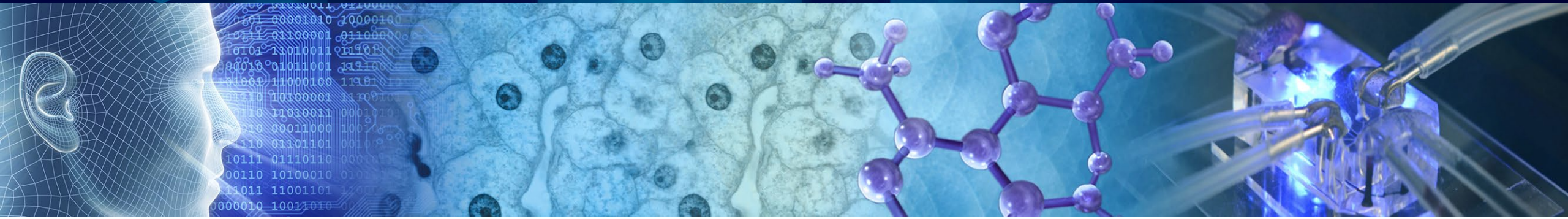
National Institute of Environmental Health Sciences
Division of Translational Toxicology



The SARA-ICE Model for Probabilistic Skin Sensitization Risk Assessment

Nicole Kleinstreuer, PhD

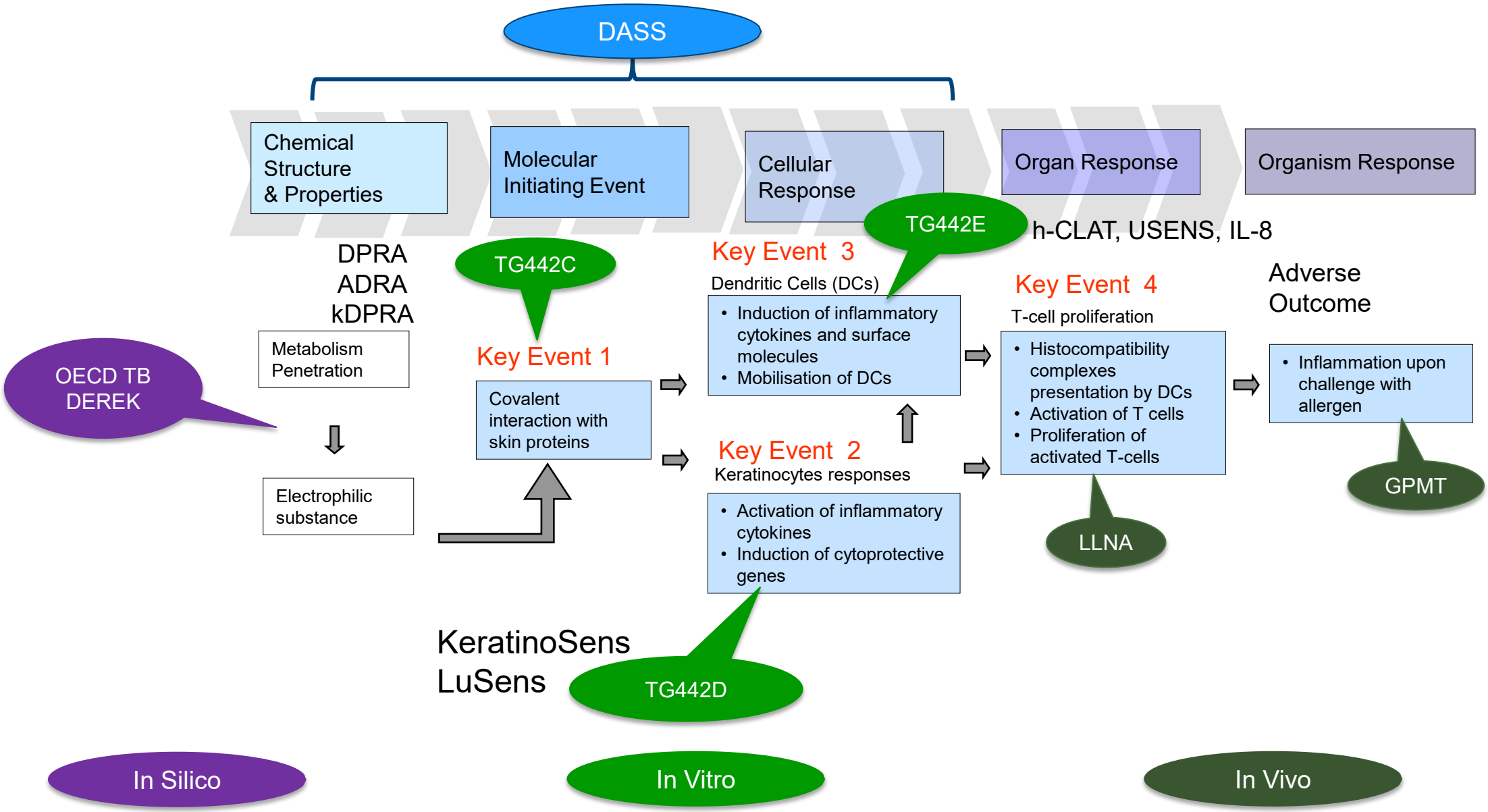
National Toxicology Program Interagency Center for the Evaluation of Alternative Test Methods (NICEATM)
NIEHS Division of Translational Toxicology

EPA/NICEATM ProbRA Workshop

7-8 October 2024

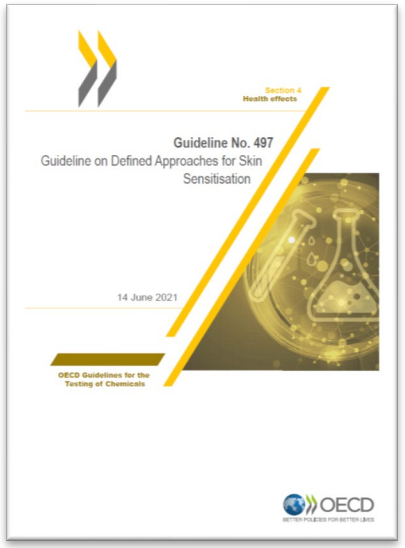
National Institute of Environmental Health Sciences
Division of Translational Toxicology

