Organisation for Economic Co-operation and Development

ENV/CBC/MONO(2025)2/ANN5

Unclassified

English - Or. English

11 September 2025

ENVIRONMENT DIRECTORATE
CHEMICALS AND BIOTECHNOLOGY COMMITTEE

Annex 5: Impact of Log P on the performance of in chemico/in vitro assays and ITSv1, ITSv2, 2o3 Defined Approach on Skin Sensitisation

Series on Testing and Assessment
No. 336

(Second edition)

E-mail: ehscont@oecd.org

JT03570713

Please cite this publication as:

OECD (2025), *Supporting document to Test Guideline 497 on Defined Approaches for Skin Sensitisation, Annex 5: Impact of Log P on the performance of in chemico/in vitro assays and ITSv1, ITSv2, 2o3 Defined Approach on 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/ANN5/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).

This publication was developed in the IOMC context. The contents do not necessarily reflect the views or stated policies of individual IOMC Participating Organizations.

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

TABLES

Table 9. Performance metrics of the DPRA, KeratinoSens™, and h-CLAT assays, Derek Nexus and OECD Toolbox in silico tools, and ITSv1, ITSv2, 2o3 DASS against human data in a subset of the OECD DASS reference dataset containing only chemicals with a logP > 3.5.

1 Part 1. Summary

1.1 Purpose and background

1. The present work was undertaken to investigate the relationship between *in chemico* and *in vitro* assay results and the proposed defined approaches for skin sensitisation (DASS) (ITSv1, ITSv2, 2o3) and logP (log Kow). 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 work was conducted as a joint effort between the UK and Denmark, and its outcome presented by Denmark in its first version at the virtual F2F meeting held on 22nd and 23rd June 2020 and in its final version by the UK during the DASS EG teleconference held on 26th October 2020.

3. The presentation given to the OECD DASS EG was based on an older version of the OECD DASS reference dataset. The analysis presented in this report was carried out using OECD_invitro_invivo_insilico_DA_2020-12-10.txt and the analysis conducted on 10th December 2020. The conclusions from each analysis remain the same.

    - Paragraph 4 of Annex I: *In vitro* skin sensitisation: human cell line activation test (h-CLAT) in OECD TG 442E[1] states that "*Test chemicals with a Log Kow greater than 3.5 tend to produce false negative results (14). Therefore negative results with test chemicals with a Log Kow greater than 3.5 should not be considered. However, positive results obtained with test chemicals with a Log Kow greater than 3.5 could still be used to support the identification of the test chemical as a skin sensitiser.".* Hence this analysis was inspired by OECD TG 442E[1] and reference 14 therein (Takenouchi et al 2013[2]), who reported a marked reduction in sensitivity of the h-CLAT vs the LLNA for 31 (27 sensitisers, 4 non-sensitisers) chemicals with log P > 3.5 compared to 112 (78 sensitisers, 34 non-sensitisers) chemicals with log P ≤ 3.5 when investigating a dataset of 143 chemicals. The sensitivities were 52% and 94%, respectively, i.e. there was a reduction of 42% in sensitivity for chemicals with logP > 3.5. This led to OECD TG 442E stating that negative h-CLAT results should not be used.
        - Note: the current analysis has been carried out on the highly curated OECD DASS reference dataset containing 196 chemicals (168 with LLNA data; 66 with human data) where particular attention was made to ensure that the assignment of negative and positive LLNA and human results were robust and highly reliable (insert citation of relevant section of the Supporting Info). In comparison, less information is available on how the h-CLAT and LLNA results were derived in the dataset used in Takenouchi et al (2013), although it is likely that it has not been curated to the same extent as the OECD DASS reference dataset.

4. Based on the statement in OECD TG 442E, the following analyses were conducted:

    a. Analysis of the following and their performance using a logP cut-off of 3.5:
        i. The *in chemico/in vitro* assays; DPRA, KeratinoSens™ and h-CLAT.
        ii. The *in silico* tools used in the ITS DASS; Derek Nexus and OECD Toolbox.
        iii. ITSv1 (which uses Derek Nexus) and ITSv2 (which uses OECD Toolbox) DASS.
        iv. 2o3 DASS

1.2 Data analysis, results, and conclusions

Data analysis

1. All analysis was undertaken using the OECD DASS reference dataset of 196 chemicals (168 with LLNA data; 66 with human data) (see Supporting Info) and each information source (*in chemico/in vitro* assays and *in silico* tools) and DASS (ITSv1, ITSv2, 2o3) was assessed against the LLNA and human data therein. The analysis used all chemicals in the dataset, a subset of chemicals with logP ≤ 3.5, and a subset of chemicals with logP > 3.5.

