Estimated Likelihood For Under- and Over-Classification for a Sequential Draize Rabbit Eye Test

NY Choksi[1,2], JK Haseman[3], DG Allen[1,2], RR Tice[2], WS Stokes[2].

[1]Integrated Laboratory Systems, Inc., Research Triangle Park, North Carolina, USA, [2]National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM), National Institute of Environmental Health Sciences, NIH/DHHS, Research Triangle Park, North Carolina, USA, [3]Consultant, Raleigh, North Carolina, USA

Abstract

Recently, the ICCVAM evaluated four *in vitro* test methods for their ability to identify ocular corrosives/severe irritants. Ideally, these evaluations would have included an assessment of the ability of the Draize rabbit eye test to correctly identify human ocular corrosives/severe irritants, but comparative human and rabbit data are not available. Therefore, using an *in vivo* rabbit eye test database of 180 severe irritant studies, the estimated likelihood of underpredicting a corrosive/severely irritating response as a nonsevere response was determined to evaluate the performance of the Draize test according to the UN Globally Harmonized ocular hazard classification system. The distribution of individual rabbit responses within each severity class was used to estimate the likelihood of under- and/or over-classification for a sequential 1 to 3 rabbit testing strategy. Based on three assumptions about the variability in individual rabbit responses among substances within each classification category, the estimated likelihood for underclassifying corrosives/severe irritants as nonsevere irritants or nonirritants ranged from 4% to 14%. The database of severe irritants showed strong evidence of heterogeneity in response among substances, and when this heterogeneity was taken into account, the estimated underclassification likelihood for the *in vivo* rabbit eye test method ranged from 11% to 14%. Analyses based on test substance physical form suggested that the underclassification likelihoods for solids were lower than liquids (2.8%-8.3% vs. 5.7%-16.2%, respectively), although these differences were not statistically significant. Estimated likelihoods for underclassification were higher when corrosive/severe irritant classification was based solely on persistent lesions present at observation day 21. By chemical class, carboxylic acids had the highest underclassification likelihood (37-39%). Likelihoods for overclassification of substances as corrosive/severe irritants, based on 610 nonsevere and nonirritant studies, were estimated to be 6.5%-7.6% for Category 2A, 3.2%-5.2% for Category 2B, and 0.3%-0.5% for nonirritant substances. With the relative distribution of studies in the nonsevere/nonirritant classification was taken into account the overall likelihood of overclassifying these substances, as corrosives/severe irritants, was 1.2% to 1.7%.
Supported by NIEHS contract N01-ES 35504.

Introduction

Accidental eye injury is the leading cause of visual impairment in the United States (BLS 2004). In 2003, eye injuries from chemicals and their products (6,080) accounted for 16% of all eye injuries (36,940) reported as the cause of Days Away From Work for employees. The ocular irritation or corrosion potential of substances to which humans may be exposed has been evaluated since 1944 using the Draize rabbit eye test (Draize et al. 1944). In recent years, there have been widespread efforts to develop and validate *in vitro* alternatives that might reduce or replace the use of animals for ocular irritancy assessments. ICCVAM is evaluating four *in vitro* test methods for their ability to detect ocular corrosives/severe irritants in a weight-of-evidence tiered testing strategy. Ideally, this evaluation would assess the ability of each of the four *in vitro* test method versus the Draize eye test to correctly predict human ocular toxicity. However, the lack of appropriate human data only allows for a determination of how well the alternative *in vitro* assays predict the rabbit response. In evaluating the performance of alternative assays, information on the reliability of the Draize eye test would be useful but the paucity of repeat test data precludes a formal evaluation. However, Draize eye test results can be used to estimate the likelihood of underclassifying an ocular corrosives/severe irritant or overclassifying a nonsevere irritant/nonirritating substance according to the sequential testing scheme implemented by the United Nations (UN) Globally Harmonized System (GHS) for Classification and Labeling of Chemicals (UN 2003).

