Organisation for Economic Co-operation and Development

ENV/CBC/MONO(2025)2/ANN11

Unclassified

English - Or. English

11 September 2025

ENVIRONMENT DIRECTORATE
CHEMICALS AND BIOTECHNOLOGY COMMITTEE

Annex 11: Evaluation of the SARA-ICE ED01 as a point-of-departure (PoD) - Supporting Document of Test Guideline (TG) 497 on Defined Approaches for Skin Sensitisation

Series on Testing and Assessment
No. 336

(Second edition)

E-mail: ehscont@oecd.org.

JT03570727

Please cite this publication as:

OECD (2025), *Supporting document to Test Guideline 497 on Defined Approaches for Skin Sensitisation, Annex 11: Evaluation of the SARA-ICE ED01 as a point-of-departure (PoD)*, OECD Series on Testing and Assessment, No. 336, OECD Environment, Health and Safety, Paris, https://one.oecd.org/document/ENV/CBC/MONO(2025)2/ANN11/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 11: Evaluation of the SARA-ICE ED01 as a point-of-departure (PoD).

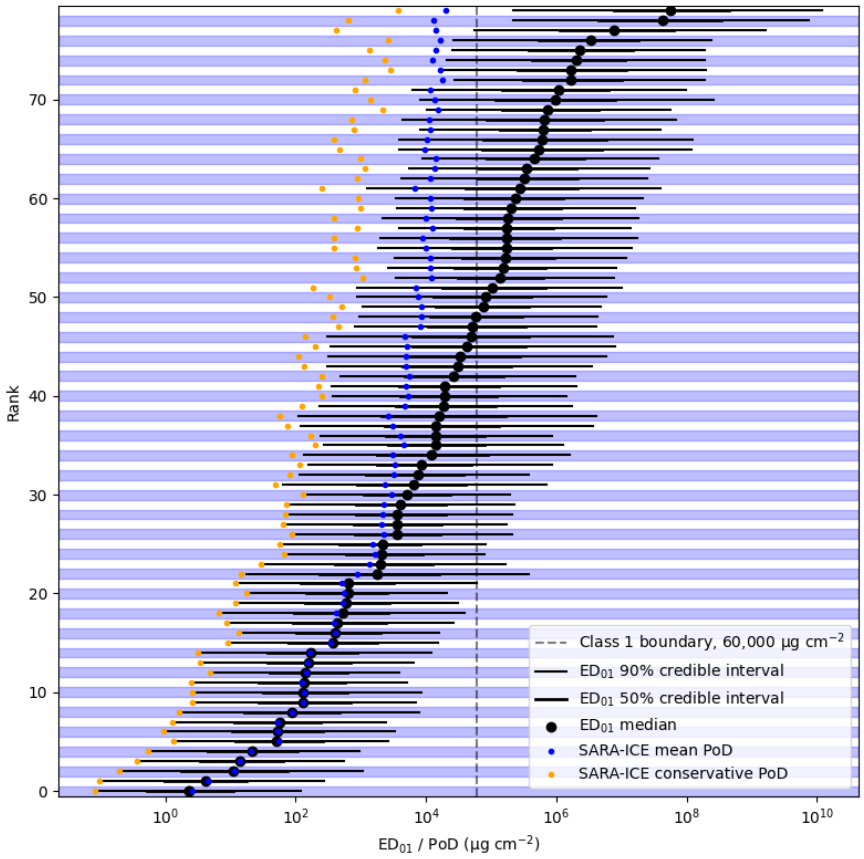
Introduction

The primary output of the SARA-ICE model is an estimate of dermal dose in a HPPT at which there is a 1% chance of sensitising a randomly selected individual from a HPPT-eligible test population (ED_{01}). This quantity is estimated as a probability distribution, the variance of which is a measure of the uncertainty in the estimate. The distribution of the ED_{01} may be converted to a point-of-departure PoD for use in exposure-based risk assessment. The calculation of SARA-ICE PoDs from the distribution of the ED_{01} is defined in TG497 Appendix 2.
In this document, SARA-ICE PoDs are calculated for 194 chemicals using only NAM inputs into the SARA-ICE model. Specifically, SARA-ICE inputs are limited to a single DPRA, single KeratinoSens and single h-CLAT sourced from [1] and, if available, a single kDPRA input sourced from [2]. These are compared against various sets of reference values including *weight-of-evidence* estimates of the DSA1+ published in [3], reference LLNA *median-like-location-parameters* published in [1] and a *reference chemical potency list* (RCLP) published in [4]. Two possible PoD types are considered throughout this document: a mean PoD and conservative PoD defined as the 5th percentile of the distribution of the ED_{01} conditional on the assumption that a chemical is a sensitiser.

