Organisation for Economic Co-operation and Development

ENV/CBC/MONO(2025)2/ANN10

Unclassified

English - Or. English

11 September 2025

ENVIRONMENT DIRECTORATE
CHEMICALS AND BIOTECHNOLOGY COMMITTEE

Annex 10: SARA-ICE Technical Specification - Supporting Document of Test Guideline 497 on Defined Approaches for Skin Sensitisation

Series on Testing and Assessment
No. 336

(Second edition)

E-mail: ehscont@oecd.org.

JT03570725

Please cite this publication as:

OECD (2025), *Supporting document to Test Guideline 497 on Defined Approaches for Skin Sensitisation, Annex 10: SARA-ICE Technical Specification,* OECD Series on Testing and Assessment, No. 336, OECD Environment, Health and Safety, Paris, https://one.oecd.org/document/ENV/CBC/MONO(2025)2/ANN10/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 10: SARA-ICE Technical Specification

The SARA-ICE database

Model Input Types

The SARA-ICE Defined Approach (DA) utilises data from the following sources:
Human patch test data; HPPT (Human Repeat Insult Patch Test (HRIPT) [1] and human maximization test (HMT [2, 3]). Only studies in which the cohort size, number sensitised and applied dose are reported are admissible as data into the approach.
Local Lymph Node Assay (LLNA, OECD TG 429, TG 442A and TG 442B). EC_3 (effective concentration (EC) inducing a stimulation index of 3) values obtained from individual studies are admissible data. Representative or averaged values are not admissible.
Direct Peptide Reactivity Assay (DPRA, OECD TG 442C). Maximum percentage depletion with either cysteine and/or lysine peptides (%).
Kinetic DPRA (OECD TG 442c). log Kmax value (typical units of $M^{-1}s^{-1}$ must be converted to $gL^{-1}s^{-1}$).
KeratinoSens™ Assay (OECD TG 442D). Reported $EC_{1.5}$ value (typical units of µM must be converted to µg ml^{-1}).
Human cell line activation test (h-CLAT Assay, OECD TG 442E). Reported CD54 EC200, CD86 EC150 and CV75 (units µg ml^{-1}).
U-Sens™ Assay (OECD TG 442E). Reported CD86 EC150 (units µg ml^{-1}).
The SARA-ICE DA may be run with subsets of these information sources, e.g. with *in vitro* assay data only.

Data curation

The curated SARA-ICE database consists of 434 distinct chemicals, identified using CAS number. The number of study results per study type tallies at 871 HPPT (365 HRIPT, 506 HMT), 536 LLNA, 650 DPRA, 361 kDPRA, 972 KeratinoSens™, 431 h-CLAT and 164 U-Sens™. These are divided heterogeneously between each CAS number, with the minimum number of studies per chemical equal to two. The database is freely available for download on the NICEATM website (https://ntp.niehs.nih.gov/whatwestudy/niceatm/test-method-evaluations/skin-sens/da/SARA-ICE) and will not be modified without OECD review.
HPPT data were curated in accordance to the methodologies described in the OECD Guideline 497 project (see Annex IV in [4]). Briefly, the documentation of each test was reviewed to determine protocol (HMT or HRIPT), protocol subtype, identification of the test substance, applied dose, number of subjects tested, and number of subjects responding. Because they were

assumed to be significantly different from the standard tests, tests with an unspecified subtype, "other," were excluded unless they had been used in published safety assessments from the Research Institute for Fragrance Materials (RIFM) or the Scientific Committee on Consumer Safety (SCCS). Additional exclusions included tests without a dermal surface area (DSA) value, tests for which subject numbers or doses were reported as ranges, and tests of substances of variable or ill-defined composition, such as natural extracts and other mixtures.

LLNA data were also curated in accordance to the methodologies described in the OECD Guideline 497 project (see Annex III in [4]). LLNA studies with three or four days of exposure to topical application of test substance on both ears of the mice were accepted. Acceptable tests also used a radiolabeled marker and measured lymphocyte proliferation in the lymph nodes draining the site of test substance application during the induction phase of skin sensitisation in either pooled or individual animals. Tests must include a vehicle control, and reports must provide all test concentrations and corresponding stimulation index (SI) values. Tests were excluded if sodium lauryl sulfate was applied to enhance the response or if the EC_3 values were extrapolated and did not meet the following criteria [5]:

The lowest measured SI value was < 5

The extrapolated EC_3 was less than 10-fold of the closest tested concentration.

The slope ratio was ≤ 2 and non-negative. This value is the ratio of the slope from the high dose to the mid-dose, to the slope from the mid-dose to the lowest dose.

In vitro OECD TG-based method data (DPRA, KeratinoSens™, h-CLAT and U-Sens™) were first obtained from the ICE database. Additional studies were obtained from [6] and [7]. Compiled databases were checked for duplicate studies via manual inspection of study results and references.

Once all data had been sourced, chemicals with a single study in total across all study types were removed. The computational burden of using the SARA-ICE model scales with the number of chemicals in the core SARA-ICE database. The primary purpose of the core SARA-ICE database is to estimate chemical-agnostic model variables. Chemicals with a single study in total across all study types provide no information towards learning chemical-agnostic model variables, therefore were removed for computational efficiency.

The SARA-ICE database is supplied as a supplementary workbook "SARA-ICE_database.xlsx". The first sheet lists all compounds in the database and each additional sheet lists all study data. Each data type is included on a data type-specific sheet and each row corresponds to a single study.

Assay results and conversion to observations of SARA-ICE variables

The SARA-ICE model describes, probabilistically, relationships between data from HPPTs, LLNAs and *in vitro* assays including the DPRA, kDPRA, KeratinoSens™, h-CLAT and U-Sens™. Prior to use within the model, data from each of the study types described undergo a deterministic transformation. The purpose of this transformation is to: a) convert data points to a unit scale for computational stability, and b) ensure it is reasonable to assume linear relationships between model variables corresponding to each study type. Details of the applied transformations are provided below.

HPPT

The information taken from a HPPT study includes the dermal induction dose (μg cm^{-2}), the total number of test subjects and the total number of subjects sensitised following challenge. The dermal induction dose is converted to its base-10 logarithm before use in the model.

LLNA

The information taken from an LLNA is the EC_3. The EC_3 is first converted to units of µg cm^{-2} and then the base-10 logarithm is calculated. For negative LLNAs in which an EC_3 is unavailable, the input is $> x$ where x is the base-10 logarithm of the maximum tested concentration in units of µg cm^{-2}. The core SARA-ICE database contains EC_3s calculated via extrapolation to a value less than the minimum concentration tested. However, in instances where this cannot be done reliably the SARA-ICE input may be provided as $< x$ where x is the base-10 logarithm of the minimum concentration tested.

DPRA

From the DPRA, we determine the maximum depletion of either the cysteine or lysine peptides. Assuming the maximum depletion is expressed as a percentage y, and the value is between 1% and 99%, this is converted to the value

$$x = \log\left(\frac{100\% - y}{y}\right),$$

where the base of the logarithm is Euler's constant. This conversion is essentially the logit transformation and serves to map the percentage depletion to a real number.
If the maximum depletion is less than 1%, which includes studies in which the reported maximum percentage depletion is negative, then the percentage depletion is expressed as <1%. In such cases, the input to SARA-ICE is $> x_{\max}$, where

$$x_{\max} = \log\left(\frac{100\% - 1\%}{1\%}\right) \approx 4.595\ .$$

Alternatively, if the maximum depletion is greater than 99%, the input to SARA-ICE is $< x_{\min}$ where

$$x_{\min} = \log\left(\frac{100\% - 99\%}{99\%}\right) \approx -4.595\ .$$

Censoring of inputs at 1% and 99% is performed automatically.

