Performance statistics of defined approaches for eye hazard identification of liquids to distinguish between the three UN GHS categories

Poster #1043

Karsten R Mewes[1], Takayuki Abo[2], Els Adriaens[3], Dan Bagley[4], Jalila Hibatallah[5], Nathalie Alépée[6], Arianna Giusti[7]

[1] Henkel AG & Co. KGaA, [2] Kao Corporation, [3] Adriaens Consulting bvba, [4] Colgate-Palmolive, [5] Chanel Parfums Beauté, [6] L'Oréal Research & Innovation, [7] Cosmetics Europe - The Personal Care Association

Introduction

Cosmetics Europe (CE) developed two defined approaches (DALs) for eye hazard identification, i.e. addressing serious eye damage (UN GHS Cat. 1), eye irritation (UN GHS Cat. 2) and (the absence thereof; UN GHS No Cat.), for non-surfactant liquid test substances. In both DALs, the Bovine Corneal Opacity and Permeability (BCOP) test method with the laser light-based opacitometer (LLBO) is used, however only the opacity is used. The performance of the DALs to distinguish between the three UN GHS categories was compared against minimum performance values for each category proposed by CE and discussed with the OECD Expert Group on Eye/Skin Irritation/Corrosion and Phototoxicity (see table below).

Performance metrics for assessment of the predictivity of a DA for eye hazard identification

UN GHS	Defined Approach		
	Cat. 1	Cat. 2	No Cat.
Cat. 1	≥ 75%	≤ 25%	≤ 5%
Cat. 2	≤ 30%	≥ 50%	≤ 30%
No Cat.	≤ 5%	≤ 30%	≥ 70%

Materials and methods

For the selection of the reference chemicals from the CE Draize Eye Test Reference Database (DRD), the key criteria as identified by Barroso and co-workers were applied (2017). The reference set for each DA contained chemicals classified based on the most important drivers of *in vivo* Cat. 1 and Cat. 2 classification (see results section "Performance of DALs with respect to the driver of classification / subgroups No Cat."). Note that for chemicals that were classified based on the driver CO persistence on day 21 or CO = 4, this effect was present in at least 60% of the animals as advised by Barroso and co-authors (2017).

The **predictive capacity of the DAs** to distinguish between the three UN GHS categories were compared against the proposed minimum performance values. For overall measure of performance, the balanced accuracy (average of the proportion of the correct predictions of each category) was reported for each DA since it takes imbalances of the dataset into account whereas accuracy does not.

Furthermore, the **class-specific performance metrics** (sensitivity = true within class, specificity = true outside class and balanced accuracy) based on a one-vs-all other classes approach was provided for each DA. Details on the calculations and interpretation of the statistics are provided in poster #1037.

In order to gain further insight in explaining the performance in the context of the applicability domain of the individual DAs, the predictivity was also assessed based on the **drivers of classification and the subgroups** for No Cat. chemicals.

Conclusions

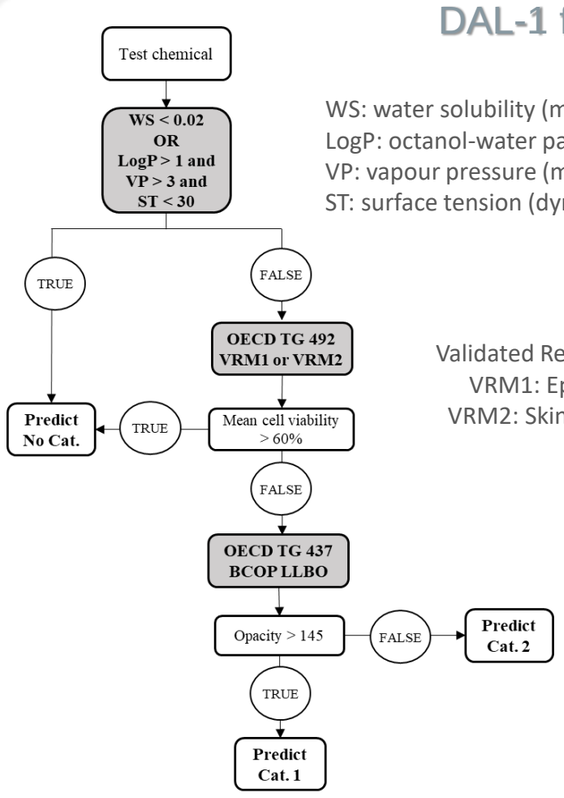
- The percentage of correct predictions for each DA was greater than the proposed minimum values for correct predictions of 75% for Cat. 1, 50% for Cat. 2 and 70% for No Cat.
- The true within and outside class performance of Cat. 1 and No Cat. resulted in a balanced accuracy which ranges from 82.2-89.9%. For the class-specific performance of Cat. 2, the different DAs have a similar balanced accuracy (66.1-74.0%), but different ranges of correct predictions within the class (56.3-68.7%) and outside the class (73.1-84.9%).
- Regarding the drivers of Cat. 1 and Cat. 2 classification, liquids that were classified *in vivo* based on CO severity (Cat. 1: mean CO ≥ 3 or CO = 4; Cat. 2: mean CO ≥ 1) were mostly correctly predicted with the DAs. No Cat. liquids from the subgroup CO = 0 had the highest predictivity.
- DAL-1 and DAL-2 have shown to successfully distinguish between the three UN GHS categories for eye hazard identification.
- Both DAs were submitted to support OECD acceptance and currently an OECD draft test guideline is under review to support regulatory acceptance.