Relationship between SARA-ICE ED_{01} estimates and SARA-ICE PoDs

The SARA-ICE PoD is calculated from the distribution of the ED_{01} subject to the additional assumption that the compound in question is sensitising. This constrains PoD estimates to be less than 60,000 µg cm^{-2} (threshold for GHS NC). This is illustrated in Figure 1. SARA-ICE estimates of the ED_{01} were computed using NAM inputs into the model and both SARA-ICE mean PoDs and conservative PoDs (5th percentile) were computed from the distribution of the ED_{01} conditional on GHS class 1. Compounds towards the top of Figure 1 are estimated to have a high probability of being non-sensitisers (observe that the bulk of the distribution of the ED_{01} exceeds 60,000 µg cm^{-2} for several compounds). SARA-ICE PoDs approach the maximum possible value of 60,000 µg cm^{-2} for these compounds. At the other end of the potency scale, when the majority of the probability mass of the distribution of the ED_{01} is less than 60,000 µg cm^{-2}, SARA-ICE PoDs are practically the same as if calculated from the unconditional distribution of the ED_{01}.

Figure 1. Comparison of ED_{01} estimates, represented as centred 50% and 90% credible intervals of the distribution against SARA-ICE PoDs computed from the conditional distribution of the ED_{01} assuming class 1.



Comparisons of SARA-ICE PoDs against reference values

This section presents pairwise comparisons between

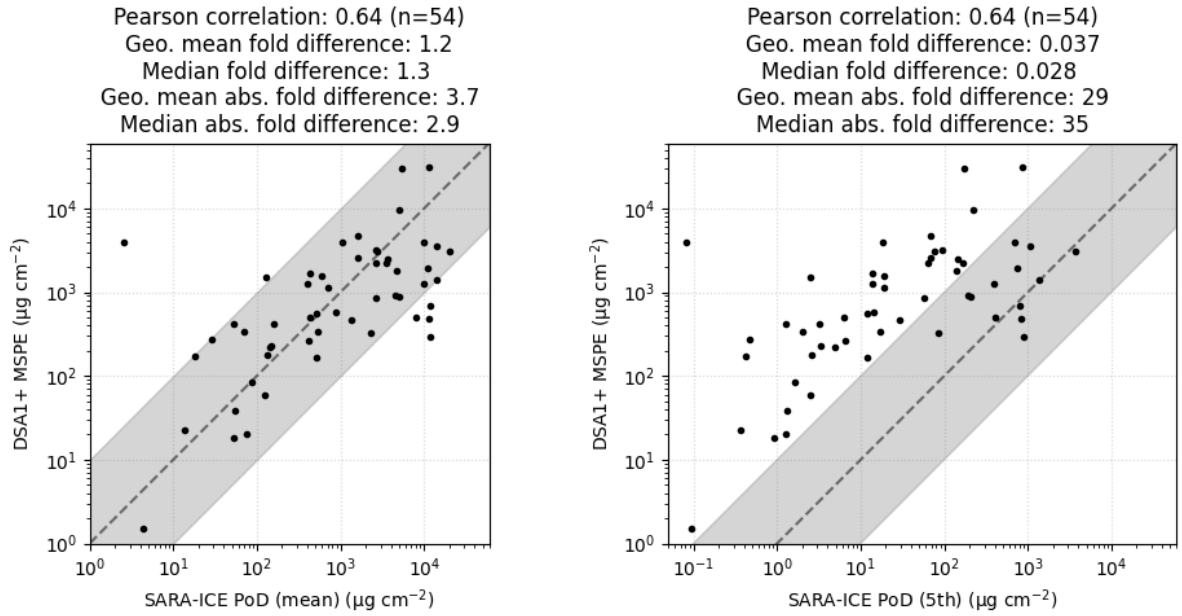
1) SARA-ICE PoDs obtained using NAM inputs versus human reference DSA1+ values,
2) LLNA MLLPs versus human reference DSA1+ values,
3) SARA-ICE PoDs obtained using NAM inputs versus LLNA MLLPs,
4) SARA-ICE PoDs obtained using NAM inputs versus a published RCPL,
5) SARA-ICE PoDs obtained using NAM inputs versus a published list of NESILs.

Comparisons of the SARA-ICE PoDs against DSA1+ values

In this section we compare PoD estimated generated from *in vitro* data only against reference DSA1+ values reported in [3]. Specifically, PoD estimates are computed for the 196 chemicals in the DASS evaluation set using a) NAM data also reported in the evaluation set [1] (single input for DPRA, KeratinoSens and h-CLAT) and b) a single kDPRA sourced from [2]. Reference DSA1+ values are extracted from the supplementary information of [3]. The reference value is termed the *median sensitization potency estimate* (MSPE) and is reported as weight-of-evidence potency metric calculated from human data. A value has been defined for 54 / 196 chemicals in