Skin Sensitization: Biology-Mapped Methods

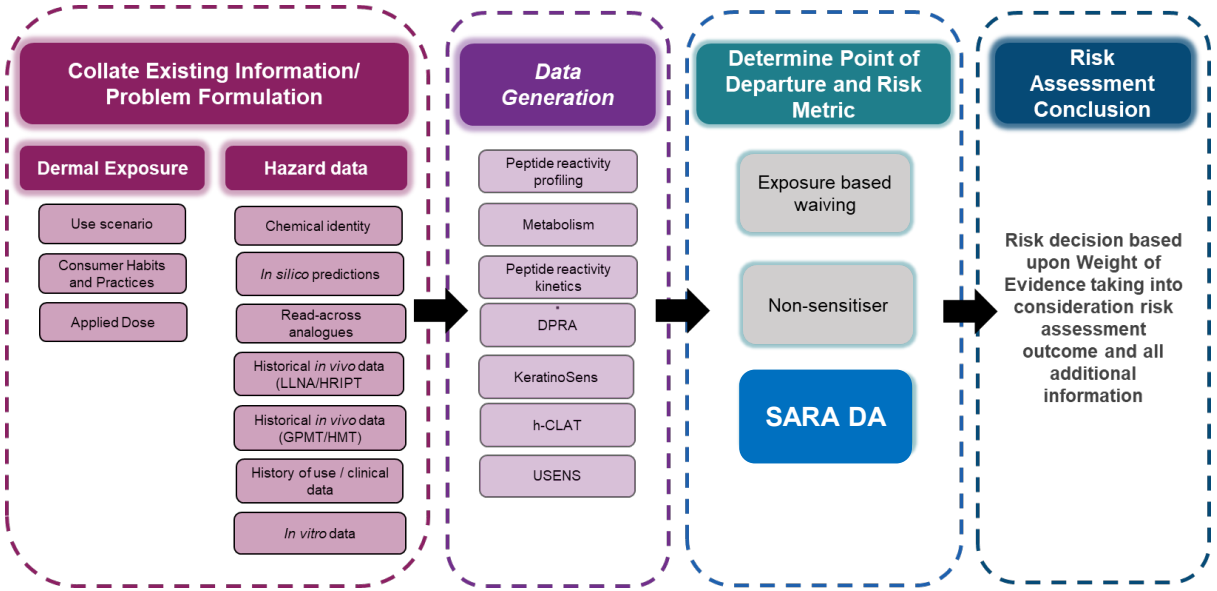


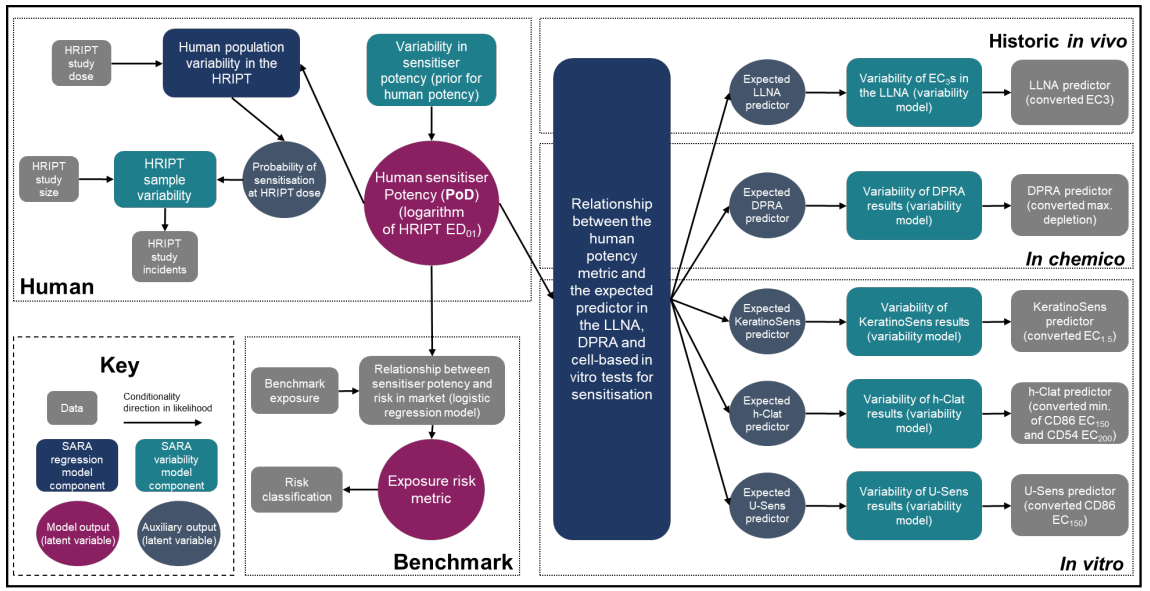
OECD Defined Approaches for Skin Sensitization Guideline Project

- Extensive curation efforts undertaken to build LLNA (168 substances) and human (66 substances) reference databases
- Applicability domain and DA confidence were defined
- The resulting Guideline 497 was adopted in 2021
- It meets regulatory requirements of:
  - DAs that discriminate between sensitizers and non-sensitizers
  - DAs that discriminate strong from weak/moderate sensitizers (i.e., GHS potency categories)
- Ongoing: DAs that address regulatory needs of quantitative risk assessment
  - US and UK leading a project under OECD for evaluating a defined approach that can provide a point of departure for quantitative risk assessment



Skin Allergy Risk Assessment Defined Approach (SARA DA) was developed for application as part of a tiered, WoE NGRA framework





- **Unilever NGRA framework for Skin Allergy** was designed to use a WoE based upon all available information, accommodate range of consumer product exposure scenarios and provide a quantitative point of departure and risk metric → **SARA DA**

The use-case of the SARA DA is to estimate:

1. ED_{01}, the dose at which there is a 1% chance of sensitization in an HPPT-eligible population
2. Probability that a consumer exposure to some chemical is 'low risk', conditional on the available data and the model

Reynolds et al 2022 https://pubmed.ncbi.nlm.nih.gov/35835397/

Development history of the SARA-ICE model

2017-2019

A prototype Bayesian statistical model was developed at Unilever to estimate a no-effect-dose from HPPT data. This model was published in 2019.

2019-2021

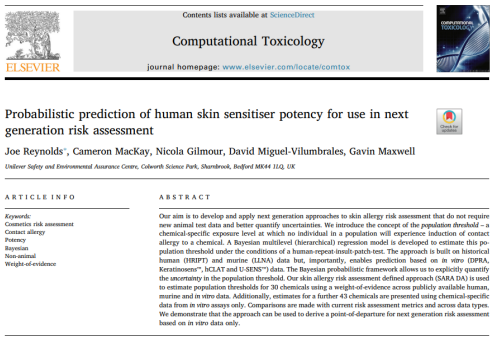
The model and underlying database are revised and expanded. Unilever performs an internal review to endorse for use in risk assessment.

2021-2022

The revised model is published within a set of three papers which the model and explore its use in case study risk assessment scenarios.

2021 - present

Unilever begins working with NICEATM to adapt the model for regulatory use. The SARA database is merged with the ICE database and the SARA-ICE model is developed.