2. The total number of chemicals with results for each *in chemico/in vitro* assay, *in silico* tool, and ITSv1, ITSv2, 2o3 DASS prediction may differ slightly for several reasons: many chemicals lack human data; some lack LLNA data; a few chemicals lack data for one or more *in chemico/in vitro* assay; for one chemical a logP could not be calculated/was unavailable (Kathon CG); for some chemicals deriving an *in silico* result was not possible.

3. The performance metrics used are balanced accuracy, sensitivity, and specificity.
    - Balanced accuracy = (Sensitivity + Specificity / 2)
    - Sensitivity = True Positives / (True Positives + False Negatives)
    - Specificity = True Negatives / (True Negatives + False Positives)

4. The OECD DASS reference dataset is highly biased towards positives (LLNA - 135/168, 80% sensitisers; human - 55/66, 83% sensitisers), quite unlike the chemical universe (analysis of REACH registrations found that around 20% may be positive[3,4]) which can have a significant effect on the performance metrics e.g. it was difficult to get robust figures for specificity due to small number of negatives and the ensuing disproportionate impact relatively few false positives can make to this metric. Thus, the following analyses focus on sensitivity. Sensitivity is important because to protect human health, ideally there would be very few false negatives generated from *in chemico/in vitro/in silico* results or any given DASS utilising these information sources.

5. The assessment in this document is purely statistical and does not take into account other considerations (e.g. the chemistry, additional evidence besides reference LLNA and human classifications).

6. Note that the analysis was performed with all the chemicals in the dataset with agreed *in vivo* classifications, including inconclusive results of the DAs, borderline results in the *in chemico/in vitro* methods and out of domain *in silico* predictions.

Results

7. ***In silico, in chemico/in vitro* and DASS approaches vs LLNA**: The performance of the DPRA, KeratinoSens™, h-CLAT, Derek Nexus, OECD Toolbox, and ITSv1, ITSv2, 2o3 DAs were assessed against the LLNA data in the OECD DASS reference dataset containing chemicals with a log P > 3.5 (n = 39) and chemicals with logP ≤ 3.5 (n = 127-128)**.** The main results of the analysis on the DASS reference chemicals were a marked reduction in sensitivity for the *in chemico/in vitro* assays and the DASS approaches vs the LLNA for the 39 chemicals with log P > 3.5 relative to the 128 chemicals with log P ≤ 3.5 (Table 1, Table 2). The difference in sensitivities were DPRA: 29%; KeratinoSens™: 17%; h-CLAT: 23%; ITSv1: 11%; ITSv2: 12%; 2o3: 21%. The reduction of the sensitivity for the 2o3 vs the LLNA for chemicals with a log P > 3.5 was larger than the reduced sensitivity of either ITS for these chemicals. The reason for ITS having lower reduction in sensitivity than 2o3 for chemicals with a log P > 3.5 is plausibly due to the fact that the *in silico* tools used in the ITS (both of which use information from the LLNA and the GPMT) were to some extent mitigating the impact of the reduced sensitivity of the *in chemico/in* vitro assays vs. the LLNA on the overall ITS prediction.

Table 1. Summary of performance metrics of the DPRA, KeratinoSens™, and h-CLAT assays, Derek Nexus and OECD Toolbox in silico tools, and ITSv1, ITSv2, 2o3 DASS against LLNA data in the OECD DASS reference dataset.

	Sensitivity (%)			Specificity (%)	
LogP	**≤ 3.5** (n = 127-128)	**> 3.5** (n = 39)	**Difference**	**≤ 3.5** (n = 127-128)	**> 3.5** (n = 39)
DPRA	77	48	-29	85	83
KeratinoSens™	75	58	-17	74	50
h-CLAT	87	64	-23	70	50
Derek Nexus	90	85	-5	74	67
OECD Toolbox	89	88	-1	70	83
ITSv1	90	79	-11	78	50
ITSv2	91	79	-12	74	50
2o3	79	58	-21	85	83

*Specificity calculations for log P > 3.5 have a significant amount of uncertainty due to the small number of negative chemicals (TN and FP) in each subset (see Table 2).

Table 2. Summary of outcomes of the DPRA, KeratinoSens™, and h-CLAT assays, Derek Nexus and OECD Toolbox in silico tools, and ITSv1, ITSv2, 2o3 DASS against LLNA data in the OECD DASS reference dataset.