Kinetic DPRA

To use the kinetic DPRA as an input type, the log Kmax value is first converted to have units g L^{-1} s^{-1}. This is readily achieved by subtracting the base-10 logarithm of the chemical's molecular weight from the log Kmax value when in typical units of M^{-1} s^{-1}. For input into SARA-ICE, the converted log Kmax value is multiplied by -1. This transformation is performed to ensure a positive correlation between the input and the ED_{01}, i.e., low values of the input correspond to low ED_{01} values.

KeratinoSens™

The information used within SARA-ICE from a KeratinoSens™ assay is the reported $EC_{1.5}$ value and the IC_{50} (concentration reducing cellular viability by 50%). These values are first converted to units of µg ml^{-1} following which the base-10 logarithm of each is calculated. Since the usual reported units for a KeratinoSens™ assay are µM, it is usually necessary to know the molecular weight of a substance to use a KeratinoSens™ result in the SARA-ICE DA.
If the KeratinoSens™ assay exhibits no concentration-response, e.g., if the luciferase expression is less than 1.5-fold greater than control at all test concentrations, then the input into SARA-ICE is $> x$ where x is the maximum concentration tested within the assay and the conversion above is similarly followed. The SARA-ICE model may also accept inputs of the form $< x$ which may occur if there was more than 1.5-fold increase in luciferase activity at all tested concentrations.

h-CLAT

The information taken from the h-CLAT assay is the minimum of either reported CD86 EC150 or CD54 EC200, and the CV75 (75% cellular viability) value. If only one of CD86 EC150 and CD54 EC200 is reported, the reported value is used. If neither value is reported, the input is $> x$ where x is the maximum concentration tested. This is inferred as 1.2 times the CV75 if not provided explicitly. Units of inputs are assumed to be µg ml^{-1} and base-10 logarithms are calculated. If either of the CD86 EC150 or CD54 EC200 are reported as $< x$, this also forms a valid input.

U-Sens™

The information taken from the U-Sens™ assay is the CD86 EC150 and CV_{70} (70% cellular viability) in units of µg ml^{-1}. Similarly to above, both inputs are transformed to their base-10 logarithm. If the U-Sens™ study is negative, the input is $> x$ where x is the maximum concentration tested. The input $< x$ may also be supplied, for example, if the CD86 EC150 was determined to be lower than the minimum concentration tested.

The SARA-ICE model

In Bayesian statistics, there is no distinction between model variables and data (a data point is simply an observation of one of the random variables within the model). In the case of the SARA-ICE DA, the number of variables within the model scales with the number of chemicals in the database. There are 21 variables reserved for each chemical, which for the core database results in a total of 21 x 434 = 9,114 chemical-specific variables. Three chemical-specific variables (one location-like, two scale-like) are reserved for the HPPT, whilst two variables (one location-like and one scale-like) are reserved for each of the other input types. Location-like variables are used to describe the first moment (average) of some quantity whilst scale-like variables are used to describe the second moment (variance) of the same distribution. The HPPT has two scale-like chemical-specific variables, one for population variability in individual thresholds for sensitisation and the second for variability in HPPT test conditions.
The SARA-ICE model also includes a far smaller number of “chemical-agnostic” variables. These variables are not specific to any chemical and form the core set of variables of the model. All of these variables are present in the model irrespective of the number of chemicals. A subset of these describes the relationships between the chemical-specific location-like variables. A second subset is used to describe how chemical-specific scale-like parameters vary across chemical space. In what follows, the subscript i is used to index a chemical within the model. If a variable is written with subscript i, this implies the variable is chemical-specific. Alternatively, if a variable is written without subscript i, it can be assumed to be chemical-agnostic.

Modelling HPPT data

The primary variable of interest within the SARA-ICE DA is the chemical-specific random variable θ_i, defined as the base-10 logarithm of the HPPT dose expected (ED) to sensitise 1% of a HPPT-eligible population (ED_{01}). Within the SARA-ICE model, both HRIPT and HMT studies are treated identically. Each θ_i is assigned the prior distribution

$$\theta_i \sim \mathrm{Normal}(\theta_{\mathrm{loc}}, \theta_{\mathrm{scale}}) ,$$

with $\theta_{\mathrm{loc}} = 3.738$ and $\theta_{\mathrm{scale}} = 2.413$. This prior choice is obtained from assuming

$$\Pr\left(\theta_i \leq \log_{10}(500)\right) \approx \frac{1}{3} ,$$

$$\Pr\left(\theta_i \geq \log_{10}(60{,}000)\right) \approx \frac{1}{3}.$$

A potency threshold of 500 µg cm^{-2} is used to dichotomise sensitiser potency into subcategories 1A and 1B under the *Globally Harmonized System of Classification and Labelling of Chemicals* (GHS) [8]. The higher threshold of 60,000 µg cm^{-2} is applied to ED_{01} estimates to enable discrimination between class 1 and NC compounds. Thus, this prior is constructed under the assumption of a uniform distribution over the classification set $\{1A, 1B, NC\}$.

An individual from a HPPT-eligible population (indexed by j), is assumed to possess a personal threshold for sensitisation to chemical i, denoted $\hat{\theta}_{i,j}$, such that, for study dose d, if $\log_{10}(d) < \hat{\theta}_{i,j}$, the individual does not present with contact allergy upon challenge and if $\log_{10}(d) \geq \hat{\theta}_{i,j}$, then the individual will be counted as sensitised. The distribution of thresholds within a population is assumed to be normally distribution on the logarithmic scale and the probability density function for this is re-parameterised in terms of the 1st percentile (θ_i), rather than the mean. The cumulative density function for this distribution of personal thresholds for induction of sensitisation to chemical i is expressed as

$$\Pr\left(\hat{\theta}_{i,j} \leq x \mid \theta_i,\, s_i\right) = F(x \mid \theta_i,\, s_i) = \Phi\left(\frac{x - \theta_i}{s_i} + \Phi^{-1}(0.01)\right),$$

where $\Phi(\cdot)$ is the cumulative density function (CDF) of a standard normal distribution. Whilst individual thresholds for sensitisation may vary in time, the overall distribution across a population is assumed stationary with respect to time. The variable s_i gives the standard deviation of individual thresholds for induction of sensitisation (on the log-10 scale). Estimates of this variable are regularised using the prior structure

$$s_i \sim \text{Weibull}(5, s_{\text{scale}}),$$
$$s_{\text{scale}} \sim \text{HalfNormal}(0,1).$$

The probability of sensitisation within a HPPT study (indexed by k) is assumed to vary in response to experimental choices such as the vehicle used to apply the chemical. Given the nominal study induction dose $d_{i,k}$ for chemical i, the probability of sensitisation at $d_{i,k}$ is modelled as

$$p_{i,k} = \text{logit}^{-1}\left(\text{logit}\left(\Phi\left(\frac{\log_{10}(d_{i,k}) - \theta_i}{s_i} + \Phi^{-1}(0.01)\right)\right) + \lambda_{i,k}\right)$$

where $\lambda_{i,k}$ is a study-specific offset (on the logit scale) such that study-to-study variability in results can be adequately accounted for. Estimates of $\lambda_{i,k}$ are regularised with the prior structure

$$\lambda_{i,k} \sim \text{Weibull}(5,\, \lambda_{\text{scale}}),$$
$$\lambda_{\text{scale}} \sim \text{HalfNormal}(0, 2).$$

Given probability of sensitisation $p_{i,k}$, the number of subjects sensitised, $n_{i,k}$ out of those tested $N_{i,k}$ is assumed to be binomially distributed such that

$$n_{i,k} \sim \text{Binomial}\left(N_{i,k}, p_{i,k}\right).$$

Numerical stability when calculating the normal distribution CDF

The use of the CDF of a standard normal distribution may result in numerical instability when performing floating point arithmetic. To counter this, we use the analytical approximation to the standard normal CDF provided in the documentation for the Stan programming language [9]. For numerical stability, we let

$$\Phi(x) \approx \text{logit}^{-1}(0.07056x^3 + 1.5976x).$$

Modelling correlations between study types

Each chemical in the model is assigned ten location-like variables, one of which is θ_i as introduced in the previous section. The other nine variables describe the (latent) mean for chemical i in terms of each of the other inputs. These include the LLNA mean EC_3, DPRA mean maximum depletion, kinetic DPRA mean reactivity rate, KeratinoSens™ mean $EC_{1.5}$ and mean IC_{50}, h-CLAT mean minimum of the CD86 EC150 and CD54 EC200 biomarkers and mean CV75, and U-Sens™ mean CD86 EC150 and mean CV_{70}. As a more specific example, considering the input transformations defined above, let $\bar{y}_{i,\mathrm{L}}$ be the location-like variable for chemical i in the LLNA. Then, after reversing the transformation defined above, $10^{\bar{y}_{i,\mathrm{L}}}$ is the average EC_3 (µg cm^{-2}) for chemical i in the LLNA.