Contents lists available at ScienceDirect

Computational Toxicology

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

Probabilistic prediction of human skin sensitiser potency for use in next generation risk assessment

Joe Reynolds*, Cameron MacKay, Nicola Gilmour, David Miguel-Vilumbrales, Gavin Maxwell

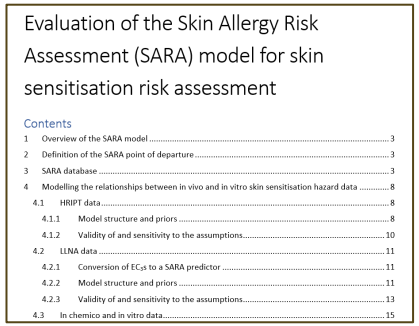
Unilever Safety and Environmental Assurance Centre, Colworth Science Park, Sharnbrook, Bedford MK44 1LQ, UK

ARTICLE INFO

Keywords:
Cosmetics risk assessment
Contact allergy
Potency
Bayesian
Non-animal
Weight-of-evidence

ABSTRACT

Our aim is to develop and apply next generation approaches to skin allergy risk assessment that do not require new animal test data and better quantify uncertainties. We introduce the concept of the *population threshold* – a chemical-specific exposure level at which no individual in a population will experience induction of contact allergy to a chemical. A Bayesian multilevel (hierarchical) regression model is developed to estimate this population threshold under the conditions of a human-repeat-insult-patch-test. The approach is built on historical human (HRIPT) and murine (LLNA) data but, importantly, enables prediction based on *in vitro* (DPRA, Keratinosens™, hCLAT and U-SENS™) data. The Bayesian probabilistic framework allows us to explicitly quantify the *uncertainty* in the population threshold. Our skin allergy risk assessment defined approach (SARA DA) is used to estimate population thresholds for 30 chemicals using a weight-of-evidence across publicly available human, murine and *in vitro* data. Additionally, estimates for a further 43 chemicals are presented using chemical-specific data from *in vitro* assays only. Comparisons are made with current risk assessment metrics and across data types. We demonstrate that the approach can be used to derive a point-of-departure for next generation risk assessment based on *in vitro* data only.


Evaluation of the Skin Allergy Risk Assessment (SARA) model for skin sensitisation risk assessment

Contents


Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

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

Decision making in next generation risk assessment for skin allergy: Using historical clinical experience to benchmark risk

J. Reynolds*, N. Gilmour†, M.T. Baltazar, G. Reynolds, S. Windebank, G. Maxwell

ARTICLE INFO

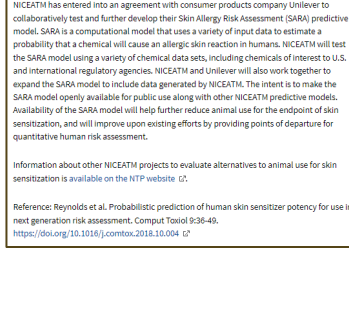
Handling Editor: Dr. Martin Van den berg

Keywords:
Skin sensitisation
Allergic contact dermatitis
Next generation risk assessment
Non-animal alternatives
New approach methodologies
Consumer exposure
SARA model
Defined approach
Risk management
Uncertainty analysis

ABSTRACT

Our aim is to develop and apply next generation approaches to skin allergy risk assessment that do not require new animal test data and better quantify uncertainties. Quantitative risk assessment for skin sensitisation uses safety assessment factors to extrapolate from the point of departure to an acceptable human exposure level. It is currently unclear whether these safety assessment factors are appropriate when using non-animal test data to derive a point-of departure.

Our skin allergy risk assessment model Defined Approach uses Bayesian statistics to infer a human-relevant metric of sensitiser potency with explicit quantification of uncertainty, using any combination of human repeat insult patch test, local lymph node assay, direct peptide reactivity assay, Keratinosens™, h-CLAT or U-SENS™ data. Here we describe the incorporation of benchmark exposures pertaining to use of consumer products with clinical data supporting a high/low risk categorisation for skin sensitisation. Margins-of-exposure (potency estimate to consumer exposure level ratio) are regressed against the benchmark risk classifications, enabling derivation of a risk metric defined as the probability that an exposure is low risk. This approach circumvents the use of safety assessment factors and provides a simple and transparent mechanism whereby clinical experience can directly feed-back into risk assessment decisions.


NICEATM to collaborate with Unilever on development of predictive model for skin sensitization

NICEATM has entered into an agreement with consumer products company Unilever to collaboratively test and further develop their Skin Allergy Risk Assessment (SARA) predictive model. SARA is a computational model that uses a variety of input data to estimate a probability that a chemical will cause an allergic skin reaction in humans. NICEATM will test the SARA model using a variety of chemical data sets, including chemicals of interest to U.S. and international regulatory agencies. NICEATM and Unilever will also work together to expand the SARA model to include data generated by NICEATM. The intent is to make the SARA model openly available for public use along with other NICEATM predictive models. Availability of the SARA model will help further reduce animal use for the endpoint of skin sensitization, and will improve upon existing efforts by providing points of departure for quantitative human risk assessment.

Information about other NICEATM projects to evaluate alternatives to animal use for skin sensitization is available on the NTP website.

Reference: Reynolds et al. Probabilistic prediction of human skin sensitizer potency for use in next generation risk assessment. Comput Toxiol 9:36-49. https://doi.org/10.1016/j.comtox.2018.10.004

Modification of SARA to create SARA-ICE DA for Regulatory Application

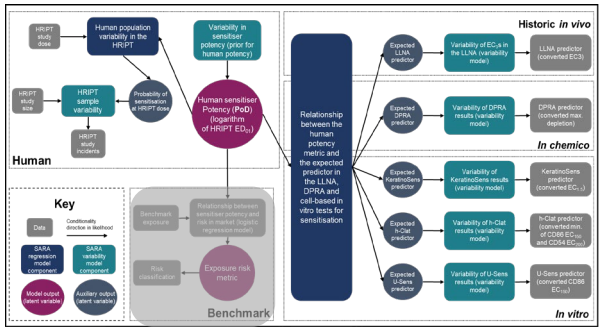
Database
Aim to expand the core dataset underpinning the model using data in the ICE database (relaxing the constraint that chemicals be limited to cosmetic ingredients).

Risk benchmarking
De-emphasize the risk benchmarking component of the model – previous set of benchmarks limited to use of consumer goods. Use the model for human PoD estimation for quantitative risk assessment.

GHS classification
Add functionality to predict GHS potency classification (estimated as a class probability to communicate uncertainty in classification).

