Regulatory Toxicology and Pharmacology 105 (2019) 69–76

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

journal homepage: www.elsevier.com/locate/yrtph

Eye hazard classification according to UN GHS / EU CLP and the severity of eye symptoms caused by accidental exposures to detergents and cleaning products

Roberto Scazzola[a,*], Geert Boeije[b], Ellen Bøtker Pedersen[c], William Brenneman[d], Blažena Cagáňová[e], Anna Celentano[f], Els De Coninck[g], Herbert Desel[h], Niels Ebbehøj[c], Elke Färber[i], Fabrizio Sesana[f], Sergej Zacharov[j]

[a] A.I.S.E. (International Association for Soaps, Detergents and Maintenance Products), Belgium
[b] Boeije Consulting bvba, Belgium
[c] Danish Poisons Centre, Bispebjerg Hospital, København, Denmark
[d] Procter & Gamble, USA
[e] National Toxicological Information Centre, University Hospital Bratislava, Slovakia
[f] Centro Antiveleni di Milano, Azienda Ospedaliera Niguarda Ca' Granda, Italy
[g] Procter & Gamble, Belgium
[h] BfR (German Federal Institute for Risk Assessment), Germany, Formerly GIZ-Nord Poisons Center, University Medical Center Göttingen, Germany
[i] GIZ-Nord Poisons Centre, University Medical Center Göttingen, Germany
[j] Toxicological Information Centre, Department of Occupational Medicine, First Faculty of Medicine, Charles University and General University Hospital Prague, Czech Republic

ARTICLE INFO

Keywords:
EU CLP
UN GHS
EU DPD
Eye hazard classification category
Poisoning severity score (PSS)
Accidental eye exposure
Detergents and cleaning products

ABSTRACT

The use of lower cut-off values/concentration limits for the calculation of mixture classification in UN GHS/EU CLP versus the previous regulatory scheme (EU Dangerous Preparations Directive, DPD), has resulted in an increased number of classifications in the highest eye hazard category. Herein, a semi-quantitative categorisation of severity of eye effects, following accidental human exposures to detergents, was compared to the classification category of the products. Three schemes were evaluated: EU DPD; EU CLP (based on all available data and information, including weight of evidence); and EU CLP (based entirely on the calculation method). As reported by four EU Poison Centres, the vast majority of exposures had caused minor or no symptoms. Classification was a poor predictor of effects in man subjected to accidental exposure. Note however that this is also because effects are not only driven by the intrinsic hazard (as reflected in the classification), but also by the exposure conditions and mitigation (i.e. rinsing). EU CLP classification using all available data and information was more predictive of medically relevant symptoms than the EU CLP calculation method. The latter led to a poorer differentiation between irritating products versus products potentially causing serious eye damage.

1. Introduction

In the European Union's chemicals control framework, hazards are determined by means of a classification scheme as defined by EU CLP, Regulation (EU) No 1272/2008 on Classification, Labelling and Packaging of Substances and Mixtures (European Union, 2008). EU CLP is based on a standard developed at UN level (UN GHS, United Nations Globally Harmonized System, United Nations, 2002). Before the introduction of EU CLP for mixtures on June 1st 2015, the classification of consumer products was governed by the Dangerous Preparations Directive (1999/45/EC, EU DPD) (European Union, 1999).

Under the EU DPD, the primary basis for eye hazard classification of mixtures was to apply additivity of the eye irritancy classification of a product's ingredients ("conventional calculation method"). If available, data from the standard animal test for eye irritation, the OECD (Organisation of Economic Co-operation and Development) TG 405 "Draize" test (OECD, 2017a; Draize et al., 1944), overruled the calculation's outcome. However, animal welfare considerations and the limited relevance for effects in man precluded the use of such tests on detergents and cleaning products. EU DPD also allowed the use of

* Corresponding author. A.I.S.E., Boulevard du Souverain, 165, B-1160, Auderghem, Belgium.
E-mail address: roberto.scazzola@aise.eu (R. Scazzola).

https://doi.org/10.1016/j.yrtph.2019.04.004
Received 7 December 2018; Received in revised form 26 March 2019; Accepted 5 April 2019
Available online 09 April 2019

Abbreviations

EU CLP	European Union Regulation (EC) No 1272/2008 on classification, labelling and packaging of substances and mixtures
EU DPD	European Union Dangerous Preparations Directive 1999/45/EC
Cat1	Category 1
Cat2	Category 2
NC	not classified
PC	Poison Centre
PSS	Poisoning severity score
UN GHS	United Nations Globally Harmonized System

scientifically valid case studies (e.g. data from poison centres) to classify according to the effects on man, and antagonistic effects (e.g. between different types of surfactants in a mixture) were allowed to be taken into consideration if justifiable to avoid overestimation of the toxicological hazard. Based on this "furthermore clause" (Article 6.3 of the EU DPD), a practical approach to determine the EU DPD classification, making optimal use of all available data, was put forward by the European detergents association (A.I.S.E., 2008). This included bridging from "reference formulations" for which test data (including relevant but non-standard tests such as the Low Volume Eye Test, LVET - Griffith et al., 1980) are available.

