Organisation for Economic Co-operation and Development

ENV/CBC/MONO(2025)2/ANN7

Unclassified English - Or. English

11 September 2025

ENVIRONMENT DIRECTORATE
CHEMICALS AND BIOTECHNOLOGY COMMITTEE

# Annex 7: Impact of borderline results on the performances of the 2o3 Defined Approach for Skin Sensitisation

**Series on Testing and Assessment**
**No. 336**

**(Second edition)**

E-mail: ehscont@oecd.org.

**JT03570719**

**Please cite this publication as:**

OECD (2025), *Supporting document to Test Guideline 497 on Defined Approaches for Skin Sensitisation, Annex 7: Impact of borderline results on the performances of the 2o3 Defined Approach for Skin Sensitisation*, OECD Series on Testing and Assessment, No. 336, OECD Environment, Health and Safety, Paris, https://one.oecd.org/document/ENV/CBC/MONO(2025)2/ANN7/en/pdf

**Contact us**

**OECD Environment Directorate,**
**Environment, Health and Safety Division**
**2 rue André-Pascal**
**75775 Paris Cedex 16**
**France**

**E-mail: ehscont@oecd.org**

# About the OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation in which representatives of 38 countries in North and South America, Europe and the Asia and Pacific region, as well as the European Union, meet to co-ordinate and harmonise policies, discuss issues of mutual concern, and work together to respond to international problems. Most of the OECD's work is carried out by more than 200 specialised committees and working groups composed of member country delegates. Observers from several Partner countries and from interested international organisations attend many of the OECD's workshops and other meetings. Committees and working groups are served by the OECD Secretariat, located in Paris, France, which is organised into directorates and divisions.

The Environment, Health and Safety Division publishes free-of-charge documents in twelve different series: **Testing and Assessment**; **Good Laboratory Practice and Compliance Monitoring**; **Pesticides**; **Biocides**; **Risk Management**; **Harmonisation of Regulatory Oversight in Biotechnology**; **Safety of Novel Foods and Feeds; Chemical Accidents; Pollutant Release and Transfer Registers; Emission Scenario Documents; Safety of Manufactured Nanomaterials;** and **Adverse Outcome Pathways.** More information about the Environment, Health and Safety Programme and EHS publications is available on the OECD's World Wide Web site (https://www.oecd.org/en/topics/chemical-safety-and-biosafety.html).



The Inter-Organisation Programme for the Sound Management of Chemicals (IOMC) was established in 1995 following recommendations made by the 1992 UN Conference on Environment and Development to strengthen co-operation and increase international co-ordination in the field of chemical safety. The Participating Organisations are FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank, Basel, Rotterdam and Stockholm Conventions and OECD. The purpose of the IOMC is to promote co-ordination of the policies and activities pursued by the Participating Organisations, jointly or separately, to achieve the sound management of chemicals in relation to human health and the environment.

# Table of contents



## FIGURES

No table of figures entries found.



## TABLES

Table A B.1. Performances of the 2o3 DASS when including or excluding h-CLAT negative results for chemicals having log P > 3.5, and considering DPRA TG 442 BR vs. VS BR (in addition to KS & h-CLAT VS BR), and of use of option 1.1 vs. option 1.2. 39

# 1 Purposes & background

1. The present work was undertaken to investigate the relationship between *in chemico* and *in vitro* borderline results and the zones of (un)certainty leading to (in)conclusive predictions made by the "2 out of 3" defined approach for determining skin sensitisation hazard (2o3 DASS). It was conducted as part of the ongoing OECD work and related expert group (EG) discussions, to develop a Test Guideline (TG) on Defined Approaches for Skin Sensitisation (DASS).

2. The concept of borderline ranges was first discussed in the DASS EG in Dec 2018. The work presented in this report was conducted as a joint effort between Germany, Switzerland and BIAC, and its outcome was presented during the DASS EG teleconferences held on 21st September and 5th October 2020.

3. Based on a preliminary assessment, it was observed that paragraph 24 of the TG 442C on the DPRA test method described ranges close to the thresholds used to discriminate between positive and negative results where additional testing is recommended. However, it was unclear how these ranges were derived. Furthermore, no such a range was defined neither for the KeratinoSens (TG 442D) nor for the h-CLAT (TG 442E) test methods.

4. Furthermore, it was observed that the in the initial OECD DASS database from 21st June summarizing available *in silico*, *in chemico*, *in vitro*, and *in vivo* data did not include data on multiple runs for KeratinoSens and h-CLAT, which is required within the TG442D and 442E to make a conclusive negative or positive prediction. These data were therefore kindly provided by Givaudan for KeratinoSens, and by Kao, Shishedo, BASF SE and RIFM for h-CLAT.

5. Based on the above preliminary assessment, the following analyses were conducted and will be discussed here after:

    i. Borderline ranges have been statistically calculated for each of the test methods composing the 2o3 DASS (i.e., DPRA, KeratinoSens and h-CLAT) based on the dataset from the formal validation studies of these assays.

    ii. These ranges, as well as the ranges described within TG 442C for DPRA, were used to identify borderline results within the DASS database from 21st June and from 22nd Dec 2020, as to assess their impact on the performances of the 2o3 DASS.

iii. The final goal was to propose a decision tree for determining conclusive and inconclusive (e.g. low certainty) predictions of the 2o3 DASS, based on the borderline and non-borderline results of the individual test methods. With this approach, no modifications were made to the 2o3 DASS data interpretation procedure, nor to the prediction model used for the individual test methods.

# 2 Determination of borderline ranges based on validation data sets

6. The borderline ranges of each of the test methods composing the 2o3 DASS were determined from the respective validation study datasets.

## 2.1. Methodology

7. Statistics: For each test method, only test chemicals with at least three test runs were considered for the borderline range determination. From the validation study dataset, the log pooled median absolute deviation (MAD) was calculated for each participating laboratory as described in Gabbert el al., 2020. ***The overall borderline range for each test method was then determined based on the arithmetic mean of the log pooled MADs of the individual laboratories that participated in the validation studies.***

## 2.2. DPRA

8. Dataset available: In the DPRA OECD TG 442C adopted in June 2020 (OECD, 2020), a single test chemical concentration is assessed after a single exposure time (24 h) in triplicate. Generally, a single test run was required within this OECD TG 442C to derive a prediction (unless values close to the cut-offs are obtained)[1]. OECD TG 442C provides two prediction models, one based on the mean peptide depletion (calculated as the mean of the Cys- and Lys peptide depletions) and a second prediction model based on the Cys-depletion alone e.g. for test chemicals co-eluting with the Lys-peptide during the HPLC analysis (OECD, 2020). For both prediction models, several cut-offs for different "reactivity classes" are provided in the TG but only one of the cut-offs per prediction model is decisive for the hazard information (6.38% for the mean peptide depletion and 13.89% for the Cys-only depletion). The different datasets described here below were used to determine and compare DPRA borderline ranges.

    - DPRA log pooled MADs were statistically determined using the dataset from the EURL ECVAM coordinated validation study that was followed by independent peer review by the EURL ECVAM Scientific Advisory Committee

[1] TG 442C is currently under revision to modify this requirement (see paragraph 29).

(ECVAM TSAR Entry Direct Peptide Reactivity Assay, https://tsar.jrc.ec.europa.eu/test-method/tm2009-06). The validation dataset comprised 24 chemicals having one or three test runs in each of the three participating laboratories. In the present analyses only test chemicals having three test runs in a participating laboratory were considered for the borderline range determination (n = 13 to 14).

- DPRA log pooled MADs were also statistically determined using historical data from one individual laboratory (n=385 substances) as published in (Gabbert et al., 2020).
- Experimental borderline ranges were determined by repeatedly testing (n=27) ethylene glycol dimethylacrylate (EGDMA) at 8 µM concentration (Gabbert et al., 2020) to support the above-described statistically determined borderline ranges.

9. Outcome: The borderline ranges determined based on the log pooled MADs in the DPRA validation study are summarized in table 2.1.

**Table 2.1. DPRA borderline ranges determined based on the log pooled median absolute deviations.**

| Data source | Mean peptide depletion [%] (cut-off 6.38) | Cysteine-only depletion [%] (cut-off 13.89%) |
|---|---|---|
| Validation study lab 1 (n[a]=13) | 4.81 - 8.46 | 10.53 - 18.31 |
| Validation study lab 2 (n[a]=14) | 5.49 - 7.42 | 12.17 - 15.85 |
| Validation study lab 3 (n[a]=14) | 4.54 - 9.08 | 8.97 - 21.25 |
| **Validation study mean (lab 1-3)** | **4.95 -8.32** | **10.56 - 18.47** |
| BASF SE historical data[b] (n = 385) | 5.29 - 7.69 | 11.62 - 16.61 |
| BASF SE experimental data[b] (n = 27) | 5.45 – 7.31 | 11.89 – 15.90 |

[a] n: number of test chemicals out of the 24 chemicals assessed in the validation study for which at least three test runs were available.

[b] Published in Gabbert *et al.* (2020).

For the three laboratories participating in the validation study and only considering test chemicals for which at least three test runs were available (i.e. 13 to 14), the borderline range around the 6.38% mean peptide depletion cut-off varied between 4.54 to 5.49 (lower boundary) and 7.42 to 9.08% (upper boundary). ***The DPRA mean borderline range of all participating laboratories was 4.95 to 8.32%***, which was comparable to the range of

5.29 to 7.69% derived from historical data of a routine testing lab (assessing 385 substances) and to the experimentally determined range of 5.45 to 7.31% when repeatedly testing a single substance (i.e. EGDMA) (the latter two published in Gabbert et al., 2020). Likewise, the different borderline ranges determined for the cysteine-only depletion model were also very comparable with each other.

10. Discussion: OECD TG 442C provides ranges for the mean peptide depletion (3-10%) and Cys-only depletion (9-17%) in which a repetition of the test run is proposed (OECD TG 442C adopted 26 June 2020). The range provided in OECD TG 442C for the DPRA is not described as a borderline range as the ones calculated above, but rather as a basis to instruct when an additional experimental run shall be considered. After consultation with the OECD, EURL ECVAM and the test method developer it remains unclear however, based on which data and how the provided ranges were derived. Of note, with regard to the mean peptide depletion range (3-10%) to be considered for run repetitions provided in OECD TG 442C, particularly the lower boundary of 3% appears to be hardly applicable. This is due to the fact that in the DPRA, it is the remaining peptide that is determined during the HPLC analysis. Hence, a peptide depletion of 3% corresponds to 97% recovery of the HPLC analysis, which is close to the analytical precision of the method. Indeed, when analyzing the reference controls (vehicle controls) included at the beginning and at the end of the analysis sequence to verify their stability over the analysis time (OECD, 2020), a mean loss of 4.3% of Cys-peptide is observed based on the data from 244 DPRA runs conducted at BASF SE. Furthermore, the OECD TG 442C itself allows for a coefficient of variation of < 15% in the reference controls.