the DASS evaluation set.
In the left plot of Figure 2, mean SARA-ICE PoDs are compared against reference MSPEs. A positive correlation exists and 47 / 54 (87%) of SARA-ICE PoDs are within 10-fold of the reference value. Geometric mean and median fold differences are 1.2 and 1.3-fold respectively, implying SARA-ICE PoDs are slightly higher than the reference value, on average. The SARA-ICE PoD is lower than the reference value for 25 / 54 (47%) of the test substances. The mean absolute fold-difference is 3.7-fold and median is 2.9-fold, the former number being skewed by the large difference observed for tetramethylthiuram disulfide where the SARA-ICE PoD is approximately 1,500-fold lower than the reference value.
In the right panel of Figure 2 conservative PoDs are compared against the reference value. Conservative SARA-ICE PoDs in this case are lower than the reference value for 50 / 54 (93%) of compounds. This is roughly the same as the nominal level of conservativeness assumed when choosing the 5th percentile to define a conservative PoD. On average, conservative PoDs are 27 / 36 (geometric mean / median) fold lower than the reference value at this level of conservativeness.

Figure 2. Left: Comparison of SARA-ICE mean PoDs estimated from NAM inputs against reference DSA1+ MSPE values. Right: SARA-ICE PoDs computed as the 5th percentile of the distribution of the ED_{01} conditional on it being less than 60,000 µg cm^{-2}.

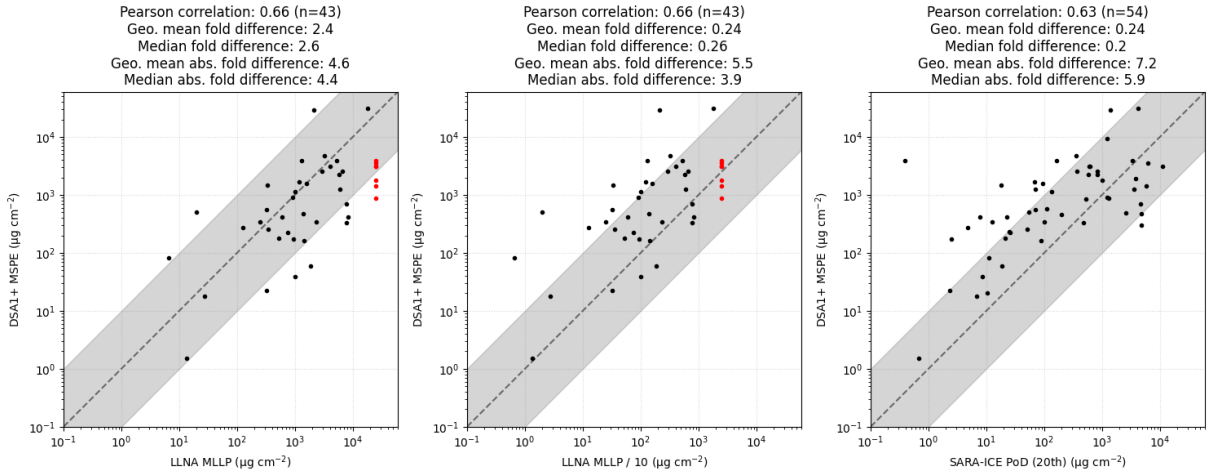


LLNA MLLP versus DSA1+

To put the above comparisons into context, LLNA EC3s are compared against the reference DSA1+ reference values (LLNA EC3s, summarised in terms a 'median-like-location-parameter' were sourced from OECD Annex II [1]. There are 43 compounds for which both a MLLP and a reference DSA1+ MSPE could be defined. For compounds classified as negative in the LLNA, an MLLP of 100% (25,000 µg cm^{-2}) was assumed. This is the case for 7 / 43 (16%) compounds. The LLNA MLLP has a stronger bias towards overestimation of the reference DSA1+. For this compound set, the LLNA MLLP overestimates the human reference value (suggests lower potency) for 31 / 43 (72%) of compounds. Correspondingly, the average fold-difference is higher.

The absolute fold-difference is also higher which implies a higher variance in the difference in estimates. Firm conclusions are tentative due to the limited number of compounds available for comparison, but the SARA-ICE mean PoDs calculated from NAM data is a marginally more accurate predictor of the reference value, on average, than the LLNA MLLP.
In the right plot of Figure 3, the MLLP is divided by a factor of 10 as a nominal safety factor for cross-species extrapolation. This shifts the MLLP enough that it more oft than not returns a value that is lower (more conservative) than the reference value. Nevertheless, there remain 9 / 43 (21%) compounds for which the reference value remains overestimated. The nominal level of conservativeness of an LLNA MLLP is approximately 80%. At this level of conservativeness, the LLNA MLLP is around 4-fold lower than the reference value, on average. If a SARA-ICE PoD is calculated at the 20th percentile (right plot, Figure 3), then the empirical proportion of SARA-ICE PoDs less than the reference value is 77% which roughly matches the empirical conservativeness of the MLLP / 10. Furthermore, the average fold-difference is 4-5 (mean-median) which is also similar to the MLLP. Thus, a SARA-ICE PoD calculated at the 20th using the NAM inputs here is roughly equivalent to an LLNA MLLP divided by a cross-species uncertainty factor of 10.