We use the subscript $s \in \mathrm{S} = \{\mathrm{L,D,kD,K,Kc,H,Hc,U,Uc}\}$ to distinguish the nine location-like variables $\bar{y}_{i,s}$ for chemical i for each input type. The vector $\bar{y}_i$ is composed from the collection of all $\bar{y}_{i,s}$ for the non-HPPT input types. The role of the study result conversions defined above is to ensure it is reasonable to assume each $\bar{y}_{i,s}$ is linearly associated with θ_i. Conditional on θ_i, it is assumed that

$$\bar{y}_i \sim \mathrm{MultivariateNormal}(\boldsymbol{\alpha} + \boldsymbol{\beta}\,(\theta_i - \theta_{\mathrm{loc}}), \Sigma)\ ,$$

where $\boldsymbol{\alpha} = [\alpha_{\mathrm{L}}, \ldots, \alpha_{\mathrm{Uc}}]$ and $\boldsymbol{\beta} = [\beta_{\mathrm{L}}, \ldots, \beta_{\mathrm{Uc}}]$ are vectors of intercept and slope variables, respectively. The covariance matrix Σ is defined as

$$\Sigma = \mathrm{diag}(\gamma_{\mathrm{L}}, \ldots, \gamma_{\mathrm{Uc}})\ \Omega\ \mathrm{diag}(\gamma_{\mathrm{L}}, \ldots, \gamma_{\mathrm{Uc}})^T$$

where γ_s, $s \in S$ are scale parameters defining the marginal standard deviations of the regression residuals and Ω is an 9×9 correlation matrix with elements $\rho_{s,t}$ for all possible pairs $s \in S$, $t \in S$. Each scale parameter is assigned the prior distribution

$$\gamma_s \sim \mathrm{HalfNormal}(0,2)$$

and the correlation matrix Ω is assigned a Lewandowski-Kurowicka-Joe prior distribution with shape parameter equal to 2 [10]:

$$\Omega \sim \mathrm{LKJ}(2)\ .$$

Prior distributions for the intercept and slope parameters are as follows:

$$\begin{aligned}
&\alpha_{\mathrm{L}} \sim \mathrm{Normal}(3,1)\ ,\\
&\alpha_{\mathrm{D}} \sim \mathrm{Normal}(0,2)\ ,\\
&\alpha_{\mathrm{kD}} \sim \mathrm{Normal}(5,2)\ ,\\
&\alpha_s \sim \mathrm{Normal}(2,1) \quad \text{for } s \in \{\mathrm{K}, \ldots, \mathrm{Uc}\}\ ,\\
&\beta_{\mathrm{L}} = 1 \quad (\text{This variable is fixed at a constant})\ ,\\
&\beta_s \sim \mathrm{Normal}(2,1) \quad \text{for } s \in \{\mathrm{D,kD}\}\ ,\\
&\beta_s \sim \mathrm{Normal}(1,1) \quad \text{for } s \in \{\mathrm{K}, \ldots, \mathrm{Uc}\}\ .
\end{aligned}$$

Notice that the LLNA slope variable is fixed to a constant equal to one. This enforces the assumption that murine sensitiser potency is proportional to human sensitiser potency (when expressed on a linear axis).

Modelling input variability

Given location-like variables $\bar{y}_{i,s}$, converted study results $y_{i,j,s}$ (where j indexes repeat studies) are assumed to be normally distributed such that

$$y_{i,j,s} \sim \mathrm{Normal}\left(\bar{y}_{i,s}, \sigma_{i,s}\right)$$

where $\sigma_{i,s}$ is the standard deviation of converted study results for chemical i for input type $s \in S$. These variables are regularised using the prior structure

$$\begin{aligned}
&\sigma_{i,s} \sim \mathrm{Weibull}\left(5, \sigma_{\mathrm{scale},s}\right)\ ,\\
&\sigma_{\mathrm{scale},s} \sim \mathrm{HalfNormal}(0,1) \quad \text{for } s \in S\ .
\end{aligned}$$

The idea here is that each input variability is distinct for different chemicals and input types. The

Weibull prior serves to shrink variability parameters estimates towards each other. If a chemical has no repeat studies for a particular assay, the estimate of the $\sigma_{i,s}$ will be distributed as its prior distribution $\mathrm{Weibull}\left(5, \sigma_{\mathrm{scale},s}\right)$.

Censored data

If an assay has a negative result, the input into the SARA-ICE model may be expressed as an inequality (< or >). These are examples of *censored data*. In such cases we integrate over the assumed sampling distribution for this observation from the censoring point. For example, if the converted datapoint is $> y_{\mathrm{obs}}$, then the probability of this event occurring is

$$\Pr\left(y_{i,j,s} > y_{\mathrm{obs}} \mid \bar{y}_{i,s}, \sigma_{i,s}\right) = 1 - \Pr\left(y_{i,j,s} \leq y_{\mathrm{obs}} \mid \bar{y}_{i,s}, \sigma_{i,s}\right) = 1 - \Phi\left(\frac{y_{\mathrm{obs}} - \bar{y}_{i,s}}{\sigma_{i,s}}\right).$$

Similarly, if the converted datapoint is $< y_{\mathrm{obs}}$, then

$$\Pr\left(y_{i,j,s} < y_{\mathrm{obs}} \mid \bar{y}_{i,s}, \sigma_{i,s}\right) = \Pr\left(y_{i,j,s} \leq y_{\mathrm{obs}} \mid \bar{y}_{i,s}, \sigma_{i,s}\right) = \Phi\left(\frac{y_{\mathrm{obs}} - \bar{y}_{i,s}}{\sigma_{i,s}}\right).$$

The use of the standard Gaussian CDF may result in numerical instability. Therefore, we use the approximation recommended in the Stan programming language technical manuals (see section on computational approximation above).

Inferences from the SARA-ICE model

The SARA-ICE model defined in the previous section describes the probability distribution of (transformed) observations from the HPPT, LLNA, DPRA, kDPRA, KeratinoSens™, h-CLAT and U-Sens™ conditional on the assumed prior structure. Let Θ be the joint distribution of all model variables and $\mathcal{D}$ be the joint distribution of all data. Then the previous section defines the density

$$p(\mathcal{D} \mid \Theta)\, p(\Theta)\,.$$

Bayes theorem defines the conditional probability,

$$p(\Theta \mid \mathcal{D}) = \frac{p(\mathcal{D} \mid \Theta)\, p(\Theta)}{p(\mathcal{D})}\,,$$

the left side of which expresses the joint probability distribution of all model variables, conditional on a set of observations $\mathcal{D}$, such as those in the core SARA-ICE database. Estimates of quantities of interest, such as the HPPT 1% sensitising dose for a particular chemical with data in $\mathcal{D}$, are obtainable as marginal distributions of $p(\Theta \mid \mathcal{D})$.

Computation

The distribution $p(\Theta \mid \mathcal{D})$ is analytically intractable (cannot be evaluated directly) therefore it is evaluated by drawing samples of model parameters from it. Sampling is achieved using Monte Carlo Markov Chain (MCMC) approaches. Whilst any state-of-the-art sampler may be used for this purpose, the Python API to the programming software Stan [9] has been used to realise the model in a computational environment. The number of draws from the posterior distribution is typically chosen to be 10,000, obtained by running 4 chains of 3,000 iterations each and discarding the first 500 samples as burn-in.