Integrated Chemical Environment

ICE: Integrated Chemical Environment (nih.gov)



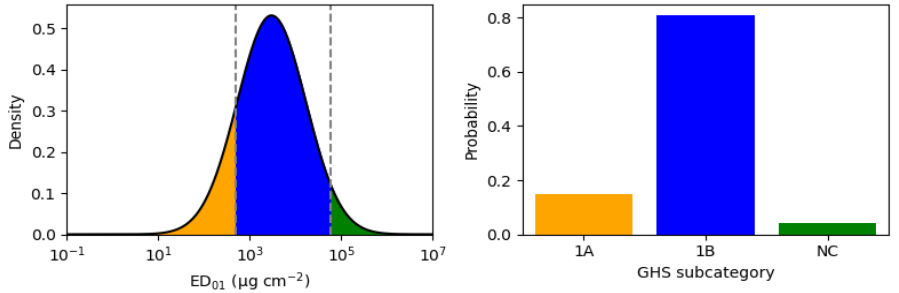
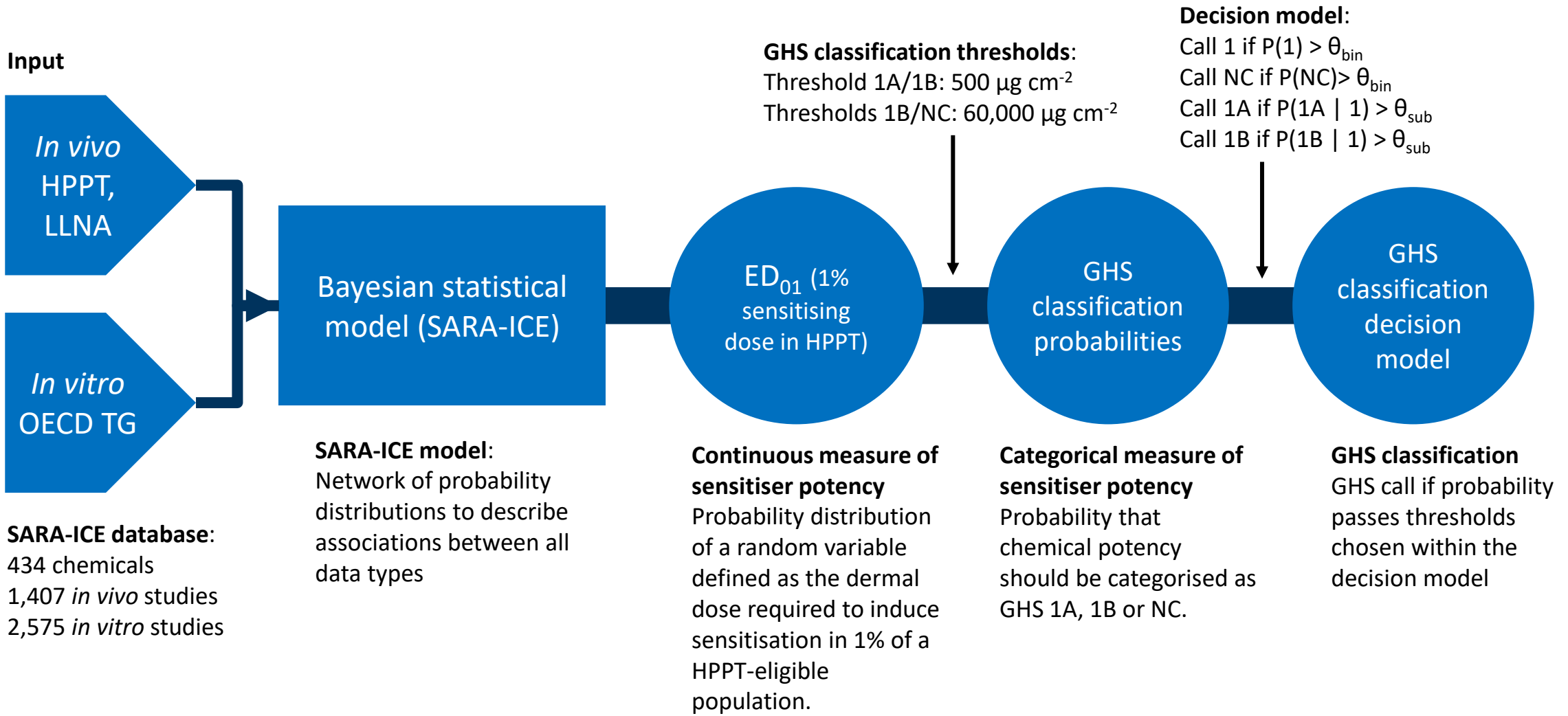


Figure (a) Example estimate of ED_{01} distribution with overlay of GHS subcategories 1A, 1B and NC defined thresholds, (b) probability of each GHS subcategory from ED_{01} distribution

National Institute of Environmental Health Sciences
Division of Translational Toxicology

SARA-ICE DA: Skin Allergy Risk Assessment - Integrated Chemical Environment Defined Approach



SARA-ICE database:
434 chemicals
1,407 *in vivo* studies
2,575 *in vitro* studies

SARA-ICE model:
Network of probability distributions to describe associations between all data types

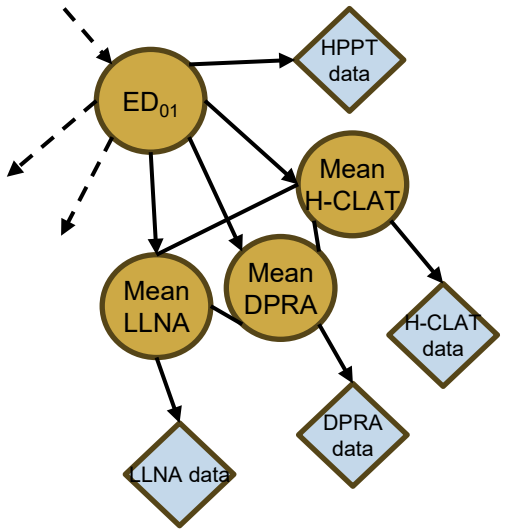
Continuous measure of sensitiser potency
Probability distribution of a random variable defined as the dermal dose required to induce sensitisation in 1% of a HPPT-eligible population.

Categorical measure of sensitiser potency
Probability that chemical potency should be categorised as GHS 1A, 1B or NC.

GHS classification
GHS call if probability passes thresholds chosen within the decision model

The SARA-ICE model

The SARA-ICE model is a high dimensional probability distribution built from a set of assumptions around conditional probability relationships.