Figure 3. Left: Comparison of LLNA MLLP values against reference DSA1+ values. Red points indicate compounds that are negative in the LLNA for which a MLLP of 100% (25,000 µg cm^{-2}) was assumed. Middle: LLNA MLLP divided by a cross-species extrapolation factor of 10 compared against reference DSA1+ values. Right: SARA-ICE PoDs calculated at the 20th percentile compared against reference DSA1+ values.

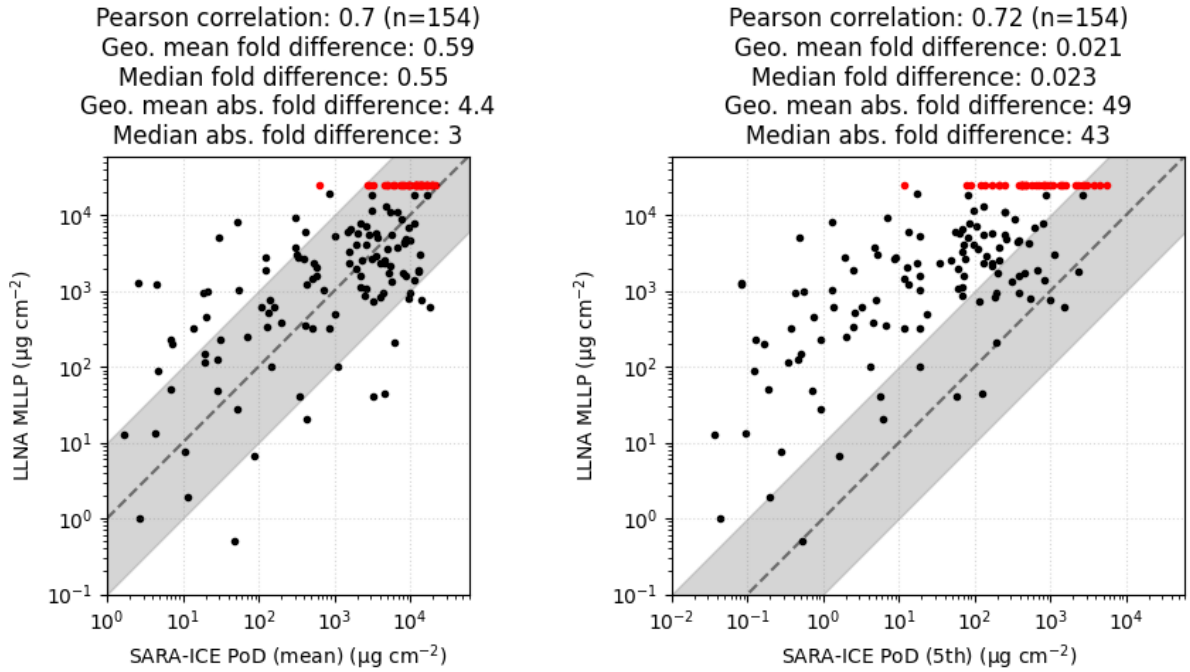


Comparison of SARA-ICE PoDs against LLNA MLLPs

NAM inputs for the SARA-ICE DA and LLNA MLLPs were both sourced from [1]. There are 154 compounds for which a comparison of the SARA-ICE PoD against the MLLP can be made. There are 38 / 154 (25%) compounds which are negative in the LLNA. To enable comparison for these chemicals, an LLNA MLLP of 100% (25,000 µg cm^{-2}) was assumed.
The average fold-difference for mean SARA-ICE PoDs is 2-fold lower than the reference LLNA MLLP. The average absolute fold-difference is 3-fold by median and 4.5-fold by mean. The latter statistic inflated by some very large differences. The largest of which is again the compound tetramethylthiuram disulfide for which the SARA-ICE PoD is ~500 times lower than the reference MLLP. These statistics are slightly larger than those observed with the human reference values and the bias towards being 2-fold lower than the reference MLLP is consistent with reference

MLLP being roughly 2-fold higher than the human reference value observed above.
Conservative SARA-ICE PoDs are compared against reference LLNA MLLP in the right plot of Figure 4. The conservative SARA-ICE PoD is less than the reference value for 148 / 154 (96%) of compounds. This again matches the nominal level of conservativeness assumed when using the 5th percentile to calculate the conservative SARA-ICE PoD. Average fold-differences are between 40 to 50-fold more conservative. If the LLNA MLLP is divided by a safety factor of 10 to account for cross-species extrapolation, the SARA-ICE conservative PoD is less than the reference value on average by about 4-fold.

Figure 4. Left: Comparison of SARA-ICE mean PoDs against reference LLNA MLLPs. Red points indicate compounds which are negative in the LLNA. Right: Comparison of SARA-ICE conservative PoDs (5th percentile) against LLNA MLLPs.

