Organisation for Economic Co-operation and Development

ENV/CBC/MONO(2026)16/ANN12

Unclassified

English - Or. English

4 September 2026

ENVIRONMENT DIRECTORATE
CHEMICALS AND BIOTECHNOLOGY COMMITTEE

Cancels & replaces the same document of 8 July 2026

Annex 12.c to the Supporting Document of Test Guideline 497: Prediction of Point of Departure (PoD) for skin sensitization based on Defined Approach with regression models: Evaluation vs. reference data

OECD Series on Testing and Assessment

E-mail: ehscont@oecd.org.

JT03591922

This work was approved and declassified by the Chemicals and Biotechnology Committee on 22/06/2026.

Please cite this publication as:

OECD (2026), *Supporting document to Test Guideline 497 on Defined Approaches for Skin Sensitisation, Annex 12.c to the Supporting Document of Test Guideline 497: Prediction of Point of Departure (PoD) for skin sensitization based on Defined Approach with regression models: Evaluation vs. reference data,* OECD Series on Testing and Assessment, No. 438, OECD Environment, Health and Safety, Paris, https://one.oecd.org/document/ENV/CBC/MONO(2026)16/ANN12/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.

Annex 12.c. Prediction of Point of Departure (PoD) for skin sensitization based on Defined Approach with regression models: Evaluation vs. reference data

Background

The regression models were developed to derive a quantitative PoD and they are independent of the hazard prediction model of the individual tests and take only the quantitative data from the in vitro tests into account [1, 2]. Thus, a PoD is also derived if e.g. only one test is positive. In this regard they are similar to the SARA model, which is also independent of the hazard ID prediction model of the independent test. To perform both hazard assessment and risk assessment, the PoD models can be combined in a tiered assessment with, for example, the "2 out of 3" DA, as outlined in [3].
The predictivity of the regression models trained on LLNA data (EQ 1,4,5 and 6) to predict LLNA EC3 values has been assessed in detail [1]. The predictivity of the regression models trained on potency values (EQ 1d, 4d, 5e and 6d) to predict potency values was analysed, too, for details see [2].

These models were now combined with a fixed workflow as Defined Approach (DA): For each set of data (two assays or three assays conducted), the appropriate equation is automatically selected and the lowest PoD from the EC3 prediction and the PV prediction is taken forward as the final PoD defined by the DA (Figure 1).

For all chemicals in the published database [1], this DA was applied and the results and the calculations are given for all chemicals in the file "Appendix 4.2A_DA_regression models_database and results.xls".

Note: For non-cytotoxic chemicals negative in all assays, the mathematically calculated PoD would only be dependent on the molecular weight. Thus, chemicals with a MW < 150 and negative in all assays would have a predicted PoD which is < 100% (<25'000 µg/cm^2), while for chemicals with a MW > 150 the calculated PoD can also be > 100% (>25'000 µg/cm^2). However, in practice calculation of a PoD for chemicals negative in all assays is not actually required.

The database contains:

a) All chemicals with at least two assays (n = 220)
   In addition, the database was analysed for the following subsets:
b) All chemicals with all three assays (n = 188)
c) All chemicals with OECD MLLP LLNA data [4] and *in vitro* data (n = 146)
d) All chemicals with Potency values as derived by [2] and *in vitro* data (n = 139)
e) All chemicals with Potency values as derived by [5] and *in vitro* data (n = 29)

For a), b) and c), the comparison vs. reference data is based on the LLNA EC3, as for these chemicals

only LLNA EC3 data are available for all chemicals.
For d) and e) comparison vs. reference data can be done both against LLNA EC3 values and Potency values (PV).