    Finally, the impact of using 3% mean peptide depletion as lower boundary of a borderline range becomes evident when assessing the incidences of chemicals resulting in mean peptide depletions between 3 and 6.38% in routine testing (table 2.2). During routine testing at Givaudan, 112 chemicals were tested in the DPRA and of those 23 (20%) resulted in mean peptide depletion values falling in the range between 3% and 6.38%. Likewise, during routine testing at BASF SE, 395 chemicals were tested in the DPRA and of those 50 (13%) resulted in mean peptide depletions between 3 and 6.38%. To compare, using the borderline range below the cut-off, which was determined from the validation study dataset (i.e. 4.95 to 6.38%), only 10 of 112 (i.e. 8.9%) routinely tested chemicals at Givaudan, and 16 of 395 (i.e. 4.1%) routinely tested chemicals at BASF resulted in mean peptide depletions between 4.95 and 6.38%, respectively (table 2.2). Furthermore, the frequency of chemicals in the borderline ranges below the cut-off (i.e. borderline negatives) appears to be higher than the percentage of chemicals in the borderline range above the cut-off (borderline positives), which is also true for the OECD DASS reference database from 22$^{nd}$ December 2020 (table 2.2).

**Table 2.2. Frequency of mean peptide depletion borderline calls in the OECD DA SS reference database as well as in routine testing conducted at BASF SE and Givaudan**

| Mean peptide depletion [%] | BASF SE n = 395 | Givaudan n = 112 | OECD DASS DB n = 168 |
|---|---|---|---|
| 3 - 4.95 | 34 (8.6%) | 13 (11.6%) | 11 (6.6%) |
| 4.95 - 6.38 | 16 (4.1%) | 10 (8.9%) | 8 (4.8%) |
| 6.38 - 8.32 | 4 (1.0%) | 5 (4.5%) | 6 (3.6%) |
| 8.32 - 10 | 9 (2.3%) | 5 (4.5%) | 2 (1.2%) |
| Calls outside of the 3-10 range | 332 (84.1%) | 79 (70.5%) | 141 (83.9%) |

## 2.3. KeratinoSens

11. Dataset available: In the KeratinoSens (OECD TG 442D (OECD 2018a)), a test chemical is assessed in multiple concentrations (i.e. 12) at a single exposure time (48 hours) in triplicate within each independent run. If a concordant result is obtained between the first two runs, no additional run is required. Besides evaluating whether luciferase activity is induced to 1.5-fold or above the vehicle control level, statistical significance is assessed in this assay. Furthermore, only concentrations showing a viability higher than 70% are evaluated (for both positive and negative predictions). A Givaudan coordinated validation study, followed by an independent peer review by the EURL ECVAM Scientific Advisory Committee, was conducted using 28 chemicals tested in five laboratories (Natsch *et al.* 2011). In the present analyses only test chemicals having at least three test runs in a participating laboratory were considered for the borderline range determination (n = 26 to 28).

    In addition to the borderline range that was statistically derived from the KeratinoSens validation study data based on the log pooled MAD method, an experimental borderline range was determined from repeatedly testing the positive control cinnamic aldehyde at 16 µM (n=123).

12. Outcome: The borderline ranges determined based on the log pooled MADs in the KeratinoSens validation study are summarized in table 2.3.

**Table 2.3. KeratinoSens™ borderline ranges determined based on the log pooled median absolute deviations.**

| Data source | Luciferase induction (cut-off 1.5) |
|---|---|
| Validation study lab 1 ($n^a$ = 28) | 1.37 – 1.64 |
| Validation study lab 2 ($n^a$ = 28) | 1.33 – 1.69 |
| Validation study lab 3 ($n^a$ = 28) | 1.33 – 1.69 |
| Validation study lab 4 | 1.35 – 1.67 |

| | |
|---|---|
| (n[a] = 28) | |
| Validation study lab 5 (n[a] = 26) | 1.37 – 1.65 |
| **Validation study mean (lab 1-5)** | **1.35 – 1.67** |
| Givaudan experimental data on positive control (n = 123) | 1.40 – 1. 60 |

[a] n: number of test chemicals out of the 28 chemicals assessed in the validation study for which at least three test runs were available.

13. Discussion: Based on the five laboratories participating in the validation study and considering chemicals for which at least three runs were available (i.e. 26 to 28), the borderline range around 1.5 luciferase fold-induction cut-off varied between 1.33 to 1.37 (lower boundary) and 1.64 to 1.69 (upper boundary). ***The KeratinoSens mean borderline range of all participating laboratories was 1.35 to 1.67***, which was comparable to the historical data range of 1.40 to 1.60 that was experimentally derived in the routine testing lab.

## 2.4. h-CLAT

14. Dataset available: In the h-CLAT (OECD TG 442E, (OECD 2018b)), a test chemical is assessed in multiple concentrations (i.e. 8) at a single exposure time (24 hours) in a single replicate within each independent run. If a concordant result is obtained between the first two runs, no additional run is conducted. A test chemical is predicted to activate dendritic cells in case the relative expression levels (measured as fluorescence intensity relative to the concurrent vehicle control) of the cell surface markers CD54 and CD86 surpasses 200 (for CD54) and/or 150 (for CD86), respectively in any test chemical concentration showing a cell viability higher or equal to 50% in at least two independent experiments. For test chemicals tested positive, the concentrations at which a relative fluorescence intensity (RFI) of 200 (the $EC_{200}$ for CD54) or 150 (the $EC_{150}$ for CD86) are surpassed can optionally be determined, respectively. A negative result is reported if the markers are not induced above their cut-off values whereas cell viability is reduced by at least 10% at the highest concentration tested, or when the highest concentration tested is the limit test concentration of a test chemical (based on solubility and chosen solvent) even if the cell viability is not reduced by at least 10%.

    An EURL ECVAM coordinated validation study, followed by an independent peer review by the EURL ECVAM Scientific Advisory Committee, was conducted with 24 chemicals tested in three runs by four laboratories (ECVAM TSAR Entry Human Cell Line Activation Test, https://tsar.jrc.ec.europa.eu/test-method/tm2008-05).

    In addition, for the h-CLAT the log pooled MAD was also determined based on historical data from one individual laboratory as published in (Gabbert *et al.*, 2020).

15. Outcome: The borderline ranges determined based on the log pooled median MADs in the h-CLAT validations study are summarized in table 2.4.

**Table 2.4. h-CLAT borderline ranges determined based on the log pooled median absolute deviations.**

| Data source | RFI CD54 (cut-off 200) | RFI CD86 (cut-off 150) |
|---|---|---|
| Validation study lab 1 (n[a] = 24) | 152 - 264 | 125 - 181 |
| Validation study lab 2 (n[a] = 24) | 153 - 261 | 125 - 181 |
| Validation study lab 3 (n[a] = 24) | 161 - 248 | 115 - 196 |
| Validation study lab 4 (n[a] = 24) | 162 - 247 | 125 - 180 |
| **Validation study mean (lab 1-4)** | **157 - 255** | **122 - 184** |
| BASF SE historical data[b] (n = 136) | 170 - 235 | 132 - 170 |

[a] n: number of test chemicals of the 24 chemicals assessed in the ring trial for which at least three test runs were available.

[b] Published in Gabbert *et al.* (2020).

16. Discussion: For the four laboratories participating in the h-CALT validation study and considering only chemicals for which at least three runs were available (i.e. 24) the borderline range around the RFI 200 cut-off for CD54 varied between 152 to 162 (lower boundary) and 247 to 264 (upper boundary). ***The CD54 mean borderline range of all participating laboratories was 157 to 255***, which was somewhat wider than the historical data range of 170 to 235 experimentally derived in a routine testing lab (assessing 136 chemicals; published in Gabbert *et al*., 2020). Likewise, the borderline range around the RFI 150 cut-off for CD86 varied between 115 to 125 (lower boundary) and 180 to 196 (upper boundary). ***The CD86 mean borderline range of all participating laboratories was 122 to 184***, which was also somewhat wider than the historical data range of 132 to 170 experimentally derived in a routine testing lab (assessing 136 chemicals; published in Gabbert *et al*., 2020).

## 2.5. Conclusions

17. In most toxicological test methods used for regulatory purposes, continuous data is translated into a e.g. binary classification (“positive” or “negative”) using cut-off values irrespective of the data’s variability. Any test result is however subject to variation and these variations increase the uncertainty of a test result in particular close to a (classification) cut-off (in the borderline range). The uncertainty/borderline calls for chemicals may come from:

    - Technical variance (e.g. pipetting errors, equipment variability, biological variability).

- "Real" borderline chemicals with reproducible results that fall very close to the cut offs for hazard categories

18. Therefore, during method development and validation, besides defining cut-offs for binary classifications, the predictions' certainty close to the cut-off in the borderline range should be taken into consideration. As a method's reproducibility is usually assessed in formal validation studies, it is hence proposed to determine the variability close to the classification cut-off based on the validation study datasets, in which usually at least 3 laboratories participate and assess multiple test chemicals in a coded manner and in multiple independent runs. As the data generated during validation studies nevertheless remains limited (for the three assays evaluated here 24 to 28 chemicals were assessed in three to five laboratories), we propose to use the arithmetic means of the borderline ranges determined by the log pooled MADs for the individual participating laboratories. Table 2.5 shows the proposed borderline ranges based on the arithmetic means of the log pooled MADs of the validation study laboratories.

**Table 2.5. Proposed borderline range based on the arithmetic means of the log pooled median absolute deviations in the validation study laboratories.**

| | Endpoint | Cut-off | TG borderline range | Validation study mean |
|---|---|---|---|---|
| **DPRA (OECD TG 442C)** | Mean peptide depletion [%] | 6.38 | 3-10[a] | 4.95 - 8.32 |
| | Cysteine-only depletion [%] | 13.89 | 9-17[a] | 10.56 - 18.47 |
| **KeratinoSens (OECD TG 442D)** | Luciferase induction (fold-change) | 1.5 | n/a | 1.35 - 1.67 |
| **h-CLAT (OECD TG 442E)** | Relative fluorescence intensity CD54 | 200 | n/a | 157 - 255 |
| | Relative fluorescence intensity CD86 | 150 | n/a | 122 - 184 |

[a] The range provided in OECD TG 442C (adopted 26 June 2020) for the DPRA is not described as borderline range as such. Instead, when a result in this range is obtained it is proposed that additional independent experiments are considered.

# 3 Impact of borderline results on the performance of the 2o3 DASS

19. In order to assess the impact of borderline ranges on the performance of the 2o3 DASS, two aspects were considered as described in chapter 3.1 and 3.2 here below.

- First, an assessment was conducted on how the currently existing TG on the individual test methods composing the 2o3 DASS address the issue of borderline / discordant outcomes.
- Secondly, an assessment was conducted on the impact of these borderline outcomes on the (un)certainty of 2o3 DASS predictions used in a stand-alone manner for skin sensitisation hazard classification.