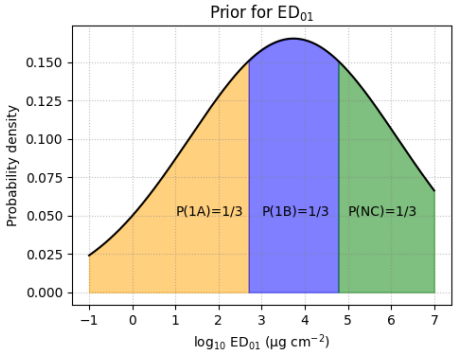


Parameters of the model are "learnt" using Bayesian updating.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Bayes theorem is applied to calculate the conditional probability distribution of each parameter given the available data.

The primary variable of interest includes the ED_{01}, defined as the HPPT dermal dose at which there is a 1% sensitisation rate.



The ED_{01} is converted to GHS classification probabilities for classification and labelling.

Model assumptions

HPPT

1. There is a dermal dose at which there is a 1% chance of inducing sensitisation in a randomly selected individual from a HPPT-eligible population.
2. The probability of inducing sensitisation in a HPPT increases with dose.
3. Each individual within a HPPT-eligible population has a personal threshold for sensitisation to any given chemical. This threshold may be greater than the maximum possible dose.
4. The distribution of the base-10 logarithm of personal thresholds has a Gaussian shape. The standard deviation is chemical-specific; different chemicals have different variabilities within the human population with respect to sensitivity to induction of sensitisation.
5. The number of individuals sensitised in a HPPT study follows a logit-normal-binomial compound distribution.

Model assumptions

Non-HPPT data

1. Data from the LLNA, DPRA, kDPRA, KeratinoSens, h-CLAT and U-Sens assays can be transformed such that it is reasonable to model variability in chemical-specific data in terms of a normal distribution (transformations mostly involve logarithms).

2. The same transformations put data on a scale in which it is reasonable to assume linear relationships between the average transformed datapoint on the base-10 logarithm of the ED_{01}.

3. The relationships between the average results can be described by a multivariate Gaussian distribution.

4. Variability in each test is chemical-specific. There is a latent variable for each test and each chemical which defines the variance of the chemical in the particular test.

5. Chemical-specific variance parameters can be estimated using partial pooling. The population of variances for each tested can be learnt and used to regularise chemical-specific estimates when limited data is available.

The SARA-ICE database

Study type	HPPT	LLNA	DPRA	kDPRA	KeratinoSens	h-CLAT	U-Sens
Inputs into SARA-ICE	Dermal dose, number tested, number sensitised	EC_3 or maximum concentration tested if no response observed	% depletion of cysteine and lysine peptides	Log Kmax	$EC_{1.5}$ or maximum concentration tested IC50 or maximum concentration tested	CD86 EC_{150}, CD50 EC_{200} or maximum concentration tested CV_{75} or maximum concentration tested	CD86 EC_{150} or maximum concentration tested CV_{75} or maximum concentration tested
Number of studies in database	871	536	650	361	972	428	164
Number of unique CASRN with this study type	276	195	251	185	258	211	90

434 distinct CASRN

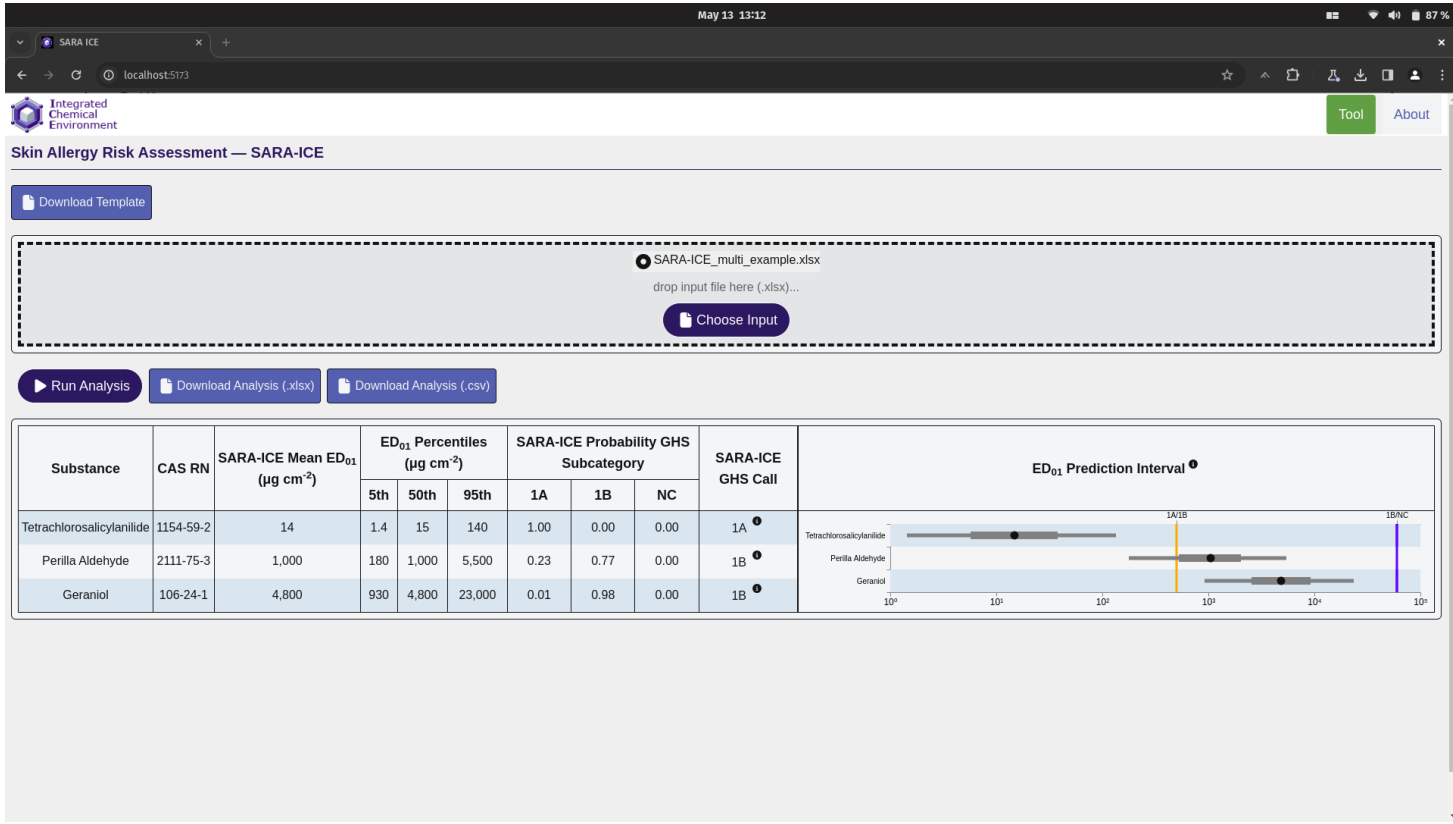
Computation

The SARA-ICE model is a mathematical model; it's assumptions and equations are expressible with pen and paper.