To compare vs reference values, for each chemical the absolute fold difference is calculated according to the formula:

$$Absolute\ fold\ difference = 10^{|Log\ (Predicted\ PoD) - Log(Reference\ value)|}$$

This gives the fold absolute deviation of the predicted PoD from the reference value in either direction. The results on the individual chemicals are given in the file "Annex 11.a A_DA_regression models_database and results.xls" and are summarised below graphically and as summary figures in the form of the geometric mean and the median value.
In parallel for each chemical the fold difference is calculated according to the formula

$$Fold\ difference = 10^{(Log\ (Predicted\ PoD) - Log(Reference\ value))}$$

A geometric mean of the fold difference close to 1 indicates, that the model is balanced and has no bias towards under- or over-prediction. A value below 1 indicates a model biased for conservative predictions and a value above one indicates an overall bias for underprediction.

Reference data and PoD values are expressed both in % in a test vehicle which is applied at a dose of 25 mg/cm^2 (i.e. the applied amount in the LLNA) or in µg/cm^2. The % values can be multiplied by 250 to arrive at the µg/cm^2 value.

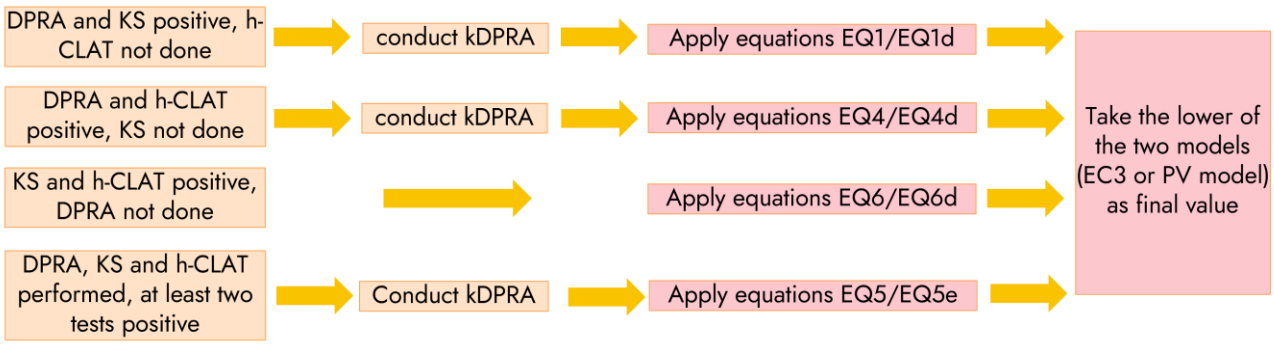

Figure 1. Combination of the individual regression models to arrive at a fixed PoD based on a given set of input data in the Defined Approach (DA).

Results

a) All chemicals with at least two assays (n = 220)

When evaluated against the LLNA EC3, for all chemicals, the Geomean and Median absolute fold difference are 2.6 – 3.5, while the Geomean fold difference is slightly below 1 indicating a balanced prediction.

Geo. mean fold difference: 0.86
Geo. mean abs. fold difference: 3.46
Median abs. fold difference: 2.56

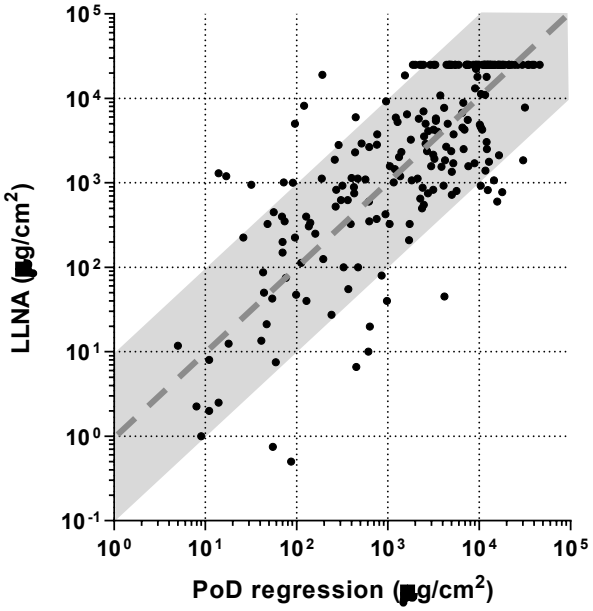

Figure 2. All chemicals with at least two tests available, compared versus OECD MLLP LLNA EC3 data or other LLNA data according [1] in case no OECD MLLP EC3 is available. Grey dashed line: line of identity. Grey area – area of less than 10-fold misprediction. Chemicals negative in the LLNA are plotted with the default value of 25'000, i.e. 100% LLNA test concentration.