## 3.1. Considering borderline/discordant results within individual Test Guidelines

20. Regarding the **DPRA test method**, paragraph 24 of TG 442C adopted in June 2020 (OECD, 2020) establishes that:

*"…in cases of results close to the threshold used to discriminate between positive and negative results (i.e. borderline results), additional testing may be necessary. If situations where the mean percent depletion falls in the range of 3% to 10% for the cysteine 1:10/lysine 1:50 prediction model or the cysteine percent depletion falls in the range of 9% to 17% for the cysteine 1:10 prediction model, a second run **may** be considered, as well as a third one in case of discordant results between the first two runs."*

21. Although the initial OECD DASS database (version 21st June) did not contain data on multiple DPRA runs, we found existing data on multiple runs for 6 and 3 chemicals having DPRA borderline results vs. LLNA and human (i.e. HDSG) data, respectively. It was observed that conducting multiple runs can result in conclusive predictions and should therefore be considered when obtaining borderline results. However, TG 442C adopted in June 2020 (OECD, 2020) did not require additional runs to be conducted, but only state that additional runs 'may' be considered.

22. For the data with a single run as reported in the DASS database, borderline cases in the DPRA were identified straightforward based on the borderline range for the mean peptide depletion as described above. To decide on repeated testing based on the refined requirement for repeated runs and to then determine positives, negatives and borderline final outcomes, the prediction model in Appendix 1, Figure 1.1 was used. This PM

introduces a third outcome (BL) to be used within the 2o3 DA, based on the same decision cut-offs of the PM described in OECD TG 442C. Thus, a negative in the original PM can only become negative or BL, while a positive from the original PM can only become positive or BL.

**Table 3.1. Chemicals from the DASS database (from 22 Dec 2020) having DPRA outcome within the TG 442 borderline range that have multiple runs identified from other data sources. In orange: positive predictions; in green: negative predictions; in dark orange: non-borderline positive predictions; in dark green: non borderline negative predictions.**

| Chemical name | CAS | LLNA.GHS. SUB | HU.GHS. SUB | DA.2of3. Call | DPRA | KS | hCLAT | DASS reported DPRA mean | Additional source | DPRA mean peptide depletion - different runs from additional source | | | | | | | | | DPRA Cyst depl. runs | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cinnamic alcohol | 104-54-1 | 1B | 1B | 1 | + | + | + | 7.55 | Urbish et al. (2015) | 7.55 | 12.00 | | | | | | | | | |
| Penicillin G | 61-33-6 | 1B | 1B | 1 | + | - | + | 7.15 | DPRA prevalidation | 12.35 | | | | | | | | | 14.30 | 18.50 |
| Cyclamen aldehyde | 103-95-7 | 1B | n.a. | 1 | + | + | - | 9.95 | DPRA prevalidation | 9.30 | 36.90 | 29.95 | 24.30 | | | | | | | |
| Aniline | 62-53-3 | 1B | 1B | 0 | - | - | + | 4.85 | Urbish et al. (2015) | 4.85 | 1.19 | | | | | | | | | |
| R(+) Limonene | 5989-27-5 | 1B | n.a. | 0 | - | - | + | 3.10 | ECVAM VS (2012) | 7.90 | 1.90 | 6.60 | 7.10 | 10.20 | 9.20 | 26.00 | 16.80 | 17.40 | | |
| Dihydroeugenol | 2785-87-7 | 1B | n.a. | 0 | - | + | - | 6.15 | ECVAM VS (2012) | 4.40 | 4.0Lys | 9.20 | | | | | | | | |

23. Regarding **KeratinoSens**, TG 442D requires that at least two independent runs (named 'repetitions' within TG 442D, where each run/repetition is composed of triplicates) should be conducted to derive a prediction, and a third run/repetition should be performed in case the first two runs/repetitions are not concordant. The DASS database did not include data on individual runs/repetitions, however Givaudan had access to this data and analysed the multiple runs/repetitions to verify whether a definitive or borderline result was obtained.

24. For the assessment whether the outcome of repeated runs yields a positive, negative or borderline final outcome in KeratinoSens™, the prediction model in Appendix 1, Figure 1.2 is applied. This PM introduces a third outcome (BL) to be used within the 2o3 DA, based on the same decision cut-offs of the PM described in OECD TG 442D. Thus, a negative in the original PM can only become negative or BL, while a positive from the original PM can only become positive or BL.

25. Similarly, for **h-CLAT,** TG 442E requires that at least two runs (where each repetition is comprised by single replicates) are conducted. Although the DASS database did not encompass the results of individual runs, such data was kindly provided by **Kao, Shiseido and RIFM** and analysed to verify whether a definitive or borderline result was obtained.

26. For the assessment whether the outcome of repeated runs yields a positive, negative or borderline final outcome in the h-CLAT, the prediction model in Appendix 1, Figure 1.3 is applied. This PM introduces a third outcome (BL) to be used within the 2o3 DA, based on the same decision cut-offs of the PM described in OECD TG 442E. Thus, a negative

in the original PM can only become negative or BL, while a positive from the original PM can only become positive or BL.

27. As a consequence, for both KeratinoSens and h-CLAT, at least two concordant runs/repetitions are needed for a prediction to be made. Furthermore, the two assays require that 12 and 8 concentrations are tested and, in the case of KeratinoSens, that triplicates of each concentration are tested in each run/repetition, respectively.

28. In view of the positive impact multiple runs can have on reducing uncertainty of results, and in order to harmonise the requirements of TG 442C with the ones of TG 442D and 442E, it was proposed to the OECD EG on SS to modify the wording of paragraph 24 of TG 442C (which is currently open for revision) as follows:

*"24. A single HPLC analysis for both the cysteine and the lysine peptide should be sufficient for a test chemical when the result is unequivocal. However, in cases of results close to the threshold used to discriminate between positive and negative results (i.e. borderline results), additional testing should be conducted. If situations where the mean percent depletion falls in the range of 3% to 10% for the cysteine 1:10/lysine 1:50 prediction model or the cysteine percent depletion falls in the range of 9% to 17% for the cysteine 1:10 prediction model, a second run should be conducted, as well as a third one in case of discordant results between the first two runs*."

29. During the 2020-2021 revisions of the OECD TG 442C, the following modifications have been agreed by the EG[2]:

"*24. A single HPLC analysis for both the cysteine and the lysine peptide should be sufficient for a test chemical when the result is unequivocal. However, in cases of results close to the threshold used to discriminate between positive and negative results (i.e. in the range of 3% to 10% for the cysteine 1:10/lysine 1:50 prediction model or the cysteine percent depletion falls in the range of 9% to 17% for the cysteine 1:10 prediction model), additional testing is recommended. In particular, in case of negative results in these ranges (i.e. 3% to 6.38% for the cysteine 1:10/lysine 1:50 prediction model or 9% to 13.89% for the cysteine 1:10 prediction model), a second run should be conducted, as well as a third one in case of discordant results between the first two runs."*

## 3.2. Considering borderline/inconclusive results in 2o3 DASS predictions

### *3.2.1. Methodology*

30. In order to investigate the impact of borderline results from individual methods on the 2o3 DASS predictions for hazard classification, borderline ranges (BR) were calculated based on validation studies (VS BR) for DPRA, KeratinoSens (KS) and h-CLAT as

[2] Currently under consultation at the OECD level.

described in chapter 2. To take into account also the ranges currently provided in TG 442C (i.e. TG 442C BR), the following possible borderline ranges were used in the present analyses:

- **DPRA TG 442C BR:** *mean peptide depletion: 3-10; Cys-only depletion: 9-17*
  KS VS BR: Imax: 1.35-1.67
  h-CLAT VS BR: RFI CD86: 122-184; RFI CD54: 157-255
- **DPRA VS BR:** *mean peptide depletion: 4.95 - 8.32, Cys-only depletion: 10.56 - 18.47*
  KS VS BR: Imax: 1.35-1.67
  h-CLAT VS BR: RFI CD86: 122-184; RFI CD54: 157-255

31. The *principle for deciding on borderline results for the tests with multiple runs* (KeratinoSens and h-CLAT) followed the *majority principle*, i.e. two concordant runs needed to determine a positive or negative borderline or non-borderline outcome (see Appendix 1).

32. When the result of a method was found to be within the above borderline ranges, two options were considered as follows:

- **Option 1.1: All borderline predictions (positive and negative) were considered of low confidence**. In this case:
  i. If at least two test method results were positive AND non-borderline: the 2o3 DASS UN GHS Cat. 1 prediction was made.
  ii. If at least two test method results were negative AND non-borderline: the 2o3 DASS UN GHS No Cat. prediction was made.
  iii. In all other cases, the 2o3 DASS prediction was considered ***inconclusive.***
- **Option 1.2: Only negative borderline predictions were considered of low confidence, whereas positive borderline predictions were considered as a positive outcome.** In this case:
  i. The 2o3 DASS prediction UN GHS Cat. 1 was made if at least two test method results were positive AND
     1. non-borderline, or
     2. borderline positive (one or more test methods).
  ii. The DASS prediction UN GHS No Cat. was made if at least two test method results were negative AND non-borderline.
  iii. In all other cases, the DASS prediction was considered ***inconclusive***.

### *3.2.2. Use of the DPRA range described within TG 442C (TG 442C BR)*

33. Table 3.2 shows the number of borderline chemicals identified (n) for each test method individually and in combination, based on the latest version of the OECD DASS database from 22nd Dec 2020 and taking into account the additional data obtained for h-CLAT individual runs/repetitions (see section 3.1)[3]. Interestingly, with regard to LLNA data, only 20% (14/71) of these borderline chemicals were borderlines in two or three methods at the same time (see footnote of table 3.2), suggesting that the three test methods cover complementary mechanisms that lead to different types of borderlines chemicals.

Table 3.2 also shows the proportion of the borderline chemicals that resulted in 2o3 DASS false negative (FN) and false positive (FP) predictions, vs. both LLNA and human (i.e. HDSG) reference data as reported in the OECD DASS database from 22nd Dec 2020[4]. When considering all chemicals having borderline results and 2o3 DASS predictions (last row of table 3.2), it can be seen that 60% (21/35) of the FN of the overall 2o3 DASS database had borderline results in at least one test method when comparing to the LLNA dataset (first row). Furthermore, when comparing to the human dataset, 78% (7/9) of the FN of the overall DASS database had borderline results in at least one test method. ***As a consequence, borderline outcomes have a high probability of leading to 2o3 DASS FN predictions.***

[3] This is due to the fact that the data originally provided within the OECD DASS database was not sufficient to assess whether an h-CLAT fell into the borderline range. In contrast, for DPRA and KeratinoSens the same data entry from the OECD database that was used to establish the 2o3 DA performances, was used to determine whether or not a result was within the borderline range (e.g. use of average Imax for KeratinoSens). In a separate analysis, the impact of DPRA and KeratinoSens individual runs/repetitions was further assessed as shown in table 3.4.

[4] For example, according to table 3.2 a total of 27 chemicals were found to have DPRA results falling within the OECD TG 442C BR. For these 27 chemicals, 10 false negative predictions were identified out of 23 LLNA sensitisers, and 0 false positives predictions were identified out of 4 LLNA non-sensitisers.

