effects
- **Eye irritation**: *the production of changes in the eye following the application of a test substance to the anterior surface of the eye, which are fully reversible within 21 days of application (UN 2003) which are not considered severe in nature*



# Background

---

- **2003 Globally Harmonized System of Classification and Labelling of Chemicals(GHS)**
  - *Tiered-testing approaches incorporating the use of valid and accepted in vitro methods for eye irritation should be considered*
- **1980s to present: Many non-animal alternative methods developed and proposed for assessing eye irritation**
  - ICCVAM evaluating 4 methods proposed for identifying severe or irreversible ocular irritants and corrosives:  
Ø ICE, IRE, BCOP, HETCAM
- **Estimates of the underprediction of the reference test method would assist with interpreting the usefulness and limitations of in vitro test methods**



# GHS Tiered-Testing Strategy

**Valid and accepted *in vitro* ocular corrosion test**

Severe damage → Category 1

Not a severe/irreversible eye irritant/corrosive

**Valid and accepted *in vitro* eye irritation test\***

Irritant → Category 2

Negative response or no data

***In vivo* dermal corrosion (in vivo or in vitro)**

Corrosive response → Not evaluated on the eye

Negative response or no data

**1 animal tested for eye irritation potential**

Severe/irreversible damage → Category 1

Negative response or no data

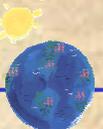
**1 or 2 additional animals tested**

Severe/irreversible damage → Category 1

**Not an eye irritant**

Irritant → Category 2

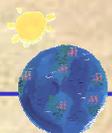
\* Must be capable of identifying false negative substances from an *in vitro* severe/irreversible ocular irritation test method



# Current GHS Testing for Eye Irritation/Corrosion

---

- **New Zealand White rabbits**
  - 6 rabbits per substance up until 1990s
  - current OECD (2003) guidelines:
    - Ø 1 to 3 animals; tested sequentially
    - Ø 0.1 mL or 0.1 g of test substance in conjunctival sac
- **Scoring at 24, 48, 72 hours; mean scores calculated for:**
  - Corneal Opacity
  - Iris
  - Conjunctiva: (1) redness and (2) swelling (chemosis)
- **Observations for 21 days to determine persistence or delayed effects**
  - Euthanasia permitted earlier if considered irreversible
- **Stopping rules:**
  - 1 animal with an irreversible/severe response
  - 2 animals with concordant irritant or non-irritant responses



# Ocular Irritation Scoring (1)

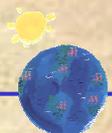
---

- **Cornea Opacity**

- ∅ 1 = Scattered or diffuse areas - details of iris clearly visible
- ∅ 2 = Easily discernible translucent areas, details of iris slightly obscured
- ∅ 3 = Opalescent areas, no details of iris visible, size of pupil barely discernible
- ∅ 4 = Opaque, iris invisible

- **Iris**

- ∅ 1 = Folds above normal, congestion, swelling, circumcorneal injection (any one or all of these), iris still reacting to light
- ∅ 2 = No reaction to light, hemorrhage, gross destruction



# Ocular Irritation Scoring (2)

---

- **Conjunctiva**

- **Redness**

- ∅ 1 = Vessels definitely injected above normal

- ∅ 2 = More diffuse, deeper crimson red, individual vessels not easily discernible

- ∅ 3 = Diffuse beefy red

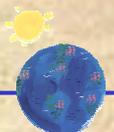
- **Chemosis**

- ∅ 1 = Any swelling above normal

- ∅ 2 = Obvious swelling with partial eversion of the lids

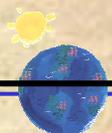
- ∅ 3 = Swelling with lids about half closed

- ∅ 4 = Swelling with lids about half closed to completely closed



# Irritancy Classification - GHS

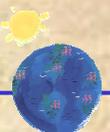
Category	Criteria for Classification (based on up to a 3 animal test)
<p><b>Category 1:</b> <i>Irreversible effects on the eye/ serious eye damage</i></p>	<ul style="list-style-type: none"> <li>- At least 1 animal with corneal opacity score of 4 at any time (NICEATM Cat 1A)</li> <li>- At least 1 animal with effects not expected to reverse or that do not fully reverse within 21 days (NICEATM Cat 1A)</li> <li>- At least 2 of 3 animals with mean corneal opacity score <math>\geq 3</math> and/or iritis score <math>\geq 1.5</math> ( NICEATM Cat 1B)</li> </ul>
<p><b>Category 2A</b> <i>Irritating to eyes</i></p>	<p>At least 2 of 3 animals with mean scores for one of more of the following:</p> <ul style="list-style-type: none"> <li>- corneal opacity <math>\geq 1</math></li> <li>-Iritis <math>\geq 1</math></li> <li>-Redness <math>\geq 2</math></li> <li>-Chemosis <math>\geq 2</math></li> </ul> <p><u>and</u> the effects fully reverse within 21 days</p>
<p><b>Category 2B</b> <i>Mildly irritating to eyes</i></p>	<p>When the effects listed for Cat2A fully reverse within 7 days</p>



