ABSTRACT ID #1616

Critical Review of the UN GHS decision criteria for chemicals causing irreversible effects on the eye/serious damage to the eye: Cosmetics Europe Analysis

Martina Klaric[1], Uwe Pfannenbecker[2], Els Adriaens[3], Nathalie Alépée[4], João Barroso[1,5], Jalila Hibatallah[6], Karsten R. Mewes[7], Pauline McNamee[8]

[1]Cosmetics Europe; [2]Beiersdorf AG; [3]Adriaens Consulting; [4]L'Oréal Research & Innovation ; [5]European Commission Joint Research Centre, Institute for Health and Consumer Protection, EURL ECVAM / Systems Toxicology Unit; [6]Chanel Parfums Beauté; [7]Henkel AG & Co. KGaA; [8]The Procter & Gamble Company

Cosmetics Europe
the personal care association

Introduction

A thorough understanding of which of the effects assessed in the *in vivo* Draize eye test are responsible for driving UN GHS classification is critical for an appropriate selection of chemicals to be used in the development and/or evaluation of *in vitro* methods and for properly assessing test method predictive capacity and limitations. For this reason, Cosmetics Europe has compiled a database of Draize data (Draize eye test Reference Database, DRD) from external lists that were created to support past validation activities. The key goals for compiling the DRD were: i) to enable a comprehensive analysis and understanding regarding *in vivo* drivers of classification based on the Draize eye test, ii) to further evaluate the variability of the Draize eye test based on data obtained from repeat studies, iii) to make available an extensive list of chemicals with TG 405 *in vivo* data, beyond those generally used historically, for further method development and validation, and iv) to provide guidance for selecting reference chemicals based on understanding ocular tissue effects that drive classification in the *in vivo* rabbit Draize eye test. The work presented here describes one part of the overall analysis which is focused on Category 1 (Cat 1) studies classified based on persistence of effects in the minority/majority of animals (77 studies) or on corneal opacity (CO) = 4 (34 studies). Furthermore, data on chemicals tested in multiple *in vivo* studies were analyzed to determine the concordance of classification between repeat studies and the consistency for drivers of classification.

Strategy for Developing the DRD

The DRD was primarily compiled using different sources of historical *in vivo* Draize eye test data i.e. ECETOC, ZEBET, Laboratoire National de la Santé (Gautheron), NICEATM, EURL ECVAM, which were created to support past validation activities. These data were produced according to OECD Test Guideline 405 using proprietary and commercially available chemicals. The studies were classified according to the serious eye damage/eye irritation classification criteria defined by UN GHS and EU CLP which implemented UN GHS in the EU. According to these classification systems, there are several criteria derived from the four ocular tissue effects assessed in the Draize eye test, namely corneal opacity (CO), iritis (IR), conjunctival redness (CR) and conjunctival chemosis (CC) with CR and CC identified together here as conjunctival effects or "Conj". Each ocular tissue effect can independently drive the classification of a chemical. Therefore, a chemical can be classified based on a single or multiple drivers of classification. These drivers of classification are described in Table 1. A full identification of all drivers of classification observed in each individual *in vivo* study are reported in the DRD. Selection of the main driver of classification in each study was performed according to the driver appearing in the largest number of animals and, in case of equal number of animals, according to the prioritisation scheme identified in Table 1 in moving from the left to right for each UN GHS/EU CLP category. Overall, the DRD contains 681 independent *in vivo* studies on 634 individual chemicals representing a wide range of chemical classes and different physical states.

Results: Distribution According to Drivers of Classification

Table 1: List of the *in vivo* drivers of UN GHS classification for the chemicals requiring classification and subgroups for the chemicals not requiring classification. This table also contains the proportion (%) and number (n) of studies according to main driver of classification or according to the subgroups (No Cat).

Category 1 28.1% (n=165)					Category 2 [a] 13.5% (n=79)			No Category 58.4% (n=343)			
Severity [b] *in ≥ 60% of the animals* 27.3% (n=45)	Persistence on Day 21 *in at least one animal* 46.7% (n=77)			Severe CO *in at least one animal* 20.6% (n=34)	Severity [b] *in ≥ 60% of the animals*			*in at least one observation time in at least one animal*	*in all observation times in all animals*		
CO mean ≥ 3 73.3% (n=33)	**IR mean > 1.5** 26.7% (n=12)	**CO** 80.5% (n=62)	**Conj** 19.5% (n=15)	**IR** 0% (n=0)	**CO=4** 100% (n=34)	**CO mean ≥ 1** 60.8% (n=48)	**Conj mean ≥ 2** 38% (n=30)	**IR mean ≥ 1** 1.3% (n=1)	**CO > 0 **** 8.7% (n=30)	**CO > 0** 13.1% (n=45)	**CO = 0 **** 1.7% (n=6)	**CO = 0** 76.4% (n=262)