**Table 3.2. Impact of the different test method borderline results on 2o3 DASS predictions.**

| | 2o3 vs. LLNA | | | 2o3 vs. HDSG | | |
|---|---|---|---|---|---|---|
| | FN | FP | n | FN | FP | n |
| DASS database | 35/135 (26%) | 5/33 (15%) | | 9/54 (17%) | 2/11 (18%) | |
| DPRA borderlines (TG 442C BR: mean PM: 3-10, Cys PM: 9-17) | 10/23 (43%) | 0/4 (0%) | 27/168 (16%) | 2/9 (22%) | 1/2 (50%) | 11/65 (17%) |
| KS borderlines (VS BR: Imax: 1.35-1.67) | 8/11 (73%) | 0/3 (0%) | 14/168 (8%) | 5/7 (71%) | 1/2 (50%) | 9/65 (14%) |
| hCLAT borderlines (VS BR: RFI CD86: 122-184; RFI CD54: 157-255) | 9/35 (26%) | 3/11 (27%) | 46/167*** (28%) | 3/11 (27%) | 0/3 (0%) | 14/64 (22%) |
| DPRA TG 442C BR + KS & hCLAT VS BR | 21/55 (38%) | 3/16 (19%) | 71/168* (42%) | 7/21 (33%) | 1/6 (17%) | 27/65** (42%) |

FN: false negatives; FP: false positives; HDSG: human effects based on the evaluation of the OECD DASS human data subgroup ; KS: KeratinoSens; LLNA: Local Lymph Node Assay (OECD TG 429, 2010); n: number of borderlines; PM: prediction model; VS BR: borderline range determined from validation study datasets as described in chapter 2; TG442C BR: borderline range defined within OECD TG 442C (2020).

*13 chemicals with 2, and 1 chemical with 3 test methods borderline results; ** 7 chemicals with 2 test method borderline results. ***h-CLAT data missing for one substance (2-hexylidenecyclopentanone)

34. Table 3.3 shows the distribution of the borderline results based on the predictions from individual test methods. The majority, i.e. 69% (49/71) of chemicals having borderline outcomes had at least one (out of 3) discordant test method predictions against the LLNA in the 2o3 DASS. In particular, 58% or 11 out of a total of 19 chemicals that were found to have the D-,K-,h+ combination in the DASS database and to be false negatives when compared to the LLNA, had at least one test method borderline outcome, suggesting this is a combination of test methods with a majority of borderline outcomes.

**Table 3.3. Distribution of DASS reference chemicals having borderline results according to the 2o3 DASS predictions and combination of individual methods prediction.**

| DPRA TG 442C BR + KS & hCLAT VS BR | 2o3 vs. LLNA | | | | 2o3 vs. HDSG | | | |
|---|---|---|---|---|---|---|---|---|
| | FN | FP | TP | TN | FN | FP | TP | TN |
| D+, K+, h+ | 0 | 1 | 14 | 0 | 0 | 0 | 8 | 0 |
| D-, K+, h+ | 0 | 0 | 9 | 0 | 0 | 0 | 3 | 0 |
| D+, K-, h+ | 0 | 1 | 7 | 0 | 0 | 1 | 3 | 0 |
| D+, K+, h- | 0 | 1 | 4 | 0 | 0 | 0 | 0 | 0 |
| D-, K-, h+ | 11 | 0 | 0 | 2 | 2 | 0 | 0 | 2 |
| D-, K+, h- | 4 | 0 | 0 | 5 | 1 | 0 | 0 | 1 |
| D+, K-, h- | 4 | 0 | 0 | 1 | 3 | 0 | 0 | 0 |
| D-, K-, h- | 2 | 0 | 0 | 5 | 1 | 0 | 0 | 2 |
| Total | 71/168 (42%) | | | | 27/65 (42%) | | | |

D: DPRA; FN: false negatives; FP: false positives; h: h-CLAT; HDSG: human effects based on the evaluation of the OECD DASS human data subgroup; K: KeratinoSens; LLNA: Local Lymph Node Assay (OECD TG 429, 2010); TN: true negatives; TP: true positives; +: positive outcome; - negative outcome.

35. Table 3.4 shows the impact in the 2o3 DASS performances of: i) taking into account multiple runs/repetition[5] leading to conclusive prediction of borderline chemicals, and ii) using options 1.1 and 1.2 (see section 3.2.1) to decide on the confirmed borderline results. Albeit the information available on multiple runs that lead to a conclusive prediction is limited (n=7), we see that taking such information into account can decrease the number of false negatives from 3 to 2 out of 6 when comparing to the LLNA reference data (second row of Table 3.4). Furthermore, applying options 1.1 and 1.2 to decide on inconclusive predictions for chemicals having at least one test method borderline outcome, lead to reduced false negative predictions when compared to e.g. LLNA data

[5] As described above, OECD TG 442C (adopted 26 Jun 2020) did not call for multiple DPRA runs. Hence additional DPRA runs from literature were considered as described in section 3.1 and a new column entitled 'DPRA_MR' added to the database). In contrast, OECD TGs 442D and 442E require 2 concordant results to come to a prediction for KS and h-CLAT. The original OECD database did however not contain this level of detail. For example, in the case of KeratinoSens only average Imax values were given but not the Imax of individual runs. In this case, an average Imax value may be within the borderline range (indicated in the column 'KS.Call_Imax_VSBR' in the database), but when taking into account the individual runs, a clear prediction may be made according to Appendix 1, leading to a "conclusive multiple runs/repetitions". This is the correct way to apply the borderline prediction model as shown in Figure 1.2 in Appendix 1 and this information was added to the updated OECD database with a column entitled 'KS.Call_IR_VSBR'. Regarding h-CLAT in contrast, the database does not contain a parameter from which it can be determined whether a substance should be rated borderline. Therefore, only predictions already based on multiple runs were considered (column entitled 'hCLAT.Call_IR_VSBR').

(9-13% vs. 38%) and showed a similar tendency for human data.

**Table 3.4. Impact of multiple runs/repetition prediction of borderline chemicals in the 2o3 DASS predictions, and of using options 1.1 and 1.2 to decide on (in)conclusive predictions of confirmed borderline chemicals.**

| DPRA TG 442C BR + KS & hCLAT VS BR | 2o3 vs. LLNA | | | 2o3 vs. HDSG | | |
|---|---|---|---|---|---|---|
| | FN | FP | *Inconclusive* | FN | FP | *Inconclusive* |
| DPRA TG 442C BR + KS & hCLAT VS BR | 21/55 (38%) | 3/16 (19%) | 71/168 (42%) | 7/21 (33%) | 1/6 (17%) | 27/65 (42%) |
| DPRA & KS with conclusive multiple runs/repetition prediction | 3/6 -> 2/6 | 0/1 -> 0/1 | | 2/3 -> 2/3 | - | |
| DPRA TG 442C + (KS & hCLAT VS) confirmed borderlines | 18/49 | 3/15 | | 5/18 | 1/6 | |
| Option 1.1: All *in vitro* borderlines | 3/23 (13%) | 2/8 (25%) | 34/168 (20%) | 1/10 (10%) | 0/3 (0%) | 9/65 (14%) |
| Option 1.2: *In vitro* borderlines positives considered positive | 3/35 (9%) | 3/9 (33%) | 21/168 (13%) | 1/13 (8%) | 1/4 (25%) | 5/65 (8%) |

FN: false negatives; FP: false positives; HDSG: human effects based on the evaluation of the OECD DASS human data subgroup; KS: KeratinoSens; LLNA: Local Lymph Node Assay (OECD TG 429, 2010); VS BR: borderline range determined from validation datasets as described in chapter 2; TG442C BR: borderline range defined within OECD TG 442C (2020).

36. Finally, table 3.5 shows the impact of using options 1.1 and 1.2 on the overall false negative and false positive rates of the 2o3 DASS when taking into account the entire OECD DASS dataset, i.e. including also the test chemicals that did not have borderline outcomes and taking into account the additional data obtained for DPRA, KeratinoSens and h-CLAT multiple individual runs/repetitions (see section 3.1 and table 3.4). ***Considering borderline outcomes within options 1.1 and 1.2 results in an overall decreased FN rate of the 2o3 DASS, with only a slight increase in the FP rate. However, it also implies that additional data and/or information is needed for up to 20% chemicals identified as inconclusive.***

**Table 3.5. . Impact on the performances of the 2o3 DASS when using options 1.1 and 1.2 to decide on (in)conclusive predictions based on individual test method borderline results.**

| DPRA TG 442C BR + KS & hCLAT VS BR | 2o3 vs. LLNA | | | 2o3 vs. HDSG | | |
|---|---|---|---|---|---|---|
| | FN | FP | Inconclusive | FN | FP | Inconclusive |
| DASS database | 35/135 (26%) | 5/33 (15%) | | 9/54 (17%) | 2/11 (18%) | |
| DASS database, excluding confirmed borderlines and including DPRA/KS multiple runs/repetitions | 16/85 | 2/18 | | 5/38 | 1/5 | |
| Option 1.1: All *in vitro* borderlines excluded | 19/108 (18%) | 4/26 (15%) | 34/168 (20%) | 6/48 (13%) | 1/8 (13%) | 9/65 (14%) |
| Option 1.2: *In vitro* borderlines positives considered positive | 19/120 (16%) | 5/27 (19%) | 21/168 (13%) | 6/51 (12%) | 2/9 (22%) | 5/65 (8%) |

FN: false negatives; FP: false positives; HDSG: human effects based on the evaluation of the OECD DASS human data

subgroup; KS: KeratinoSens; LLNA: Local Lymph Node Assay (OECD TG 429, 2010); VS BR: borderline range (BR) determined from validation datasets as described in chapter 2; TG442C BR: BR defined within OECD TG 442C (2020).

### *3.2.3. Use of the DPRA borderline range calculated from the validation studies (VS BR)*

37. Table 3.6 shows the number of borderline chemicals identified (n) for each test method individually and in combination, based on the latest version of the OECD DASS database from 22nd Dec 2020 and taking into account the additional data obtained for h-CLAT multiple individual runs/repetitions (see section 3.1)[6]. Interestingly, with regard to LLNA data, only 14% (9/64) of these borderline chemicals were borderlines in two or three methods at the same time (see footnote of table 3.6), suggesting that the three test methods cover complementary mechanisms that lead to different types of borderlines chemicals. As compared to Table 3.2, a lower number of chemicals having DPRA borderline results was observed when using the VS BR as compared to the TG 442 BR (14/168 vs. 27/168). Similarly, a lower overall number of chemicals having borderline results was observed when taking into account the three test methods (64/168 vs. 71/168) and comparing the VS BR versus the TG 442 BR.

    Table 3.6 also shows the proportion of the borderline chemicals that resulted in 2o3 DASS false negative (FN) and false positive (FP) predictions, vs. both LLNA and human (i.e. HDSG) reference data as reported in the OECD DASS database from 22nd Dec 2020. When considering all chemicals having borderline results and 2o3 DASS predictions (last row of table 3.6), it can be seen that 46% (16/35) of the FN of the overall 2o3 DASS database had borderline results in at least one test method when comparing to the LLNA dataset (first row). Furthermore, when comparing to the human dataset, 78% (7/9) of the FN of the overall DASS database had borderline results in at least one test method. ***As a consequence, borderline outcomes have a high probability of leading to 2o3 DASS FN predictions.***

[6] This is due to the fact that the data originally provided within the OECD DASS database was not sufficient to assess whether an h-CLAT fell into the borderline range. In contrast, for DPRA and KeratinoSens the same data entry from the OECD database that was used to establish the 2o3 DA performances, was used to determine whether or not a result was within the borderline range (e.g. use of average Imax for KeratinoSens). In a separate analysis, the impact of multiple runs for DPRA and KeratinoSens was further assessed as shown in table 3.4.