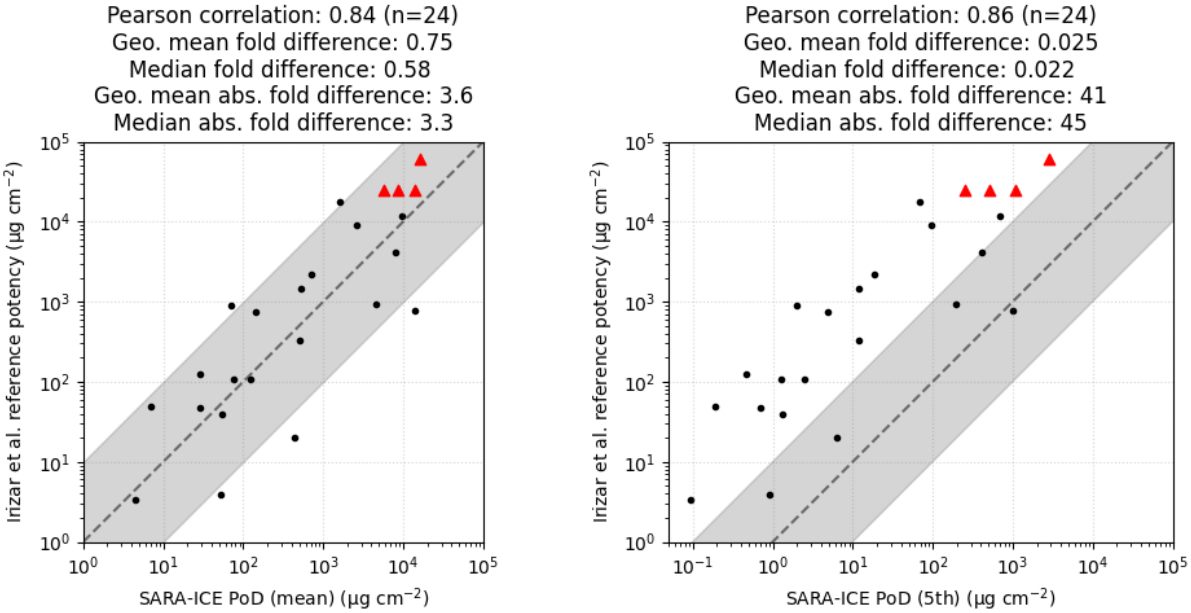


Comparison against Irizar et al. RCPL

Irizar et al. have published a *reference chemical potency list* (RCPL) whose objective is to serve as a benchmark set of skin sensitisation potency values for evaluating new approach methodologies. This dataset overlaps with the OECD evaluation set for 24 compounds. SARA-ICE PoDs generated from NAM data are compared against this list in Figure 5.
Overall bias and average error for the best-guess PoD is similar to that observed against reference DSA1+ values. Conservative PoDs are less than the reference value for 23/24 (96%) compounds. On average, the conservative PoDs are 40-fold below the reference value.

Figure 5. Left: Comparison of SARA-ICE mean PoDs estimated from NAM inputs against Irizar et al. reference potency values. Red triangles indicate reference potencies reported as a censored

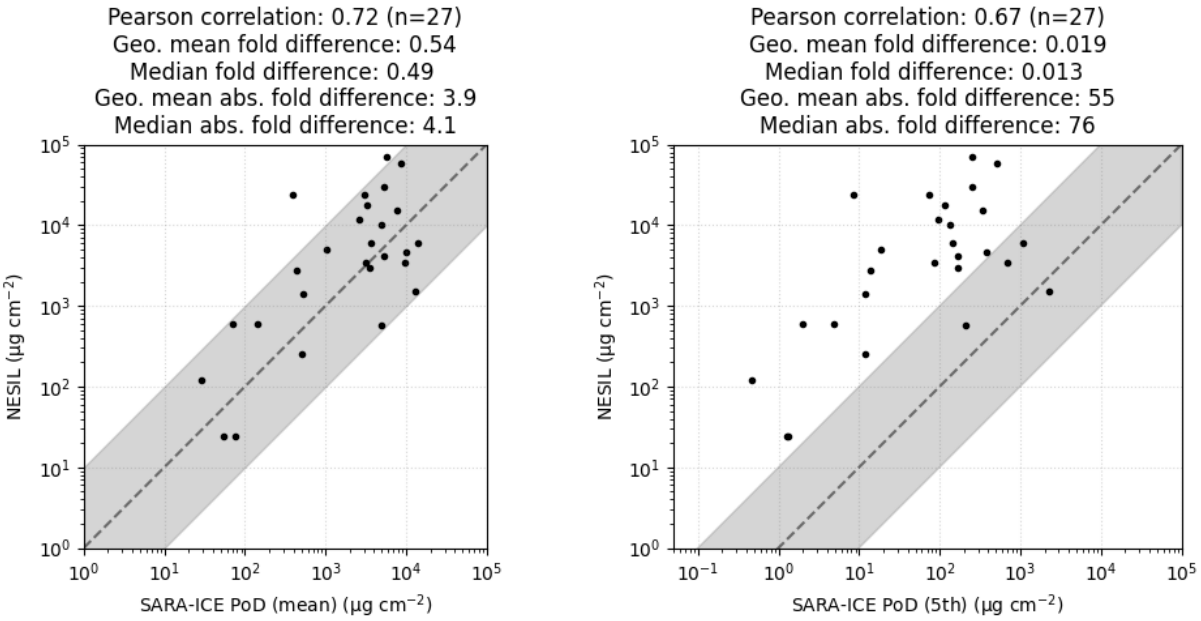
values, i.e., of the form >X for some number X. Right: comparison with SARA-ICE conservative PoDs.