Under EU CLP, the primary basis for eye hazard classification is data on the mixture itself. This includes the standard *in vivo* animal test (OECD TG 405), validated animal test alternatives such as *in vitro* studies (within the scope of their applicability), epidemiological data and experience on the effects on humans, such as data from accident databases (Article 6.1 of CLP). If data on the actual mixture are not available, data on similar mixtures can be used, by applying bridging principles. In absence of appropriate data for bridging, a weight-of-evidence expert judgment, based on all available data, can be developed. This is described in OECD's Integrated Approach on Testing and Assessment (OECD, 2017b). If no conclusive expert classification decision can be made, the EU CLP calculation method must be applied. This method, introduced by UN GHS (United Nations, 2002), is similar to the "conventional calculation method" of the EU DPD, but with a lower (more conservative) threshold for serious eye damage categorisation (i.e., Eye Category 1). Under the EU DPD, up to 10% of ingredients with the highest eye hazard classification were allowed in the formula before the product itself was assigned the highest classification. Under EU CLP, this threshold is lower at 3%. Cazelle et al. (2014) and Corvaro et al. (2017) pointed out that this revised limit might be over-conservative when compared to the standard *in vivo* test for household products and agrochemical end-use formulations, respectively.

The current study investigates the predictivity of a product's regulatory classification for eye hazard, following reported accidental human exposures to common household detergent and cleaning products. Three hazard classification approaches are evaluated:

(1) the EU DPD classification according to the A.I.S.E. (2008) approach, as used in practice before 2015;
(2) the EU CLP classification based on all available data including bridging principles, and weight-of-evidence expert judgment as needed (further referred to as "actual EU CLP classification"); and finally
(3) the EU CLP classification based purely on the calculation method, disregarding available data.

The underlying human dataset for the current assessment was obtained from a prospective multicentre poison centre (PC) study, that reported eye effects caused by accidental exposures to detergents, cleaning and maintenance products. This study, MAGAM[1] II DISC (Denmark, Italy, Slovakia and the Czech Republic) (Färber and Desel, 2015; Färber et al., 2016), was conducted between 2013 and 2015, in parallel with a second study branch MAGAM II DEAT (Germany and Austria) (Hermanns-Clausen et al., 2019). The PCs participating in both MAGAM II studies covered approximately one-third of the EU population. In total, 1235 accidental human exposures were documented. The poisoning severity in more than 90% of the registered cases was determined according to the Poisoning Severity Score (PSS) (Persson et al., 1998). This was carried out in a harmonised way by means of follow-up telephone interviews and reviews of medical reports where relevant. No symptoms were observed in 9% of the reported exposures. 82% of the cases led to minor symptoms. Moderate effects were reported in 9% of the cases. Signs of eye irritation (such as redness, a burning sensation and increased lacrimation) were noted most frequently, and healing was nearly always reported within hours or a few days. Two cases with residual eye damage after 21 days were recorded: a child with persistent sensitivity to light, and an adult with the same symptom accompanied by reduced vision. The authors concluded that most patients only experienced minor symptoms and that serious eye damage (i.e., long-lasting potentially irreversible symptoms) only occurred very rarely after accidental ocular exposure to detergents or cleaning agents. It should be noted that early eye irrigation may have been a mitigating factor that prevented a higher severity. For example Färber et al. (2016) reported that eyes were rinsed in 96% of the cases, nearly always immediately after the exposure.

2. Methods

2.1. Product identification and regulatory classification

Of the 657 accidental exposures recorded in Denmark, Italy, Slovakia and the Czech Republic (MAGAM II DISC) (Färber and Desel, 2015; Färber et al., 2016), information about the product identification and the severity of ocular responses was available for the current study. For 598 cases, a detailed follow-up was successful with a conclusive PSS determination. 185 of these exposures occurred with unambiguously identified products for which it was possible to obtain the regulatory classification directly from the manufacturing company. This data subset (n = 185) forms the basis for the assessment reported in this study.

The manufacturers provided the classification of the involved products at the time of the incident. As this was before mid 2015, the EU DPD regulatory framework was still in place. The actual EU DPD eye classification of these products was retrievable for 91% of the cases (n = 169). The manufacturing companies were also asked to determine the EU CLP classification for each product. They provided the product classification as it would actually have been under the EU CLP regulation (implemented as of June 1st, 2015). Depending on each individual situation, product classification could have been based on available (historical) animal test data, on the use of the permitted variation or substantially similar mixtures bridging principles, on the use of *in vitro* data, or on weight-of-evidence and expert judgment (cf. OECD, 2017b). Or, product classification could also have been based on the calculation method, in the absence of more appropriate information. Separately, for the purpose of the current assessment, the producers were additionally asked to provide the EU CLP classification using only the calculation method. Both options for the EU CLP classification were determined for all products in the data subset (n = 185).

[1] Multinationale Analyse von Daten der Giftinformationszentren zur Frage korrosiver Augenläsionen durch feste Maschinengeschirrspülmittel und andere Wasch-, Pflege-und Reinigungsmittel

2.2. Severity of the symptoms

All signs and symptoms were reported and recorded either during the initial call to the PC, or during the follow-up interview. To facilitate a harmonized PSS grading, according to Persson et al. (1998), each reported symptom was assigned a PSS score by means of a standardized list used by all participating PCs (Färber and Desel, 2015) - as shown in Table 1. The severity grading assigned to a case was determined by the most severe symptom(s) or sign(s) observed.

2.3. Statistical analysis

For comparison of proportions, (e.g. to compare the proportions of products with two different classifications within a subset of accidents with a given severity) the Fisher's Exact Test was applied with alpha = 0.05 as a threshold for significance. This method was also applied for the assessment of the predictivity of classification for eye effects. The Fisher's Exact Test does not rely on large sample theory and is generally valid for both small and large sample sizes. The 2-tailed p-value was calculated as defined in Agresti (1992).