[a] sub-categorised in two categories: Category 2A (irritant to eyes) when any of the eye effects in any animal is not fully reversible within 7 days of observation (i.e. CO, IR, CR and/or CC > 0 at 7 ≤ day < 21) and 2B (mildly irritant to eyes) when all observed eye effects are fully reversible within 7 days of observation (i.e. CO, IR, CR and CC = 0 on day 7 and beyond); [b] Mean scores calculated from gradings at 24, 48, and 72 hours after instillation of the test chemical; ** at least one animal with a mean score of days 1-3 above the classification cut-off for at least one endpoint

Results: Variability of the Draize eye test

Table 2. Main drivers of classification for studies requiring classification and distribution of chemicals not requiring classification by subgroup: details for replicate Draize eye test studies. N corresponds to the number of repeated studies. Studies between squared brackets [e.g. UN GHS (main driver)^{SCNM}] indicate studies for which study criteria allowing an unambiguous classification were not met (SCNM: study criteria not met) because the study was terminated before day 21 without full reversibility of all endpoints and in the absence of any other effects driving a Cat 1 classification.

Chemical	N	UN GHS (main driver)	Agree
Pyridine	2	Cat 1 (CO mean ≥ 3); Cat 2 (CO mean ≥ 1)	No
2-Benzyl-4-chlorophenol	2	Cat 1 (CO mean ≥ 3); Cat 1 (CO pers D21)	Yes (≠ main driver)
Dibenzoyl-L-tartaric acid	2	Cat 1 (CO mean ≥ 3); Cat 1 (CO mean ≥ 3)	Yes
Imidazole	2	Cat 1 (CO mean ≥ 3); Cat 1 (CO pers D21)	Yes (≠ main driver)
Promethazine HCL	2	Cat 1 (CO mean ≥ 3); Cat 1 (CO mean ≥ 3)	Yes
Triton X-100 (100%)	3	Cat 1 (IR mean > 1.5); Cat 1 assumed (CO pers D 21)^{SCNM}]	Yes (≠ main driver)
Quinacrine	2	Cat 1 (IR mean > 1.5); Cat 1 (CO pers D21)	Yes (≠ main driver)
Butoxyethanol	3	Cat 1 (CO pers D21); Cat 1 (Conj pers D21); Cat 2 (Conj mean ≥ 2)	No
Ethanol (100%)	4	Cat 1 (CO pers D21); Cat 2 (CO mean ≥ 1); [at least Cat 2A (CO mean ≥ 1) ^{SCNM}]; Cat 2 (Conj mean ≥ 2)	No
Benzalkonium chloride (1%)	2	Cat 1 (CO pers D21); Cat 1 (CO pers D21)	Yes
Cetyltrimethyl ammonium bromide (10%)	2	Cat 1 (CO pers D21); Cat 1 (CO pers D21)	Yes
Sodium lauryl sulphate (10%)	3	Cat 1 (CO pers D21); Cat 1 (CO pers D21); Cat 2 (CO mean ≥ 1)	No
Sodium oxalate	2	Cat 1 (CO pers D21); Cat 1 (CO pers D21)	Yes
(3-Aminopropyl)triethoxy silane	2	Cat 1 (CO = 4); Cat 1 (CO = 4)	Yes
n-Butanol (100%)	2	Cat 1 (CO = 4); Cat 2 (CO mean ≥ 1)	No
iso-Butanol	2	[Cat 1 assumed (CO pers D 21) ^{SCNM}]; Cat 2 (CO mean ≥ 1)	No
gamma-Butyrolactone	2	Cat 2 (CO mean ≥ 1); Cat 2 (CO mean ≥ 1)	Yes
Methyl acetate	2	Cat 2 (CO mean ≥ 1); Cat 2 (CO mean ≥ 1)	2A/2B
Methyl N,N,N-trimethyl-4... (30%, aqueous) ¶	2	Cat 2 (CO mean ≥ 1); Cat 2 (Conj mean ≥ 2)	Yes (≠ main driver)
n-Octanol	2	Cat 2 (CO mean ≥ 1); Cat 2 (CO mean ≥ 1)	2A/2B
Tetra aminopyrimidine sulphate	2	Cat 2 (CO mean ≥ 1); No Cat (CO > 0)	No
Triton X-100 (5%)	2	[Cat 2B (CO mean ≥ 1) *]; Cat 2 (Con) mean ≥ 2]	Yes (≠ main driver)
Toluene	2	[at least Cat 2 (Conj mean ≥ 2)^{SCNM}]; No Cat (**)	No
o-Phenylenediamine	2	[at least Cat 2A (CO mean ≥ 1) ^{SCNM}]; [at least Cat 2A (CO mean ≥ 1^{SCNM}]	Uncertain
Methyl amyl ketone	2	No Cat (CO > 0**); No Cat (CO > 0)	Yes (≠ group)
Phosphoric acid, tributyl ester	2	No Cat (CO > 0**); No Cat (CO = 0)	Yes (≠ group)
1,2,3-Trichloropropane	2	No Cat (CO > 0); No Cat (CO = 0)	Yes (≠ group)
Methyl iso-butyl ketone	2	No Cat (CO > 0); No Cat (CO > 0)	Yes
Triethanolamine (100%)	2	No Cat (CO > 0); No Cat (CO = 0)	Yes (≠ group)
Sodium lauryl sulphate (1%)	2	No Cat (CO > 0); No Cat (CO > 0)	Yes
Xylene	2	No Cat (CO = 0**), No Cat (CO = 0)	Yes (≠ group)
3-Phenoxy benzaldehyde (100%)	2	No Cat (CO = 0); No Cat (CO = 0)	Yes
gamma-Glycidyloxypropyltrimethoxy silane	2	No Cat (CO = 0); No Cat (CO = 0)	Yes
gamma-Mercaptopropyl trimethoxy silane (100%)	2	No Cat (CO = 0); No Cat (CO = 0)	Yes
Glycerol (100%)	2	No Cat (CO = 0); No Cat (CO = 0)	Yes
Kronitex TXP	2	No Cat (CO = 0); No Cat (CO = 0)	Yes
Perfluoro-n-hexane	2	No Cat (CO = 0); No Cat (CO = 0)	Yes
Polyethylene glycol 400 (100%)	2	No Cat (CO = 0); No Cat (CO = 0)	Yes
Tricresyl phosphate	2	No Cat (CO = 0); No Cat (CO = 0)	Yes
Tween 20	2	No Cat (CO = 0); No Cat (CO = 0)	Yes
PEG-40 hydrogenated castor oil	2	No Cat (CO = 0); No Cat (CO = 0)	Yes
Tetrabromobisphenol A	2	No Cat (CO = 0); No Cat (CO = 0)	Yes