Comparison of SARA-ICE PoDs against NESILs

The final set of comparisons compares SARA-ICE PoDs to a published list of *no-expected-sensitisation-induction-levels* (NESILs) for fragrance ingredients [5]. The overlap between this list and the OECD reference compound list is 27 compounds. SARA-ICE PoDs are plotted against NESILs in Figure 6. The mean SARA-ICE PoD is on average 2-fold lower than the NESIL, exhibiting a similar absolute bias as observed in all the previous comparisons. Average absolute fold-differences are around 4. The conservative PoD is lower than the NESIL for 26 / 27 (96%) of compounds – matching the nominal level of conservativeness assumed. For this set of comparisons, the conservative PoD is on average 53 (mean) / 77 (median) fold more conservative than the NESIL.

Figure 6. Left: Comparison of SARA-ICE mean PoDs against reference NESILs. Right: Comparison of SARA-ICE conservative PoDs against NESILs.



Prediction accuracy when using fewer inputs

The comparisons presented in the previous section use all available NAM data from sources [1] and [2]. This section explores prediction accuracy of SARA-ICE PoDs against DSA1+ and LLNA benchmarks when PoDs are computed using fewer inputs. The following subsets of inputs are considered:

1. (n=1) Either a single DPRA, kDPRA, KeratinoSens or h-CLAT input.
2. (n=2) Any two inputs from distinct key events in the skin sensitisation AOP.

Including the combination of all available data, subsetting the data in this manner results in up to 10 SARA-ICE PoDs for each benchmark chemical. Note that since kDPRA is not available for all chemicals, some chemicals may not feature in specific subdata combinations. Prediction accuracy against benchmark values is summarised in terms of mean and median fold-difference (which serves as a measure of the bias in the prediction) and mean and median absolute-fold difference (which serves as a measure of size of the unsigned error in the prediction). Accuracy statistics against DSA1+ MSPE and LLNA MLLP benchmarks are presented in Table 1. .

Moving from all available inputs to two inputs results in a small increase in the bias towards underestimation of the reference value for most input combinations. The magnitude of the bias is notably larger for some of the single input comparisons. The maximum bias is approximately 5-fold under the reference LLNA MLLP, obtained using a single kDPRA input when the bias is represented as the median.

The average absolute fold-difference does not appear to be correlated with the number of inputs when comparing against human reference values. However, it tends to be larger when comparing against LLNA reference values when fewer input values are used.

Surprisingly, the input combination of a single DPRA has a smaller average absolute fold-differences than when using all available inputs when comparing against DSA1+ values (see the top row in Figure 7). However, relative to the LLNA benchmarks, average absolute fold-differences are notably larger.

The “best” combination of inputs against these benchmarks, as determined by the average of the absolute fold-difference statistics is a single DPRA and a single h-CLAT. The data combinations are ranked by this measure in Figure 8. Interestingly, it can be observed that h-CLAT inputs are present in the top 5 five positions in this ranking and are absent from the bottom 5 rankings. This tentatively suggests that this input type should be favoured if one is in the position of choosing with data to generate for input into the model.

Table 1. Summary of distribution of prediction errors against DSA1+ and LLNA benchmarks for SARA-ICE PoDs computed using different combinations of inputs.

	DSA1+ MSPE					LLNA MLLP				
Input data combination	Num. comparisons	Geo. Mean fold difference	Median fold difference	Geo. Mean abs. fold difference	Median abs. fold difference	Num. comparisons	Geo. Mean fold difference	Median fold difference	Geo. Mean abs. fold difference	Median abs. fold difference
1xDPRA, 1xkDPRA (if available) 1xKeratinoSens, 1xh-CLAT	54	1.1	1.1	4	3.1	154	0.58	0.59	4.5	3.1
1xDPRA, 1xKeratinoSens	54	0.94	0.9	3.6	3.5	154	0.43	0.47	5.4	3.9
1xDPRA, 1xh-CLAT	53	1.2	0.99	3.6	2.9	153	0.55	0.54	4.4	3
1xKeratinoSens, 1xh-CLAT	53	0.99	0.84	4.2	3.3	153	0.45	0.38	4.7	3.7
1xkDPRA, 1xKeratinoSens	49	0.89	0.81	4.4	3.3	148	0.45	0.42	5.1	3.5
1xkDPRA, 1xh-CLAT	48	1.1	1	4.5	3.3	147	0.53	0.51	4.4	3
1xDPRA	54	0.79	1.3	3.4	2.4	154	0.35	0.26	6.2	4.2
1xkDPRA	49	0.82	0.99	4.8	3.6	148	0.35	0.21	6.2	5.6
1xKeratinoSens	54	0.78	0.71	4	3.4	154	0.32	0.26	6.5	5.2
1xh-CLAT	53	1	1.1	4.5	3.2	153	0.43	0.36	4.8	3.6

