

Estimated Under- and Over-Classification Rates for a 1-3 Rabbit Sequential Draize Rabbit Eye Test

J Haseman¹, D Allen^{2,3}, N Choksi^{2,3}, R Tice^{2,3*}, W Stokes³

¹Consultant, Raleigh, NC; ²Integrated Laboratory Systems, Inc., RTP, NC; ³National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM), National Institute of Environmental Health Sciences, RTP, NC

Introduction

The ocular irritation or corrosion potential of substances to which humans may be exposed has been evaluated since 1944 by the Draize rabbit eye test (Draize et al. 1944). There have been widespread efforts to develop and validate *in vitro* alternatives that might reduce or replace the use of rabbits for ocular irritancy assessments. ICCVAM is evaluating four *in vitro* assays 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* assays and 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 assessing the performance of alternative assays, information on the Draize eye test reliability would be useful but the paucity of repeat test data precludes an accurate estimate of inter- and intra-laboratory reproducibility. However, Draize eye test results can be used to estimate the likelihood of underclassifying a positive substance or overclassifying a negative substance using the Globally Harmonized System (GHS) sequential testing scheme (UN 2003).

Individual rabbit data from Draize eye testing using from 1 to 6 rabbits was obtained for 1004 studies from publications, U.S. Federal regulatory agencies, and scientists and organizations. Studies were conducted in accordance with Organisation for Economic Co-operation and Development (OECD) Test Guideline 405. Ocular irritation categories were assigned based on the United Nations (UN) Globally Harmonized System for Classification and Labeling of Chemicals (UN 2003) (Table 1). As some studies used more than three rabbits, rather than the 1 to 3 rabbits currently needed for the GHS ocular hazard classification system, rules were established to classify the substances tested in these studies according to the 3-rabbit 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:

- 777 total studies classified¹
- 181 studies classified as Category 1
- 60 studies classified as Category 2A
- 51 studies classified as Category 2B
- 185 studies classified as Nonirritant
- 15 studies were not classifiable²

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	<ul style="list-style-type: none"> • At least 1 animal with a corneal opacity score of 4 at any time (NICEATM Cat 1A) • At least 1 animal with effects not expected to reverse or that do not fully reverse within 21 days (NICEATM Cat 1A) • At least 2 animals with a mean corneal opacity score ≥ 3 and/or an iris score ≥ 1.5 (NICEATM Cat 1B)
Category 2A: Irritating to eyes	<ul style="list-style-type: none"> • At least 2 of 3 animals with mean scores for one of more of the following: <ul style="list-style-type: none"> - 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)

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

Category	Criteria Necessary for Classification
Category 1	<ul style="list-style-type: none"> ≥ 1 of 3 (≥ 2 of 6) rabbits have a NICEATM Cat 1A response ≥ 1 of 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	<ul style="list-style-type: none"> ≥ 2 of 3 (≥ 4 of 6) rabbits have a Cat 2A responses 1 of 3 (2 of 6) rabbits have a Cat 2A responses and 1 of 3 (2 of 6) rabbits have a Cat 2B responses
Category 2B	≥ 2 of 3 (≥ 4 of 6) rabbits have a Cat 2B responses
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

¹ 230 studies could not be assigned a GHS classification due to study criteria not 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).

² These substances (less than 2% of the total) showed variable responses and were not included in the statistical analysis because they could not be easily assigned to any classification. Because of their low frequency, inclusion of these substances in the analyses does not appreciably change the underclassification or overclassification estimates.

Underclassification Analysis

Classification Rules

In estimating the underclassification rate for the *in vivo* rabbit eye test, sampling 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, the study is complete
 - 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., 1B, 2A, 2B, or non-irritant)
 - 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"
- When the same substance was tested at multiple concentrations, the dose response for ocular toxicity was 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)

Calculations Conducted

- **Calculation 1:** Homogeneity of response within a given category
 - Assumed that rabbits 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 rabbits 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 rabbits have similar pattern of response for chemicals within response categories
 - **Strong responders.** Category 1 substances that always 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% of the rabbits 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.