# Retrospective Classification for the Data Analysis (1)

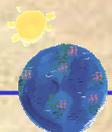
---

- **Some studies used 4 or 6 animals, rather than the 1-3 currently needed for GHS classification**
- **Rules were established for GHS classification of test outcome using 4-6 animals**



# Additional Classification Rules (2)

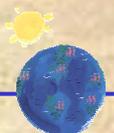
Category	Criteria Necessary for Classification
Category 1	$\geq 2$ of 6 animals have a NICEATM Cat 1A response $\geq 1$ of 6 animals has a NICEATM Cat 1A response <i>and</i> $\geq 1$ of 6 animals has a NICEATM Cat 1B response $\geq 4$ of 6 animals have NICEATM Cat 1B response
Category 2A	$\geq 4$ of 6 animals have Cat 2A responses
Category 2B	$\geq 4$ of 6 animals have Cat 2B responses
Nonirritant	$\geq 4$ of 6 animals have nonirritant response



# Rules Used for Analysis

---

- **Sampling from data based on sequential testing and stopping rules as follows:**
- **First Animal:**
  - If first animal is Cat 1A, then substance classified as Cat 1
  - If not, then test second animal
- **Second Animal:**
  - If second animal is Cat 1A, then substance classified as Cat 1
  - If lesions for first and second animals are same category, study is complete (i.e., 1B, 2A, 2B, NI)
  - If not, test third animal
- **Third Animal:**
  - If third animal is Cat 1A, then substance classified as Cat 1
  - If lesions for 2 of 3 animals are same category, then classified as that category (i.e., 1B, 2A, 2B, or non-irritant)
  - If 1 animal is Cat 2A, 1 animal is Cat 2B, and the third animal is Cat 1B or nonirritant, then the study is classified as Cat 2A
  - If all animals have different classifications (e.g., Cat 1B, NI, and 2A or 2B) then chemical is classified as “unknown”



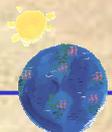
# Prior Analysis of the Reproducibility of the Rabbit Ocular Irritation Test

---

- **Weil and Scala (1971)**
  - Evaluated the reproducibility of the Draize rabbit eye test method among 24 laboratories for 10 substances
- **Conclusions**
  - Moderate intra-laboratory reproducibility
  - Low inter-laboratory reproducibility
  - Primary reasons for the low inter-laboratory reproducibility attributed to the subjective nature of the visual observations
- **Limitations**
  - Good Laboratory Practice (GLP) Guidelines had not yet been established (impact unknown)
  - Prior to publication of EPA Ocular Effects Atlas
  - Substances classified according to FHSA classification system
  - Not possible to apply GHS, EPA, or EU classification systems
    - ∅ Individual animal data were not available

---

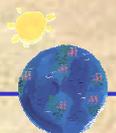
Weil CS, Scala RA. 1971. Study of intra- and interlaboratory variability in the results of rabbit eye and skin irritation tests. *Toxicol. App. Pharmacol.* 19:276-360.



# NICEATM Analysis

---

- **Study Objective**
  - Evaluate the likelihood of underpredicting an ocular corrosive or severely irritating substance as a nonsevere irritant/nonirritant in the current rabbit eye irritation test, according to GHS
- Data may assist in establishing an acceptable false-negative rate for severe/irreversible effects for *in vitro* test methods proposed as complete replacements for the rabbit eye test
  - i.e., those tests where no *in vivo* confirmation would be performed



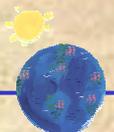
# *In Vivo* Ocular Database

---

- Data compiled from studies using the rabbit ocular test method protocol (e.g., EPA, OECD TG 405)
- Data requested from public and Federal agencies via an FR notice<sup>1</sup> in March 2004
- Current database: 505 studies on 448 substances
  - 79 formulations and 369 chemicals
- Many of the substances in the database are commercial products with unknown formulations and chemical composition

---

<sup>1</sup>FR Notice (Vol. 69, No. 57, pp. 13859-13861, March 24, 2004). Available: <http://iccvam.niehs.nih.gov/docs/FR/6842067.htm>



