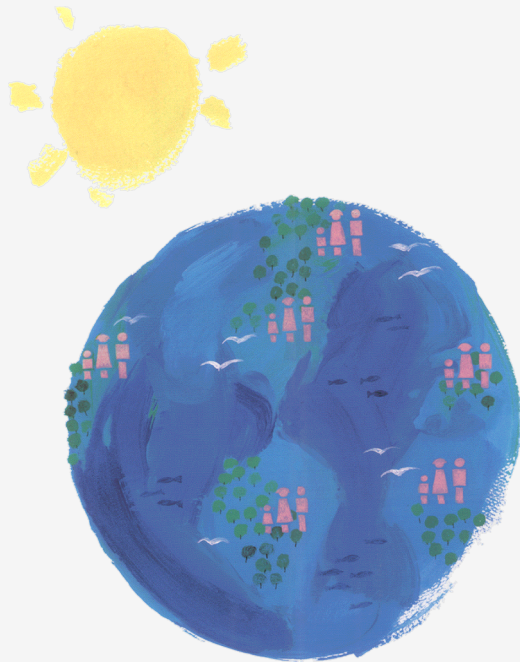


Evaluation of the Under- and Over- Classification Likelihood of the *In Vivo* Rabbit Ocular Irritation Test Method



Dr. Joseph Haseman

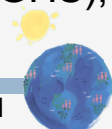
**SACATM Meeting
December 12, 2005
Alexandria, Virginia**

Study Objectives

Using the United Nations Globally Harmonized System of Classification and Labelling of Chemicals (GHS)¹, evaluate the likelihood of:

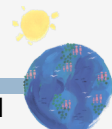
- (a) underclassifying an ocular corrosive or severely irritating substance as a nonsevere irritant/nonirritant in the current rabbit eye irritation test
- (b) overclassifying an ocular nonsevere irritant or nonirritant as a corrosive or severely irritating substance

¹UN (2003) Globally Harmonized System of Classification and Labelling of Chemicals (GHS), New York & Geneva: United Nations Publications.



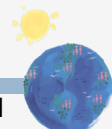
General Approach (1)

- Form a database of substances based on observed rabbit effects using GHS classification system
 - Category 1: Irreversible effects on the eye
 - Category 2A: Irritating to eyes
 - Category 2B: Mildly irritating to eyes
 - Nonirritant
- Devise classification strategy for three rabbit sequential test
- Based on the database and classification strategy,
 - estimate the underclassification rate for Category 1 substances
 - estimate the overclassification rate for Category 2 and nonirritant substances



General Approach (2)

- For this evaluation, GHS Category 1 substances were separated into two “unofficial” subcategories:
 - **Category 1A:** (1) Rabbit with corneal opacity score of 4 at any time or (2) effects not expected to reverse or that do not fully reverse (all endpoints = 0) within 21 days (1 of 3 rabbits required for classification)
 - **Category 1B:** Rabbit with mean (over observation days 1, 2, and 3) corneal opacity score ≥ 3 and/or iritis score ≥ 1.5 (2 of 3 rabbits required for classification)
- Additional GHS rabbit classifications:
 - **Category 2A:** Rabbit with (a) mean scores for one of more of the following: corneal opacity ≥ 1 , iritis ≥ 1 , redness ≥ 2 , chemosis ≥ 2 , and (b) the effects fully reverse within 21 days;
 - **Category 2B:** Rabbit with (a) mean scores for one of more of the following: corneal opacity ≥ 1 , iritis ≥ 1 , redness ≥ 2 , chemosis ≥ 2 , and (b) the effects fully reverse within 7 days;
 - **Nonirritant:** Rabbit with corneal opacity, iritis, redness, an chemosis mean scores < 1



Data Used for Analysis

- 1005 *in vivo* rabbit eye studies with individual rabbit data
 - Studies performed consistent with OECD TG 405
 - 213 studies did not meet test method acceptance criteria (e.g., not able to evaluate persistence out to day 21)
 - 181 Category 1 studies
 - 596 nonsevere (Category 2A, Category 2B, Nonirritant) studies
 - 15 studies which could not be classified due to lack of a consensus response among the tested rabbits