	True positives (TP)	False negatives (FN)	True negatives (TN)	False positives (FP)

LogP	≤ 3.5	> 3.5	≤ 3.5	> 3.5	≤ 3.5	> 3.5	≤ 3.5	> 3.5
DPRA	78	16	23	17	23	5	4	1
KeratinoSens™	76	19	25	14	20	3	7	3
h-CLAT	87	21	13	12	19	3	8	3
Derek Nexus	91	28	10	5	20	4	7	2
OECD Toolbox	90	29	11	4	19	5	8	1
ITSv1	90	26	10	7	21	3	6	3
ITSv2	91	26	9	7	20	3	7	3
2o3	80	19	21	14	23	5	4	1

8. ***In silico, in chemico/in vitro* and DASS approaches vs human data:** The performance of the DPRA, KeratinoSens™, h-CLAT, Derek Nexus, OECD Toolbox, and ITSv1, ITSv2, 2o3 DASS were assessed against the human data in a subset of the OECD DASS reference dataset containing chemicals with a log P > 3.5 and chemicals with logP ≤ 3.5 (Table 3). The number of chemicals with log P > 3.5 and human data was however so low that no conclusions regarding sensitivity and specificity for this subset of chemicals can be drawn (n = 12). There were 53 chemicals with a logP ≤ 3.5 and human data, although depending on the information source or DASS approach, the number of chemicals with log P ≤ 3.5 ranged from 43-53 chemicals. A sufficiently high number of chemicals with positive human results were available in this subset to give robust sensitivity measures, but a too low number of chemicals with negative human data were available to make conclusions regarding specificity. The results indicate for substances with log P ≤ 3.5, that LLNA can predict human sensitisers with a high sensitivity of 93%, and that ITSv1 and ITSv2 DASS had a similar (2% lower) sensitivity whereas the 2o3 DASS had a clear lower sensitivity (10%).

Table 3. Performance metrics of the DPRA, KeratinoSens™, and h-CLAT assays, Derek Nexus and OECD Toolbox in silico tools, and ITSv1, ITSv2, 2o3 DASS against human data in a subset of the OECD DASS reference dataset containing only chemicals with a logP ≤ 3.5.

	BA	Se	Sp*	TP	FP	TN	FN	*n* chemicals
LLNA	63	93	33	37	2	1	3	43
DPRA	83	85	80	40	1	4	7	52
KeratinoSens™	76	73	80	35	1	4	13	53
h-CLAT	83	85	80	40	1	4	7	52
Derek Nexus	77	94	60	45	2	3	3	53
OECD Toolbox	93	85	100	41	0	5	7	53
ITSv1	86	91	80	42	1	4	4	51
ITSv2	86	91	80	42	1	4	4	51
2o3	81	83	80	39	1	4	8	52

*Specificity calculations are all highly uncertain/inconclusive due to the small number of negative chemicals (n = 4-5) in the dataset.

Conclusions

9. It is clear from the analysis on the OECD DASS reference dataset that skin sensitising chemicals (as identified by the LLNA) with a LogP ≤ 3.5 can generally be identified by the DASS approaches analysed (ITSv1, ITSv2, 2o3), although ITSv1 and ITSv2 have higher sensitivities than 2o3. However, for substances with a Log P > 3.5 within this dataset there is marked decreased sensitivity for both ITSv1, ITSv2 and 2o3 DASS (Table 1, Table 2), again using LLNA data as the benchmark.

10. For chemicals with a logP ≤ 3.5 which also have human sensitisation data, both ITSv1 and ITSv2 DASS have similar sensitivities as the LLNA, whereas the sensitivity of the 2o3 DASS with respect to human data is clearly lower (10%) (Table 3).

11. The very low number of DASS reference chemicals with logP > 3.5 and negative LLNA data means that the specificity measures for the *in chemico/in vitro* tests, *in silico* tools and DASS are inconclusive (see Table 1, Table 2 and Part II of this report).

12. The number of DASS reference substance data with human data and logP > 3.5 was low meaning that the performance measures of the *in chemico/in vitro* assays, *in silico* tools and DASS approaches against human data are inconclusive (see Part II).

2 Part 2. Detailed analysis

2.1 Analysis of *in chemico/in vitro* assays, *in silico* tools, and ITSv1, ITSv2, 2o3 DASS using the OECD DASS reference dataset

2.1.1 Compared against LLNA data

1. The performance of the DPRA, KeratinoSens, h-CLAT, Derek Nexus, OECD Toolbox, and ITSv1, ITSv2, 2o3 DASS were assessed against the LLNA data in the OECD DASS reference dataset (n = 168). Depending on the information source or DASS used the number of chemicals assessed ranged from 167-168.