# Data Considered for Analysis

---

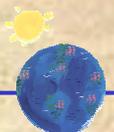
<b>Data Source</b>	<b>Total Studies (# of Chemicals)</b>
<b>Access Business Group</b>	<b>14 (14)</b>
<b>CTFA</b>	<b>56 (46)</b>
<b>ECETOC</b>	<b>149 (132)</b>
<b>FDA</b>	<b>168 (168)</b>
<b>ISOPA</b>	<b>8 (7)</b>
<b>EPA TSCA</b>	<b>48 (42)</b>
<b>NIHS-Y.Ohno</b>	<b>62 (39)</b>
<b>Total</b>	<b>505 (448)</b>



# Analysis (1)

---

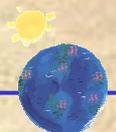
- **Assessed the underprediction rate of the *in vivo* ocular test method using the GHS classification system**
  - Including stopping rules and sequential testing (1-3 animals)
- **An initial preliminary analysis conducted on the ECETOC database**
  - 143 studies: 123 chemicals, 31 severe irritants
  - Some studies from ECETOC not used because definitive GHS classifications could not be made (e.g., study ended before 21 days with effects still present in animals)
  - ECETOC studies were in accordance with OECD TG 405 and GLP guidelines



# Analysis (2)

---

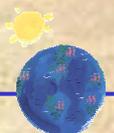
- **A second preliminary analysis conducted using the current database: ECETOC plus submitted data**
  - **464 studies: 305 chemicals, 115 severe irritants**
  - **Some studies not used because definitive GHS classifications could not be made (e.g., study ended before 21 days with effects still present in animals)**
  - **Studies performed consistent with OECD TG 405**
  - **Most studies performed in accordance with GLP guidelines**



# Plans for Future Analyses

---

- Continue to seek high quality data to add to the database, and perform reanalysis
  - *Federal Register* request for existing animal data (March, 2004)
  - ECVAM additional call for data
  - EPA TSCATS database
    - ∅ Currently collaborating with EPA OPPTS to obtain microfiche reports for ~2400 commercially available chemicals that may have ocular test results
    - ∅ To date; 638 reports received, 138 reports reviewed, 39 reports (51 substances) with suitable data identified
  - Animal data requested from authors of publications evaluating the 6-animal Draize test
- Evaluate interlaboratory performance for substances tested in multiple laboratories when sufficient high-quality data are available
- Evaluate the relative underprediction rate for each GHS decision criteria used to classify a substance as an ocular corrosive or severe irritant (e.g., severity, persistence, severity and persistence, tissue type)



# NICEATM

National Toxicology Program Interagency  
Center for the Evaluation Of Alternative  
Toxicological Methods

# ICCVAM

Interagency Coordinating Committee  
on the Validation of Alternative  
Methods



## Preliminary Evaluation of the Underprediction Rate of *In Vivo* Ocular Irritation Test Method Part II: Data Analysis

Joseph Haseman, Ph.D.

Scientific Advisory Committee on Alternative  
Toxicological Methods

October 20, 2004

Research Triangle Park, NC



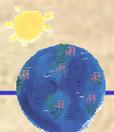
ICCVAM  
NICEATM



# General Approach

---

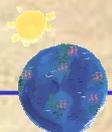
- **The underprediction rate depends on**
  - the distribution of animal responses for substances assigned to a specific classification category
  - the strategy that is used to assign a test substance to a classification category
- **Substances tested as multiple doses are assumed to be Category 1 at and above the minimum dose producing a Category 1 response**



# Assumptions

---

- **Assumption 1: Homogeneity of response within a given category**
  - It is assumed that animals have the same pattern of response for all chemicals within a given classification category
  - Requires only one calculation but may underestimate the underprediction rate
- **Assumption 2 - Heterogeneity of response within a given category**
  - It is assumed that animals have a different pattern of response for all chemicals within a given classification category
  - Leads to higher misclassification rates than Assumption 1, but may overestimate the underprediction rate
- **The true underprediction rate is likely between these two estimates, and probably closer to the Assumption 2 estimate**