Figure 7. Top row: Comparison of PoDs generated using a single DPRA input against benchmark DSA1+ MSPE references values (left) and benchmark LLNA MLLP (right). The cluster of vertically arranged points on the left of each plot are due to the DPRA input having a maximum depletion greater than 99% for all chemicals in the cluster, resulting in a PoD of approximately 4 µg cm^{-2}. Bottom row: comparisons against the same benchmarks when using a single DPRA and single h-CLAT study to inform predictions.

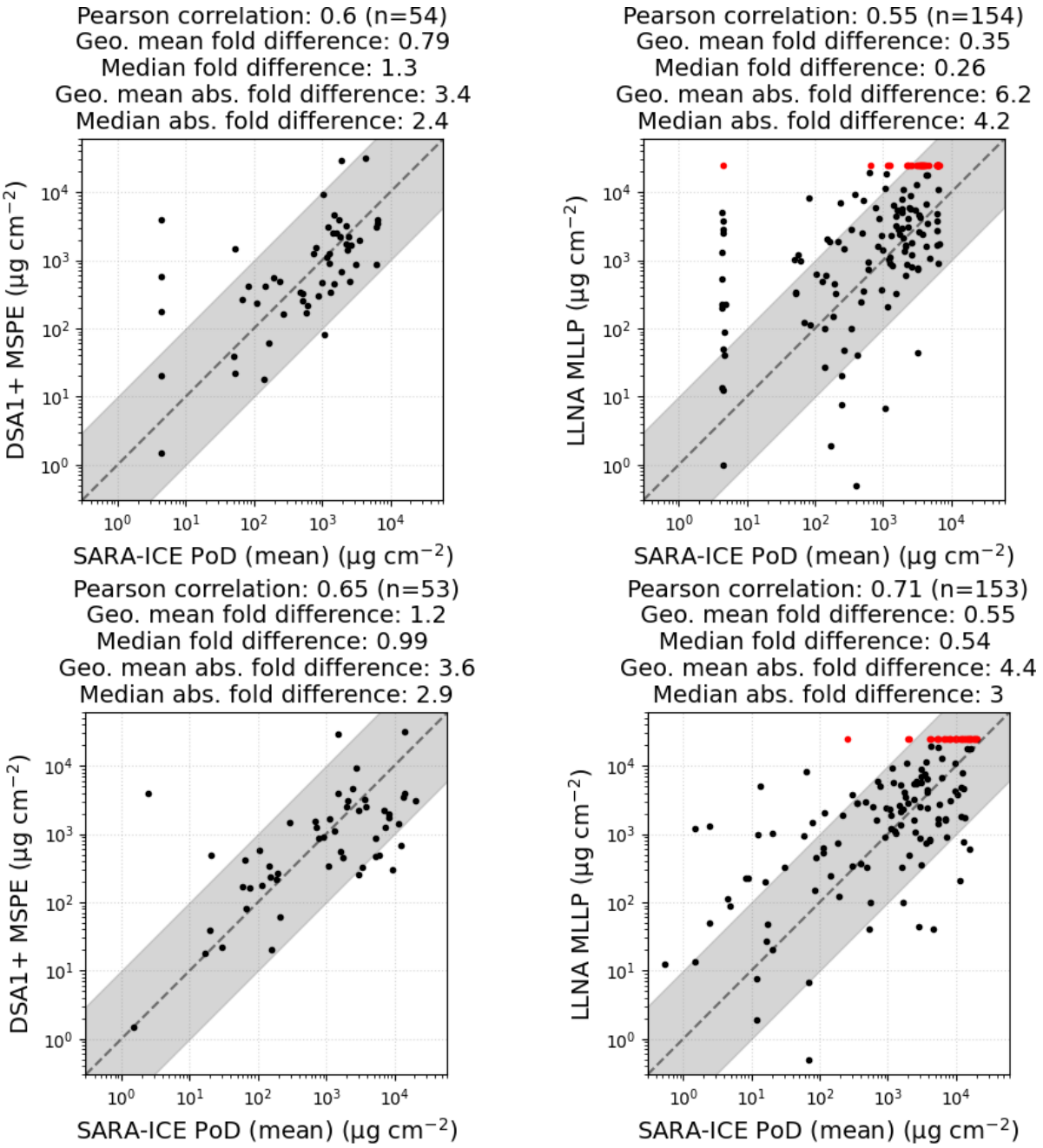
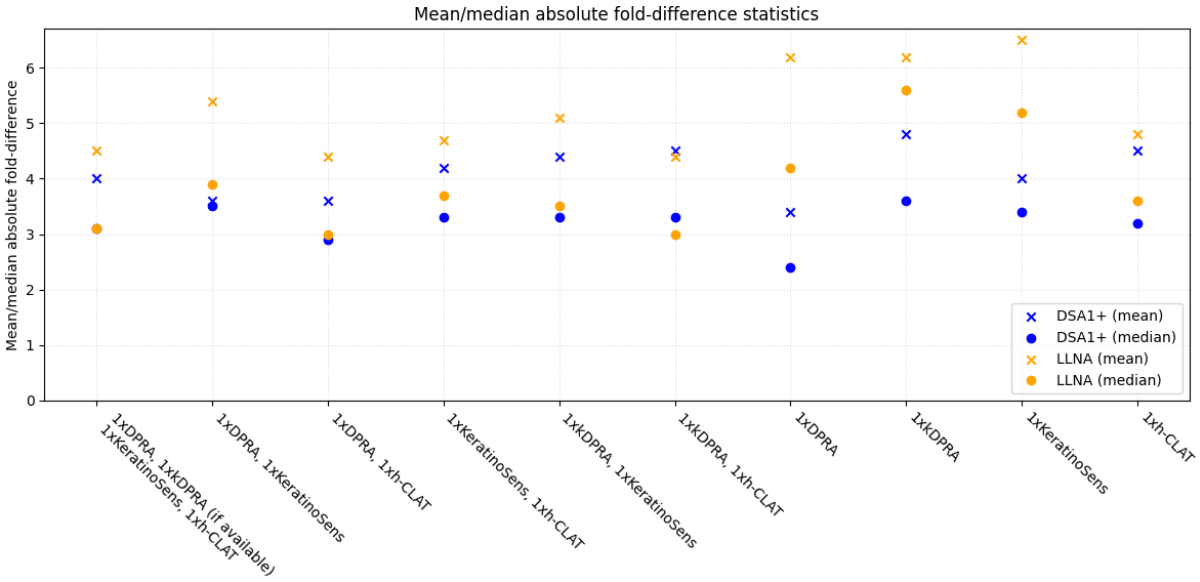