Samples of model variables may be used to estimate the marginal densities of variables of interest (or even transformations of model variables of interest). For example, the density of the ED_{01} for some chemical of interest can be visualised in the form of a histogram using samples $10^{\theta_{i,k}}$ where $\theta_{i,k}$ is the kth draw of variable θ_i. Other quantities of interest such as posterior expectations, variances etc. can be obtained by computing sample means and variances from

the set of draws of model variables. Using the current SARA-ICE database and model as a benchmark, the computational time using a 4-core CPU is around one hour to obtain 10,000 samples from the posterior distribution.
It is important to note that MCMC techniques which are used to obtain parameter estimates are probabilistic methods. This implies estimates have a degree of variation to them. The magnitude of the variation can be controlled by adjusting by the number of draws, but this comes at the cost of an increased computational load. The default choice of 10,000 draws is usually sufficient to ensure posterior quantities, such as mean estimates, are reproducible to two significant figures. As a defined approach, this implies that SARA-ICE DA inferences will not produce identical outputs even if the DA is run on the same dataset multiple times (unless random number generators used are initialised with the same seed). The minor variance in SARA-ICE DA estimates is judged to represent an acceptable level of reproducibility for regulatory use (e.g. if benchmarked against the level of variance in experimental results from the assays used to inform SARA-ICE predictions), however, it could be further reduced with increased computation time, if needed.

Inferences for chemicals not in the core SARA-ICE database

Suppose we are interested in obtaining inferences for a chemical not in the SARA-ICE database, but for which we have some assay results. To obtain inferences for this chemical, its data must be merged with SARA-ICE database to create a new dataset $\widetilde{\mathcal{D}}$ following which we evaluate the distribution $p\left(\Theta, \Theta_{\text{chemical}} \mid \widetilde{\mathcal{D}}\right)$, where Θ_{chemical} are the additional chemical-specific model variables. The joint distribution of chemical-specific variables can be obtained as

$$p\left(\Theta_{\text{chemical}} \mid \widetilde{\mathcal{D}}\right) = \int_{\Theta} p\left(\Theta, \Theta_{\text{chemical}} \mid \widetilde{\mathcal{D}}\right) \, \mathrm{d}\Theta \, .$$

The posterior distribution in this case needs to be evaluated using the computational techniques discussed above. Because the full SARA-ICE database, plus some extra datapoints are being used, the computational time sample from the posterior distribution is roughly the same at 1 hour using a 24-core CPU. This can be prohibitively slow if deploying the model within a production environment. An approach to address this is provided in the following section.

Ensuring rapidly obtainable results for a chemical not in the core SARA-ICE database

Given a small set of study results for some chemical of interest which is not in the core SARA-ICE database, it is desirable to implement a computational environment that enables estimation of the chemical-specific variables for this chemical in seconds, not hours.
This is achieved by first evaluating the posterior distribution $p(\Theta \mid \mathcal{D})$ conditional on the core SARA-ICE database $\mathcal{D}$. The set of model variables Θ is split into those which are chemical-agnostic, Θ_{agnostic} and those which are chemical-specific, Θ_{specific}. We set up an analytical approximation to the joint distribution of chemical-agnostic variables

$$p(\widehat{\Theta}) \approx \int_{\Theta_{\text{specific}}} p\left(\Theta_{\text{specific}}, \Theta_{\text{agnostic}} \mid \mathcal{D}\right) \, \mathrm{d}\Theta_{\text{specific}}$$

and replace the original prior in the SARA-ICE with this approximation. Letting $\mathcal{D}_{\text{chemical}}$ be the data for the chemical of interest, the joint distribution of model variables for this chemical is approximated as

$$p(\Theta_{\text{chemical}} \mid \mathcal{D}_{\text{chemical}}) = \int_{\widehat{\Theta}} \frac{p(\mathcal{D}_{\text{chemical}} \mid \Theta_{\text{chemical}}) p\left(\Theta_{\text{chemical}} \mid \widehat{\Theta}\right) p(\widehat{\Theta}) a}{p(\mathcal{D}_{\text{chemical}})} \, \mathrm{d}\widehat{\Theta} \, .$$

MCMC techniques are used to evaluate this distribution, however, the number of data points is small, and correspondingly, the number of chemical-specific model variables is small. This

computation can be performed in a matter of seconds as opposed to the hours when performed over the 1,000s of datapoints in the core SARA-ICE database.
This SARA-ICE model with the prior distribution replaced by an approximation of the marginal posterior conditioned on the SARA-ICE database is referred to as the *SARA-ICE production* model. The method of obtaining an analytical approximation to the full posterior is outlined in the next section.

Analytical approximation of the SARA-ICE posterior distribution

The chemical-agnostic variables are split into two groups: the first group consists of all γ_s, $s \in S$ and $\rho_{s,t}$, $s,t \in S$ making up the covariance matrix Σ defined above. The second group consists of everything else which includes α_s, β_s and $\sigma_{\text{scale},s}$ for $s \in S$ (except β_{L} which is constant) and lastly s_{scale} and λ_{scale}.
The marginal posterior distribution of Σ is approximated using a Wishart distribution. The statistical model

$$\begin{aligned}
&\Sigma \mid \mathcal{D} \sim \text{Wishart}(\Psi, \nu)\ , \\
&\nu \propto 1\ , \\
&\Psi = \text{diag}(\omega)\Omega\text{diag}(\omega)\ , \\
&\omega \sim \text{HalfNormal}(0,1)\ , \\
&\Omega \sim \text{LKJ}(2)\ ,
\end{aligned}$$

is assumed and samples from the density $p(\Psi, \nu \mid \Sigma, \mathcal{D})$ are obtained using Stan. The posterior expectations

$$\begin{aligned}
\overline{\Psi} &= \text{E}[\,\Psi \mid \Sigma, \mathcal{D}\,]\ , \\
\overline{\nu} &= \text{E}[\,\nu \mid \Sigma, \mathcal{D}\,]
\end{aligned}$$

are calculated from which we define an approximate sampling distribution for the posterior distribution of Σ as

$$\Sigma \sim \text{Wishart}(\overline{\Psi}, \overline{\nu})\ .$$

The marginal distribution of the second group of variables is approximated using a multivariate Gaussian. Some of these random variables are strictly positive, therefore the multivariate Gaussian is used to approximate the sampling distribution for the logarithm of these variables. Define the vector

$$\boldsymbol{\theta} = \left[\log(s_{\text{scale}}), \log(\lambda_{\text{scale}}), \alpha_{\text{L}}, \ldots, \alpha_{\text{Uc}}, \beta_{\text{D}}, \ldots, \beta_{\text{Uc}}, \log\left(\sigma_{\text{scale,L}}\right), \ldots, \log\left(\sigma_{\text{scale,L}}\right)\right]^T\ ,$$

then the parameters statistical model

$$\boldsymbol{\theta} \sim \text{MultivariateNormal}(\boldsymbol{\mu}_\theta, \Sigma_\theta)$$

are estimated using maximum likelihood estimates $\hat{\boldsymbol{\mu}}_\theta$ and $\hat{\Sigma}_\theta$. These then define the approximation to the marginal posterior distribution of $\boldsymbol{\theta}$ as

$$\boldsymbol{\theta} \sim \text{MultivariateNormal}\left(\hat{\boldsymbol{\mu}}_\theta, \hat{\Sigma}_\theta\right)\ .$$

The prior distribution of chemical agnostic parameters is then defined as the product

$$p\left(\widehat{\Theta}\right) \propto p(\overline{\Psi}, \overline{\nu}) p\left(\hat{\boldsymbol{\mu}}_\theta, \hat{\Sigma}_\theta\right)\ .$$

This approximation will preserve correlations between variables within Σ and between variables within $\boldsymbol{\theta}$, however, any correlations between these two are lost within the approximation. Practically, this loss of information results in small differences between ED_{01} estimates obtained when using this prior and only chemical-specific information.