### Table 3.6. Impact of the different test method borderline results on 2o3 DASS predictions.

| | 2o3 vs. LLNA | | | 2o3 vs. HDSG | | |
|---|---|---|---|---|---|---|
| | FN | FP | n | FN | FP | n |
| DASS database | 35/135 (26%) | 5/33 (15%) | | 9/54 (17%) | 2/11 (18%) | |
| DPRA borderlines (VS BR: mean PM: 4.95-8.32, Cys PM: 10.56-18.47) | 3/12 (25%) | 0/2 (0%) | 14/168 (8%) | 1/5 (20%) | 1/2 (50%) | 7/65 (11%) |
| KS borderlines (VS BR: Imax: 1.35-1.67) | 8/11 (73%) | 0/3 (0%) | 14/168 (8%) | 5/7 (71%) | 1/2 (50%) | 9/65 (14%) |
| hCLAT borderlines (VS BR: RFI CD86: 122-184; RFI CD54: 157-255) | 9/35 (26%) | 3/11 (27%) | 46/167*** (28%) | 2/11 (18%) | 0/3 (%) | 14/64 (22%) |
| DPRA + KS + hCLAT VS BR | 16/49 (33%) | 3/15 (20%) | 64/168* (38%) | 7/19 (37%) | 1/6 (17%) | 25/65** (39%) |

FN: false negatives; FP: false positives; HDSG: human effects based on the evaluation of the OECD DASS human data subgroup; KS: KeratinoSens; LLNA: Local Lymph Node Assay (OECD TG 429, 2010); n: number of borderlines; PM: prediction model; VS BR: borderline range determined from validation study datasets as described in chapter 2.

*8 chemicals with 2, and 1 chemical with 3 *in vitro* borderline results. ** 5 chemicals with 2 *in vitro* borderline results. ***h-CLAT data missing for one substance (2-hexylidenecyclopentanone)

38. Table 3.7 shows the distribution of the borderline results based on the predictions from the individual test methods. The majority, i.e. 68% (43/64) of chemicals having borderline outcomes had at least one (out of 3) discordant test method predictions against the LLNA in the 2o3 DASS.

### Table 3.7. Distribution of DASS reference chemicals having borderline results according to the 2o3 DASS predictions and combination of individual methods prediction.

| DPRA + KS + hCLAT VS BR | 2o3 vs. LLNA | | | | 2o3 vs. HDSG | | | |
|---|---|---|---|---|---|---|---|---|
| | FN | FP | TP | TN | FN | FP | TP | TN |
| D+, K+, h+ | 0 | 1 | 14 | 0 | 0 | 0 | 7 | 0 |
| D-, K+, h+ | 0 | 0 | 8 | 0 | 0 | 0 | 2 | 0 |
| D+, K-, h+ | 0 | 1 | 7 | 0 | 0 | 1 | 3 | 0 |
| D+, K+, h- | 0 | 1 | 4 | 0 | 0 | 0 | 0 | 0 |
| D-, K-, h+ | 7 | 0 | 0 | 2 | 2 | 0 | 0 | 2 |
| D-, K+, h- | 3 | 0 | 0 | 5 | 1 | 0 | 0 | 1 |
| D+, K-, h- | 4 | 0 | 0 | 1 | 3 | 0 | 0 | 0 |
| D-, K-, h- | 2 | 0 | 0 | 4 | 1 | 0 | 0 | 2 |
| Total | 64/168 (38%) | | | | 25/65 (39%) | | | |

D: DPRA; FN: false negatives; FP: false positives; h: h-CLAT; HDSG: human effects based on the evaluation of the OECD DASS human data subgroup; K: KeratinoSens; LLNA: Local Lymph Node Assay (OECD TG 429, 2010); TN: true negatives; TP: true positives; +: positive outcome; - negative outcome.

39. Table 3.8 shows the impact in the 2o3 DASS performances of: i) taking into account multiple runs/repetition leading to conclusive prediction of borderline chemicals, and ii) using options 1.1 and 1.2 (see section 3.2.1) to decide on the confirmed borderline results. Albeit the information available on multiple runs that lead to a conclusive prediction is limited (n=5), we see that taking such information into account can decrease the number of false negatives from 3 to 2 out of 4 when comparing to the LLNA reference data (second row of Table 3.8). Furthermore, applying options 1.1 and 1.2 to decide on inconclusive predictions for chemicals having at least one test method borderline outcome, lead to reduced false negative predictions when compared to e.g. LLNA (9-13% vs. 33%) and showed a similar tendency for human data.

**Table 3.8. Impact of multiple runs/repetition prediction of borderline chemicals in the 2o3 DASS predictions, and of using options 1.1 and 1.2 to decide on (in)conclusive predictions of confirmed borderline chemicals.**

| DPRA + KS + hCLAT VS BR | 2o3 vs. LLNA | | | 2o3 vs. HDSG | | |
|---|---|---|---|---|---|---|
| | FN | FP | Inconclusive | FN | FP | Inconclusive |
| DPRA + KS + hCLAT VS BR | 16/49 (33%) | 3/15 (20%) | 64/168 (38%) | 7/19 (37%) | 1/6 (17%) | 25/65 (39%) |
| DPRA with conclusive multiple runs/repetition prediction | 3/4 -> 2/4 | 0/1 -> 0/1 | - | 2/3 -> 2/3 | - | - |
| DPRA + KS + hCLAT VS confirmed borderlines | 13/45 | 3/14 | | 5/16 | 1/6 | |
| Option 1.1: All *in vitro* borderlines | 3/23 (13%) | 2/7 (29%) | 29/168 (17%) | 1/9 (11%) | 0/3 (0%) | 10/65 (15%) |
| Option 1.2: *In vitro* borderlines positives considered positive | 3/35 (9%) | 3/8 (38%) | 16/168 (10%) | 1/12 (8%) | 1/4 (25%) | 6/65 (9%) |

FN: false negatives; FP: false positives; HDSG: human effects based on the evaluation of the OECD DASS human data subgroup; KS: KeratinoSens; LLNA: Local Lymph Node Assay (OECD TG 429, 2010); VS BR: borderline range determined from validation datasets as described in chapter 2.

40. Finally, table 3.9 shows the impact of using options 1.1 and 1.2 on the overall false negative and false positive rates of the 2o3 DASS when taking into account the entire OECD DASS dataset, i.e. including also the test chemicals that did not have borderline outcomes, and taking into account the additional data obtained for DPRA, KeratinoSens and h-CLAT multiple individual runs/repetitions (see section 3.1 and table 3.8). ***Considering borderline outcomes within options 1.1 and 1.2 results in an overall decreased FN rate resulting in increased sensitivity of the 2o3 DASS, with only a slight increase in the FP rate i.e., decreased specificity. Furthermore, it implies that additional data and/or information is needed for up to 17% inconclusive chemicals.***

**Table 3.9. Impact on the performances of the 2o3 DASS when using options 1.1 and 1.2 and using the validation study BR for all three tests to decide on (in)conclusive predictions based on individual test method borderline results.**

| DPRA + KS + hCLAT VS BR | 2o3 vs. LLNA | | | 2o3 vs. HDSG | | |
|---|---|---|---|---|---|---|
| | FN | FP | Inconclusive | FN | FP | Inconclusive |
| DASS database | 35/135 (26%) | 5/33 (15%) | | 9/54 (17%) | 2/11 (18%) | |
| DASS database, excluding confirmed borderlines and including DPRA/KS multiple runs/repetitions | 21*/90 | 2/19 | | 4/38 | 1/5 | |
| Option 1.1: All *in vitro* borderlines excluded | 24/113 (21%) | 4/26 (15%) | 29/168 (17%) | 5/47 (11%) | 1/8 (13%) | 10/65 (15%) |
| Option 1.2: *In vitro* borderlines positives considered positive | 24/125 (19%) | 5/27 (19%) | 16/168 (10%) | 5/50 (10%) | 2/9 (22%) | 6/65 (9%) |

FN: false negatives; FP: false positives; HDSG: human effects based on the evaluation of the OECD DASS human data subgroup; KS: KeratinoSens; LLNA: Local Lymph Node Assay (OECD TG 429, 2010); VS BR: borderline range determined from validation datasets as described in chapter 2.

*Includes one FN non borderline that became TP based on DPRA multiple runs (limonene)

### *3.2.4. Comparison of different options*

41. Table 3.10 summarises the sensitivity, specificity, balanced accuracy, and fraction of inconclusive chemicals obtained when using the DPRA TG 442 BR vs. DPRA VS BR (in addition to KS & h-CLAT VS BR) when using either option 1.1 or 1.2 to decide on borderline results. ***In general, consideration of borderline outcomes (i.e. rating those as inconclusive) results in more reliable predictions of the 2o3 DASS.*** In particular, the borderline ranges seem to have a larger impact on the FN as compared to the FP rate. This may be due to the dynamic range of the test methods, i.e. the range of results for concluding that a chemical is negative is quite small compared to the range for positive calls for a chemical (and even smaller if the borderline is considered). Note: the present evaluation was conducted considering h-CLAT non-borderline negative outcomes for chemicals having log P > 3.5 as negative. The impact of considering h-CLAT non-borderline negative outcomes for chemicals having log P > 3.5 as being of low confidence can be found in Appendix 2.

42. The use of the DPRA TG 442C BR results in a slightly higher sensitivity of 2o3 DASS as compared to using the DPRA VS BR but also to a higher number of inconclusive substances. However, it is unclear how the DPRA TG 442C borderline ranges were derived. Furthermore, the lower boundary of the DPRA TG 442C BR of 3% is close/within the HPLC variability since the TG allows for 15% CV of solvent control. Finally, as stated in chapter 3.1, TG 442C is currently under revisions to require that when results fall within the TG 442C BR in particular in case of negative results, at least two concordant runs are needed for a prediction to be made. Taking into account these upcoming

requirements for DPRA multiple runs may further improve the predictions shown in table 3.10. Appendix 3 shows the additional false negative chemicals vs. LLNA obtained when using the DPRA VS BR instead of the TG 442C BR. Collected evidence from different sources shows that 3 of the 5 additional FN, seem to be human non-sensitisers or borderline skin sensitisers (i.e. tocopherol, salicylic acid and linalool).

43. Finally, considering borderline outcomes was also found to prevent UN GHS subcategory 1A mispredictions. Indeed, all three LLNA subcategory 1A and one human subcat. 1A that were mispredicted by 2o3 DASS had borderline results from at least one test method. All would become 'inconclusive' when taking into account the borderline outcomes, independent of the borderline ranges and options applied (options 1.1 and 1.2 with both DPRA TG 442C BR and VS BR).