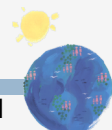
Data Sources

Source	Number of Acceptable Corrosive or Severe Irritant Studies	Number of Acceptable Nonsevere Irritant or Nonirritant Studies
Access Business Group	6	1
CTFA	17	36
ECETOC	30	105
ExxonMobil	2	8
EPA TSCA	33	101
GlaxoSmithKline	8	17
ISOPA	0	8
Laboratoire Nationale de la Sante	6	41
NIHS	18	34
SC Johnson	6	9
TNO	5	76
ZEBET	23	72
Total	181	596



Sequential Testing Classification Strategy

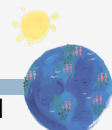
- First Rabbit:
 - If first rabbit is Cat 1A, then substance classified as Cat 1
 - If not, then test second rabbit
- Second Rabbit:
 - If second rabbit is Cat 1A, then substance classified as Cat 1
 - If lesions for first and second rabbits are same category, study is complete (i.e., 1B, 2A, 2B, NI)
 - If not, test third rabbit
- Third Rabbit:
 - If third rabbit is Cat 1A, then substance classified as Cat 1
 - If lesions for 2 of 3 rabbits are same category, then classified as that category (i.e., 1B, 2A, 2B, or non-irritant [NI])
 - If 1 rabbit is Cat 2A, 1 rabbit is Cat 2B, and the third rabbit is Cat 1B or nonirritant, then the study is classified as Cat 2A
 - If all rabbits have different classifications (e.g., Cat 1B, NI, and 2A or 2B) then chemical is classified as “undetermined”



Subset Data Evaluations for GHS Category 1

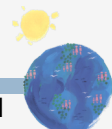
Three subsets of the total database also were evaluated

- Subset 1: Criteria used to classify Category 1 substances
 - Criterion 1: Substances produced a persistent lesion through 21 days in at least one of three rabbits
 - Criterion 2: Substances produced a severe response in at least two of three rabbits
 - Criterion 3: Substances produced a persistent lesion through 21 days in at least one of three rabbits and produced a severe response in at least two of three rabbits
 - Criterion 4: Substances produced a corneal opacity of 4 in at least one of three rabbits
- Subset 2: Physical form of the test substance (liquids/gel vs. solids)
- Subset 3: Chemical class



Sources of Response Variability that Affect Over- and Under-Classification Rates

- Variability among animals within a given study
 - Extensive data exist to evaluate this source of variability
- Variability among studies of a given substance
 - Very limited data exist to evaluate this source of variability
- Some substances showed a non-monotonic dose response
 - To be conservative, substances tested as multiple doses are assumed to be Category 1 at and above the minimum dose producing a Category 1 response



Assumptions Concerning Variability in Response (1)

- Calculation 1 - Homogeneity of response
 - It is assumed that all substances have the same pattern of rabbit response within a irritancy category
 - Requires only one calculation but may underestimate the underclassification potential
- Calculation 2 - Heterogeneity of response
 - It is assumed that substances have a different pattern of rabbit response within an irritancy category
 - Leads to higher misclassification rates than Assumption 1, but may overestimate the underclassification potential

Assumptions Concerning Variability in Response (2)

- Calculation 3 - Homogeneity/Heterogeneity of response combination
 - Assumed that Category 1 substances have similar patterns of rabbit response within 3 categories
 - *Strong responders.* Substances that produce a Cat 1A or 1B response in all rabbits
 - *Moderate responders.* Substances that produce a Cat 1 response in at least 50% but not 100% tested rabbits
 - *Weak responders.* Substances that produced a nonsevere (i.e., Cat 2A, 2B) or nonirritant response in >50% tested rabbits
 - Should provide the most reasonable estimate of the underclassification potential

Calculation 1: Distribution of Rabbit Responses per GHS Studies

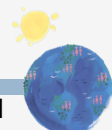
Study Rabbits	Category 1 Classification (181 Studies)	Nonsevere Classification (596 studies)	Equivocal (15 studies)	Total (792 studies)
Category 1A	499	9	9	517
Category 1B	50	8	0	58
Category 2A	96	202	11	309
Category 2B	58	225	23	306
Nonirritant	20	2037	45	2102
Total Rabbits	723	2481	88	3292

Calculation 1: Example Calculation (Category 1 Substance being Classified as Category 2A)