2. The balanced accuracies (BA) for all *in chemico/in vitro* assays were high (70%-78%) with a reasonably fair balance between sensitivity (Se) and specificity (Sp).

3. Both *in silico* tool options in the ITS (Derek Nexus - ITSv1; OECD Toolbox - ITSv2) have high sensitivity (89% for both) and specificity (73% for both) when compared against the LLNA, as do the ITS which use these as an information source. Contrary to this, the 2o3 has lower sensitivity (74%) and higher specificity (85%) (Table 4).

Table 4. Performance metrics of the DPRA, KeratinoSens™, and h-CLAT assays, Derek Nexus and OECD Toolbox in silico tools, and ITSv1, ITSv2, 2o3 DASS against LLNA data in the OECD DASS reference dataset.

	BA	Se	Sp	TP	FP	TN	FN	*n* chemicals
DPRA	78	70	85	95	5	28	40	168
KeratinoSens™	70	71	70	96	10	23	39	168
h-CLAT	74	81	67	109	11	22	25	167
Derek Nexus	81	89	73	120	9	24	15	168
OECD Toolbox	81	89	73	119	9	24	15	167
ITSv1	80	87	73	117	9	24	17	167
ITSv2	79	88	70	118	10	23	16	167
2o3	79	74	85	100	5	28	35	168

2.1.2 Compared against human data

1. The performance of the LLNA, DPRA, KeratinoSens™, h-CLAT, Derek Nexus, OECD Toolbox, and ITSv1, ITSv2, 2o3 DASS were assessed against the human data in the

OECD DASS reference dataset (n = 66). Depending on the information source or DASS used the number of chemicals assessed ranged from 56-66 which is approximately a third of the size of the OECD DASS reference dataset of 168 chemicals with *in chemico/in vitro* data and LLNA data (Table 5).

2. The LLNA can identify human sensitisers well, with a sensitivity of 94%, however, the specificity is extremely low at 22%. The conclusion regarding specificity is highly uncertain because the number of DASS reference chemicals with conclusive negative human data and LLNA data is only 9. The conclusion regarding sensitivity is much more certain as the number of chemicals with positive human data and LLNA data is 47.
3. The DPRA and KeratinoSens™ assays have a sensitivity of 81 % and 71%, respectively. For h-CLAT the sensitivity was very high (87%), (Table 5).

4. The *in silico* tools and both ITS perform comparably to the LLNA – with high sensitivities generally above 90% whereas 2o3 has a lower but still high sensitivity of 83%.

Table 5. Performance metrics of the DPRA, KeratinoSens™, and h-CLAT assays, Derek Nexus and OECD Toolbox in silico tools, and ITSv1, ITSv2, 2o3 DASS against human data in the OECD DASS reference dataset.

	BA	Se	Sp*	TP	FP	TN	FN	*n* chemicals
LLNA	58	94	22	44	7	2	3	56
DPRA	82	81	82	44	2	9	10	65
KeratinoSens™	76	71	82	39	2	9	16	66
h-CLAT	71	87	55	47	5	6	7	65
Derek Nexus	69	93	45	51	6	5	4	66
OECD Toolbox	74	85	64	46	4	7	8	65
ITSv1	73	92	55	49	5	6	4	64
ITSv2	73	92	55	49	5	6	4	64
2o3	83	83	82	45	2	9	9	65

*Specificity calculations all have a significant amount of uncertainty due to the small number of negative chemicals with human data (n = 11) in the dataset.

2.2 Analysis of *in chemico/in vitro* assays, *in silico* tools, and ITSv1, ITSv2, 2o3 DASS using a subset of the OECD DASS reference dataset only containing chemicals with a logP ≤ 3.5.

2.2.1 Compared against LLNA data

1. The performance of the DPRA, KeratinoSens™, h-CLAT, Derek Nexus, OECD Toolbox, and ITSv1, ITSv2, 2o3 DASS were assessed against the LLNA data in a subset of the OECD DASS reference dataset containing only chemicals with a logP ≤ 3.5 (n = 128).

Depending on the information source or DASS used the number of chemicals assessed ranged from 127-128 chemicals.