Omitting the non-sensitizers (n=66), the following figures are obtained:

Geo. mean fold difference: 1.18
Geo. mean abs. fold difference: 3.90
Median abs. fold difference: 2.88

b) All chemicals with all three assays conducted (n = 188)

When evaluated against the LLNA EC3, for all chemicals for which all three tests are available, the

Geomean and Median absolute fold difference are 2.6 – 3.6, while the Geomean fold difference is slightly below 1 indicating a balanced prediction.
Geo. mean fold difference: 0.88
Geo. mean abs. Fold difference: 3.57
Median abs. fold difference: 2.63

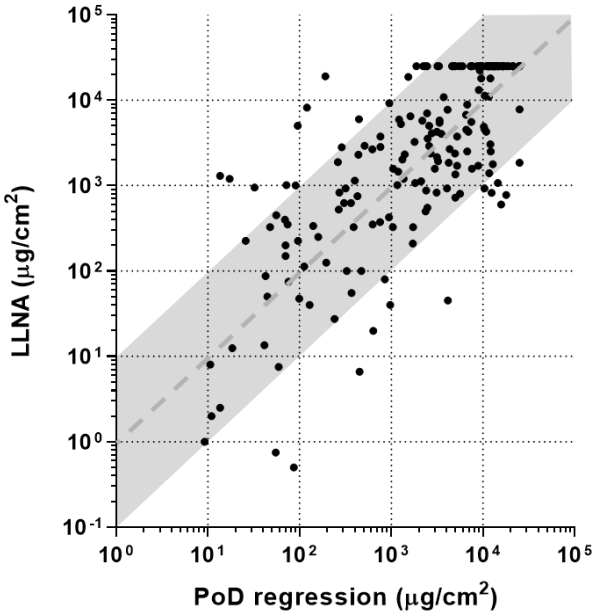

Figure 3. All chemicals with all three tests available, compared versus OECD MLLP LLNA EC3 data or other LLNA data according [1] in case no OECD MLLP EC3 is available. Grey dashed line: line of identity. Grey area – area of less than 10-fold misprediction. Chemicals negative in the LLNA are plotted with the default value of 25'000, i.e. 100% LLNA test concentration.

Omitting the non-sensitizers (n=52), the following figures are obtained:

Geo. mean fold difference: 1.16
Geo. mean abs. fold difference: 4.05
Median abs. fold difference: 2.98

c) All chemicals with OECD MLLP LLNA data and in vitro data (n = 146)

Next, the subset of chemicals, for which a curated MLLP LLNA EC3 is available [4] was separately assessed. When evaluated against the LLNA EC3, the Geomean and Median absolute fold difference are 2.5 – 3.5, while the Geomean fold difference is slightly below 1, again indicating a balanced prediction for this subset too.

This result is overall very similar to the result of the evaluation of the full database or the evaluation on the chemicals with all three tests available (Figures 2 and 3).