For the predictivity of classification for eye effects, several goodness of fit metrics are reported that are commonly used with Generalized Linear Models (McCullagh and Nelder, 1989). These metrics include Entropy RSquare (one minus the ratio of the negative log-likelihoods from the fitted model and the constant probability model), Generalized RSquare (a measure based on the likelihood function), Mean Abs Dev (the average of the absolute values of the differences between actual and predicted responses), AICc (the Akaike information criterion, which estimates the quality of each model, dealing with the trade-off between the goodness of fit of the model and the simplicity of the model), BIC (the Bayesian information criterion, based in part on the likelihood function and closely related to the AICc), Mean -Log p (the average of negative log of the fitted probability associated with the event that occurred), and RMSE (the root mean square error). The models and 'goodness of fit' metrics were all computed with the software JMP® Pro 13.2.1. Each 'goodness of fit' measure was identified as either "larger is better" or "smaller is better".

In addition, Cooper descriptive statistics (false positives, false negatives and concordance) for predictivity of moderate or severe ocular effects (PSS ≥ 2) by the regulatory classification are reported, and the F_1 statistic was determined. F_1 is the harmonic average of precision (in this case, percentage of cases with PSS ≥ 2 if the classification is in the highest category) and recall (1 - false negatives), and thus balances these two measures. An F_1 score reaches its best value at 1 and worst at 0.

3. Results and discussion

3.1. Poisoning severities related to product classification according to EU DPD

For 169 cases the product's regulatory classification under EU DPD could be retrieved. Nearly two thirds of the products (64%) were classified as hazardous to the eyes: 44% as R41 (Risk of serious damage to eyes) and 20% as R36 (Irritating to eyes). The remaining 36% of the products were not classified (NC) for eye hazard. Following exposure, 4.7% of cases were asymptomatic, a clear majority (85%) led to minor ocular symptoms, and 9.5% caused moderate ocular symptoms. One accidental exposure in the study subset led to severe effects.

The poisoning severity of cases with respect to the EU DPD eye hazard classification category is shown in Fig. 1.

Moderate or severe symptoms (one case) were reported in 15% of the accidental exposures with products classified as R41. Of the products classified as R36, less exposures with moderate symptoms (9%) were recorded. Of non-classified products, moderate symptoms were seen in only 5% of the exposures. However, these differences were not statistically significant (respectively p = 0.75 and p = 0.088 for R41 vs R36 and for R41 vs NC).

Overall, the differences between the classification subgroups are relatively small, with "minor" (PSS = 1) by far the predominant severity score irrespective of the classification.

3.2. Poisoning severities related to product classification according to EU CLP

For 185 cases, the EU CLP classification of the involved product could be determined. The actual EU CLP classification was based on all available data (including non-animal data, as well as bridging principles, weight-of-evidence and expert judgment), or the calculation method when needed. Separately, the hypothetical EU CLP classification when based only on the calculation method was also determined. Under EU CLP (irrespective of the approach used), about 50% more products are classified as hazardous for the eyes than under the EU DPD. Based on actual EU CLP implementation, 52% of the products are classified as Category 1 and 41% as Category 2, with only 7% of non-classified products (compared to 36% under the EU DPD). Applying only the calculation method under EU CLP, a clear majority (70%) of the Category 2 products would be more severely classified - leading to 82% of products in Category 1, and only 12% remaining in Category 2. Whilst application of the calculation method hardly impacts the proportion of non-classified products (6.5%).

The overall distribution of severity scores of ocular effects of the EU CLP case data set is nearly identical to the EU DPD subset. In the CLP data set, there were 5.4% asymptomatic exposures, 85% cases with minor symptoms, and 8.6% with moderate effects - in addition to one accidental exposure with severe effects. Fig. 2 shows the poisoning severities per classification group under the actual EU CLP implementation. Fig. 3 shows this when only the calculation method is applied.

The products classified as Category 1 under the actual EU CLP implementation were associated with moderate or severe symptoms in 13% of the exposures. This is virtually identical to what was found for the EU DPD R41 classification (p = 0.83). This is not a surprise, because the subgroup of products classified as Category 1 under EU CLP has a

Table 1
Assignment of reported ocular symptoms to Poisoning Severity Scores (PSS).

PSS score	Symptom description
Asymptomatic (PSS = 0)	no symptoms observed
Minor (PSS = 1)	irritation, eye redness, burning eyes, pain in the eyes, (minor) sensation of foreign body, itching eyes, sensation of dry eyes, blinking, lacrimation, swelling/mild palpebral edema, blurred or decreased vision, photophobia, eyelid secretion, minor eyelid corrosion, conjunctivitis; persistent crying, uneasiness
Moderate (PSS = 2)	intense irritation, severe foreign body sensation, blepharospasm, severe pain, purulent conjunctivitis, chemosis, erosion of conjunctival epithelium, moderate conjunctival corrosion, erosion of corneal epithelium, punctate damage of the corneal epithelium, corrosive eye injury not further specified, moderate corrosion
Severe (PSS = 3)	corneal ulcers (other than punctate), perforation, permanent damage, or any of the symptoms listed for PSS = 1 or PSS = 2 when healing had not been achieved within 21 days after exposure