Individual rabbit data from Draize eye tests using from 1 to 6 rabbits was obtained for 1005 studies from different sources (e.g., publications, U.S. Federal regulatory agencies, individual scientists, private and public organizations/companies). Studies were conducted in accordance with the current or previous version of Organisation for Economic Co-operation and Development (OECD) Test Guideline 405 (OECD 2002). Ocular irritation categories were assigned based on the GHS ocular hazard classification system (UN 2003) (**Table 1**). As some studies used more than three rabbits, rather than the 1 to 3 rabbits currently needed for the GHS classification system, rules were established by NICEATM to classify the substances tested in these studies according to the 3-animal GHS classification rules (**Table 2**).

Based on the classification rules described in **Table 2**, the distribution of *in vivo* rabbit eye test results in the NICEATM database, by GHS ocular hazard classification, was:

- 790 total studies classified[1]
- 180 studies classified as Category 1
- 61 studies classified as Category 2A
- 59 studies classified as Category 2B
- 490 studies classified as Nonirritant

[1]215 studies could not be assigned a GHS classification due to minimal study criteria not being met (e.g., one or more rabbits exhibited an ocular response on the last observation day but the observation period was not carried out to 21 days)

Table 1. GHS Ocular Hazard Classification System

Category	Criteria for Classification (based on a 1 to 3 animal test)
Category 1: Irreversible or serious eye damage	• At least 1 animal with a corneal opacity score of 4 at any time (NICEATM Cat 1A[1]) • At least 1 animal with effects not expected to reverse or that do not fully reverse within 21 days (NICEATM Cat 1A[1]) • At least 2 animals with a mean corneal opacity score ≥ 3 and/or an iritis score ≥1.5 (NICEATM Cat 1B[1])
Category 2A: Irritating to eyes	• At least 2 of 3 animals with mean scores for one of more of the following: – -Corneal opacity ≥1 – -Iritis ≥ 1 – -Redness ≥ 2 – -Chemosis ≥ 2 and the effects fully reverse within 21 days
Category 2B: Mildly irritating to eyes	When the effects listed for Category 2A fully reverse within 7 days
Nonirritating	When the substance does not meet the criteria for Category 1, 2A, or 2B

GHS = United Nations Globally Harmonized System for Classification and Labeling of Chemicals (UN 2003).
[1]NICEATM categories were developed internally for the purpose of this study. These are not formal classifications and are not part of the GHS classification system (UN 2003).

Table 2. Retrospective Classification of *In Vivo* Rabbit Eye Test Results

Category	Criteria Necessary for Classification
Category 1	≥ 1 of 3 (≥ 2 of 6) rabbits has a NICEATM Cat 1A response ≥ 1of 6 rabbits has a NICEATM Cat 1A response and ≥ 1 of 6 rabbits has a NICEATM Cat 1B response ≥ 2 of 3 (≥ 4 of 6) rabbits have a NICEATM Cat 1B response
Category 2A	≥ 2 of 3 (≥ 4 of 6) rabbits have a Cat 2A response 1 of 3 (2 of 6) rabbits have a Cat 2A response and 1 of 3 (2 of 6) rabbits have a Cat 2B response
Category 2B	≥ 2 of 3 (≥ 4 of 6) rabbits have a Cat 2B response
Nonirritant	The substance cannot be classified as GHS Category 1, 2A, or 2B, no rabbits were shown to have a NICEATM Cat 1A or 1B response, and ≥ 2 of 3 (≥ 4 of 6) rabbits have a nonirritant response

Cat = Category; GHS = United Nations Globally Harmonized System for Classification and Labeling of Chemicals (UN 2003).

Underclassification Analysis

Classification Rules

The estimated underclassification rates for the *in vivo* rabbit eye test was based on GHS sequential testing and the following rules:

First Rabbit:

- If the first rabbit is Category 1A, then the substance is classified as Category 1
- If not, then test the second rabbit

Second Rabbit:

- If the second rabbit is a Category 1A, then the substance is classified as Category 1
- If lesions for the first and second rabbits are in the same classification category, then the substance is classified in that category (i.e., 1, 2A, 2B, or nonirritant)
- If neither of the above apply, then test a third rabbit

Third Rabbit:

- If the third rabbit is Category 1A, then the substance is classified as Category 1
- If lesions for 2 of the 3 rabbits are in the same classification category, then the substance is classified in that category (i.e., 1, 2A, 2B, or nonirritant)
- If 1 rabbit is Category 2A, 1 rabbit is Category 2B, and the third rabbit is Category 1B or nonirritant, then the substance is classified as Category 2A
- If all rabbits have different classifications (e.g., Cat 1B, NI, and 2A or 2B), then the chemical is classified as "variable responder"

Calculations Conducted

Calculation 1: Homogeneity of response within a given category

- 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 underclassification rate

Calculation 2: Heterogeneity of response within a given category

- Assumed that animals have a different pattern of response for all chemicals within a given classification category
- Leads to higher misclassification rates than Calculation 1, but may overestimate the underclassification rate

Calculation 3: Homogeneity/heterogeneity of response combination

- Assumed that animals have similar pattern of response for chemicals within three response categories
  - Strong responders. Category 1 substances that produce either a Cat 1A or 1B response in all tested rabbits
  - Moderate responders. Category 1 substances that produced a Cat 1A or 1B response in at least 50% but not 100% tested
  - Weak responders. Category 1 substances that produced a nonsevere (i.e., Cat 2A, 2B) or nonirritant response in more than half of the rabbits tested

The underclassification rate based on Calculation 3 is likely the best estimate of the actual underclassification rate.

Underclassification Analysis

Limitations

- No human data to confirm the irritancy classifications of the tested substances
- Using the *in vivo* rabbit eye test method as the basis for assigning substances to ocular irritancy categories and then evaluating the performance of that same test method for under- and over-classifying substances to those categories
- The reproducibility of the *in vivo* rabbit eye test results from study to study cannot be assessed since there is little repeat data for substances tested at the same concentration within the database
- Dose response curves for substances tested at multiple concentrations were assumed to saturate or increase with increasing dose (i.e., once a severe response was detected, all higher dose levels were assumed to induce a severe response regardless of the actual outcome); however, there are data suggesting that, for some substances, there is less ocular damage when tested at high compared to low levels

Calculation 1: Homogeneity of Response

Table 3 provides the distribution of rabbit responses for those substances that were classified as Category 1 based on the classification rules described above.

Table 3. Distribution of Individual Rabbit Responses in 180 GHS Category 1 Studies

GHS Ocular Hazard Classification	Number of Rabbits
Category 1A	499 (70%)
Category 1B	44 (6%)
Category 2A	96 (13%)
Category 2B	58 (8%)
Nonirritant	20 (3%)
Total	717 (100%)

Tables 4-7 provide the calculations used to estimate the underclassification rate of Category 1 substances as Category 2A, Category 2B, Nonirritant, and Variable Responder.

Table 4. Likelihood that a Category 1 Substance will be Classified as Category 2A

Potential Outcome	Probability Calculation	Contribution to Underclassification Rate
2A-2A	(96/717)x(96/717)	0.01793
(2A-X[a]-2A)	[(96/717)x(122/717)x(96/717)]x2	0.00610
(2A-2B-Nonirritant)	[(96/717)x(58/717)x(20/717)]x6	0.00181
(2A-2B-Category 1B [b])	[(96/717)x(58/717)x(44/717)]x6	0.00399
Total		0.0298 **(2.98%)**

[a]X refers to an outcome of either Category 1B, Category 2B, or nonirritant
[b]Refers to a rabbit classified based on severity of opacity or iris effects

Table 5. Likelihood that a Category 1 Substance will be Classified as Category 2B

Potential Outcome	Probability Calculation	Contribution to Underclassification Rate
2B-2B	(58/717)x(58/717)	0.00654
(2B-2B-X[a])	[(58/717)x(58/717)x(160/717)]x2	0.00292
Total		0.0095 **(0.95%)**

[a]X refers to an outcome of either Category 1B, Category 2A, or nonirritant

Table 6. Likelihood that a Category 1 Substance will be Classified as Nonirritant

Potential Outcome	Probability Calculation	Contribution to Underclassification Rate
NI-NI	(20/717)x(20/717)	0.0008
(NI-NI-X[a])	(20/717)x(20/717)x(196/717)x2	0.0004
Total		0.0012 **(0.12%)**