Potential Outcome	Probability Calculation	Contribution to Underprediction Rate
2A-2A	$(96/723) \times (96/723)$	0.0176
2A-X-2A ^a	$(96/723) \times (128/723) \times (96/723)$	0.0031
X-2A-2A	$(128/723) \times (96/723) \times (96/723)$	0.0031
2A-2B-Nonirritant	$[(96/723) \times (58/723) \times (20/723)] \times 6$	0.0018
2A-2B-Category 1B ^b	$[(96/723) \times (58/723) \times (50/112)] \times 6$	0.0044
Total		0.0300 (3.00%)

^aX refers to an outcome of either Category 2B, nonirritant, or Category 1B

^bRefers to rabbit classified based on severity of opacity or iris effects



Calculation 2: Heterogeneous 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, based on prevalence within the database, to produce an overall underclassification rate

Calculation 2: Example Calculation

- Example: Category 1 irritant has 4 rabbits classified as Category 1A and 2 rabbits classified as Category 2A
- Likelihood of this irritant being classified as Category 2A is $(2/6) \times (2/6) = 0.1111$ (11.11%)
- The likelihood of classification as Category 2B or nonirritant is estimated to be zero
- Calculations are carried out for the other Category 1 irritants and the rates averaged to estimate underclassification rate

Calculation 3: Underclassification Rate

Underclassification rate for each subgroup determined using Calculation 1

Category	Strong Responders (n=108)	Moderate Responders (n=32)	Weak Responders (n=41)	Overall ¹ (n=181)
Category 2A	0.00%	5.87%	29.34%	7.68%
Category 2B	0.00%	1.49%	11.26%	2.81%
Nonirritant	0.00%	0.10%	1.88%	0.44%
Not Determined	0.00%	0.16%	1.12%	0.28%
Total	0.00%	7.62%	43.60%	11.21%

¹Overall underclassification likelihood calculated as an average based on the prevalence of substances within the three responder categories.



Estimated Overall Underclassification Likelihood

GHS Underclassification	Calculation 1	Calculation 2	Calculation 3
as Category 2A	3.00%	7.51%	7.68%
as Category 2B	0.94%	4.29%	2.81%
as Nonirritant	0.12%	1.44%	0.44%
as Undetermined	0.24%	0.00%	0.28%
Total	4.30%	13.24%	11.21%

Subset 1: Estimated Underclassification Likelihood (based on GHS classification criteria)

GHS Under-classification	Criterion 1 (%) (≥1/3 rabbit with persistence only) (n=61)			Criterion 2 (%) (≥2/3 rabbits with severe response) (n=8)			Criterion 3 (%) (persistence plus 2/3 rabbits with severe response) (n=3)			Criterion 4 (%) (corneal opacity = 4 in ≥1/3 rabbits) (n=88)		
	Calc 1	Calc 2	Calc 3	Calc 1	Calc 2	Calc 3	Calc 1	Calc 2	Calc 3	Calc 1	Calc 2	Calc 3
as Category 2A	5.28	9.07	9.0	0.58	1.56	2.99	0.00	0.00	0.00	0.42	2.36	2.32
as Category 2B	3.08	7.47	5.16	0.20	3.24	1.24	0.00	0.00	0.00	0.10	1.59	0.97
as Nonirritant	0.14	1.08	0.36	0.00	0.00	0.00	0.00	0.00	0.00	0.07	1.35	0.59
as Un-determined	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.03
Total	8.52	17.62	14.52	0.78	4.80	4.23	0.00	0.00	0.00	0.61	5.30	3.91



Subset 2: Estimated Underclassification Rates (Based on Physical Form, Total Database)

GHS Under-classification	Liquid/Gel (n=100)			Solid (n=30)		
	Calc 1	Calc 2	Calc 3	Calc 1	Calc 2	Calc 3
as Category 2A	3.26%	8.62%	8.40%	2.74%	7.04%	7.13%
as Category 2B	1.79%	5.61%	4.55%	0.04%	1.08%	0.13%
as Nonirritant	0.11%	1.56%	0.33%	0.01%	0.19%	0.10%
as Undetermined	0.20%	0.00%	0.35%	0.10%	0.00%	0.15%
Total	5.36%	15.79%	13.63%	2.89%	8.31%	7.51%



Subset 3: Chemical Class Underclassification Evaluation (Calculation 1)*

Chemical Class	Number of Studies	Number of Rabbits
Alcohols	23	89
Amines	22	81
Organic Salts and Chemicals	21	68
Carboxylic Acids	19	66
Ethers	11	39
Heterocyclics	12	37
Esters	10	37
Esters	10	37
Onium Compounds	9	29
Inorganic Salts and Chemicals	7	29
Phenols	6	31