Geo. mean fold difference: 0.86
Geo. mean abs. fold difference: 3.50
Median abs. fold difference: 2.47

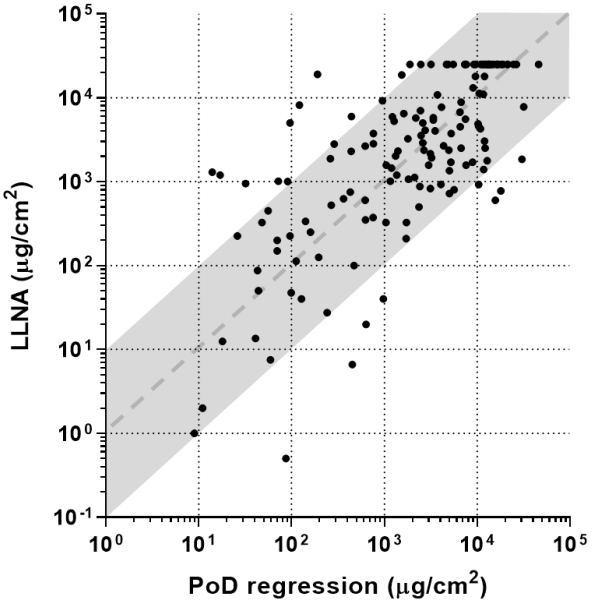

Figure 4. All chemicals with all three tests available, compared versus OECD MLLP LLNA EC3 data. Grey dashed line: line of identity. Grey area – area of less than 10-fold misprediction. Chemicals negative in the LLNA are plotted with the default value of 25'000, i.e. 100% LLNA test concentration.

Omitting the non-sensitizers (n=33), the following figures are obtained:

Geo. mean fold difference: 1.04
Geo. mean abs. fold difference: 3.93
Median abs. fold difference: 2.86

d) All chemicals with Potency values and in vitro data (n = 139)

The subset of chemicals in the database with potency values [2] was further evaluated. First the PoD by the DA was compared vs. the LLNA EC3 values (Figure 5). The Geomean and Median absolute fold difference are 2.4 – 3.4, while the Geomean fold difference is slightly below 1, again indicating a balanced prediction for this subset too.
This result, again comparing to LLNA EC3 values, is overall very similar as the result of the evaluation of the other data-subsets (Figures 2 - 4).
Geo. mean fold difference: 0.96
Geo. mean abs. fold difference: 3.42
Median abs. fold difference: 2.42

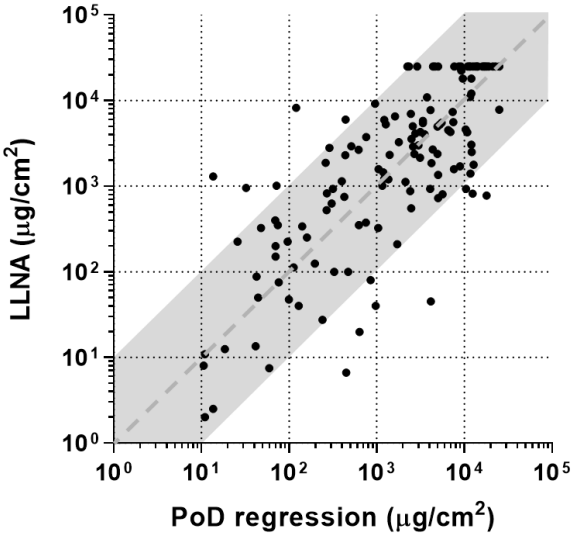

Figure 5. All chemicals with all three tests available and with a potency value (PV) according to [2, 5], compared versus OECD MLLP LLNA EC3 data, other LLNA data according [1] were used if OECD MLLP EC3 not available. Grey dashed line: line of identity. Grey area – area of less than 10-fold misprediction. Chemicals negative in the LLNA are plotted with the default value of 25'000, i.e. 100% LLNA test concentration.

Omitting the non-sensitizers (n=35), the following figures are obtained:

Geo. mean fold difference: 1.17
Geo. mean abs. fold difference: 3.84
Median abs. fold difference: 2.88

The same evaluation was then performed against the potency values. The prediction against PV is slightly less accurate with Geomean and Median absolute fold difference are 2.7 – 3.7, while the Geomean fold difference is slightly below 1, again indicating a balanced prediction for this subset too.