Learning model parameters requires numerical computation: the model is realised numerically using the programming language Stan. Python is used to process model inputs and outputs.

Computation requires many CPU cycles; however, a production version of the model has been developed to alleviate this limitation.

A standalone, downloadable version of the model has been created by NICEATM.

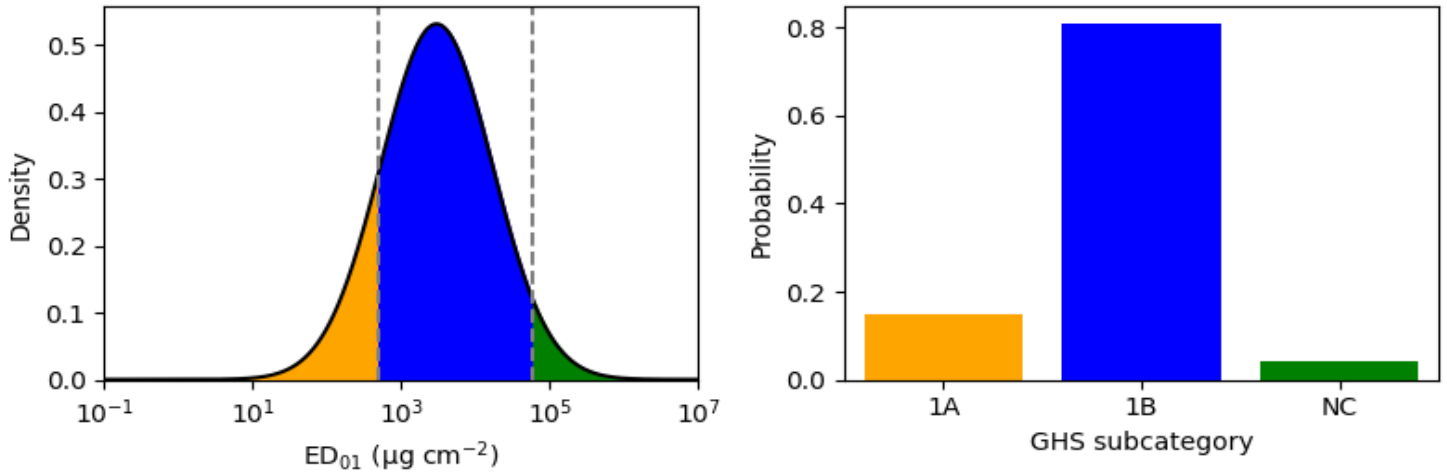


GHS classification

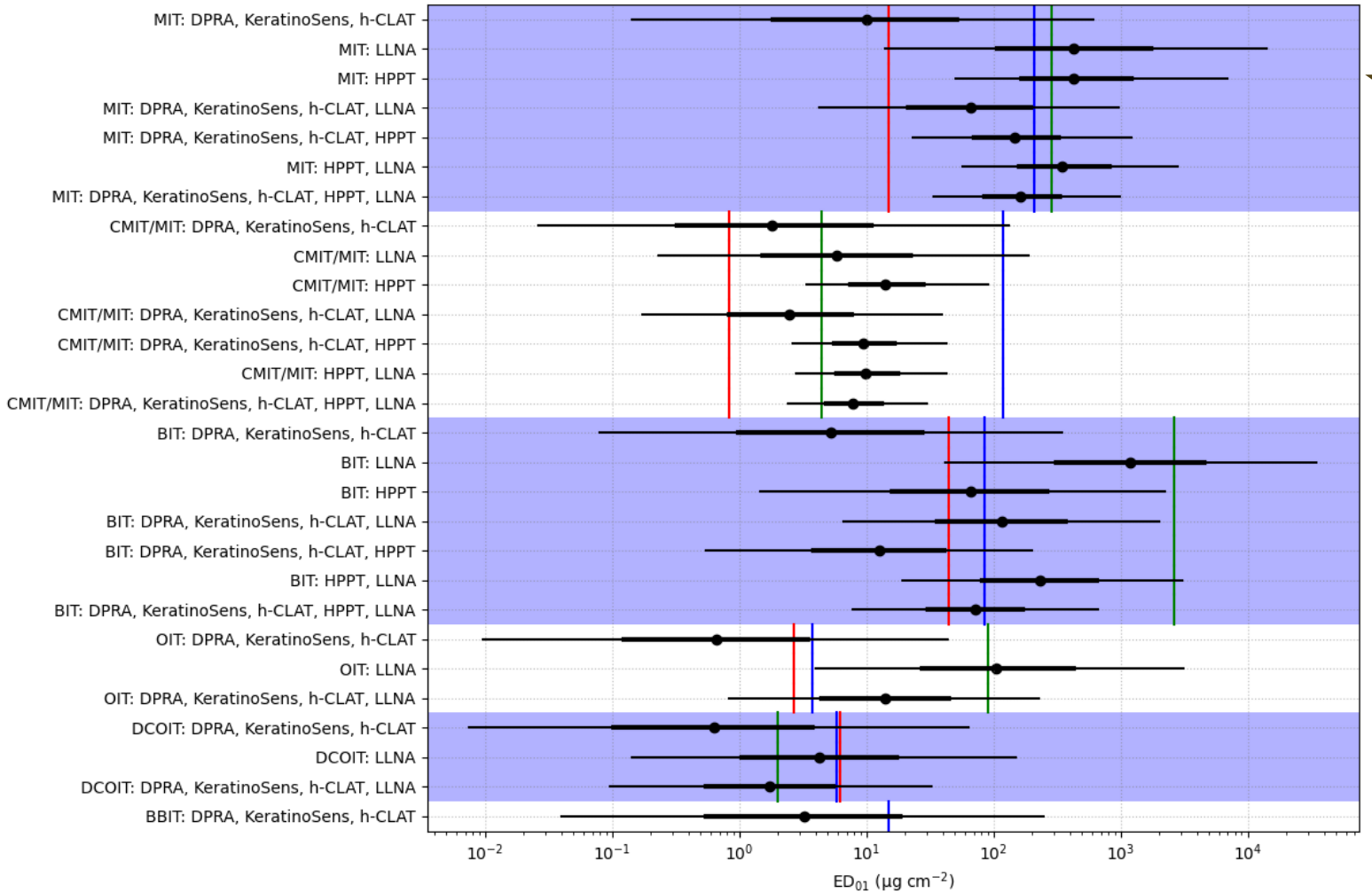
The distribution of the ED_{01} is used to defined GHS classification probabilities:

1. A threshold of 60,000 cm^{-2} (maximum possible HPPT dose under standard volume and patch size) is used to define the boundary between binary categories 1 and NC.
2. A threshold of 500 µg cm^{-2} is used to define the boundary between subcategories 1A and 1B.