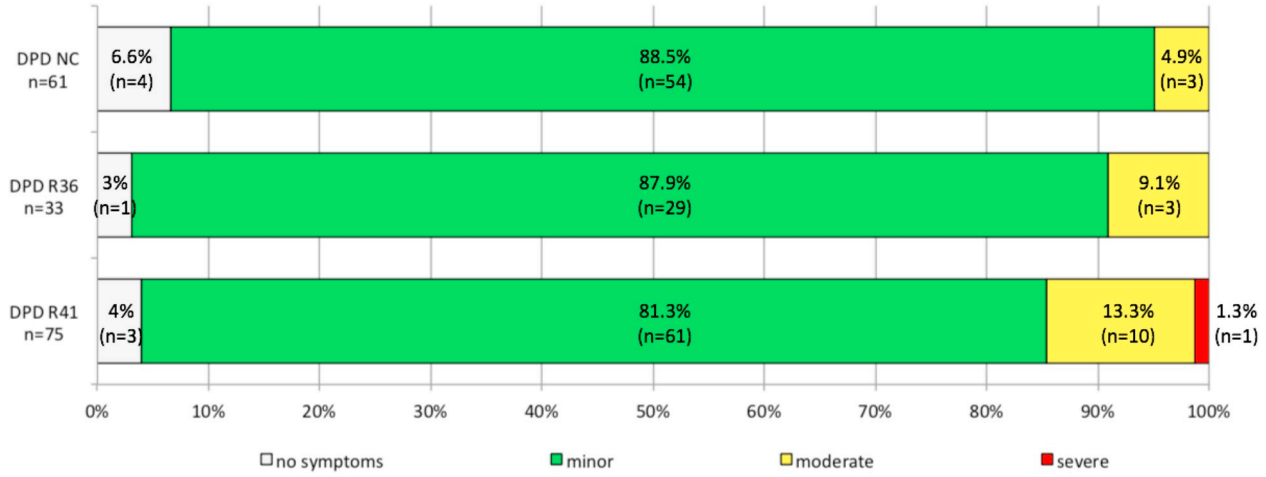

Fig. 1. Number and proportion of cases per severity score, for products with different EU DPD classifications.

large overlap with the EU DPD R41 subgroup. Among the larger number of products that would fall in Category 1 based strictly on the EU CLP calculation method, 11% caused moderate or severe symptoms (not significantly different from the actual EU CLP approach, $p = 0.55$).

Of the products classified as Category 2 under EU CLP based on all available data, only 5% caused moderate symptoms. This is less than half of the 13% found for Category 1, but nevertheless not statistically significant ($p = 0.12$). Among the smaller number of products that would be classified as Category 2 based on the EU CLP calculation method, one exposure caused moderate symptoms (equivalent to 5%).

For the products not classified as hazardous to the eyes under EU CLP (both approaches), all exposures were associated with minor symptoms.

Importantly, as under the EU DPD scheme, symptoms were predominantly "minor" irrespective of the product classification under EU CLP.

3.3. Classification as a predictor for severity of ocular effects in humans

The objective of chemical heath hazard classification is to identify, describe and rank the intrinsic hazards of a substance or mixture. Hazardous agents only lead to adverse health effects under specific exposure conditions. To cause a health effect fully linked to the assigned hazard the exposure must be sufficiently high with respect to dose, concentration and duration. For eye exposures this implies that early eye irrigation may mitigate or even entirely prevent effects. Consequently, not all accidental exposures to classified agents will cause symptoms with a severity that is fully consistent with the hazard classification.

Nevertheless, it can be expected that there should be a correlation between the severity of the ocular effects observed in reality and the level of classification. In other words, one may expect a higher frequency of moderate or severe ocular effects (PSS ≥ 2) with accidental exposures to products that have a Category 1 (or R41 under EU DPD) hazard classification.

3.3.1. Predictivity of medically relevant symptoms

From a public health point of view, accidental exposures leading to no or minor symptoms are of a lower concern. The highest medical relevance is with cases that have symptoms scored for moderate or high severity. This is reflected by the ToxIndex (Desel, 2013), defined as the percentage of cases leading to effects with PSS ≥ 2. If classification is to be predictive of such symptoms, the subgroup of products with the highest classification level (i.e. R41/Category 1) should be associated with the highest ToxIndex. In this respect, the EU DPD approach performed marginally better than the actual EU CLP approach, which in turn was better than the calculation method EU CLP approach (Table 2, top row). One would also expect products with the lower classifications (i.e. those either not classified or classified R36/Category 2) to be linked with a lower percentage of PSS ≥ 2 cases. In this respect, the EU DPD approach had the poorest performance and the EU CLP calculation method was the best (Table 2, bottom row).

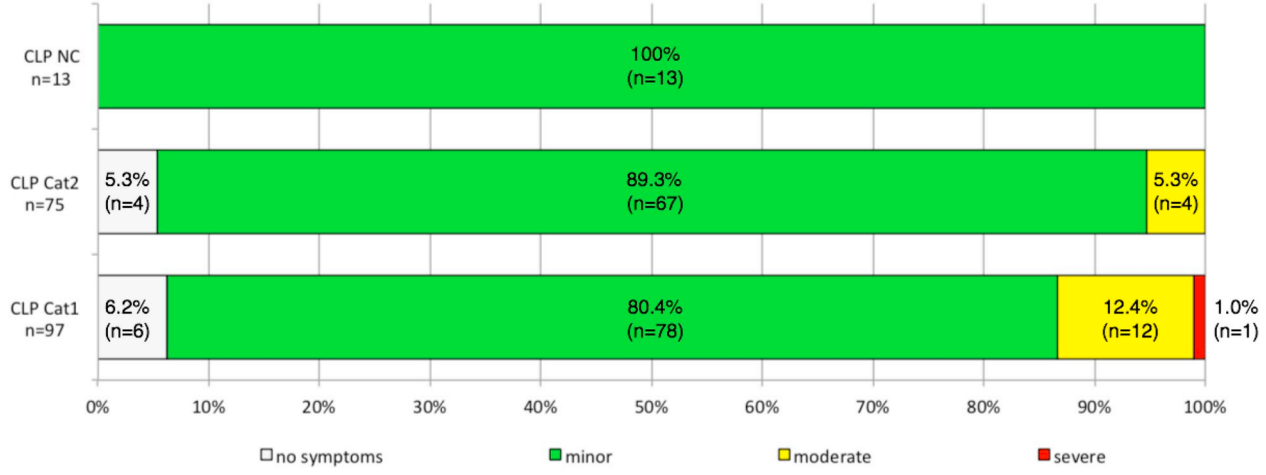

Fig. 2. Number and proportion of cases per severity score, for products with different EU CLP classifications, based on the actual EU CLP implementation.