Geo. mean fold difference: 0.99
Geo. mean abs. fold difference: 3.67
Median abs. fold difference: 2.70

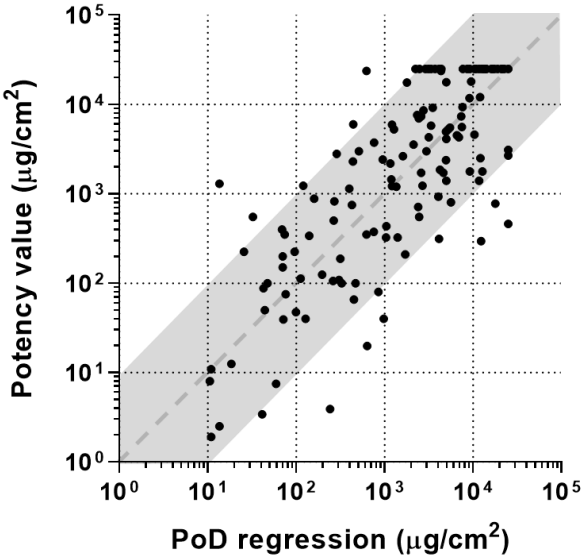

Figure 6. All chemicals with all three tests available and with a potency value (PV)* value according to [2, 5], compared versus the PV. Grey dashed line: line of identity. Grey area – area of less than 10-fold misprediction. Chemicals rated with a PV > 25.000 are plotted with a default value of 25'000.

Omitting the non-sensitizers (n=35), the following figures are obtained:

Geo. mean fold difference: 1.15
Geo. mean abs. fold difference: 4.3
Median abs. fold difference: 3.24

* Note: Additional PV next to the data from Irizar *et al.* were derived as described in Irizar [5], with the exception that next to the Human Repeat Insult Patch Test (HRIPT) data also human maximization tests (HMT) were taken into account, i.e. all positive human data from the OECD DASS database [6] were used.

e) All chemicals with Potency values as derived by Irizar et al. and with in vitro data (n = 29)

The chemicals for which Irizar *et al.* derived PV was separately assessed, here only evaluation vs. PV is shown. The prediction against PV gives a Geomean and Median absolute fold difference of 2.6 – 3.5, while the Geomean fold difference is close to 1.1, again indicating a balanced prediction for this subset too.

Geo. mean fold difference: 1.11
Geo. mean abs. fold difference: 3.51
Median abs. fold difference: 2.61

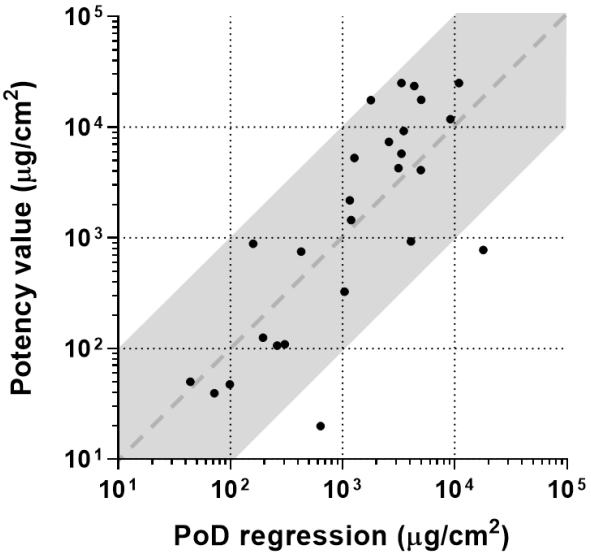

Figure 7. All chemicals with all three tests available and with a poteny value (PV) value according to[5] compared versus the PV. Grey dashed line: line of identity. Grey area – area of less than 10-fold misprediction

Omitting the non-sensitizers (n=3), the following figures are obtained:

Geo. mean fold difference: 1.30
Geo. mean abs. fold difference: 3.53
Median abs. fold difference: 2.71

Analysis of chemicals with > 10-fold underprediction

Of most concern for risk assessment are chemicals with a strong underprediction. To better understand the underlying reason of significant misprediction, Table 1 lists all these chemicals with a > 10-fold underprediction vs. LLNA EC3 and/or potency values from the full database.
While for some chemicals the discrepancy cannot be fully explained, there are a few key observations to note:

a) Some chemicals are outside of the predictive applicability domain of the kDPRA as listed in APPENDIX III, Annex 1 of TG442C. Thus, the potency of chemicals with strong reactivity towards amine groups can be underestimated due to the specific detection of cysteine reactivity by the kDPRA.
b) For some chemicals, the sensitization potential is not well characterised and is relying only on a single LLNA value. For some of these chemicals there is no clear structural alert which would explain the potency measured in the LLNA
c) For three chemicals there is evidence that the LLNA overestimates the potency or is false-positive.

Reasons for significant mispredictions for individual chemicals are discussed in more detail in ESM3 in Natsch and Gerberick, 2022, Part II accessible under https://www.altex.org/index.php/altex/article/view/2422/2386, and these are often associated with limitations of the applicability domain of individual methods or limitations in in vivo data.

Discussion

Accuracy of the potency assessment is best characterised by the absolute fold misprediction. Analysis of the different subsets of the full database gives overall a similar picture of the accuracy of the potency evaluations based on the PoD derived by the DA. This holds true for both evaluations vs. LLNA data and vs. PV data.

The geometric mean of the fold misprediction is close to 1 for all datasets. This indicates that the PoD does predict either the EC3 for the LLNA or the DSA04 for human data, the latter being the key measure of human potency used to derive PV values. Both these reference values are not a no-observed-effect level – but both the 3-fold stimulation in the LLNA and the DSA04 are concentrations where an effect becomes clearly manifest in the experimental setting (e.g. one positive reaction in a maximization test on 25 individuals), and this concentration has been characterized as an 'inflection point' or lower end of the dose-response. To derive a no-effect concentration an added factor may then be applied in safety assessments based on this PoD, which is beyond the scope of this document. It is still interesting to note, that in the analysis of the relationship between DSA1+ values vs. LLNA, the team developing SARA (See Figure 3 in Annex 11 to the Supporting document to TG497) noted a fold difference of 2.5 between LLNA and DSA01, indicating that the LLNA EC3 is systematically higher than the DSA1+, which is in line with the observation that the EC3 correlates better to the DSA04 than a DSA01. Thus, as SARA is trained on DSA1+ values and not EC3 / DSA04 values, a systematically lower PoD is expected based on the SARA model (factor 2 - 4). This can be considered when using the regression-based PoD vs. the SARA –based PoD in a risk assessment.

References

1. Natsch, A. and G.F. Gerberick, *Integrated skin sensitization assessment based on OECD methods (I): Deriving a point of departure for risk assessment.* ALTEX, 2022. **39**(4): p. 636-646.
2. Natsch, A., *Integrated skin sensitization assessment based on OECD methods (III): Adding human data to the assessment.* ALTEX, 2023. **40**(4): p. 571-583.
3. Natsch, A. and G.F. Gerberick, *Integrated skin sensitization assessment based on OECD methods (II): Hazard and potency by combining kinetic peptide reactivity and the "2 out of 3" Defined Approach.* ALTEX, 2022. **39**(4): p. 647-655.
4. OECD. *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation-Annex 3*. Series on Testing and Assessment No. 336

2021 9.12.2021]; Available from: https://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm.

5. Irizar, A., et al., *Reference Chemical Potency List (RCPL): A new tool for evaluating the accuracy of skin sensitisation potency measurements by New Approach Methodologies (NAMs).* Regul Toxicol Pharmacol, 2022. **134**: p. 105244.
6. OECD. *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation-Annex 4.* Series on Testing and Assessment No. 336 2021 9.12.2021]; Available from: https://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm.