The area under the curve between thresholds is the probability mass attributable to that interval. This defines the probability for the GHS classification.



Example SARA-ICE Application – Isothiazolinones



SARA-ICE – ED_{01} PoD estimates

ED01 estimates represented as centered 90% credible intervals (thin line), 50% credible intervals (thick line) and median (bullet). Red lines indicate the reference NESIL, blue lines are plotted at the EPA POD and green lines are plotted at the reference LLNA EC3.
NESILs (ECHA; Burnett et al., 2021; Novick et al., 2013; Ladics et al., 2020); EPA POD (EPA DOCKET (https://www.regulations.gov/document/EPA-HQ-OPP-2017-0720-0011); LLNA EC3 (Strickland et al., 2023)

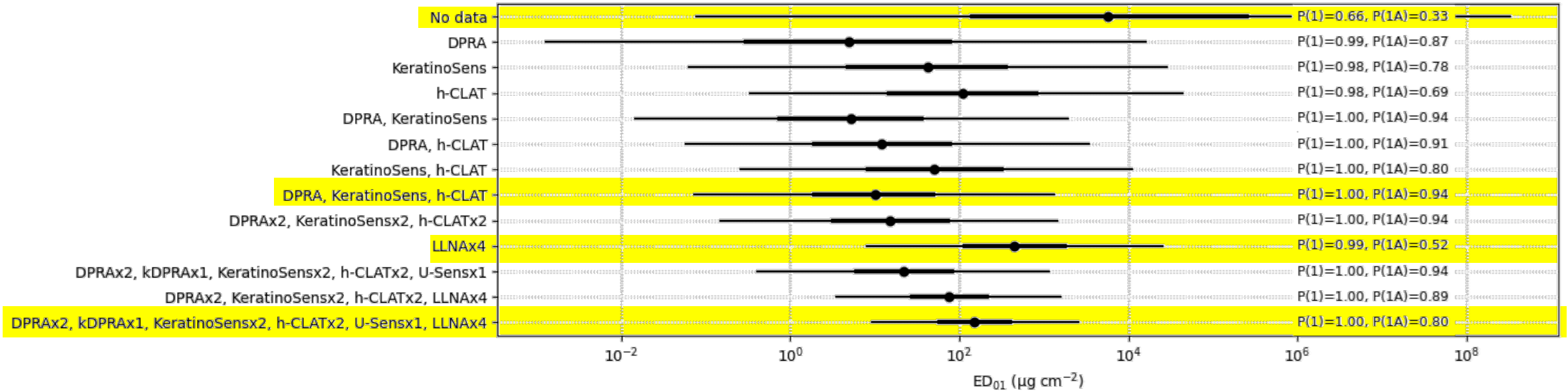
Reinke et al., 2024, under rev

SARA-ICE - MIT (2-Methyl-4-isothiazolin-3-one) – input data

Chemical	DPRA	kDPRA	KeratinoSens™	h-Clat	U-Sens™	Local Lymph Node Assay (LLNA)
MIT	Cysteine depletion: 97.9% Lysine depletion: 0% *Source*: Natsch et al., 2013	Log Kmax: -0.25 $M^{-1}s^{-1}$ *Source*: Natsch & Gerberick, 2022	$EC_{1.5}$: 11.78 µM IC_{50}: 139 µM *After unit conversion* $EC_{1.5}$: 1.4 µg ml^{-1} IC_{50}: 16 µg ml^{-1} *Source*: Natsch et al., 2013 & Urbisch et al., 2015 (Imax)	CD54 EC_{200}: 7.89 µg ml^{-1} CD86 EC_{150}: 9.23 µg ml^{-1} CV75: 24.7 µg ml^{-1} *Source*: Urbisch et al. 2015	CD86 EC_{150}: 9 µg ml^{-1} CV75: 44.3 µg ml^{-1} *Source*: Piroird et al., 2015	
	Cysteine depletion: 100% Lysine depletion: 0% *Source*: Kleinstreuer et al., 2018		$EC_{1.5}$: 9.54 µM IC_{50}: 108.25 µM After unit conversion $EC_{1.5}$: 1.1 µg ml^{-1} IC_{50}: 12 µg ml^{-1} *Source*: Kleinstreuer et al., 2018	CD54 EC_{200}: 11.6 µg ml^{-1} CD86 EC_{150}: 11.8 µg ml^{-1} CV75: 24.6 µg ml^{-1} *Source*: Kleinstreuer et al., 2018		EC_3: 2.2% EC_3: 0.4% EC_3: 0.863% EC_3: >4.5% *Source:* Kleinstreuer et al., 2018

Reinke et al., 2024, under rev

SARA-ICE - MIT example – ED_{01} PoD estimates



Summaries of ED_{01} estimates for MIT conditional on different combinations of input data. Distributions are represented as centred 95% credible intervals (thin lines), centred 50% credible intervals (thick lines) and median (bullet). Predictions are ordered, from largest (top) to smallest (bottom), with respect to the uncertainty in the estimate.

ED_{01} estimates for MIT for different SARA-ICE data inputs

Input Data	ED_{01} ($\mu g\ cm^{-2}$)	ED_{01} 2.5 percentile ($\mu g\ cm^{-2}$)	ED_{01} 25 percentile ($\mu g\ cm^{-2}$)	ED_{01} 50 percentile ($\mu g\ cm^{-2}$)	ED_{01} 75 percentile ($\mu g\ cm^{-2}$)	ED_{01} 97.5 percentile ($\mu g\ cm^{-2}$)	Prob(1A)	Prob(1B)	Prob(NC)
No data	**5,600**	0.077	140	5700	>100,000	>100,000	**0.33**	**0.33**	**0.34**
DPRA	4.7	0.0013	0.29	4.9	78	16,000	0.87	0.12	0.011
KeratinoSens	42	0.063	4.8	42	360	28,000	0.78	0.2	0.015
h-CLAT	110	0.33	15	110	820	44,000	0.69	0.29	0.02
DPRA, KeratinoSens	5.1	0.014	0.73	5.2	36	1,900	0.94	0.061	0.0008
DPRA, h-CLAT	12	0.057	1.9	12	77	3,400	0.91	0.087	0.0021
KeratinoSens, h-CLAT	52	0.26	8.3	51	320	11,000	0.8	0.19	0.0049
DPRA, KeratinoSens™ h-CLAT	**9.8**	0.072	1.9	9.9	49	1,300	**0.94**	**0.058**	**0.0004**
DPRAx2, KeratinoSensx2, h-CLATx2	15	0.15	3.2	15	73	1,500	0.94	0.064	0.0003
LLNA x4	**440**	8.1	110	440	1,800	26,000	**0.52**	**0.47**	**0.011**
DPRAx2, kDPRAx1,KeratinoSensx2, h-CLATx2,U-Sensx1	22	0.41	6	22	81	1,200	0.94	0.058	0.0001
DPRAx2, KeratinoSensx2, h-CLATx2, LLNAx4	76	3.5	28	75	210	1,600	0.89	0.11	0
DPRAx2, kDPRA KeratinoSens™x2, h-CLATx2, U-Sens™ LLNAx4	**150**	9.4	59	150	400	2,600	**0.8**	**0.2**	**0**