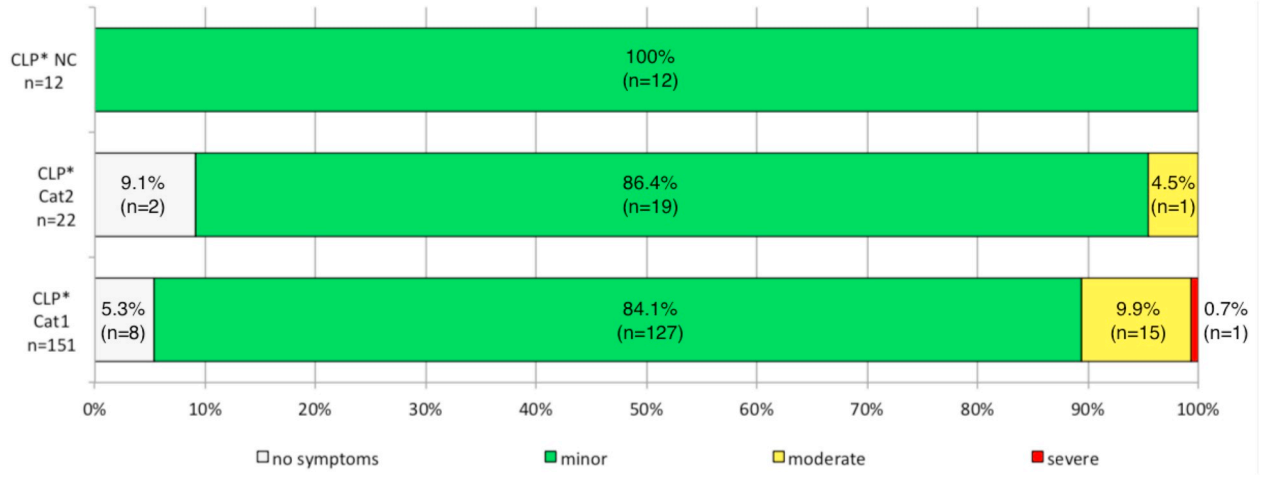

Fig. 3. Number and proportion of cases per severity score, for products with different EU CLP classifications, based only on the calculation method.

Table 2
ToxIndex (i.e., percentage of cases with moderate or severe effects) per classification method.

	EU DPD	EU CLP actual	EU CLP calculation
Cat1/R41	14.7% (11/75)	13.4% (13/97)	10.6% (16/151)
NC or Cat2/R36	6.4% (6/94)	4.5% (4/88)	2.9% (1/34)

Overall, among the three methods, the actual EU CLP approach was found to provide the best predictivity of "moderate or severe effects" - as shown by all of the applied statistical metrics (Table 3). This assessment was based on the EU DPD-subset, for which all three classification options were available (n = 169). Note that the predictive power remains limited across all classification options: the vast majority of products with the highest classification (over 85%) did not lead to symptoms beyond PSS = 1.

Based on this same data set (n = 169), Cooper descriptive statistics are provided in Table 4. It should be noted that in the context of this assessment, 'false positives' and 'false negatives' do not imply a wrong classification, but are used to indicate that the classification did not match the observed effects after accidental exposure. Thus, a Cat1 (or R41) classification associated with effects of PSS ≤ 1 would be counted as a false positive, whereas a classification as NC or Cat2 (or R36) associated with effects of PSS ≥ 2 would be seen as a false negative. Whereas the EU DPD approach had the best concordance, its false negatives rate was the highest. In contrast, the EU CLP calculation method led to a low false negatives rate. However, this came at the cost of very high false positives and a poor overall concordance. The actual EU CLP approach appeared to be the most balanced among the three options. This is confirmed by the highest F_1 score for this approach, closely followed by the EU DPD, while the lowest F_1 was determined for the EU CLP calculation method.

Table 4
Cooper descriptive statistics for predictivity of moderate or severe ocular effects by the classification.

	EU DPD		EU CLP actual		EU CLP calc.	
	NC or R36	R41	NC or Cat2	Cat1	NC or Cat2	Cat1
PSS ≤ 1	88	64	77	75	30	122
PSS ≥ 2	6	11	4	13	1	16
False negatives	35.3%		23.5%		5.9%	
False positives	42.1%		49.3%		80.3%	
Concordance	58.6%		53.3%		27.2%	
F_1 score	0.24		0.25		0.21	

3.3.2. Predictivity of symptoms severity (differentiation between none, minor, moderate + severe)

In addition, an attempt was made to assess how well the classification is predictive of overall severity of ocular symptoms - i.e. whether non-classified corresponds with no symptoms, R36/Category 2 with minor effects, and R41/Category 1 with moderate (or severe) effects. For both EU CLP approaches, 100% of the exposures to non-classified products resulted in only one class of effects (minor severity). This leads to separation in the data, and as a consequence, unstable model effect estimations. Hence, a p-value (using the Fisher's exact test) could not be determined, and the Generalized Linear Model goodness of fit metrics are not reliable in this situation.