2. This subset is relatively large and contains 128 chemicals. The results in Table 6 from the OECD DASS subset containing chemicals with a logP ≤ 3.5 suggest that ITSv1, ITSv2, and the 2o3 can discriminate well between sensitisers and non-sensitisers when compared to LLNA data. It is also seen that both ITS generally have a sensitivity around 90% but specificity around 75% whereas compared this, the 2o3 has a lower sensitivity of 79% but a higher specificity of 85%.

3. The balanced accuracy, sensitivity, and specificity are quite high for each *in chemico/in vitro* assay when only including chemicals with a logP ≤ 3.5 e.g. balanced accuracies between 75%-81% (Table 6), however the DPRA and KeratinoSens™ have a lower sensitivity than the h-CLAT.

4. Both *in silico* tool options in the ITS (Derek Nexus - ITSv1; OECD Toolbox - ITSv2) have a high sensitivity (90%; 89%) but a lower specificity (74%; 70%) when compared against the LLNA.

Table 6. Performance metrics of the DPRA, KeratinoSens™, and h-CLAT assays, Derek Nexus and OECD Toolbox in silico tools, and ITSv1, ITSv2, 2o3 DASS against LLNA data in a subset of the OECD DASS reference dataset containing only chemicals with a logP ≤ 3.5.

	BA	Se	Sp*	TP	FP	TN	FN	*n* chemicals
DPRA	81	77	85	78	4	23	23	128
KeratinoSens™	75	75	74	76	7	20	25	128
h-CLAT	79	87	70	87	8	19	13	127
Derek Nexus	82	90	74	91	7	20	10	128
OECD Toolbox	80	89	70	90	8	19	11	128
ITSv1	84	90	78	90	6	21	10	127
ITSv2	83	91	74	91	7	20	9	127
2o3	82	79	85	80	4	23	21	128

2.2.2 Compared against human data

1. The performance of the DPRA, KeratinoSens™, h-CLAT, Derek Nexus, OECD Toolbox, and ITSv1,ITSv2, 2o3 DASS were assessed against the human data in a subset of the OECD DASS reference dataset containing only chemicals with a logP ≤ 3.5 (n = 53). Depending on the information source or DASS used the number of chemicals assessed ranged from 43-53 chemicals. With respect to the conclusions below it should be noted that no clear conclusions can be drawn regarding specificity because the number of chemicals with log P ≤ 3.5 and negative human data is very low, i.e. only 5.

2. The LLNA can predict human sensitisers with a high sensitivity of 93%. KeratinoSens™ had the lowest sensitivity of 73%, whereas the DPRA, the h-CLAT, and OECD Toolbox

had sensitivities of 85%. Derek Nexus had a similar sensitivity as the LLNA, as did ITSv1 and ITSv2 DASS whereas the 2o3 DASS had a lower sensitivity (83%).

Table 7. Performance metrics of the DPRA, KeratinoSens™, and h-CLAT assays, Derek Nexus and OECD Toolbox in silico tools, and ITSv1, ITSv2, 2o3 DASS against human data in a subset of the OECD DASS reference dataset containing only chemicals with a logP ≤ 3.5.

	BA	Se	Sp*	TP	FP	TN	FN	*n* chemicals
LLNA	63	93	33	37	2	1	3	43
DPRA	83	85	80	40	1	4	7	52
KeratinoSens™	76	73	80	35	1	4	13	53
h-CLAT	83	85	80	40	1	4	7	52
Derek Nexus	77	94	60	45	2	3	3	53
OECD Toolbox	93	85	100	41	0	5	7	53
ITSv1	86	91	80	42	1	4	4	51
ITSv2	86	91	80	42	1	4	4	51
2o3	81	83	80	39	1	4	8	52

*Specificity calculations are all highly uncertain/inconclusive due to the small number of negative chemicals (n = 5) in the dataset.

2.3 Analysis of *in chemico/in vitro* assays, *in silico* tools, and ITSv1, ITSv2, 2o3 DASS using a subset of the OECD DASS reference dataset only containing chemicals with a logP > 3.5.

2.3.1 Compared against LLNA data

1. The performance of the DPRA, KeratinoSens™, h-CLAT, Derek Nexus, OECD Toolbox, and ITSv1, ITSv2, 2o3 DASS were assessed against the LLNA data in a subset of the OECD DASS reference dataset containing only chemicals with a logP > 3.5. This subset contained 39 chemicals out of which 33 have positive LLNA results and only 6 have negative LLNA results. The latter means that all statement below regarding specificity are inconclusive because they relate to such a small number of chemicals.