NI = Nonirritant
[a]X refers to an outcome of either Category 1B, Category 2A, or Category 2B

Table 7. Likelihood that a Category 1 Substance will be Classified as a Variable Responder

Potential Outcome	Probability Calculation	Contribution to Underclassification Rate
1B-2A-NI	(44/717)x(96/717)x(20/717)x6	0.0014
1B-2B-NI	(44/717)x(58/717)x(20/717)x6	0.0008
Total		0.0022 **(0.22%)**

NI = Nonirritant.

The total probability for underclassification using this approach is: (2.98% + 0.95% + 0.12% + 0.22%) = 4.27%

Calculation 2: Heterogeneity of Response

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

Example Calculation:

- A Category 1 irritant has 4 animals classified as Category 1A and 2 animals classified as Category 2A
- The likelihood of this substance being underclassified to be Category 2A is (2/6)x(2/6) = 0.1111 (11.1%)
- The likelihood of other underclassifications for this irritant is estimated to be 0%
- Similar calculations are carried out for the other Category 1 irritants and the rates averaged to produce an overall estimated underclassification rate (**Table 8**)

Table 8. Likelihood that a Category 1 Substance will be Underclassified Using Calculation 2

GHS Ocular Hazard Classification	Estimated Underclassification Rate
Category 2A	7.55%
Category 2B	4.31%
Nonirritant	1.45%
Variable Responder	0.00%
Total	**13.31%**

GHS = United Nations Globally Harmonized System for Classification and Labeling of Chemicals (UN 2003).

Calculation 3: Homogeneity/Heterogeneity of Response

- Substances with similar patterns of response are combined into three categories (strong responders, moderate responders, and weak responders)
- For each responder group, the underclassification rate was determined using the homogeneity approach (**Table 9**)
- The overall average weights the calculation by the relative proportion of strong, moderate, and weak responders in the database in that GHS category (**Table 10**)

Table 9. Distribution of Individual Rabbit Responses per Three Subgroups of GHS Category 1 Studies

GHS Category	All (n=180)	Strong Responders (n=107)	Moderate Responders (n=32)	Weak Responders (n=41)
Category 1A	499	347	104	48
Category 1B	44	32	6	6
Category 2A	96	0	32	64
Category 2B	58	0	16	42
Nonirritant	20	0	4	16
Total	717	379	162	176

GHS = United Nations Globally Harmonized System for Classification and Labeling of Chemicals (UN 2003).

Table 10. Likelihood that a Category 1 Substance will be Underclassified Using Calculation 3

GHS Category	Strong Responders (n=107)	Moderate Responders (n=32)	Weak Responders (n=41)	Weighted Overall Average (n=180)
Category 2A	0.0000	0.0587	0.2934	0.0773
Category 2B	0.0000	0.0149	0.1126	0.0283
Nonirritant	0.0000	0.0010	0.0188	0.0045
Variable Responder	0.0000	0.0016	0.0112	0.0028
Total	0.0000 (0.00%)	0.0762 (7.62%)	0.4360 (43.60%)	0.1129 (11.29%)

GHS = United Nations Globally Harmonized System for Classification and Labeling of Chemicals (UN 2003).

Underclassification Analysis Based on Data Subsets

Underclassification rates also were calculated based on specified physical and chemical classes

- Physical Form (Solids/Liquids)
- Criteria Used for GHS Classification
  - Criteria 1: Substances that produced a persistent lesion through 21 days in at least one of three rabbits
  - Criteria 2: Substances that produced a positive response in at least two of three rabbits
  - Criteria 3: Substances that produced a persistent lesion through 21 days in at least one of three rabbits and produced a positive response in at least two of three rabbits
  - Criteria 4: Substances that produced a corneal opacity of 4 at any time
- Chemical Class

Estimated underclassification rates for physical form and GHS classification criteria were evaluated using Calculations 1-3 (as described above). Results are provided in **Tables 11 and 12**, respectively.