SARA-ICE production

The *SARA-ICE production* model has the same structure as the SARA-ICE model except that the prior distribution for all chemical-agnostic parameters is replaced by the analytical

approximation discussed in the previous section. The SARA-ICE production model allows ED_{01} estimates to be generated for a single chemical at a time (with a relatively rapid computation time of only a few seconds). If the data for a chemical in the SARA-ICE database is used as inputs to the SARA-ICE production model, the output will be almost the same as having used that data alongside all other data in the SARA-ICE database in the SARA-ICE model (this is demonstrated later in this document).

One advantage of the SARA-ICE production model is that, given available data, it allows ED_{01} estimates to be rapidly generated for a chemical that *is not* in the SARA-ICE database. This makes it suitable for deployment in a production environment. A second advantage of the SARA-ICE production model is that it can be used to rapidly generate ED_{01} estimates for a single chemical using different combinations of inputs. For example, one may wish to generate ED_{01} estimates for a chemical using a) *in vivo* data only, b) *in vitro* data only and c) a combination of both data types. The SARA-ICE production model facilitates this without having to refit to the full database each time.

GHS classification probabilities

While GHS estimation has not yet been evaluated for inclusion as part of GL 497, it is possible to discretise the distribution for the ED_{01} into three intervals corresponding to GHS sensitiser categories [8]. This is done by specifying a threshold of 60,000 µg cm^{-2} to discriminate between class 1 and NC. This threshold is the maximum achievable dose in a standard HPPT (concentration at 100%, assuming 0.3 g applied within a 25 mm Hill top chamber) [1]. The SARA-ICE probability that a chemical is a sensitiser is simply the probability that the ED_{01} is less than this threshold. A second threshold of 500 µg cm^{-2} is used to discriminate between GHS subcategories 1A and 1B.

To represent this formally, let the discrete random variable $\mathcal{G}_{\mathrm{bin},i}$ represent the SARA-ICE binary GHS category for the chemical i. The outcome space for this random variable is the set {1, NC}, the elements of which correspond to GHS binary class 1 and “not requiring classification” (equivalent to a classification of a chemical not being a non-sensitiser). Then we define

$$\Pr\left(\mathcal{G}_{\mathrm{bin},i} = 1 \mid \theta_i\right) = \Pr\left(\theta_i \leq \log_{10}(60{,}000)\right),$$
$$\Pr\left(\mathcal{G}_{\mathrm{bin},i} = \mathrm{NC} \mid \theta_i\right) = \Pr\left(\theta_i > \log_{10}(60{,}000)\right).$$

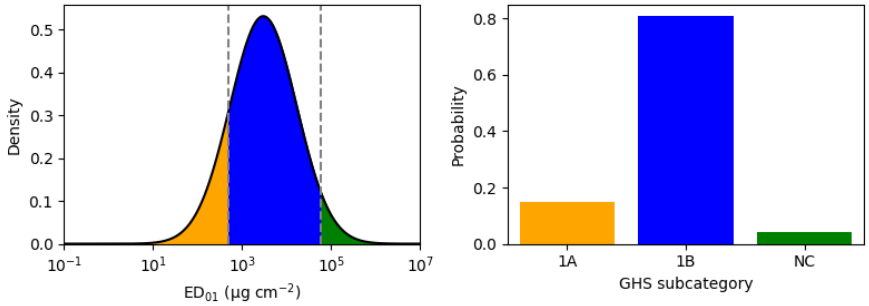
A lower threshold of 500 µg cm^{-2} is used as a boundary between GHS subcategories 1A and 1B [11]. Let the second discrete random variable $\mathcal{G}_{\mathrm{sub},i}$ represent the SARA-ICE subcategory GHS classification for chemical i. The outcome space for this random variable is the set {1A, 1B, NC} with probabilities

$$\Pr\left(\mathcal{G}_{\mathrm{sub},i} = \mathrm{1A} \mid \theta_i\right) = \Pr\left(\theta_i \leq \log_{10}(500)\right),$$
$$\Pr\left(\mathcal{G}_{\mathrm{sub},i} = \mathrm{1B} \mid \theta_i\right) = \Pr\left(\log_{10}(500) < \theta_i \leq \log_{10}(60{,}000)\right),$$
$$\Pr\left(\mathcal{G}_{\mathrm{sub},i} = \mathrm{NC} \mid \theta_i\right) = \Pr\left(\mathcal{G}_{\mathrm{bin},i} = \mathrm{NC} \mid \theta_i\right) = \Pr\left(\log_{10}(60{,}000) < \theta_i\right).$$

The translation of the continuous distribution of the ED_{01} into a discrete distribution is illustrated in **Error! Reference source not found.**. The prior distribution for θ_i is chosen such that the prior p robability of each GHS subcategory is one third. This necessarily induces a prior probability of two thirds for binary class 1 and 1/3 for binary class NC. Therefore, within the prior the SARA-ICE DA is biased towards predicting a chemical to be a sensitiser. This results in a small degree of *a priori* conservatism to estimates.

Figure 1. Example of translation of the distribution of the ED_{01} (left) to a discrete distribution for GHS subcategories.

The areas under the density in the left plot are equal to the probability that the ED_{01} lies between/beyond the chosen classification thresholds of 500 µg cm^{-2} and 60,000 µg cm^{-2}.



The SARA-ICE PoD

The ED_{01} in the SARA-ICE model is a latent representation of sensitiser potency in humans. It is statistically convenient to let the ED_{01} be above what one could physically dose in a HPPT study to enable representation of (possibly) non-sensitising compounds. Indeed, the proportion of the distribution above the threshold of 60,000 µg cm^{-2} is used to define a SARA-ICE estimate of the probability that a chemical is non-sensitising (see previous section). For risk assessment purposes, it may be desirable to have access to a single-valued representation of sensitiser potency, rather than a distribution. Such a single-valued quantity may be referred to as *point-of-departure* (PoD). A PoD can be computed from the distribution of the ED_{01} may calculating some suitable summary statistic. For example, calculating the mean of the distribution or reading off a suitable quantile. However, this may be result in a PoD that is greater than what one could physically dose – especially if the probability of GHS class NC is high.

To ensure PoDs are always calculated at a physically possible dose (ignoring solubility limits) we can choose to impose the additional assumption *that one is dealing with a potentially sensitising ingredient* before calculating a PoD. Mathematically, this amounts to conditioning on GHS class 1 and only considering the distribution of the ED_{01} truncated above at 60,000 µg cm^{-2}. Under this additional assumption, SARA-ICE PoDs are calculated as summary statistics of the truncated distribution.

Mathematically, let $\phi_i = \theta_i \mid \theta_i < \log_{10}(60000)$ and denote the cumulative density function of the random variable ϕ_i as F_{ϕ_i}. Then, two possible PoD type are defined as:

1. Mean PoD: $\mathrm{PoD} = 10^{\mathrm{E}[\phi_i]}$ (this is technically the geometric mean of the distribution of the ED_{01} assuming the compound is a sensitiser).
2. Quantile PoD: $\mathrm{PoD} = 10^{F^{-1}_{\phi_i}(q)}$ for some pre-chosen choice of quantile q.

Mean PoDs represent, in a sense, best estimates of the potency of the chemical assuming it is a sensitiser. This PoD type does not reflect any of the uncertainty inherent to its estimate. For suitably low quantiles q, e.g., 0.05, quantile PoDs represent a conservative estimate of potency. PoDs defined by this approach decouple the calculation of a PoD from hazard classification. One may always choose to calculate a SARA-ICE PoD for use in exposure-based risk assessment independently of any decisions pertaining to whether the chemical is a sensitiser or not. This may be advantageous when determination of sensitiser/non-sensitiser status is ambiguous.