3.3.3. Comparison of the different regulatory schemes

The observed differences in predictivity of the ocular symptoms' severity may be explained by the effect of the different regulatory schemes on product classification.

Under EU CLP (actual approach, based on all available data), 93% of the products would be classified as hazardous for the eyes. This is significantly higher than under the EU DPD where this is only 64%

Table 3
Statistical evaluation of predictivity of moderate or severe ocular effects by the classification.

Measure	Direction	EU DPD	EU CLP actual	EU CLP calc.
Fisher's Exact Test p-value	Smaller is better	0.1205	0.0412	0.3171
Entropy RSquare	Larger is better	0.0286	0.0431	0.0223
Generalized RSquare	Larger is better	0.0386	0.0578	0.0302
Mean Abs Dev	Smaller is better	0.1776	0.1761	0.1788
AICc	Smaller is better	111.232	109.636	111.926
BIC	Smaller is better	117.419	115.824	118.113
Mean -Log p	Smaller is better	0.317	0.3123	0.3191
RMSE	Smaller is better	0.298	0.2967	0.299

($p < 0.0001$). This is mainly because under EU CLP, 41% of the products are classified in Category 2 (Causes serious eye irritation) whereas under EU DPD only 20% were classified as the equivalent R36 (significant, $p < 0.0001$). On the other hand, the percentage of products with the highest classification level (Category 1 respectively R41) was not significantly different (52% under EU CLP, and 44% under EU DPD, $p = 0.14$).

The EU CLP calculation approach led to nearly the same proportion of products classified for eye hazard as the actual EU CLP implementation (93.5%, compared to 93%). However, using the calculation method, a significantly larger group of these products would be considered as Category 1 (82% with the calculation method, compared to 52% under actual EU CLP, $p < 0.0001$). In turn, this leads to a much lower proportion of Category 2 products (12% rather than 41%, $p < 0.0001$). The main effect of the calculation method under EU CLP is a reduced differentiation between the two categories of eye irritants, with a substantial subset of Category 2 products moving into the Category 1 classification.

The link between the actual EU CLP and the EU DPD classification approaches and the occurrence of medically relevant effects (PSS ≥ 2) in each of the classification subgroups, is shown in Fig. 4. The most notable difference between the two approaches is that 56% of the products classified as Category 2 under EU CLP, are not classified under the EU DPD. The vast majority of the products in this situation (nearly 95%) had a PSS ≤ 1. This explains why the EU DPD classification may be perceived as more in line with the reported severity of ocular effects overall. In contrast, 18% of the products classified as Category 1 under EU CLP are non-classified or R36 under EU DPD. Of this product subset, respectively 10% (for EU DPD non-classified) and 17% (for EU DPD R36) led to PSS ≥ 2. This percentage of moderate or severe effects is higher than what is typically reported for the lower classification levels, and thus points to a weakness of the EU DPD scheme in identifying all of the potentially more hazardous products. This finding explains the better predictivity of EU CLP (actual implementation) for ocular effects with PSS ≥ 2.

A similar comparison is made between the actual EU CLP implementation and the hypothetical alternative based only on the EU CLP calculation method, in Fig. 5. A large proportion (71%) of products classified as Category 2 by the actual EU CLP implementation become Category 1 when the calculation method is applied. Over one-third of the products determined to be Category 1 by the calculation method are in this situation. Only 6% of the exposures with these products led to PSS ≥ 2. This is much lower than the 13% of the products classified as Category 1 under actual EU CLP, and similar to the 5% for the products classified as Category 2 under actual EU CLP. The subset of products that would be moved from Category 2 to Category 1 when applying the EU CLP calculation method (instead of the actual approach), were associated with a severity profile that is more typical of Category 2 than of Category 1. This explains the poorer predictivity of medically relevant effects by the EU CLP approach based purely on the calculation method.

The above finding is in alignment with Cazelle et al. (2014) who reported 100% over-predictions by the EU CLP calculation method for 22 non-Category 1 detergent products (as determined by either Draize or LVET data). Further, Corvaro et al. (2017) found that among 85 agrochemical preparations that would be classified as Category 1 for eye hazard according to the EU CLP calculation method, only 23 products (27%) were actually classified like this based on standard *in vivo* test data. Hence, the high conservatism of the EU CLP calculation method leads to over-classification. This has been observed both within and outside of the detergents product category.

As a limitation of this study, it should be noted that in nearly all of the cases (as mentioned by Färber et al., 2016) the eyes were irrigated immediately after the accidental exposure. This is the recommended first-aid treatment, and it is typically the spontaneous reaction most people mention in case of an accidental eye exposure (Baert et al., 2017). Immediate rinsing might have contributed to the overall lower severity of the reported symptoms. Consequently, it might also have reduced the differentiation in terms of severity between products in different eye hazard classification groups.

4. Conclusion

The vast majority (approximately 90%) of the exposures within the scope of the current study did not cause symptoms with worse than minor severity. Also within the subgroup of products with the highest hazard classification, irrespective of the applied classification scheme, moderate or severe effects (with PSS ≥ 2) were rarely incurred (less than 15% of the exposures). This shows that, for a substantial proportion of these products, the hazards suggested by the classification level did not materialise following accidental exposure (likely mitigated by immediate eye irrigation following the exposure in most cases). Consequently, none of the classification schemes showed the highest category of regulatory classification to be a good predictor of moderate to high severity effects following accidental exposure.