Estimated underclassification rates for chemical classes with at least 25 animals tested were evaluated using Calculations 2 and 3 (as described above). Results are provided in **Table 13**.

Table 11. Estimated Underclassification Rates for GHS Category 1 Substances Based on Physical Form

Liquids/Gels (100 studies, 417 animals)			Solids (30 studies, 122 animals)		
Calculation 1	Calculation 2	Calculation 3	Calculation 1	Calculation 2	Calculation 3
5.67%	16.23%	13.89%	2.75%	8.31%	7.51%

GHS = United Nations Globally Harmonized System for Classification and Labeling of Chemicals (UN 2003).

Table 12. Estimated Underclassification Rates for GHS Category 1 Substances Based on Classification Criteria

Criteria 1 (%) (61 studies, 246 animals)			Criteria 2 (%) (7 studies, 31 animals)			Criteria 3 (%) (3 studies, 12 animals)			Criteria 4 (%) (88 studies, 360 animals)		
Calc 1	Calc 2	Calc 3	Calc 1	Calc 2	Calc 3	Calc 1	Calc 2	Calc 3	Calc 1	Calc 2	Calc 3
8.52%	17.62%	14.52%	1.08%	5.49%	4.84%	0.00%	0.00%	0.00%	0.61%	5.30%	3.91%

Calc = Calculation; GHS = United Nations Globally Harmonized System for Classification and Labeling of Chemicals (UN 2003).

Table 13. Estimated Underclassification Rates for GHS Category 1 Substances Based on Chemical Class

Chemical/ Product Class	Total Studies	Number of Animals	Calculation 2	Calculation 3
Formulations	62	306	13.75%	10.81%
Alcohol	23	89	24.34%	22.93%
Amines	22	81	6.73%	6.49%
Organic Salts	13	47	27.21%	24.20%
Heterocyclic	12	37	10.80%	9.25%
Ethers	11	39	12.37%	11.22%
Carboxylic Acid	11	35	39.35%	36.76%
Esters	10	37	16.30%	14.82%
Onium Compounds	9	29	0.00%	0.00%
Carboxylic Acid Salts	8	31	16.23%	13.89%
Phenols	5	25	0.00%	0.00%

GHS = United Nations Globally Harmonized System for Classification and Labeling of Chemicals (UN 2003).

Overclassification Analysis

Classification Rules

- Same approaches used for estimating the overclassification rate for ocular corrosive or severe irritant, used to estimate the likelihood of overclassifying a Category 2A, 2B or nonirritant as a Category 1 substance

Calculations Conducted

- As there was less evidence of heterogeneity of response and no clearly identifiable sub-groupings showing distinct patterns of response, Calculation 3 was not used
- The distribution of individual rabbit responses for studies resulting in a Category 2A, 2B, or nonirritant classification is provided in **Table 14**; the estimated underclassification rates in **Table 15**.

Table 14. Distribution of Individual Rabbit Responses by GHS Ocular Hazard Classification

GHS Category	Category 2A (61 studies)	Category 2B (59 studies)	Nonirritants (490 studies)
Category 1A	8	5	5
Category 1B	7	1	0
Category 2A	183	14	13
Category 2B	38	163	47
Nonirritant	21	49	2009
Total # of Rabbits	257	232	2074

GHS = United Nations Globally Harmonized System for Classification and Labeling of Chemicals (UN 2003).

Table 15. Estimated Overclassification Rates of GHS Category 2A, 2B or Nonirritants as Category 1 Substances

GHS Overclassification	Calculation 1 Homogeneity of Response	Calculation 2 Heterogeneity of Response
Category 2A as Category 1	7.60%	6.57%
Category 2B as Category 1	5.17%	3.20%
Nonirritant as Category 1	0.50%	0.33%
Weighted Average	1.66%	1.23%

GHS = United Nations Globally Harmonized System for Classification and Labeling of Chemicals (UN 2003).