Maximising the number of samples for PoD calculations

To calculate summary statistics of the distribution of the ED_{01} (mean, variance etc.), we draw a large sample from it (typically of size 10,000) using the computational procedures discussed above and report summary statistics calculated from the sample. However, only draws less than 60,000 µg cm^{-2} are relevant for computation of the SARA-ICE PoD and the number of draws may be far fewer than 10,000, especially when dealing with a low-potency chemical. To maximise the number of draws less than 60,000 µg cm^{-2}, this constraint may be applied *before* the sampling procedure is initialized. That is, in the SARA-ICE-production model we can draw a sample from ϕ_i directly rather than from θ_i and applying the <60,000 filter to the sample. Note that the model can be run a second time without this constraint to obtain GHS classification probabilities.

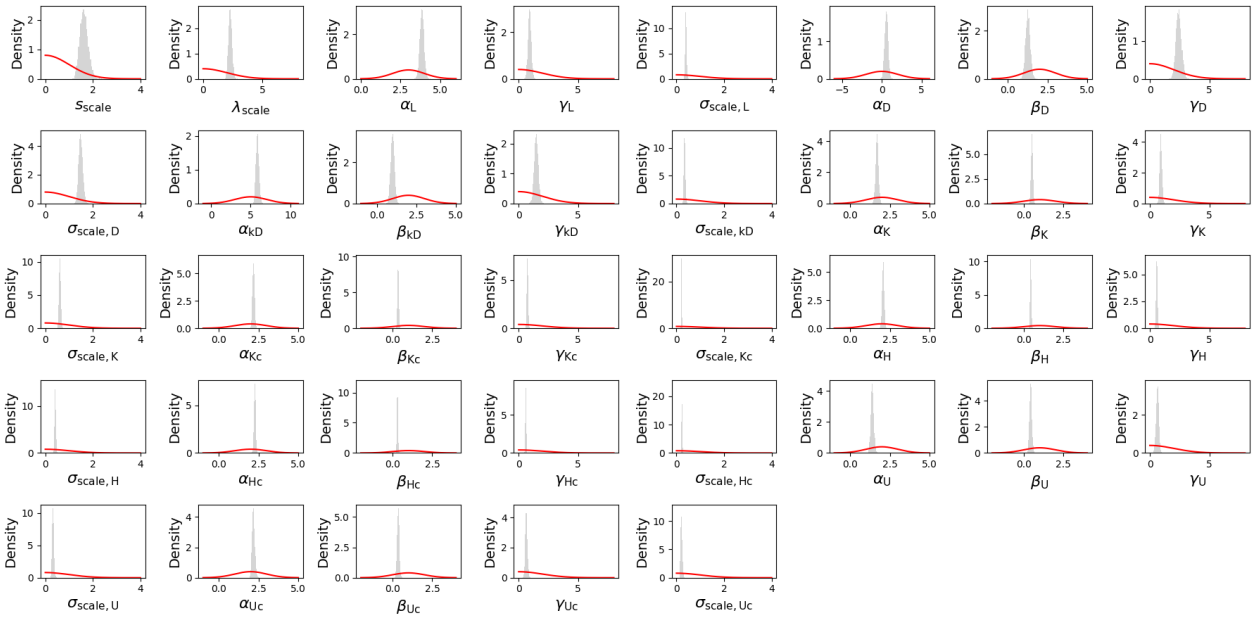
Basic model checks

Influence of prior distributions

Unless explicitly explained, such as is the case for the ED_{01}, prior distributions are chosen to be weakly regularising. This implies that we consider it undesirable if the prior distribution has a strong effect on posterior estimates. One approach to confirm this is to compare marginal posterior distributions against the prior for model parameters which we do not wish to impart too much influence. **Error! Reference source not found.** compares prior densities against h istograms of such parameters. From this figure it can be seen that the posterior distribution of each parameter is not concentrating in the extreme tail of the prior, indicating that the chosen prior is not imparting any more than a soft regularisation on the posterior estimate. From this we conclude that estimates of model parameters would be unlikely to change in any appreciable manner if prior choices were relaxed further.

Figure 2. Comparison of prior densities (red) against marginal posterior distributions of chemical-agnostic parameters.

For each parameter, observe that the posterior estimate is not concentrating in the extreme tail of the prior distribution.



Correlations between latent mean estimates

The SARA-ICE model assumes linear relationships between latent parameters representing average inputs and the base-10 logarithm of the HPPT ED_{01} (θ_i). In **Error! Reference source not found.**, posterior expectations of $\theta_i \mid \mathcal{D}$ are plotted against posterior expectation of $\bar{y}_{i,s} \mid \mathcal{D}$, $s \in S$ for chemicals with at least one data point available for the corresponding input. The strongest correlation is obtained with the LLNA input, which is expected since the model parameter β_L is fixed at a value of 1. The *in chemico* OECD TG-based input with the strongest association with human potency is the reactivity rate measure from the kinetic DPRA. For the *in vitro* OECD TG-based methods, a stronger correlation is obtained with the assay-specific output (e.g., $EC_{1.5}$ for the KeratinoSens™ assay) than is obtained with the cytotoxicity measure.
Error! Reference source not found. displays the correlations between all non-HPPT inputs. T he highest correlation is found between the h-CLAT potency input (minimum of CD86 EC150 and CD54 EC200, $\bar{y}_{i,H} \mid \mathcal{D}$) and the h-CLAT cytotoxicity input ($\bar{y}_{i,Hc} \mid \mathcal{D}$). Other pairs of inputs with notably strong correlations include DPRA depletion with kDPRA reactivity, cytotoxicity inputs from the KeratinoSens™ and h-CLAT and U-Sens™, U-Sens™ potency and cytotoxicity, finally, and h-CLAT and U-Sens™ potency. This last comparison is encouraging since these assays are considered to address the same key event (OECD TG 442E). Similarly, it is encouraging that the DPRA and kDPRA outputs are strongly associated.

Figure 3. Correlation between posterior expectations of θ_i and $\bar{y}_{i,s}$ for $s \in S$.

Each point corresponds to a single chemical. Only chemicals with at least one data point for the given study type are displayed on the plot.