This paper evaluates the predictivity of actual reported eye effects by the regulatory hazard classification. The actual reported severity is not only driven by the intrinsic hazard but also by the amount and duration of exposure, and by the mitigating action (i.e. rinsing) typically occurring quickly after exposure. It is acknowledged that the hazard classification aims to capture the intrinsic hazard “in isolation” without considering exposure or mitigation. Thus, one may argue that

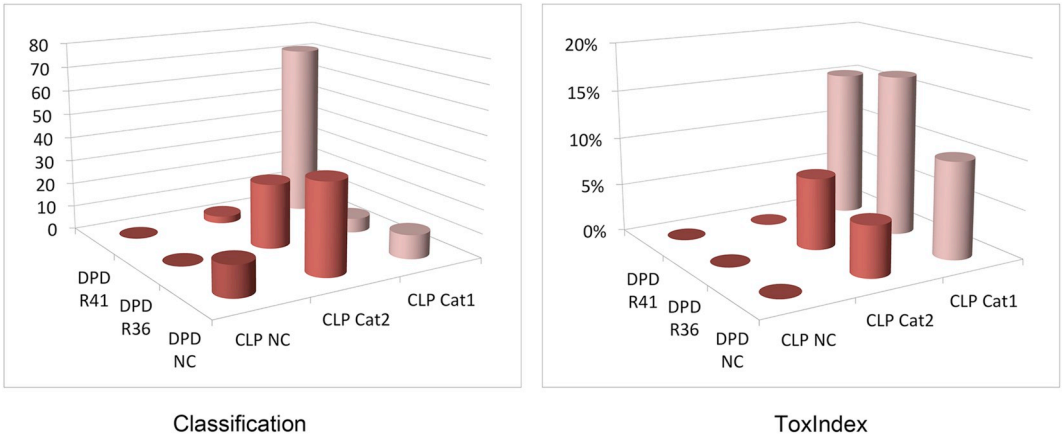

Fig. 4. Classification and severity - EU DPD versus EU CLP (actual). Left: number of cases with the different EU DPD classifications, for each EU CLP classification subgroup. Right: ToxIndex (i.e., percentage of cases with PSS ≥ 2).

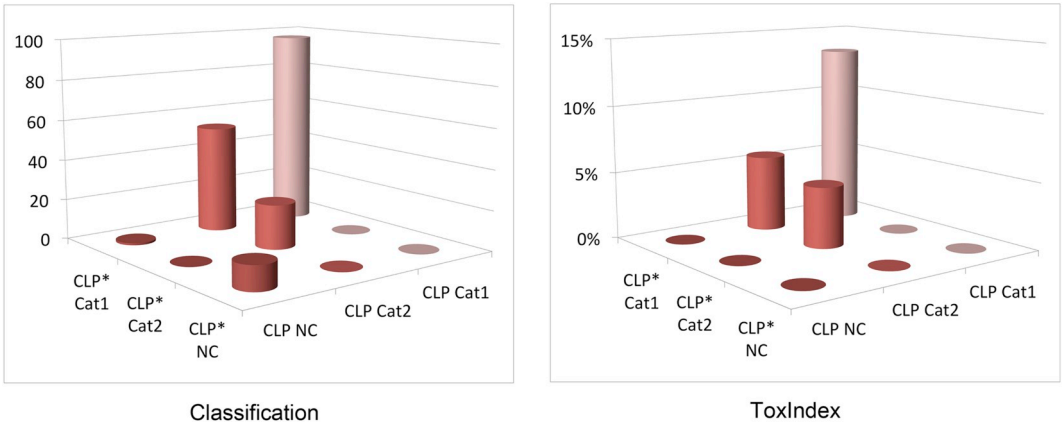

Fig. 5. Classification and severity - EU CLP (actual) versus EU CLP* (calculation method). Left: number of cases with the different EU CLP*classifications, for each EU CLP classification subgroup. Right: ToxIndex (i.e., percentage of cases with PSS ≥ 2).

both aspects are not directly comparable. Nevertheless, because hazard communication on consumer product labels is mainly determined by the regulatory classification, it is of practical relevance that there be a correlation between the two.

Intuitively, the EU DPD scheme, under which a larger proportion of products are non-classified, corresponds best with the observation that most products only caused minor (or no) symptoms. However, the actual EU CLP approach (i.e. based on all available data, including bridging, weight-of-evidence and expert judgment - or based on the calculation method in absence of any such data) was found to be the most predictive for symptoms with PSS ≥ 2 (moderate, or very sporadically severe effects). Indeed, the subset of products that were classified as R36 or non-classified under EU DPD, but that would be classified as Category 1 under EU CLP, had a severity profile more in line with the Category 1 subset. In addition, for the EU DPD approach, 5% of the non-classified products were involved with moderate symptoms.

When only applying the calculation approach under EU CLP, a much larger proportion of products would be classified for serious eye damage Category 1 (82% rather than 52%). Thus, the EU CLP calculation approach led to a poorer differentiation between products, as demonstrated by the high percentage of Category 1 products associated with only minor (or no) symptoms. This suggests that EU CLP's calculation approach can overestimate the hazard for several more products than the actual EU CLP classification approach - and thus, could result in over-classification of products. While this is in line with the precautionary principle, the limited realism of the calculation method has to be taken into account when evaluating actual risk. Indeed, the subset of products that would be classified as Category 1 by the calculation method but only as Category 2 based the actual EU CLP approach, had a severity profile in line with the Category 2 subset. These findings support the classification hierarchy under CLP, where actual data and weight-of-evidence are to be considered as primary evidence to determine the classification, and the calculation method should only be used as a fallback option.