**Table 3.10. Performance of the 2o3 DASS borderline approaches (i.e., DPRA TG BR + KS VS BR + hCLAT VS BR and of DPRA VS BR + KS VS BR + hCLAT VS BR) and using options 1.1 and 1.2.**

| 2o3 DASS performance | | 2o3 vs. LLNA | | | | 2o3 vs. HDSG | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Sensitivity | Specificity | Balanced accuracy | Inconclusive | Sensitivity | Specificity | Balanced accuracy | Inconclusive |
| DASS database | | 74% (100/135) | 85% (28/33) | 80% | - | 83% (45/54) | 82% (9/11) | 83% | - |
| DPRA TG 442C BR | Option 1.1: All *in vitro* borderlines excluded | 82% (89/108) | 85% (22/26) | 84% | 34/168 (20%) | 88% (42/48) | 88% (7/8) | 88% | 9/65 (14%) |
| | Option 1.2: *In vitro* borderlines positives considered positive | 84% (101/120) | 81% (22/27) | 83% | 21/168 (13%) | 88% (45/51) | 78% (7/9) | 83% | 5/65 (8%) |
| DPRA VS BR | Option 1.1: All *in vitro* borderlines excluded | 79% (89/113) | 85% (22/26) | 82% | 29/168 (17%) | 89% (42/47) | 88% (7/8) | 88% | 10/65 (15%) |
| | Option 1.2: *In vitro* borderlines positives considered positive | 81% (101/125) | 81% (22/27) | 81% | 16/168 (10%) | 90% (45/50) | 78% (7/9) | 84% | 6/65 (9%) |

DPRA VS BR: borderline range determined from validation datasets as described in chapter 2; DPRA TG442C BR: borderline range defined within OECD TG 442C (2020); HDSG: human effects based on the evaluation of the OECD DASS human data subgroup; LLNA: Local Lymph Node Assay (OECD TG 429, 2010).

# 4 Conclusions

44. Based on evaluation described in the present report, the following recommendations were made to the OECD:

- The wording of paragraph 24 of TG 442C (version from June 2020) has been proposed to be modified as described in paragraph 26 of this document and is currently under consultation for approval by the OECD. The revised paragraph requests that additional runs are conducted in case of DPRA negative results fall close to the threshold used to discriminate between positive and negative results.
- Feed-back was asked from DASS EG on:
  i. Preferred DPRA borderline range: TG 442C BR vs. VS BR;
  ii. Preferred option on borderlines: 1.1 (all borderlines are considered of low confidence) vs. 1.2 (only negative borderlines are considered of low confidence).

45. When asking for feed-back, the importance of distinguishing the different purposes of individual test guidelines versus the purpose of the DASS test guideline (TG) was highlighted as follows:

- Individual TGs are used to support discrimination of non-sensitisers vs. sensitisers, where the borderline ranges currently mentioned in TG 442C (2020), are used for deciding when additional runs should be conduct or not.
- In contrast, the DASS TG is intended to make a decision on classification for skin sensitisation hazard, where the proposed BR are used to define when a conclusive decision on classification can be made.
- Finally, different DASS may have different DIPs and use the individual TG results differently (e.g. the MIT derived from the h-CLAT used in the DA ITS is different from how h-CLAT results are used in the 2o3 DASS). As there is a need to take into consideration each DA specific needs and characteristics, the borderline approach presented here is not meant to be applied to other Defined Approaches than the 2o3 DA.

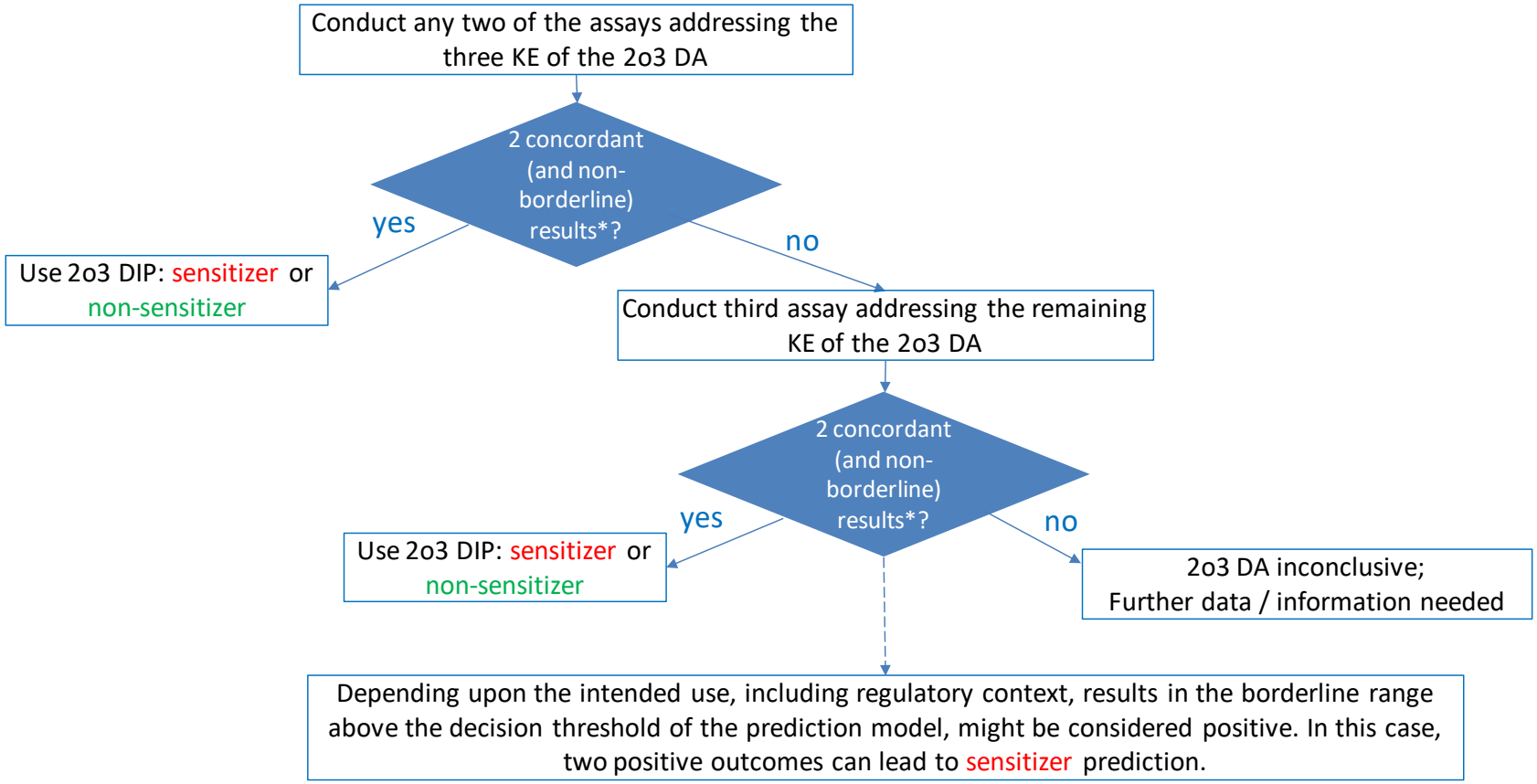
46. During a teleconference held on 16 Nov. 2020, the DASS Expert Group supported the use of the **DPRA validation study borderline** range for the 2o3 DA. Furthermore, a majority of attendants expressed preference for option 1.1, which was considered a more scientific based approach, whereas option 1.2 was considered to depend on the regulatory needs. An example mentioned was human drugs, where false positives are

not desirable. Therefore, considering the uncertainty of all borderlines like in option 1.1 was preferred.

47. During a follow-up teleconference held on 14 December 2020, it was agreed to make use of **option 1.1** as the default option for the 2o3 DASS. However, once all data for the 2o3 DASS is available, option 1.2 may still be considered as an alternative possibility, depending upon the intended use and regulatory context.

48. Based on the discussions from the DASS EG teleconference from 14 December 2020, a decision tree as shown in Figure 4.1 is suggested to derive (in)conclusive predictions of the 2o3 DASS, with no modification of the 2o3 DASS Data Interpretation Procedure.

**Figure 4.1. Decision tree to be used for the 2o3 DASS, taking into account borderline results (based on the VS BR for DPRA, KS and h-CLAT, and considering both negative and positive borderline outcomes).**




* In case a negative h-CLAT result is obtained for a chemical with log P > 3.5 (according to their related limitation as described in TG 442E (OECD, 2018b)), a 2o3 DASS prediction can only be made if the outcomes of the other two test methods composing the 2o3 DASS are concordant and are non-borderline.

# 5 References

Natsch A, Bauch C, Foertsch L, Geberick F, Norman K, Hilberer A, Inglis H, Landsiedel R, Onken S, Reuter H, Schepky A, Emter R. (2011). Toxicol In Vitro 25, 733-744 doi:10.1016/j.tiv.2010.12.014 The intra- and inter-laboratory reproducibility and predictivity of the KeratinoSens assay to predict skin sensitizers in vitro: results of a ring-study in five laboratories. Toxicol In Vitro 25, 733-744. doi: 10.1016/j.tiv.2010.12.014.

Dimitrov S, Detroyer A, Piroird C, Gomes C, Eilstein J, Pauloin T, Kuseva C, Ivanova H, Popova I, Karakolev Y, Ringeissen S, Mekenyan O (2016). Accounting for data variability, a key factor in in vivo/in vitro relationships: application to the skin sensitization potency (*in vivo* LLNA versus *in vitro* DPRA) example. J Appl Toxicol. 36, 1568-1578. doi: 10.1002/jat.3318.

ECVAM (2012). Direct Peptide Reactivity Assay (DPRA) ECVAM Validation Study Report. European Commission, Joint Research Centre, 74 pp. Available at: https://tsar.jrc.ec.europa.eu/system/files/Published/DPRA%20Validation%20Study%20Report.pdf (accessed on 05.11.2020).

ECVAM TSAR Entry Direct Peptide Reactivity Assay. Available at: https://tsar.jrc.ec.europa.eu/test-method/tm2009-06 (accessed on 15.09.2020).

ECVAM TSAR Entry Human Cell Line Activation Test. Available at: https://tsar.jrc.ec.europa.eu/test-method/tm2008-05 (accessed on 15.09.2020).

Gabbert S, Mathea M, Kolle SN, Landsiedel R (*in press*) Accounting for precision uncertainty of toxicity testing: Methods to define borderline ranges and implications for hazard assessment of chemicals. Risk Analysis *in press*

OECD (2010) Test No. 429: Skin sensitization: Local Lymph Node Assay, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris. Available at: https://www.oecd-ilibrary.org/fr/environment/test-no-429-skin-sensitisation_9789264071100-en (accessed on 05.11.2020).

OECD (2018a) Test No. 442D: In Vitro Skin Sensitisation: ARE-Nrf2 Luciferase Test Method, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris. Available at: https://doi.org/10.1787/9789264229822-en (accessed on 05.11.2020).

OECD (2018b) Test No. 442E: In Vitro Skin Sensitisation: In Vitro Skin Sensitisation assays addressing the Key Event on activation of dendritic cells on the Adverse Outcome Pathway for Skin Sensitisation, OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris. Available at: https://doi.org/10.1787/9789264264359-en (accessed on 05.11.2020).

OECD (2020) Test No. 442C: In Chemico Skin Sensitisation: : Direct Peptide Reactivity Assay (DPRA), OECD Guidelines for the Testing of Chemicals, Section 4, OECD Publishing, Paris. Available at: https://doi.org/10.1787/9789264229709-en (accessed on 05.11.2020).