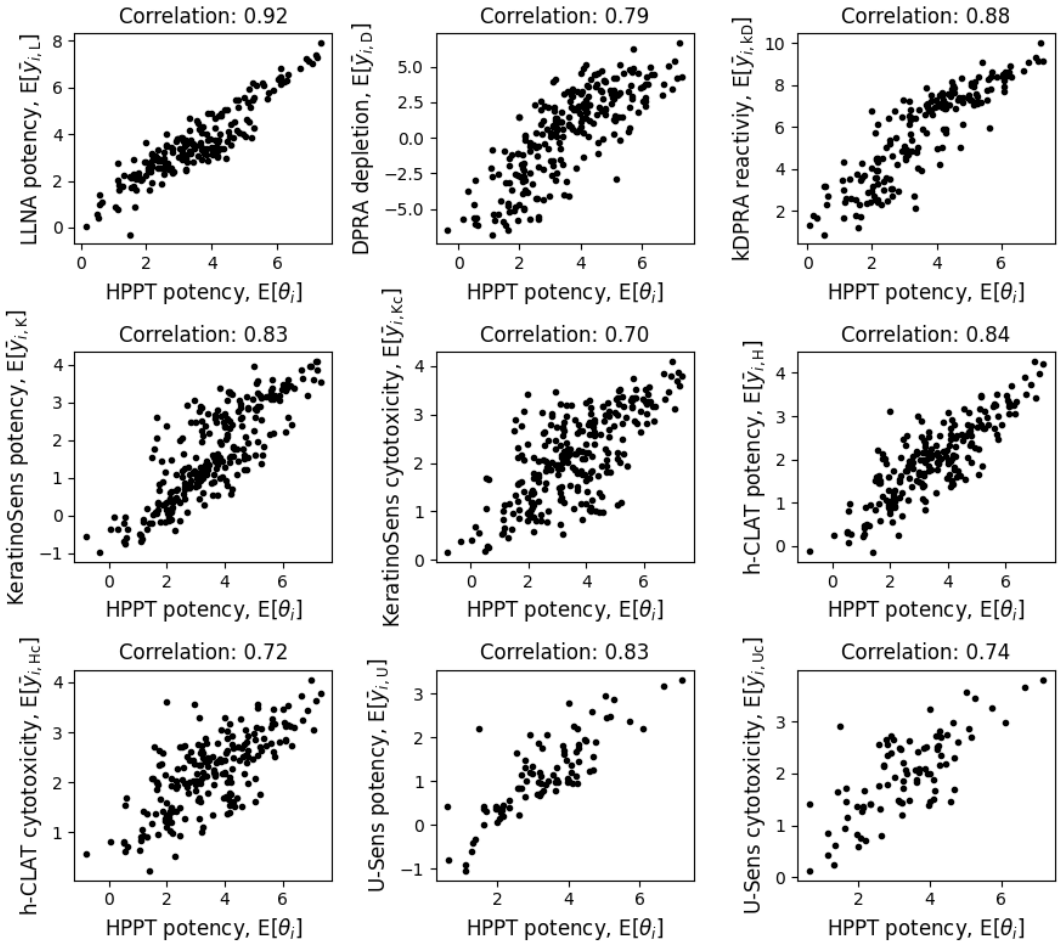
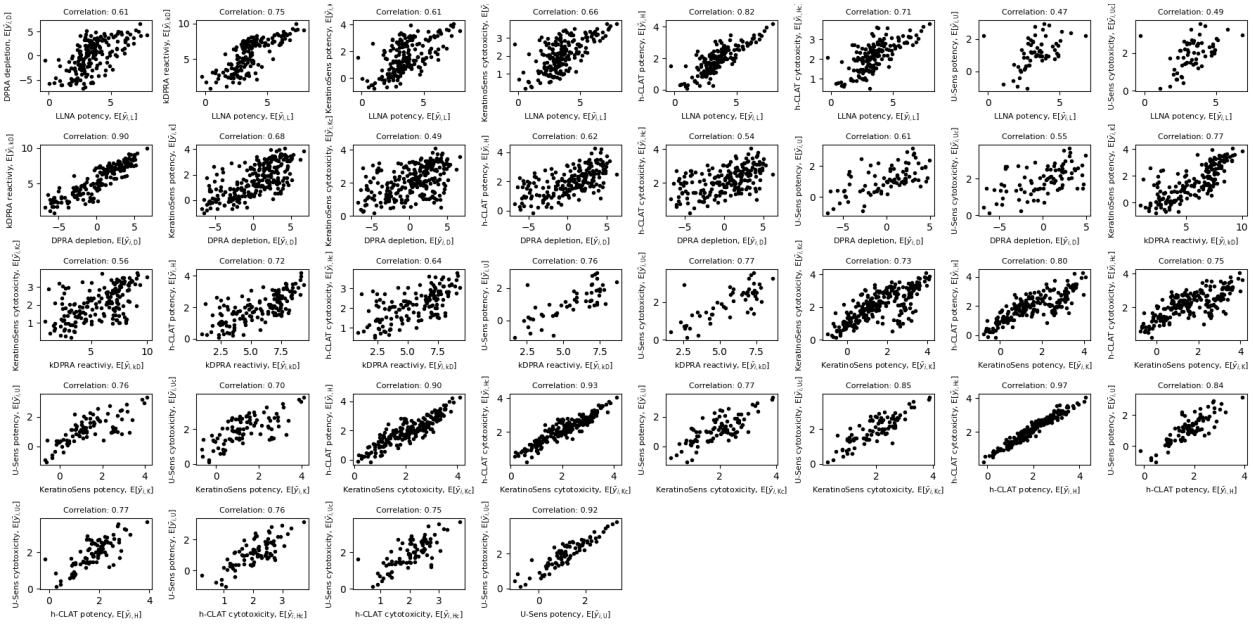


Figure 4. Correlation between each possible pair $\left(\mathrm{E}\left[\bar{y}_{i,s}\right],\mathrm{E}\left[\bar{y}_{i,t}\right]\right)$ for $s,t \in S$.



Correlations between latent mean estimates and SARA-ICE inputs

Posterior expectations of the latent mean parameters $\bar{y}_{i,s}$ and individual inputs $y_{i,j,s}$ for $s \in S$ are compared in **Error! Reference source not found.**. High correlations are obtained for all inputs i ndicating that mean estimates are largely consistent with the input data from which they are derived. The heterogenous structure of the SARA-ICE database implies some chemicals have multiple inputs available per input type. The variance of the scatter plots in the direction of the ordinate reflects the variability of the input type. This variance is noticeably smaller for cytotoxicity inputs in comparison to the specific inputs of the *in vitro* OECD TG-based methods. The kinetic DPRA also exhibits relatively low input variance, however this may be an artefact of limited repeat studies.

The informativeness of a single input for the purposes of estimating the ED_{01} is a function of both; how strong the relationship is in **Error! Reference source not found.**, in addition to the strength o f association in **Error! Reference source not found.**. To assess the net effect of these sources of variability, posterior expectations of θ_i are compared against inputs $y_{i,j,s}$ for $s \in S$ in **Error! R eference source not found.**. It is immediately noticeable that overall correlations between $\mathrm{E}[\theta_i \mid \mathcal{D}]$ and individual inputs are weaker than the interjoining correlations with latent mean estimates $\mathrm{E}\left[\bar{y}_{i,s} \mid \mathcal{D}\right]$. It is also apparent that correlations with cytotoxicity inputs are noticeably smaller, indicating that these are the least informative inputs.

Figure 5. Correlations between posterior expectation of latent mean predictors $\mathrm{E}[\bar{y}_{i,s}]$ and transformed inputs $y_{i,j,s}$ for $s \in S$.

Black points indicate regular inputs, upward pointing triangles indicate right-censored inputs and downward facing triangles indicate left-censored inputs. Correlation is measured using Spearman's ρ due to the "kinks" induced by the typical censoring points.