Table 1. List and discussion of the chemicals which are > 10-fold underpredicted vs LLNAEC3 and/or PV

Name	CAS Number	EC3 DASS database	EC3	DPRA-Cysdep	DPRA-Lysdep	PoD_DA (ug/cm2)	Abs. Fold mispred. vs LLNA	Abs. Fold mispred. vs PV	In OECD MLLP set	In PV set	Discussion
Chemicals with >10-fold underprediction vs the LLNA											
Cases with potential applicability domain limitation issue											
Phthalic anhydride	85-44-9	0.160	0.16	2	75	976	24	24	x	x	Preferential amine reactivity, outside of applicability domain of kDPRA
Glutaraldehyde	111-30-8	0.080	0.08	30	85	636	32	32	x		Preferential amine reactivity, outside of applicability domain of kDPRA
Oxazolone	15646-46-5	0.002	0.002	76	50	87	173	173	x	x	Unique potency determined by high reaction rate with lysine and protein crosslinking; not mirrored by kDPRA (Natsch et al., 2010; Natsch and Emter, 2017)
Diethylenetriamine	111-40-0	POS	3.28	45	0	12388	15	42			Primary amine, outside of applicability domain of kDPRA
Misc. Cases											
Propyl gallate	121-79-9	POS	0.32	60	27	851	11	11		x	extrapolated EC3, with associated uncertainty of the in vivo value
Diphenylcyclopropenone	886-38-4	1A	0.003	99	0	55	73	73		x	extrapolated EC3, with associated uncertainty of the in vivo value
Tetrachlorsalicylanilide	1154-59-2	0.027	0.03	37	9	450	68	7	x		Sensitizer and photosensitizer
Cases with limited information on sensitization potential											
Benzoylacetone	93-91-4		0.04			613	61				Only one LLNA value; original data not available
Methyl pyruvate	600-22-6	2	2	0	12	15701	26		x		Only one LLNA value; no other information on in vivo sensitization, no clear struct. Alert
Allyl phenoxyacetate	7493-74-5	3	3	1	4	17938	23	23	x	x	Only one LLNA value; no other information on in vivo sensitization
Squaric acid	2892-51-5	POS	4.30	47	5	14550	14				Only LLNA data, no clear struct. Alert
Likely false-positives in the LLNA											
Sodium dodecyl sulfate	151-21-3	3.70	3.70	0	0	10391	11	2	x		Known false-positive in LLNA
Hexyl salicylate	6259-76-3	POS	0.18	4	1	4149	92	6			Conflicting LLNA evidence; LLNA strongly overestimates human sensitization potency
Tocopherol	59-02-9	7.40	7.40	0	7	30388	16		x		Known non-sensitizer in human from broad use as topical cosmetic ingredient, LLNA may overestimate human potency
Chemicals with >10-fold underprediction vs the PV only											
Neomycin sulphate	1405-10-3		100.00	interference	3	34995	1	13		x	aminoglycoside antibiotics not recognized neither by LLNA nor by in vitro tests, do they have a different mechanism of sensitization?
Butylglycidylether	2426-08-6	30.90	30.90	67	12	4118	2	13	x	x	Very volatile material, whcih may also partly evaporate from in vitro tests, PoD more conservative than LLNA
Penicillin G	61-33-6	31.30	31.30	14	0	31676	4	69	x	x	Penicillin is reactive with amine groups, negative in DPRA and in vitro assays for unknown reason
1,4-Phenylenediamine	106-50-3	0.11	0.11	93	24	241	9	62	x	x	Pre-hapten, oxidation rate to0 slow to be detected as strongly reactive in kDPRA with short incubation time
1-Chloro-2,4-	97-00-7	0.05	0.05	100	15	41	3	12	x	x	Human potency underestimated