Urbisch D, Mehling A, Guth K, Ramirez T, Honarvar N, Kolle S, Landsiedel R, Jaworska J, Kern PS,

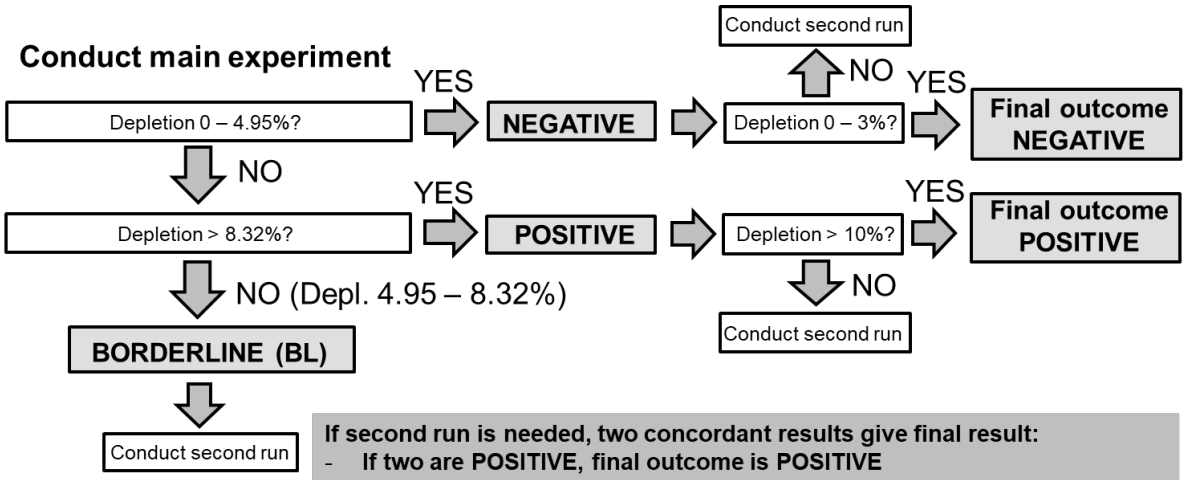
Gerberick F, Natsch A, Emter R, Ashikaga T, Miyazawa M, Sakaguchi H (2015). Assessing skin sensitization hazard in mice and men using non-animal test methods. Regul Toxicol Pharmacol. 71, 337-351. doi: 10.1016/j.yrtph.2014.12.008.

# Annex A. Prediction model for the individual in chemico/in vitro tests with multiple runs for use in 2o3 DA

The individual prediction models of h-CLAT and KeratinoSens™ require multiple runs (independent repetitions). An adaptation of the prediction model was used to determine borderline cases in the individual runs for the purpose of making predictions within the 2o3 DA. These adaptations (Figures 1.2. and 1.3.) below should be used in these methods to come to the final conclusion of the individual tests.

For the DPRA, repeated runs are required to be conducted if average depletion is within the range 3 - 10% (9 – 17% in case of Cysteine only depletion model is used). For this adaptation, the flowchart in Figure 1.1 is used to decide on run repetition and borderline assessment within the 2o3 DA.

**Figure A A.1. Flow-chart of the DPRA prediction model (mean depletion) taking into borderline ranges and multiple runs conclude on borderline results within the 2o3 DA.**



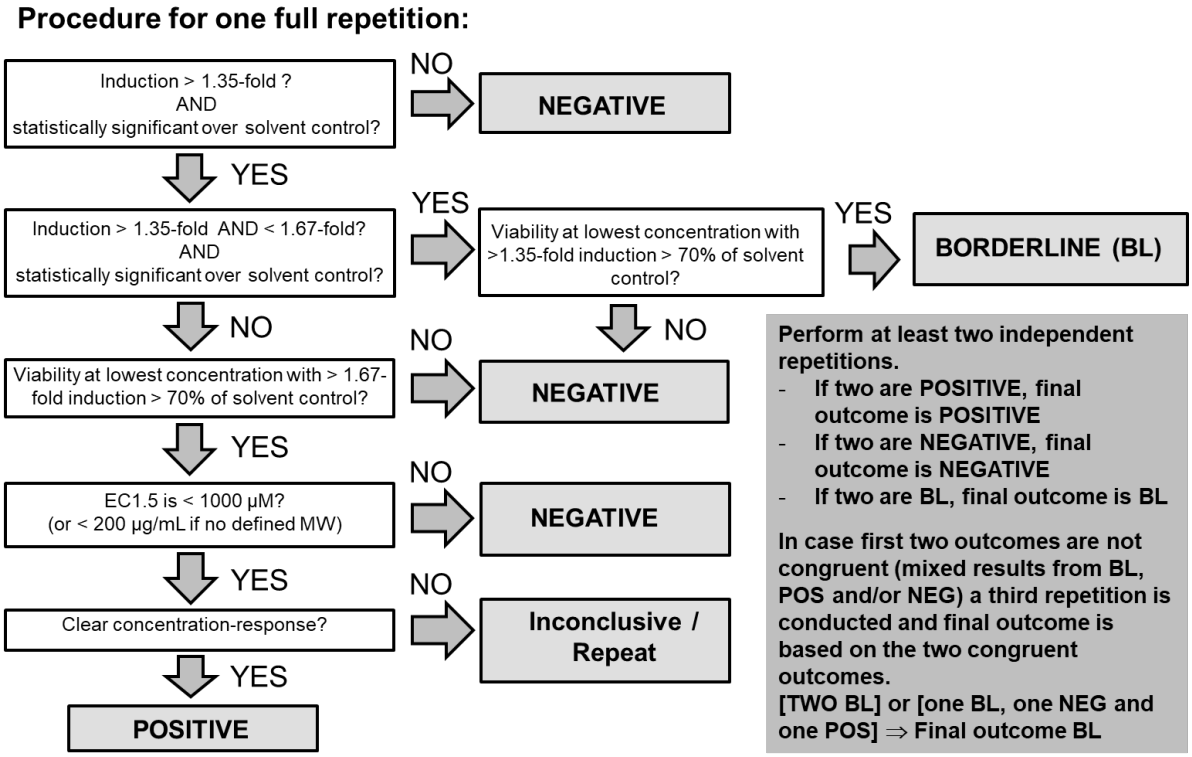

**If second run is needed, two concordant results give final result:**
- **If two are POSITIVE, final outcome is POSITIVE**
- **If two are NEGATIVE, final outcome is NEGATIVE**
- **If two are BL, final outcome is BL**

**In case first two outcomes are not congruent (mixed results from BL, POS and/or NEG) – a third run is conducted and the final outcome is based on the two congruent outcomes.**
**[TWO BL] or [one BL, one NEG and one POS] ⇒ Final outcome = BL**

The original threshold for a positive classification is 6.38%, and the statistically derived borderline range around this

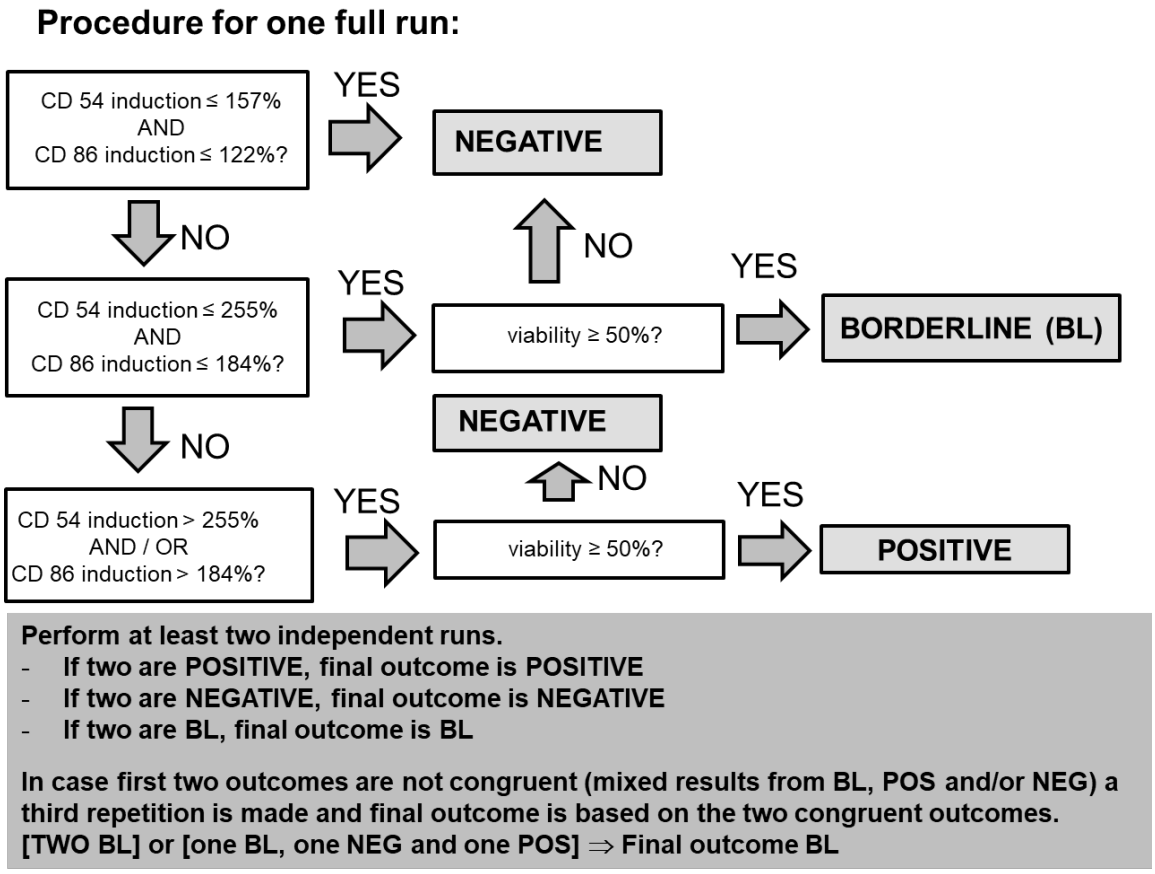
threshold is 4.95% - 8.32%. The same flowchart applies to the cysteine-only prediction model, whereby the following thresholds apply: 9% instead of 3%, >17 % instead of >10%, 10.56 % instead of 4.95% and > 18.47 % instead of >8.32%.

**Figure A A.2. Flow-chart of the KeratinoSens™ prediction model taking into account borderline ranges and multiple runs to conclude on borderline results within the 2o3 DA.**




The original threshold for a positive classification is 1.5-fold induction, and the statistically derived borderline range around this threshold is 1.35 – 1.67-fold. Note: An independent run is referred to as 'repetition' in 442D, while it is called a 'run' in 442C and 442E, these nomenclatures do mean the same thing.

**Figure A A.3. Flow-chart of the h-CLAT prediction model taking into account borderline ranges and multiple runs to conclude on borderline results within the 2o3 DA.**




The original threshold for a positive classification is 150% induction of CD86 with a statistically derived borderline range around this threshold of 122 – 184% and 200% induction of CD54 with a statistically derived borderline range around this threshold of 157 – 255%.