SARA-ICE – MIT example – Probability that an exposure is less than the ED_{01}

Input combination	Exposure (µg cm^{-2})											
	0.01	0.03	0.1	0.3	1	3	10	30	100	300	1000	3000
DPRA	0.93	0.89	0.82	0.75	0.65	0.55	0.43	0.32	0.23	0.16	0.096	0.058
KeratinoSens	0.99	0.99	0.97	0.94	0.88	0.79	0.67	0.54	0.39	0.27	0.16	0.092
h-CLAT	1	1	0.99	0.98	0.94	0.89	0.79	0.67	0.51	0.37	0.23	0.14
DPRA, KeratinoSens	0.98	0.96	0.91	0.83	0.71	0.57	0.41	0.27	0.15	0.084	0.038	0.018
DPRA, h-CLAT	0.99	0.99	0.96	0.92	0.82	0.7	0.53	0.37	0.22	0.12	0.057	0.027
KeratinoSens, h-CLAT	1	1	0.99	0.97	0.93	0.86	0.73	0.58	0.4	0.26	0.14	0.067
DPRA, KeratinoSens, h-CLAT	1	0.99	0.97	0.92	0.82	0.69	0.5	0.33	0.17	0.082	0.032	0.012
DPRAx2, KeratinoSensx2, h-CLATx2	1	1	0.98	0.95	0.88	0.76	0.58	0.39	0.21	0.096	0.035	0.012
LLNAx4	1	1	1	1	1	0.99	0.97	0.91	0.77	0.57	0.34	0.17
DPRAx2, kDPRAx1, KeratinoSensx2, h-CLATx2, U-Sensx1	1	1	1	0.98	0.94	0.84	0.66	0.43	0.22	0.091	0.029	0.0091
DPRAx2, KeratinoSensx2, h-CLATx2, LLNAx4	1	1	1	1	1	0.98	0.91	0.73	0.43	0.19	0.047	0.0095
DPRAx2, kDPRAx1, KeratinoSensx2, h-CLATx2, U-Sensx1, LLNAx4	1	1	1	1	1	1	0.97	0.88	0.61	0.32	0.095	0.019

*Comparison of ED01 estimates (based on different combinations of inputs) and probability that exposures are the less than the ED01. Thresholds of 0.2 (**orange - ≥ 80% likelihood that exposure is greater than ED_{01}**) and 0.8 (**blue - ≥80% likelihood that exposure is less than ED_{01}**).*

Conclusions

- SARA-ICE DA is being adapted for regulatory use through expanded data and functionality, and would be the first probabilistic defined approach included in an OECD TG.
- SARA-ICE DA shows good concordance with sensitizer binary and GHS sub-category classifications against OECD DASS benchmark data (82% – 95% BA)
- Case studies demonstrate benefits of SARA-ICE DA:
  - estimates human potency (ED_{01}) with uncertainty
  - estimates with in vitro and in vivo data inputs
  - estimates with incomplete and repeat datasets
- Evaluation of the SARA-ICE DA, including thresholds for conclusive predictions and performance impact, is ongoing within the OECD DASS expert group
- SARA-ICE is packaged for download for local implementation and **is available** for beta testing upon request via the NICEATM website (https://ntp.niehs.nih.gov/whatwestudy/niceatm)

National Institute of Environmental Health Sciences
Division of Translational Toxicology

Acknowledgments

The NICEATM Group

Unilever

Subscribe to NICEATM News email list

Integrated Chemical Environment

Nicole.Kleinstreuer@nih.gov

SARA-ICE NAM vs OECD DASS benchmarks

Binary classifications

Human, Θ_{bin} = 0.80	**SARA 1**	**SARA NC**	**Inconclusive**	**Total**
OECD 1	37	4	14	55
OECD NC	0	4	7	11
Total	37	8	21	66
Sensitivity: 90% Specificity: 100% **Balanced accuracy: 95%**				
LLNA, Θ_{bin} = 0.80	**SARA 1**	**SARA NC**	**Inconclusive**	**Total**
OECD 1	87	6	42	135
OECD NC	2	19	12	33
Total	89	25	54	168
Sensitivity: 94% Specificity: 90% **Balanced accuracy: 92%**				

The SARA-ICE decision model has been evaluated against OECD benchmark classifications.

Estimates of the ED01 use NAM data only (1xDPRA, 1xKeratinoSens, 1xh-CLAT, 1xkDPRA)

Sensitivity, specificity and acccuracy is computed for **conclusive** classifications only.

SARA-ICE NAM vs OECD DASS benchmarks

Subcategory classifications

Human, Θ_{bin} = 0.80, Θ_{sub}=0.55	**SARA 1A**	**SARA 1B**	**SARA NC**	**Inconclusive**	**Total**
OECD 1A	14	2	0	5	21
OECD 1B	4	9	4	14	31
OECD NC	0	0	4	7	11
Total	18	11	8	26	63
Sensitivity 1A: 88%, Specificity 1A: 81%, Balanced accuracy 1A: 84% Sensitivity 1B: 53%, Specificity 1B: 90%, Balanced accuracy 1B: 71% Sensitivity NC: 100%, Specificity NC: 88%, Balanced accuracy NC: 94% **Average balanced accuracy: 83%**					

LLNA, Θ_{bin} = 0.80, Θ_{sub}=0.55	**SARA 1A**	**SARA 1B**	**SARA NC**	**Inconclusive**	**Total**
OECD 1A	28	4	0	6	38
OECD 1B	16	22	5	42	85
OECD NC	0	1	19	13	33
Total	44	27	24	61	156
Sensitivity 1A: 88%, Specificity 1A: 75%, Balanced accuracy 1A: 81% Sensitivity 1B: 51%, Specificity 1B: 90%, Balanced accuracy 1B: 71% Sensitivity NC: 95%, Specificity NC: 93%, Balanced accuracy NC: 94% **Average balanced accuracy: 82%**					

The SARA-ICE decision model has been evaluated against OECD benchmark classifications.

Estimates of the ED01 use NAM data only (1xDPRA, 1xKeratinoSens, 1xh-CLAT, 1xkDPRA)

Sensitivity, specificity and acccuracy is computed for **conclusive** classifications only.