References

Adriaens et al., 2020. Toxicol. in Vitro 70, 105-044. doi: 10.1016/j.tiv.2020.105044.
Alépée et al., 2019a. Toxicol In Vitro; 59:100-114. doi: 10.1016/j.tiv.2019.04.011.
Alépée et al., 2019b. Toxicol In Vitro; 57:154-163. doi: 10.1016/j.tiv.2019.02.019.
Barroso et al., 2017. Arch Toxicol 91, 521–547. doi: 10.1007/s00204-016-1679-x.

Results and discussion

DAL-1 for neat non-surfactant liquids



WS: water solubility (mg/mL)
LogP: octanol-water partition coefficient
VP: vapour pressure (mm Hg)
ST: surface tension (dyne/cm)

Validated Reference Method
VRM1: EpiOcular™ EIT
VRM2: SkinEthic™ HCE EIT

Combination of four physicochemical properties with the results of two OECD adopted *in vitro* test methods (Reconstructed human Cornea-like Epithelium (RhCE) and BCOP LLBO (Alépée et al., 2019a).

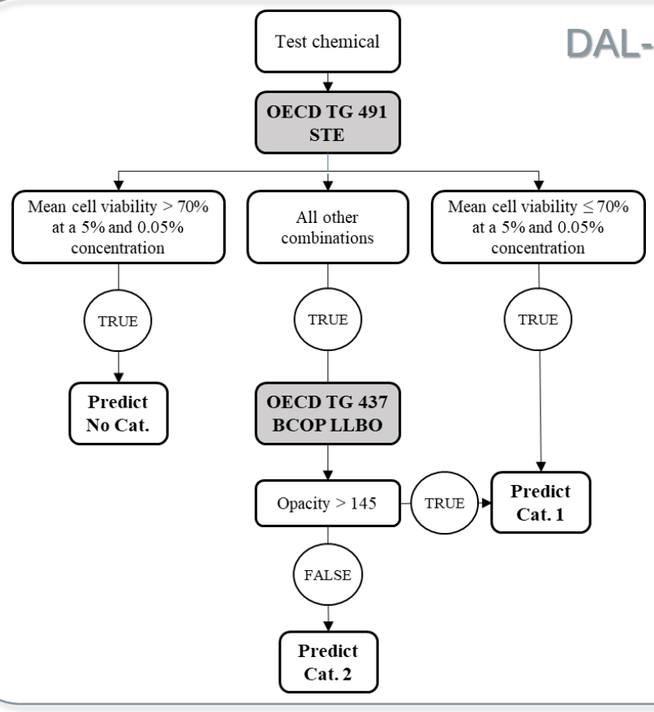
UN GHS	DAL-1 with VRM1 (N=94)		
	Cat. 1	Cat. 2	No Cat.
Cat. 1 (N=17)	76.5%	23.5%	0.0%
Cat. 2 (N=22)	27.3%	59.1%	13.6%
No Cat. (N=55)	27.9% [a]		72.1%

69.2% balanced accuracy

UN GHS	DAL-1 with VRM2 (N=86)		
	Cat. 1	Cat. 2	No Cat.
Cat. 1 (N=17)	76.5%	23.5%	0.0%
Cat. 2 (N=23)	30.4%	68.7%	0.9%
No Cat. (N=46)	19.6% [a]		80.4%

75.2% balanced accuracy

DAL-2 for non-surfactant liquids, neat and in dilution



Combination of two OECD adopted *in vitro* test methods (Short Time Exposure (STE) test method and BCOP LLBO) (Alépée et al., 2019b).