# Annex B. Impact of excluding h-CLAT negative outcomes for chemicals having log P > 3.5

In the evaluation presented below, non-borderline negative h-CLAT results obtained with chemicals having log P > 3.5 were considered to be of low confidence as recommended in the current version of OECD TG 442E (OECD, 2018). In such cases, only the outcomes of the two other methods were taken into account to decide on a 2o3 prediction, as follows:

- If the outcome of the two other test methods were non-borderline and concordant, a decision was made accordingly.
- If the outcome of the two other test methods were non-borderline and discordant, the 2o3 prediction was considered to be inconclusive.
- If the outcome of at least one of the two other test methods was borderline negative, the 2o3 prediction was considered to be inconclusive.

*No cases were found in which a h-CLAT negative outcome was obtained for chemicals having log P > 3.5, and at least one of the two other methods resulted in a borderline positive prediction. Furthermore, no cases were found of modified prediction when excluding negative h-CLAT results obtained with chemicals having log P > 3.5 when compared to human data from the HDSG.*

Overall, there are 8 chemicals with log P >3.5 in the dataset with (non-borderline) negative h-CLAT predictions. Of those, 3 chemicals were correct positive (with positive DPRA and KeratinoSens) in the 2o3 compared to LLNA. Furthermore, *5 chemicals resulted in a different prediction* when comparing to the LLNA reference data as follows:

- **Three** *false negatives having no borderline outcomes, became inconclusive* due to discordant DPRA and KeratinoSens outcomes (i.e. benzyl cinnamate, undec-10-enal, hexyl cinnamic aldehyde).
- **One** *false negative having a KeratinoSens borderline negative outcome and a DPRA non-borderline negative outcome, became inconclusive for all BR and options used* (i.e. N,N-dibutylaniline).
- When using the VS BR, **one** additional false negative having no borderline outcomes falling within the DPRA VS BR (but only within the DPRA TG BR), became inconclusive

due to discordant DPRA and KeratinoSens outcomes (i.e. tocopherol). That chemical was unaffected when using the DPRA TG BR, as it had a DPRA borderline outcome falling in the TG BR leading already to an inconclusive outcome.

Table A.1 summarises the sensitivity, specificity, balanced accuracy, and fraction of inconclusive chemicals obtained when including or excluding h-CLAT negative results obtained with chemicals having log P > 3.5, and taking into account the DPRA TG 442 BR, DPRA VS BR as well as options 1.1 and 1.2.

***When excluding h-CLAT negative results obtained with chemicals having log P > 3.5, the sensitivity and balanced accuracy increased when compared to the LLNA dataset for all BR and options considered. However, the fraction of inconclusive chemicals also increased.***

**Table A B.1. Performances of the 2o3 DASS when including or excluding h-CLAT negative results for chemicals having log P > 3.5, and considering DPRA TG 442 BR vs. VS BR (in addition to KS & h-CLAT VS BR), and of use of option 1.1 vs. option 1.2.[78]**

| 2o3 DASS performance | | 2o3 vs. LLNA including h-CLAT negatives for log P>3.5 | | | | 2o3 vs. LLNA excluding h-CLAT negatives for log P>3.5 | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Sensitivity | Specificity | Balanced accuracy | Inconclusive | Sensitivity | Specificity | Balanced accuracy | Inconclusive |
| DASS database | | 74% (100/135) | 85% (28/33) | 80% | - | 76% (100/132) | 85% (28/33) | 80% | 3/168 |
| DPRA TG 442C BR | Option 1.1: All *in vitro* borderlines excluded | 82% (89/108) | 85% (22/26) | 84% | 34/168 (20%) | 84% (87/104) | 85% (22/26) | 84% | 38/168 (23%) |
| | Option 1.2: *In vitro* borderlines positives considered positive | 84% (101/120) | 81% (22/27) | 83% | 21/168 (13%) | 85% (100/117) | 81% (22/27) | 83% | 24/168 (14%) |
| DPRA VS BR | Option 1.1: All *in vitro* borderlines excluded | 79% (89/113) | 85% (22/26) | 82% | 29/168 (17%) | 82% (89/108) | 85% (22/26) | 84% | 34/168 (20%) |
| | Option 1.2: *In vitro* borderlines positives considered positive | 81% (101/125) | 81% (22/27) | 81% | 16/168 (10%) | 84% (101/120) | 81% (22/27) | 83% | 21/168 (13%) |

[7] A total of 47% (18/38) chemicals having log P> 3.5 and LLNA data resulted in borderline results, and 6/9 (67%) of those having human data had borderline results

[8] To exemplify the effect on 'DPRA VS BR Option 1.1' for chemicals with log P > 3.5 on the false negative predictions:

vs. LLNA

Without consideration of the borderline ranges and the multiple/individual run info there is 14 2o3 FN vs. the LLNA in the data set.

With consideration of the borderline ranges and the multiple/individual run there is 11 2o3 FNs vs. the LLNA (one turns correct positive and two are inconclusive) in the data set.

With consideration of the borderline ranges and the multiple/individual run info and considering h-CLAT negatives (as inconclusive) there is 6 2o3 FN vs. the LLNA in the

DPRA VS BR: borderline range determined from validation datasets as described in chapter 2; DPRA TG442C BR: borderline range defined within OECD TG 442C (2020); LLNA: Local Lymph Node Assay (OECD TG 429, 2010).

---

data set (2 of which are HSDG negatives, two borderlines NC/1B, and one is a photosensitizer, see below).

vs. HDSG call

Without consideration of the borderline ranges and the multiple/individual run info there is 1 2o3 FN vs. the HDSG-call in the data set. This remains unchanged when doing the other analyses. However, the only FN chemical is chlorpromazine, which is a photosensitizer.

# Annex C. Discussion of additional false negatives when using DPRA VS BR instead of TG 442 BR, i.e. chemicals inconclusive with TG 442 BR

| Chemical | CASRN | LLNA GHS | LLNA MLLP | HU GHS | Bask h.pot. | Log P | Observations |
|---|---|---|---|---|---|---|---|
| A) Chemicals with evidence for borderline / no human sensitization potential (cont.) | | | | | | | |
| **Salicylic acid** | 69-72-7 | 1B | 12.2 | NA | 6 | 2.3 | - SCCS weight-of-evidence assessment concluded salicylic acid a non-sensitizer (SCCS opinion, 2019).<br>- Negative in human at 20% (close to HDSG decision threshold), see HDSG report.<br>- Very widely used up to 2% in leave-on cosmetics, but allergic reactions are not reported.<br>- Negative in Buehler test with 25% induction concentration. |
| **Tocopherol** | 59-02-9 | 1B | 7.4 | NA | 6 | 9.4 | - Very widespread use in leave-on cosmetics, and compared to frequency of use and abundant use, very low number of reactions reported.<br>- Vit. E is an endogenous compound of human skin as a natural antioxidant, sensitization to endogenous compounds generally rare (Kosari et al., 2010).<br>- Main reactions reported for esters of Tocopherol, but data should thus not be cumulated with free Tocopherol. |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Linalool** | 78-70-6 | 1B | 35.5 | NA | 4 | 3.0 | - SEQ 0.08/0.1 in Schnuch et al.(2015) indicates very low frequency compared to volume despite very high patch test concentration (10%). HRIPT NOEL of 13793 µg/cm$^2$ (Gerberick et al., 2001) and HMT NOEL of 55176 µg /cm$^2$ (Greif, 1967) with no cases of sensitization reported in both tests conducted at high concentration, thus it is close from being labelled as a human NC by HDSG criteria (20% top test concentration as compared to 25% set as threshold).<br>-Pure linalool is negative in multiple guinea pig studies in RIFM database and in Uter et al. (2004).<br>- The SCCS opinion on fragrance materials (2012) lists patch test data on 5423 subjects with only 18 reported cases globally (0.3%), despite widespread use of linalool (present in > 95% of cosmetic products investigated). SCCS opinion lists Linalool as a well-established contact allergen in humans, however this assessment is largely based on data on Linalool put under forced oxidation for several months (Christensson et al., 2012), which is not relevant for the regulatory assessment of the parent compound and also appears not to represent commercial use (Kern et al., 2014). Oxidized Linalool is positive in *in vitro* tests. |
| B) Chemical with evidence for (very) weak sensitization potential | | | | | | | |
| **Anethole** | 104-46-1 | NA | POS | NA | 5 | 3.0 | -Not widely open tested in the clinic, SCCS reports only three known positive cases.<br>- Negative in human maximization tests, tested at 2% (Kligmann 1971, 0 of 25 positive) and at 5% and at 10% on 101 and 105 subjects with 0 positive reaction in human maximization test (RIFM).<br>- Negative in closed epicutaneous guinea pig test with 10% at challenge and induction.<br>- Negative in guinea pig open epicutaneous test (OET), but only tested at 2%. |
| C) FN against available evidence or only LLNA evidence | | | | | | | |
| **3-Amino-phenol** | 591-27-5 | 1B | 3.2 | NA | NA | 0.2 | - Clear false negative with relative frequent positive patch test result confirming human sensitization potential. |

# References

Christensson JB., Andersen KE, Bruze M, Johansen JD, Garcia-Bravo B, Arnau AG, Goh CL, Nixon R, White IR (2012). Air-oxidized linalool-a frequent cause of fragrance contact allergy. Contact Dermatitis 67, 247-259.

Gerberick GF, Robinson MK, Ryan CA, Dearman RJ, Kimber I, Basketter DA, Wright Z, Marks JG (2001). Contact allergenic potency: correlation of human and local lymph node assay data. Am J Contact Dermat. 12, 156-161. S1046-199X(01)49486-3 [pii]

Greif, N. 1967. Cutaneous safety of fragrance material as measured by the maximization test. Am. Perf. Cosm. 82, 54–57.

Kern S., Dkhil H, Hendarsa P, Ellis G, Natsch A. (2014). Detection of potentially skin sensitizing hydroperoxides of linalool in fragranced products. Anal Bioanal Chem 406, 6165-6178.

Kligmann (1971). Appraisal of sensitizing powers by maximization testing in humans. Report to RIFM; Reference 1805 in RIFM database..

Kosari P, Alikhan A, Sockolov M, Feldman SR (2010). Vitamin E and allergic contact dermatitis. Dermatitis 21, 148-153.

SCCS (2012). Opinion on Fragrance allergens in cosmetic products. European Commission. SCCS/1459/11. Available at: https://ec.europa.eu/health/scientific_committees/consumer_safety/docs/sccs_o_102.pdf (accessed on 6.11.2020).

SCCS opinion on salicylic acid (CAS 69-72-7). Available at: https://op.europa.eu/en/publication-detail/-/publication/75be19bf-5390-11ea-aece-01aa75ed71a1 (accessed on 16.6.2020).

Schnuch A, Uter W, Lessmann H, Geier J (2015). Risk of sensitization to fragrances estimated on the basis of patch test data and exposure, according to volume used and a sample of 5451 cosmetic products. Flavour Frag. 30, 208-217.

Uter W, Schnuch A, Geier J, Lessmann H (2004). Isopropyl myristate recommended for aimed rather than routine patch testing. Contact Dermatitis 50, 242-244.