The actual EU CLP approach (based on all available data, including bridging, weight-of-evidence and expert judgment) provided the better predictivity for medically relevant ocular effects (i.e. with PSS ≥ 2). Nevertheless, this predictivity was relatively poor - with only 13% of the Category 1 products in this study actually associated with effects of a moderate (or worse) severity. It should be noted that this observation may be driven by the fact that UN GHS/EU CLP eye hazard classification is intrinsically based on the Draize test - which was found to be not representative of effects in humans (Roggeband et al., 2000). The low predictivity is also caused by the fundamental difference between the intrinsic hazard (as identified by the classification) and how this hazard materializes in actual effects (as observed after accidental exposures, driven by hazard, exposure conditions and mitigation - especially the immediate rinsing of the eyes after exposure).

The reported human evidence regarding severity of eye effects caused by existing products may be a valuable element as supporting evidence in the classification of specific products as “Eye Irritant Category 2” or “Not Classified” when using expert judgment and weight-of-evidence.

References

Agresti, A., 1992. A survey of exact inference for contingency tables. Stat. Sci. 7, 131–153.
A.I.S.E., 2008. CLASSIFICATION, PACKAGING and LABELLING of DANGEROUS PREPARATIONS - Guidelines for the Application of Directive 1999/45/EC. General Presentation. Update of November 2008.
Baert, W., Boeije, G., Peeters, V., Scazzola, R., Séjourné, V., Vandecasteele, B., 2017. Consumer Understanding of the Safety Label and Pictograms on Household Detergent Products. Consumer Research Report. prepared for A.I.S.E. 5 October 2017.
Cazelle, E., Eskes, C., Hermann, C., Jones, P., McNamee, P., Prinsen, M., Taylor, H., Wijnands, M., 2014. Suitability of histopathology as an additional endpoint to the Isolated Chicken Eye Test for classification of non-extreme pH detergentand cleaning products. Toxicol. Vitro 28, 657–666 2014.
Corvaro, M., Gehen, S., Andrews, K., Chatfield, R., Macleod, F., Mehta, J., 2017. A retrospective analysis of *in vivo* eye irritation, skin irritation and skin sensitisation studies with agrochemical formulations: setting the scene for development of alternative strategies. Regul. Toxicol. Pharmacol. 89, 131–147 2017.
Desel, H., 2013. ToxIndex – a tool to compare population poisoning risks for toxic agents. Clin. Toxicol. 51, 302.
Draize, J., Woodard, G., Calvery, H., 1944. Methods for the study of irritation and toxicity of substances applied topically to the skin and mucous membranes. J. Pharmacol. Exp. Ther. 82, 377–390.
European Union, 1999. Directive 1999/45/EC of the European Parliament and of the Council of 31 May 1999 Concerning the Approximation of Laws, Regulations and Administrative Provisions of the Member States Relating to the Classification, Packaging and Labelling of Dangerous Preparations.
European Union, 2008. Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on Classification, Labelling and Packaging of Substances and Mixtures, Amending and Repealing Directives 67/548/EEC and 1999/45/EC, and Amending Regulation (EC) No 1907/2006.
Färber, E., Desel, H., 2015. MAGAM II DISC – Final Report. GIZ-Nord Poisons Center, University Medical Center Göttingen (UMG), Göttingen,Germany. 20.05.2016.
Färber, E., Sesana, F., Celentano, A., Pedersen, E.B., Ebbehøjc, N.E., Plackov, S., Caganov, B., Pelclova, D., Zakharov, S., Desel, H., 2016. MAGAM II DISC: eye exposures caused by cleaning products in Denmark, Italy, Slovakia and Czech Republic. Clin. Toxicol. 54 (4), 372.
Griffith, J.F., Nixon, G.A., Bruce, R.D., Reer, P.J., Bannan, E.A., 1980. Dose-response studies with chemical irritants in the albino rabbit eye as a basis for selecting optimum testing conditions for predicting hazard to the human eye. Toxicol. Appl. Pharmacol. 55, 501–513.
Hermanns-Clausen, M., Desel, H., Färber, E., Seidel, C., Holzer, A., Eyer, F., Engel, A.,

Prasa, D., Tutdibi, E., Stürer, A., 2019. MAGAM II: prospective observational multicentre poisons centres study on eye exposures caused by cleaning products. Clin. Toxicol. https://doi.org/10.1080/15563650.2018.1560462.
McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models, second ed. Chapman and Hall 532pp.
OECD, 2017a. OECD Guidelines for the Testing of Chemicals - Section 4 - Test Nr. 405. Acute Eye Irritation/Corrosion. OECD Publishing, Paris.
OECD, 2017b. Guidance Document on an Integrated Approach on Testing and Assessment (IATA) for Serious Eye Damage and Eye Irritation. Series on Testing & Assessment No. 263. OECD Publishing, Paris.
Persson, H.E., Sjöberg, G.K., Haines, J.A., Pronczuk de Garbino, J., 1998. Poisoning severity score. Grading of acute poisoning. J. Toxicol. Clin. Toxicol. 36, 205–213.
Roggeband, R., York, M., Pericoi, M., Braun, W., 2000. Eye irritation responses in rabbit and man after single applications of equal volumes of undiluted model liquid detergent products. Food Chem. Toxicol. 38 (8), 727–734.
United Nations, 2002. Globally Harmonized System of Classification and Labeling of Chemicals. United Nations, New York and Geneva.