Conclusions

Underclassification

- The overall estimated underclassification rate for a Category 1 substance ranged from 4% to 14% (**Table 16**)
  - The most reasonable estimate of the overall underclassification likelihood for the *in vivo* rabbit eye test method is between 11% and 14%, since there is strong evidence of heterogeneity among Category 1 irritants in the database
- For the subsets based on physical and chemical classes:
  - The underclassification likelihood for liquids/gels was slightly higher than for solids, but the difference was not statistically significant
  - Underclassification likelihoods appear to vary by chemical class
- For the subset based on criteria used for GHS classification, Criterion 1 (i.e., substances that produced a persistent lesion through 21 days in at least one of three rabbits only) was shown to have a higher underclassification likelihood than Criteria 2-4
- A limitation of this analysis is that the *in vivo* rabbit eye test method is used as the basis for assigning substances to ocular irritancy categories. Then, that same test method is evaluated for under- and over-classifying substances to those categories. An alternative approach, which did not require an *a priori* assignment of substances to classification categories, also was utilized. Modifying an approach described in Springer et al. (1993), which does not require prior classification of the substances, yielded a estimated under-classification rate of 13.65% and over-classification rate of 2.59%; results which are consistent with the results of the analysis described in this poster.

Table 16. Estimated Overall Underclassification Rates of GHS Category 1 Substances to GHS Category 2A, 2B, Nonirritant, or Variable Responder Substances

GHS Under-classification	Calculation 1 Homogeneity of Response	Calculation 2 Heterogeneity of Response	Calculation 3 Homogeneity/ Heterogeneity of Response
as Category 2A	2.98%	7.55%	7.73%
as Category 2B	0.95%	4.31%	2.83%
as Nonirritant	0.12%	1.45%	0.45%
as Variable Responder	0.22%	0.00%	0.28%
Total	4.27%	13.31%	11.29%

GHS = United Nations Globally Harmonized System for Classification and Labeling of Chemicals (UN 2003).

Overclassification

- For the total database, the overall estimated overclassification rate for a nonsevere irritant or nonirritant being classified as a Category 1 substance ranged from 1.23% to 1.66%

References

BLS. 2004. US Dept. of Labor, Bureau of Labor Statistics, Case and Demographic Characteristics for Work-related Injuries and Illnesses Involving Days Away From Work Available: http://www.bls.gov/iif/oshcdnew.htm [accessed 26 October 2004]

Draize et al. 1944. J Pharmacol Exper Therapeut 82:377-390.

OECD. 2002. Test guideline 405. Acute eye irritation/corrosion, adopted April 24, 2002. In: OECD Guidelines for Testing of Chemicals. OECD:Paris. Available: http://www.oecd.org [accessed 26 August 2004].

UN. 2003. New York & Geneva: United Nations Publications. Available: http://www.unece.org/trans/danger/publi/ghs/officialtext.html [accessed 15 July 2005].

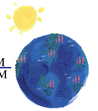
Acknowledgements

NICEATM and ICCVAM would like to thank the generous contributions of the companies and individuals who provided data for this analysis:

- Access Business Group (Luann Potts and Tom Truszkowski)
- Cosmetics, Toiletry, and Fragrance Association (Dr. Carol Eisenmann)
- ExxonMobil Biomedical Sciences, Inc. (Dr. James Freeman)
- GlaxoSmithKline (Mr. Frederick Guerriero)
- Johnson & Johnson Pharmaceutical R&D (Dr. Philippe Vanparys and Dr. Freddy Van Goethem)
- National Institute of Health Sciences (Japan) (Dr. Yasuo Ohno)
- S.C. Johnson & Son, Inc./JohnsonDiversy, Inc. (Dr. Nicole Cuellar, Dr. Judith Swanson, and Dr. Usha Vedula)
- TNO (Mr. Menk Prinsen)
- U.S. Food and Drug Administration (Ms. Donnie Lowther)
- ZEBET (Dr. med. Horst Spielmann and Dr. Manfred Liebsch)

ILS staff supported by NIEHS contract N01-ES 35504. The views expressed above do not necessarily represent the official positions of any federal agency.

More information on ICCVAM and NICEATM can be accessed at: http://iccvam.niehs.nih.gov/



U.S. Department of Health and Human Services
National Institutes of Health
National Institute of Environmental Health Sciences