CO: corneal opacity; IR: iritis; Conj: conjunctival redness (CR) and/or conjunctival chemosis (CC)¶ Methyl N,N,N-trimethyl-4-[(4,7,7-trimethyl-3-oxobicyclo[2.2.1]hept-2-ylidene)methyl]anilinium sulphate (30%, aqueous). ** At least one animal with a mean score of days 1-3 above the classification cut-off for at least one endpoint

Table 2 identifies 42 chemicals for which more than one *in vivo* study was conducted. Analysis of these studies confirms that chemicals classified as Cat 1 by persistence or CO = 4 in a single animal were classified as Cat. 2in the repeat studies where such effects were no longer observed (Figure 3).
Concordance of UN GHS classification for chemicals requiring classification in at least one of multiple studies:
-Considering Cat 1 and unified Cat 2 = 65.2 % (15/23)
-Considering Cat 1 and Cat 2A / Cat 2B = 56.5% (13/23)
-Concordance of the same main driver of classification = 39.1 % (9/23).
Concordance of subgroup for chemicals not requiring classification:
- 72.2 % (13/18) same subgroup

Conclusion: 37.5 % (6/16) of the chemicals with at least one Cat 1 study could be equally identified as Cat 2 and 28.6 % (2/7) of the Cat 2 chemicals could be equally identified as No Cat.

Results: Studies Classified Cat 1 Based Only on Persistence

Figure 1. Boxplots presenting the distribution of individual animal CO grades at 1, 2, 3, 7, 14 and 21 days after instillation of the test chemical. The symbols (+) present individual observations.

Cat. 1 studies showing CO persistence on day 21 but with CO mean < 3 and IR mean < 1.5 in the majority of the animals

Fig. 1A CO pers in majority of animals (32 studies with 116 animals)

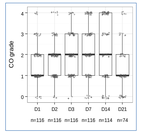
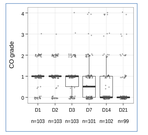
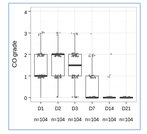


Fig. 1B CO pers in minority of animals (25 studies with 104 animals)



Cat 2A studies showing persistence of CO on day 7 in at least one animal

Fig. 1C (28 studies with 104 animals)



The analysis of UN GHS Cat 1 classification based on persistence of corneal effects in the absence of corneal severity calculated over the first three days is provided in Fig. 1A-1C. The distribution of the individual animal CO grades over time and, in particular, at Day 21 in Fig. 1A shows Cat 1 classification based only on CO persistence occurring in the majority of the animals can be distinguished from that provided in Fig. 1B which shows Cat 1 classification based only on CO persistence occurring in the minority of animals. In fact, based on the CO scores observed over the first three days, it is not possible to distinguish the Cat 1 studies with CO persistence without severity in the minority of the animals (Fig. 1B) from the Cat 2A studies with CO persistence on day 7 (Fig. 1C). This analysis supports, that when selecting chemicals for use in validation activities, chemicals classified as Cat 1 based on persistence of effects (i.e., CO > 0, IR > 0, CR > 1 or CC > 1 on day 21) or CO = 4 observed any time during the study should have application of a majority rule as currently done for effects observed on days 1-3 (severity). In fact, low level persistent effects or persistence occurring due to delayed effects, which are observed in a single animal are probably not related to the test chemical and should therefore not drive a Cat 1 classification in the absence of any other Cat 1 triggering effects in the study.