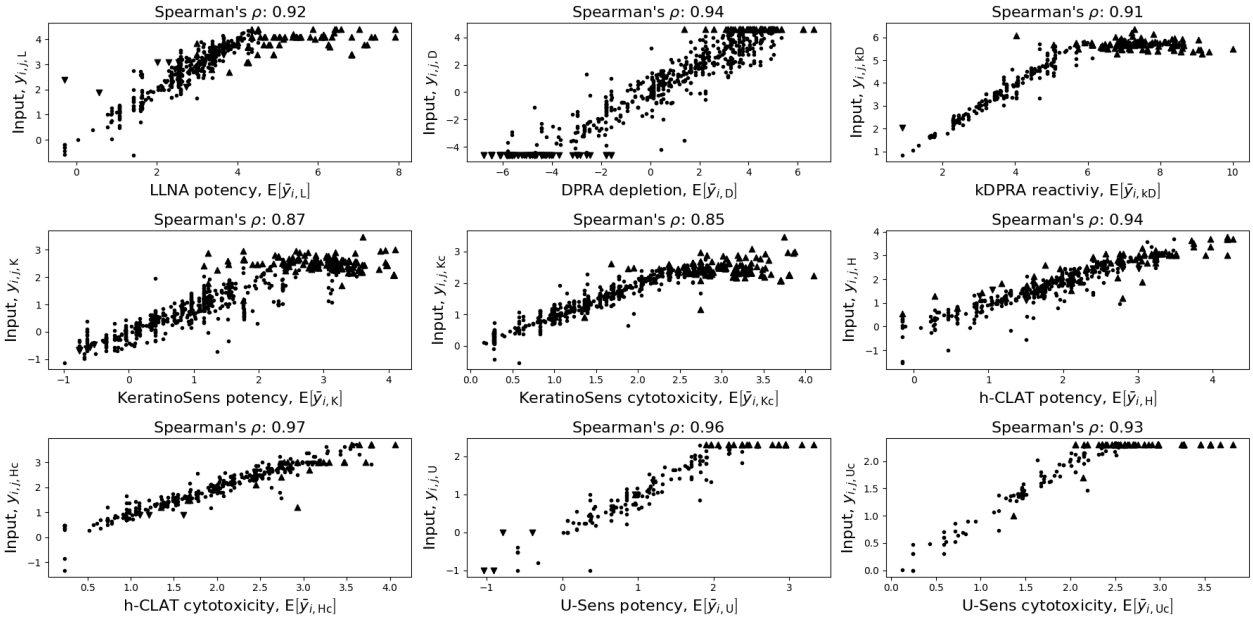
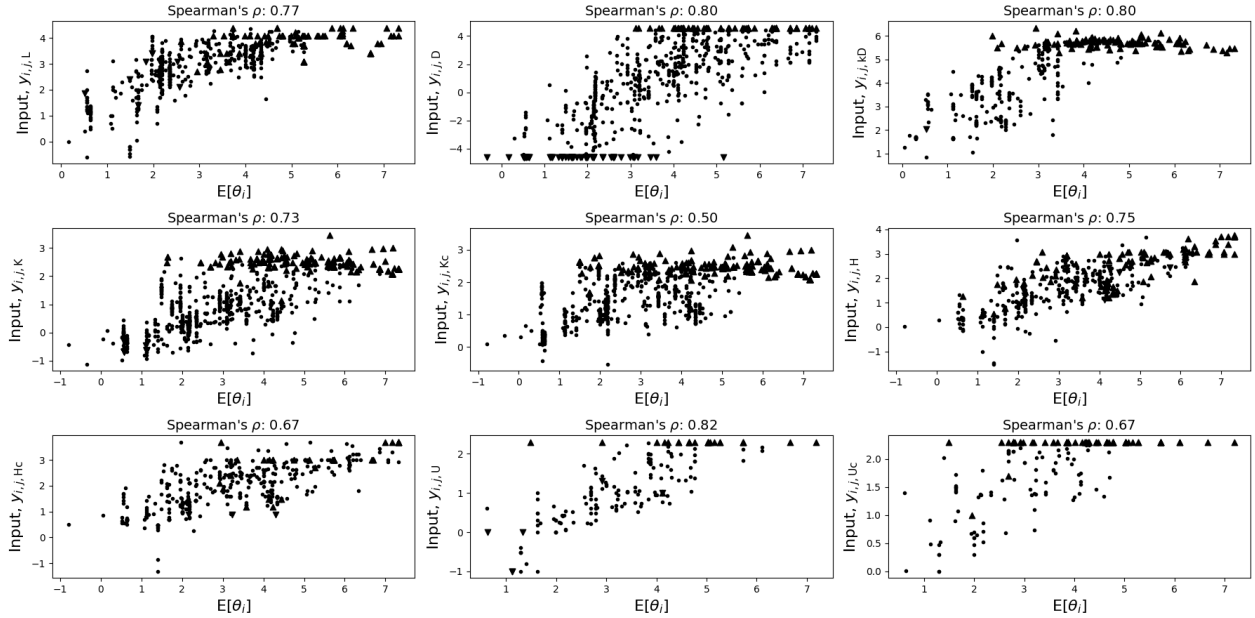


Figure 6. Correlation between posterior expectations of θ_i and inputs $y_{i,j,s}$ for $s \in S$.

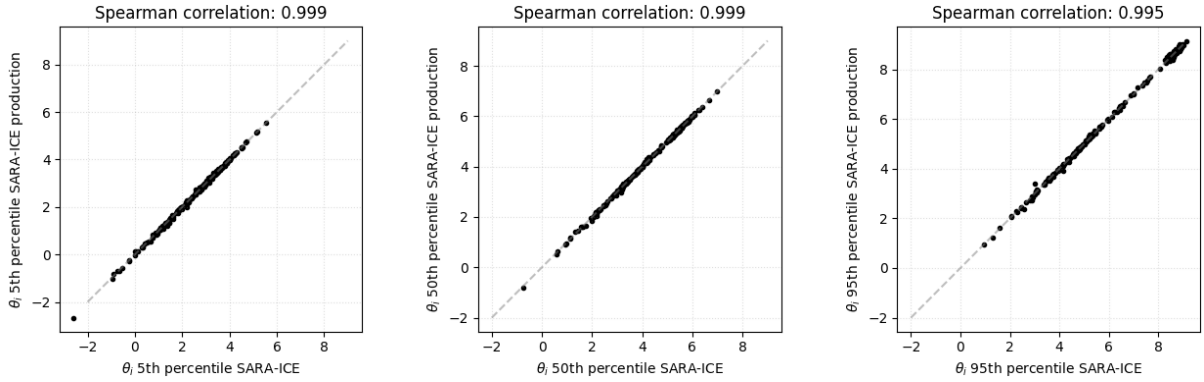
Black points indicate regular inputs, upward pointing triangles indicate right-censored inputs and downward facing triangles indicate left-censored inputs. Correlation is measured using Spearman's ρ due to the "kinks" induced by the typical censoring points.



Accuracy of the SARA-ICE production model

The accuracy of the SARA-ICE production model was checked by comparing ED_{01} estimates generated using the full model against those generated using the production model. The SARA-ICE model was fit to the full database and the 5th, 50th and 95th percentiles of the marginal posterior distribution of each θ_i, $i = 1, \ldots, 434$ were computed using a sample of 10,000 posterior draws. For each chemical in the database, we extracted the chemical-specific information and estimated $\widehat{\theta_i}$ from the SARA-ICE production model. The 5th, 50th and 95th percentiles of each $\widehat{\theta_i}$ were computed. Percentile estimates from both versions of the model are compared in **Error! Reference source not found.**. Correlation between percentile estimates is very close to 1 indicating both models for all practical purposes result in the same estimates.

Figure 7. Comparisons of percentiles (5th, 50th and 95th) of θ_i when estimated using SARA-ICE versus SARA-ICE production.



References

1. Politano, V.T. and A.M. Api, *The Research Institute for Fragrance Materials' human repeated insult patch test protocol.* Regul Toxicol Pharmacol, 2008. **52**(1): p. 35-8.
2. Kligman, A.M., *The identification of contact allergens by human assay: III. The maximization test: A procedure for screening and rating contact sensitizers.* Journal of Investigative Dermatology, 1966. **47**(5): p. 393-409.
3. Kligman, A.M. and W. Epstein, *Updating the maximization test for identifying contact allergens.* Contact Dermatitis, 1975. **1**(4): p. 231-239.
4. OECD *Supporting Document to the OECD Guideline 497 on Defined Approaches for Skin Sensitisation*. 2021.
5. Ryan, C.A., et al., *Extrapolating local lymph node assay EC3 values to estimate relative sensitizing potency.* Cutaneous and ocular toxicology, 2007. **26**(2): p. 135-145.
6. Hoffmann, S., et al., *Expansion of the Cosmetics Europe skin sensitisation database with new substances and PPRA data.* Regulatory Toxicology and Pharmacology, 2022. **131**: p. 105169.
7. Reynolds, J., et al., *Decision making in next generation risk assessment for skin allergy: Using historical clinical experience to benchmark risk.* Regulatory Toxicology and Pharmacology, 2022. **134**: p. 105219.
8. UN, *Globally Harmonised System of Classification and Labelling of Chemicals (GHS)*, U.N.N.Y.a. Geneva, Editor. 2021.
9. Carpenter, B., et al., *Stan: a probabilistic programming language.* Grantee Submission, 2017. **76**(1): p. 1-32.
10. Lewandowski, D., D. Kurowicka, and H. Joe, *Generating random correlation matrices based on vines and extended onion method.* Journal of Multivariate Analysis, 2009. **100**(9): p. 1989-2001.
11. OECD, *Annex 4: Report Of The Human Data Sub-Group On The Curation And Evaluation of the Human reference Data and the Derivation of Associated Substance Classifications. 2021*.