Figure 8. Measures of the average absolute fold-difference between SARA-ICE PoDs and benchmark DSA1+ MPSEs and LLNA MLLPs.

Data combinations along the bottom are ranked in terms of the average performance statistic.



Summary

This document sources various sets of reference values to compare against the SARA-ICE PoD when estimated using a limited set of NAM inputs. These include a weight-of-evidence DSA1+ published in [3], LLNA MLLP defined for the OECD DA evaluation set [1], a reference chemical potency list published in [4] and a reference list of NESILs for fragrances published in [5]. For all sets of comparisons, absolute bias between the SARA-ICE mean PoD and the reference value is barely more than 2-fold. Average absolute fold-differences range from 3-4.5-fold. Predictive performance of the mean PoD against reference DSA1+ is comparable to the predictive performance of the LLNA MLLP.

Conservative PoDs which are nominally 95% conservative, i.e., defined at the 5th percentile, are empirically observed to be roughly 95% conservative (empirical percentages ranges from 93%-96% across the different comparisons). The bias inherent to a conservative PoD ranges from 27-77-fold under the reference value, depending on which reference value is considered or which measure of average is used (mean or median). Lower bias may be obtained by using a nominally less conservative PoD. The LLNA MLLP divided by a cross-species extrapolation factor of 10 was observed to be ~80% conservative relative to reference DSA1+ values with an approximate average bias of 4-fold. If a SARA-ICE PoD is computed at the 20th percentile to match the empirical conservativeness of the LLNA, then this PoD also exhibits a similar level of bias relative to the reference DSA1+ value.

The analysis presented here also explores the impact of moving from a combination of NAM inputs including a single DPRA, single KeratinoSens, single h-CLAT and a single kDPRA (if available), to a smaller set of inputs. There is no appreciable drop in average prediction of SARA-ICE PoDs when using two inputs. However, average prediction error against LLNA benchmarks was found to increase for single inputs aligning with key events 1 and 2.

References

1. OECD *Supporting Document to the OECD Guideline 497 on Defined Approaches for Skin Sensitisation*. 2021.
2. Natsch, A. and G.F. Gerberick, *Integrated skin sensitization assessment based on OECD methods (I): Deriving a point of departure for risk assessment.* ALTEX-Alternatives to animal experimentation, 2022. **39**(4): p. 636-646.
3. Herzler, M., et al., *Use of human predictive patch test (HPPT) data for the classification of skin sensitization hazard and potency.* Archives of Toxicology, 2024. **98**(5): p. 1253-1269.
4. 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.
5. Api, A.M., et al., *Dermal sensitization quantitative risk assessment (QRA) for fragrance ingredients.* Regul Toxicol Pharmacol, 2008. **52**(1): p. 3-23.