* Chemical classes where 25 or more rabbits were tested; a substance could be in as many as three different chemical classes



Subset 3: Estimated Underclassification Rate (Based on Chemical Class)

Carboxylic Acids (66 rabbits)	Alcohols (89 rabbits)	Organic Salts and Chemicals (68 rabbits)	Ethers (39 rabbits)	Esters (37 rabbits)
16.64%	10.89%	9.32%	7.60%	4.85%

Heterocyclic Compounds (37 rabbits)	Amines (81 rabbits)	Inorganic Salts and Chemicals (29 rabbits)	Phenols (31 rabbits)	Onium Compounds (29 rabbits)
2.51%	2.45%	0.48%	0.00%	0.00%

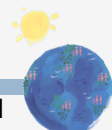


Estimated Overall Overclassification Likelihood

Overclassification as GHS Category 1	Calculation 1	Calculation 2
Category 2A (n=60)	7.70%	6.67%
Category 2B (n=51)	1.28%	0.82%
Nonirritant (n=485)	0.00%	0.00%
Total for Nonseveres (weighted average)	0.88%	0.74%

Springer et al. Analysis¹ (1)

- An important publication used to support using 3 rather than 6 rabbits per test.
- Limitation of the previous under- and over-classification analysis is that it requires an *a priori* assignment of each substance to a specific irritation category.
- Springer et al. (1993) approach does not require an *a priori* classification.
- Some limitations of the Springer method are:
 - Only two response categories (irritant and nonirritant)
 - Assumes a fixed number of rabbits (e.g., 6 rabbits tested per study)
 - Assumes a binomial distribution of responses for irritants and nonirritants (i.e., no heterogeneity of response)



Springer Analysis (2)

- Our application of the Springer method assumed Category 2A, 2B, and nonirritants to be “nonirritants” and used the 303 six-rabbit studies in the database (38% of the total database)
- Three parameters were estimated:
 - Probability of a positive response for an irritant
 - Probability of a positive response for a nonirritant
 - The proportion of nonirritants in the database
- The model fit was poor due to the heterogeneity in the irritant database

Modified Springer Analysis

- Modified Springer analysis allows for heterogeneity among irritants by assuming that the irritants are a mix of two homogeneous subcategories
 - Strong irritants
 - Weak irritants
- Modified Springer analysis estimates five parameters, which for the database were:
 - Probability of a positive response for a strong irritant = 96.4%
 - Probability of a positive response for a weak irritant = 35.3%
 - Probability of a positive response for a nonirritant = 0.89%
 - Proportion of strong irritants in the database = 13.8%
 - Proportion of nonirritants in the database = 72.6%
- Model fit was excellent, and based on these parameter values, the under-classification rate was estimated to be 13.4% and over-classification rate was estimated to 2.65%.

Model Fit For the Springer Method

Irritant Rabbits/ Total Rabbits Tested	Observed Database Distribution	Springer Analysis Fit	Modified Springer Analysis Fit
6/6	34	23.7	33.7
5/6	8	23.9	8.4
4/6	6	10.1	4.7
3/6	7	2.7	9.8
2/6	15	6.7	13.8
1/6	21	52.9	21.1
0/6	212	183.0	211.5
Total	303	303	303
P-value of fit to observed		<0.001	>0.10



Estimated Overall Under- and Over-classification Likelihood

GHS	Original Analysis			Modified Springer Analysis
	Calculation 1	Calculation 2	Calculation 3	
Under-classification	4.30%	13.24%	11.21%	13.40%
Over-classification	0.88%	0.74%	-	2.65%

Conclusions

- Two different approaches, based on different underlying assumptions and different sets of irritants and nonirritants, resulted in similar under- and over-classification likelihoods for the current rabbit eye test, when test substances are classified as ocular corrosives/severe irritants or nonsevere irritants/nonirritants according to the GHS classification system.
- Based on these analyses, the most reasonable estimate of the under-classification likelihood for the current rabbit eye test method is between 11-14%.
- For the subset analyses:
 - Criterion 1 (severity classification based on persistence) was shown to have a higher underclassification likelihood than Criteria 2-4 (severity classification based on severity)
 - 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
- The estimated overclassification likelihood for the *in vivo* rabbit eye test method is less than 3%.