2. All *in chemico/in vitro* tests have clearly lower sensitivities for chemicals with log P > 3.5. The decrease in sensitivity, compared to chemicals with a logP ≤ 3.5, was 29% for DPRA, 17% for the KeratinoSens™, and 23% for the h-CLAT.

3. Both *in silico* tools retain high sensitivity, however, Derek Nexus has a slight decrease of 5% for chemicals with log P > 3.5 compared to chemicals with a logP ≤ 3.5. Only 1% decrease in sensitivity was observed for the OECD Toolbox.

4. ITSv1 and ITSv2 DASS have a sensitivity of 79%, although this represents a 11% and 12% decrease, respectively, relative to the subset of reference chemicals with a logP <

3.5. The 2o3 DASS however, has a low sensitivity of only 58% and hence a marked decrease in sensitivity for chemicals with log P > 3.5 relative to that for chemicals with log P ≤ 3.5, i.e. a 21% lower sensitivity.

Table 8. Performance metrics of the DPRA, KeratinoSens™, and h-CLAT assays, Derek Nexus and OECD Toolbox in silico tools, and ITSv1, ITSv2, 2o3 DASS against LLNA data in a subset of the OECD DASS reference dataset containing only chemicals with a logP > 3.5.

	BA	Se	Sp*	TP	FP	TN	FN	*n* chemicals
DPRA	66	48	83	16	1	5	17	39
KeratinoSens™	54	58	50	19	3	3	14	39
h-CLAT	57	64	50	21	3	3	12	39
Derek Nexus	76	85	67	28	2	4	5	39
OECD Toolbox	86	88	83	29	1	5	4	39
ITSv1	64	79	50	26	3	3	7	39
ITSv2	64	79	50	26	3	3	7	39
2o3	70	58	83	19	1	5	14	39

*Specificity calculations are all highly uncertain / inconclusive due to the small number of negative chemicals (n = 6) in the dataset.

2.3.2 Compared against human data

1. The performance of the DPRA, KeratinoSens™, h-CLAT, Derek Nexus, OECD Toolbox, and ITSv1, ITSv2, 2o3 DASS were assessed against the human data in a subset of the OECD DASS reference dataset containing only chemicals with a logP > 3.5. This subset is extremely small and contains only 12 chemicals, of which 6 are sensitisers and 6 are non-sensitisers. Based on the low number of chemicals with a logP > 3.5 and human data, no clear conclusions can be drawn from analysis of sensitivity and specificity.

Table 9. Performance metrics of the DPRA, KeratinoSens™, and h-CLAT assays, Derek Nexus and OECD Toolbox in silico tools, and ITSv1, ITSv2, 2o3 DASS against human data in a subset of the OECD DASS reference dataset containing only chemicals with a logP > 3.5.

	BA*	Se*	Sp*	TP*	FP*	TN*	FN*	*n* chemicals
LLNA	58	100	17	6	5	1	0	12
DPRA	67	50	83	3	1	5	3	12
KeratinoSens™	67	50	83	3	1	5	3	12
h-CLAT	67	100	33	6	4	2	0	12
Derek Nexus	58	83	33	5	4	2	1	12
OECD Toolbox	58	83	33	5	4	2	1	12
ITSv1	67	100	33	6	4	2	0	12
ITSv2	67	100	33	6	4	2	0	12

2o3	83	83	83	5	1	5	1	12

*All performance metrics have a significant amount of uncertainty and are hence inconclusive due to the small number of chemicals in the dataset (n = 12).

References

1. OECD. *Key Event Based Test Guideline 442E: 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, 2018).
2. Takenouchi, O., Miyazawa, M., Saito, K., Ashikaga, T. & Sakaguchi, H. Predictive performance of the human Cell Line Activation Test (h-CLAT) for lipophilic chemicals with high octanol-water partition coefficients. *J. Toxicol. Sci.* **38**, 599–609 (2013).
3. Luechtefeld, T. *et al.* Analysis of publically available skin sensitization data from REACH registrations 2008-2014. *ALTEX* **33**, 135–148 (2016).
4. The Danish QSAR Database. (2020). https://qsardb.food.dtu.dk/db/index.html
   Date of search 15.12.2020
   A) search for number of REACH registered mono-constituent, organic substances: 11096
   B) search for number of REACH registered mono-constituent, organic substances with positive in domain battery predictions (majority of models in Leadscope, Case Ultra and SciQSAR): 2176
   C) Percentage of predicted skin sensitisers amongst 11096 REACH registered mono-constituent, organic substances in the DK QSAR Database: (217600/11096)= 19.6 %