Limitations

- No human data to confirm the irritancy classifications of the tested substances
- Inadequate repeat data for substances tested at the same concentration within the database
- Not all ocular irritants may produce the same pattern of rabbit responses
- 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

Underclassification Analysis (cont'd)

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 4 provides an example of how the underclassification rate of Category 1 substances as Category 2A were determined. Similar calculations for underclassification of Category 1 substances as Category 2B, nonirritant, and variable also are provided.

Table 3. Distribution of Individual Rabbit Responses in 181 GHS Category 1 Studies

GHS Ocular Hazard Classification	Number of Rabbits
Category 1A	499 (69%)
Category 1B	50 (7%)
Category 2A	96 (13%)
Category 2B	58 (8%)
Nonirritant	20 (3%)
Total	723 (100%)

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

Potential Outcome	Probability Calculation	Contribution to Underclassification Rate
2A-2A	$(96/723) \times (96/723)$	0.01760
2A-X-2A*	$(96/723) \times (128/723) \times (96/723)$	0.00312
X-2A-2A	$(96/723) \times (128/723) \times (96/723)$	0.00312
2A-2B-Nonirritant	$[(96/723) \times (58/723) \times (20/723)] \times 6$	0.00177
2A-2B-Category 1B*	$[(96/723) \times (58/723) \times (50/723)] \times 6$	0.00442
Total		0.0300 (3.00%)

*X refers to an outcome of either Category 1B, Category 2B, or nonirritant

*Refers to a rabbit classified based on severity of opacity or iris effects

Other Underclassification Rates:

- The likelihood of a Category 1 chemical being misclassified as Category 2B is the sum of two outcomes: $[(58/723) \times (58/723)] + [(58/723) \times (58/723) \times (166/723) \times 2] = 0.0094$ (0.94%)
- The likelihood of Category 1 chemical being misclassified as a nonirritant is the sum of two outcomes: $[(20/723) \times (20/723)] + [(20/723) \times (20/723) \times (2) \times 2] = 0.0012$ (0.12%)
- The likelihood of Category 1 chemical not clearly classified based on the results is the sum of two outcomes: $[(50/723) \times (20/723) \times (96/723) \times 6] + [(50/723) \times (20/723) \times (58/723) \times 6] = 0.0024$ (0.24%)

The total probability for underclassification is: $(3.00\% + 0.94\% + 0.12\% + 0.24\%) = 4.30\%$

Calculation 2: Heterogeneity of Response

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

Example Calculation:

- A Category 1 irritant has 4 rabbits classified as Category 1A and 2 rabbits classified as Category 2A
- The likelihood of this irritant being underclassified to be Category 2A is $(2/6) \times (2/6) = 0.1111$ (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 5)

Table 5. Likelihood that a Category 1 Substance will be Underclassified Using Calculation 2

GHS Ocular Hazard Classification	Estimated Underclassification Rate
Category 2A	7.51%
Category 2B	4.29%
Nonirritant	1.44%
Variable	0.00%
Total	13.24%

Calculation 3: Homogeneity/Heterogeneity of Response

Table 6. Distribution of Individual Rabbit Responses per Three Subgroups of GHS Category 1 Studies

GHS Category	All (n=181)	Strong Responders (n=108)	Moderate Responders (n=32)	Weak Responders (n=41)
Category 1A	499	347	104	48
Category 1B	50	38	6	6
Category 2A	96	0	32	64
Category 2B	58	0	16	42
Nonirritant	20	0	4	16
Total	723	385	162	176

Table 7. Likelihood that a Category 1 Substance will be Underclassified Using Calculation 3

GHS Category	Strong Responders (n=108)	Moderate Responders (n=32)	Weak Responders (n=41)	Weighted Overall Average (n=181)
Category 2A	0.0000	0.0587	0.2934	0.0768
Category 2B	0.0000	0.0149	0.1126	0.0281
Nonirritant	0.0000	0.0010	0.0188	0.0044
Variable	0.0000	0.0016	0.0112	0.0028
Total	0.0000	0.0762	0.4360	0.1121
		(7.62%)	(43.60%)	(11.21%)

For each subgroup, the underclassification rate was determined using the homogeneity approach (Calculation 1); the overall average weights the calculation by the relative proportion of strong, moderate, and weak responders in the database in that GHS category.