Figure 2 Boxplots presenting the distribution of individual animal CR grades at 1, 2, 3, 7, 14 and 21 days after instillation of the test chemical. The symbols (+) present individual observations.

Cat 1 studies showing CR persistence on day 21 but with CO mean < 3 and IR mean < 1.5 in the majority of the animals

Fig. 2A CR pers in majority of animals (16 studies with 55 animals)

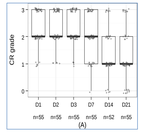
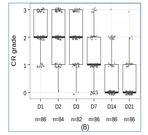
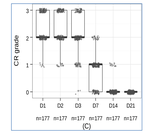


Fig. 2B CR pers in minority of animals (20 studies with 86 animals)



Cat 2A studies showing persistence of CR on day 7 in at least one animal

Fig. 2C (49 studies with 177 animals)

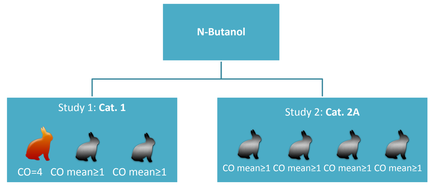


In contrast with CO (Fig. 1A), CR scores in studies showing CR persistence in the majority of the animals (Fig. 2A) decrease with time. Of note, CR ≥ 2 on Day 21 is almost always associated with CO > 0 on Day 21 and therefore these studies are generally also classified as Cat 1 based on CO persistence. Furthermore, no important difference in the distribution of the CR scores can be observed between Cat 1 studies with CR persistence in the minority of the animals (Fig. 2B) and Cat 2A studies with CR persistence on day 7 (Fig. 2C). In fact, it is not possible to distinguish the Cat 1 studies (Fig. 2A, 2B) from the Cat 2A studies (Fig. 2C) based on the distribution of the CR scores from the first three observation days, which demonstrates that CR is not useful to identify Cat 1 chemicals, at least when it comes to effects observed in the first 3 days after instillation. Similar findings were observed for CC scores (data not shown).

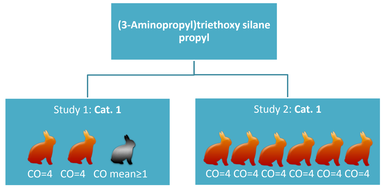
Results: Studies Classified Cat 1 Based on CO = 4

Figure 3: Examples for minority-based Cat 1 classification being not reproducible in a second study (3A) and majority-based Cat 1 classification being reproducible in a second study (3B).

3A: Reproducibility of studies CO=4 in minority of animals *



3B: Reproducibility of studies CO=4 in majority of animals *



* Representation of historic *in vivo* data derived from DRD

Conclusions:

UN GHS classification criteria define that a Cat 1 classification can be triggered based on tissue effects observed in a single animal. It is questionable whether such results should lead to a Cat 1 classification, especially in case of delayed effects observed in a single animal in the absence of any other Cat 1-triggering effects.
On the basis of the analysis provided here, implementation of the following recommendations should thus be considered:

(i) CR and CC scores of less than 2 on day 21 should be recognised as fully reversed and should therefore not drive a Cat 1 classification in the absence of any other Cat 1 triggering effects; (ii) Grade 4 CO scores and/or persistent effects appearing in a minority (<60 %) of the animals should not drive a Cat 1 classification in the absence of any other Cat 1 triggering effects;
(iii) Grade 4 CO scores that fully reverse within 21 days should not trigger a Cat 1 classification in the absence of any other Cat 1 triggering effects.

Overall, based on this analysis we suggest that chemicals identified as Cat 1 on the basis of the effects described above should not be included in validation studies of alternative methods. Such chemicals could lead to a decision of non-validity of an alternative method when validation criteria such as the absence of underclassified Cat 1 chemicals is used which, in our opinion, would not be scientifically justifiable.
Therefore, the establishment of validity criteria and chemicals selection should be carefully considered before initiating a validation study.

[#] References: Barroso et al.; Cosmetics Europe compilation of historical serious eye damage/eye irritation *in vivo* data analysed by drivers of classification to support the selection of chemicals for development and evaluation of alternative methods/strategies: the Draize eye test Reference Database (DRD); Arch. Toxicol.; 2016; in press