UN GHS	DAL-2 (N=164)		
	Cat. 1	Cat. 2	No Cat.
Cat. 1 (N=17)	81.2%	17.6%	1.2%
Cat. 2 (N=24)	30.2%	56.3%	13.5%
No Cat. (N=123)	14.7% [a]		85.3%

74.3% balanced accuracy

a *BCOP LLBO test results do not exist for 40/55 (DAL-1 VRM1), 34/46 (DAL-1 VRM2), and 108/123 (DAL-2) No Cat. Liquids, respectively. It is however very unlikely that a large number of false positives will be predicted Cat. 1 by the BCOP LLBO (Adriaens et al., 2020).*

Class-specific performance metrics

Statistic	DAL-1 with VRM1			DAL-1 with VRM2			DAL-2		
	Cat. 1 vs other Cat's	Cat. 2vs other Cat's	No Cat. vs other Cat's	Cat. 1 vs other Cat's	Cat. 2vs other Cat's	No Cat. vs other Cat's	Cat. 1 vs other Cat's	Cat. 2vs other Cat's	No Cat. vs other Cat's
True within class	76.5	59.1	72.1	76.5	68.7	80.4	81.2	56.3	85.3
True outside class	92.2	73.1	92.3	89.9	79.3	99.5	95.1	84.9	91.6
Balanced accuracy	84.4	66.1	82.2	83.2	74.0	89.9	88.1	70.6	88.5

Performance of DAs with respect to the driver of classification / subgroup for No Cat.

Cat. 1	CO mean ≥ 3		CO pers D21		CO=4	
	DAL-1 VRM1&2	DAL-2	DAL-1 VRM1&2	DAL-2	DAL-1 VRM1&2	DAL-2
N	7	7	4	5	6	5
TP (%)	85.7	85.7	50.0	56.0	83.3	100
UP (%)	14.3	14.3	50.0	40.0	16.7	0.0
FN (%)	0.0	0.0	0.0	4.0	0.0	0.0

Cat. 2	CO mean ≥ 1			Conj mean ≥ 2		
	DAL-1 VRM1	DAL-1 VRM2	DAL-2	DAL-1 VRM1	DAL-1 VRM2	DAL-2
N	17	18	17	5	5	7
OP (%)	32.4	36.1	27.9	10.0	10.0	35.7
TP (%)	60.4	62.8	55.9	54.8	90.0	57.1
FN (%)	7.2	1.1	16.2	35.2	0.0	7.1

CO: corneal opacity; IR: iritis; Conj: conjunctival redness (CR) and/or conjunctival chemosis (CC)
CO mean and Conj mean scores are calculated from gradings at 24, 48, and 72 hours after instillation of the test chemical

No Cat.	CO > 0 **			CO > 0			CO = 0 **			CO = 0		
	DAL-1 VRM1	DAL-1 VRM2	DAL-2	DAL-1 VRM1	DAL-1 VRM2	DAL-2	DAL-1 VRM1	DAL-1 VRM2	DAL-2	DAL-1 VRM1	DAL-1 VRM2	DAL-2
N	7	5	15	8	4	15	1	1	2	38	36	91
FP (%)	57.1	60.0	20.8	62.5	33.3	33.3	0.0	0.0	0.0	16.7	13.0	11.0
TN (%)	42.9	40.0	79.2	37.5	66.7	66.7	100	100	100	83.3	87.0	89.0

TP: True Prediction
OP: Over-prediction
UP: Under-prediction
FP: False Positive
FN: False Negative
TN: True Negative

CO = 0: CO scores equal to 0 in all animals and all observed time points
CO > 0: in at least 1 observation time in at least 1 animal and all animals showing mean scores of days 1–3 below the classification cut-offs for all endpoints
** Indicates at least 1 animal with a mean score of days 1–3 above the classification cut-off for at least one endpoint

Cosmetics Europe
the personal care association

Other relevant research: Combination of four physicochemical properties (Poster #383) with the results of two OECD adopted *in vitro* test methods (Reconstructed human Cornea-like Epithelium (RhCE) and BCOP LLBO (Alépée et al., 2019a) and illustration of the DALs with case studies, Poster #221