Table 8. Estimated Overall Underclassification Rates of GHS Category 1 Substances to GHS Category 2A, 2B, Nonirritant, or Variable Responder Substances

GHS Underclassification	Calculation 1 Homogeneity of Response	Calculation 2 Heterogeneity of Response	Calculation 3 Homogeneity/Heterogeneity of Response
as Category 2A	2.85%	7.51%	7.68%
as Category 2B	0.89%	4.29%	2.81%
as Nonirritant	0.12%	1.44%	0.44%
as Variable	0.24%	0.00%	0.28%
Total	4.10%	13.24%	11.21%

Underclassification Analysis (cont'd)

Underclassification Analysis Based on the Criteria Used (see Table 1) to Classify a Substance as GHS Category 1

The GHS category 1 substances were classified according to the

- 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

For each criterion, the underclassification rate was estimated using Calculations 1 and 2. The results of these analyses, totaled across the various nonsevere and nonirritant classifications, are provided in Table 9.

Table 9. Estimated Underclassification Rates for GHS Category 1 Substances Based on Classification Criteria

Criteria 1 (%)	Criteria 2 (%)	Criteria 3 (%)	Criteria 4 (%)
Calc 1	Calc 2	Calc 3	Calc 1
8.52%	17.62%	14.57%	0.78%
4.80%	2.43%	0.00%	0.00%
0.00%	0.00%	0.00%	0.61%
5.30%	4.06%		

Calc = calculation

Overclassification Analysis

The same approaches used for estimating the underclassification rate for a positive ocular corrosive or severe irritant, can be used to estimate the likelihood of overclassifying a negative substance, using the GHS 1 to 3 rabbit sequential test (see Underclassification Analysis). As there was less evidence of heterogeneity of response and as there were no clearly identifiable sub-groupings showing distinct patterns of response as was seen for Category 1 substances, Calculation 3 (homogeneity/heterogeneity of response combination) was not used to estimate the overclassification rate for nonsevere irritants/nonirritants. The distribution of individual rabbit responses for studies resulting in a Category 2A, 2B, or nonirritant classification is provided in Table 10, the resulting estimated underclassification rates in Table 11.

Table 10. Distribution of Individual Rabbit Responses by GHS Ocular Hazard Classification

GHS Category	Category 2A (60 studies)	Category 2B (51 studies)	Nonirritants (485 studies)
Category 1A	8	1	0
Category 1B	7	1	0
Category 2A	181	10	11
Category 2B	38	141	46
Nonirritant	19	31	1987
Total # of Rabbits	253	184	2044

Table 11. Estimated Overclassification Rates of 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	7.70%	6.67%
Category 2B as Category 1	1.28%	0.82%
Nonirritant as Category 1	0.00%	0.00%
Overall Average*	0.88%	0.74%*

* The overall average weights the calculation by the relative proportion of GHS Category 2A, 2B, and nonirritating substances in the database.

Conclusions

Underclassification

- For this database, the overall estimated underclassification rate for a Category 1 substance ranged from 4.10% to 13.24%
 - Calculation 3 (11.21%) likely is the most accurate estimate for this database
- For the criteria for classification:
 - The highest underclassification rate (8.52% to 17.62%) was for Criteria 1 (substances that produced a persistent lesion through 21 days in at least one of three rabbits)
 - The lowest underclassification rates (0.00% to 5.30%) were for those criteria that included lesion severity as the reason for classifying a substance as a GHS Category 1; the rates were similar among these three subgroups

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 7.49% to 8.98%

References

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

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