# Distribution of Animal Responses per GHS Category 1 Studies

<b>Animal</b> \ <b>Study</b>	<b>ECETOC Category 1 Classification<sup>a</sup></b>	<b>Total Category 1 Classification<sup>b</sup></b>
<b>Category 1A</b>	<b>55 (49%)</b>	<b>312 (63%)</b>
<b>Category 1B</b>	<b>19 (17%)</b>	<b>31 (6%)</b>
<b>Category 2A</b>	<b>19 (17%)</b>	<b>66 (13%)</b>
<b>Category 2B</b>	<b>18 (16%)</b>	<b>68 (14%)</b>
<b>Nonirritant</b>	<b>1 (1%)</b>	<b>18 (4%)</b>
<b>Total</b>	<b>112</b>	<b>495</b>

<sup>a</sup> Animal responses are for 31 Category 1 substances

<sup>b</sup> Animal responses are for 115 Category 1 substances

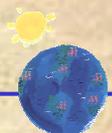


# Calculation of Likelihood that a Category 1 Substance will be Classified as Category 2A (ECETOC Database, Assumption 1)

Potential Outcome	Probability Calculation	Contribution to Underprediction Rate
2A-2A	$(19/112) \times (19/112)$	0.0288
2A-X-2A <sup>a</sup>	$(19/112) \times (38/112) \times (19/112)$	0.0098
X-2A-2A	$(19/112) \times (38/112) \times (19/112)$	0.0098
2A-2B-Nonirritant	$[(19/112) \times (18/112) \times (1/112)] \times 6$	0.0015
2A-2B-Category 1B <sup>b</sup>	$[(19/112) \times (18/112) \times (19/112)] \times 6$	0.0278
<b>Total</b>		<b>0.0777 (7.77%)</b>

<sup>a</sup>X refers to an outcome of either Category 2B, nonirritant, or Category 1B

<sup>b</sup>Refers to animal classified based on severity of opacity or iris effects



# Assumption 1: Other Underprediction Rates (ECETOC Database)

---

- The likelihood of a Category 1 chemical being misclassified as Category 2B is the sum of two outcomes:  $[(18/112) \times (18/112)] + [(18/112) \times (39/112) \times (18/112) \times 2] = 0.0438$  (4.38%)
- The likelihood of Category 1 chemical being misclassified as a nonirritant is the sum of two outcomes:  $[(1/112) \times (1/112)] + [(1/112) \times (56/112) \times (1/112) \times 2] = 0.0002$  (0.02%)
- The likelihood of Category 1 chemical not clearly classified based on the results is the sum of two outcomes:  $[(19/112) \times (1/112) \times (18/112) \times 6] + [(19/112) \times (1/112) \times (19/112) \times 6] = 0.003$  (0.3%)

The total probability for underprediction is =  
 $(7.77\% + 4.38\% + 0.02\% + 0.3\%) = 12.47\%$



## Assumption 2: Heterogeneous Response (ECETOC Database)

---

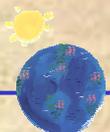
- **Distribution of animal responses is determined for each test substance in Category 1**
- **Estimated underprediction rate is calculated for each Category 1 test substance**
- **These estimated underprediction rates are averaged to produce an overall underprediction rate**



## Assumption 2: Example Calculation (ECETOC Database)

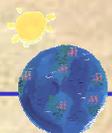
---

- Suppose a Category 1 irritant has 4 animals classified as Category 1A and 2 animals classified as Category 2A
- The likelihood of this irritant being underpredicted to be Category 2A is  $(2/6) \times (2/6) = 0.1111$  (11.1%)
- The likelihood of other underprediction classifications for this irritant is estimated to be zero
- Similar calculations are carried out for the other Category 1 irritants and the rates averaged to produce an overall estimated underprediction rate



# Estimated Underprediction Rate

GHS Classification Underprediction	ECETOC Database		Total Database	
	Assumption 1	Assumption 2	Assumption 1	Assumption 2
as Category 2A	7.77%	9.03%	3.71%	6.55%
as Category 2B	4.38%	10.8%	2.77%	8.59%
as Nonirritant	0.02%	0.4%	0.22%	2.26%
as Undetermined	0.30%	0%	0.37%	0.07%
<b>Total</b>	<b>12.47%</b>	<b>20.23%</b>	<b>7.07%</b>	<b>17.47%</b>



# Conclusions

---

- **Within the limits of the assumptions, the overall underprediction rate for a Category 1 substance ranged from 7.1% to 20.23% for the GHS classification scheme**
- **If downturns in irritation responses are biologically plausible, then these rates would be smaller**
- **The underprediction rate for misclassifying a Category 1 substance**
  - **as Category 2A ranges from 3.7% to 9.0%**
  - **as Category 2B ranges from 2.8% to 10.8%**
- **A much lower underprediction rate was calculated for misclassifying a Category 1 substance as a nonirritant (0%-2.3%)**

