Research Article

Integrated Skin Sensitization Assessment Based on OECD Methods (I): Deriving a Point of Departure for Risk Assessment

Andreas Natsch¹ and George Frank Gerberick²
¹Fragrances S&T, Ingredients Research, Givaudan Schweiz AG, Kemptthal, Switzerland; ²GF3 Consultancy, LLC, Cincinnati, OH, USA

Abstract
Three guidelines covering key events in the skin sensitization adverse outcome pathway are endorsed by the Organisation for Economic Co-operation and Development (OECD). A recent guideline covers Defined Approaches (DA) to combine data from these tests for regulatory (sub)-classification. These methods provide continuous data that could characterize the sensitization potency on a more granular scale beyond (sub)-classifications. We assembled a comprehensive database on in vitro and in vivo tests in OECD methods. Building on a previous approach using regression models, we provide quantitative models using input data from the kinetic Direct Peptide Reactivity Assay (kDPRA), the KeratinoSens™ (KS) assay and the human Cell Line Activation Test (h-CLAT) to calculate a point of departure (PoD) in the form of a predicted Local Lymph Node Assay (LLNA) EC3 value for use in risk assessment. Predictive models include results from either two or all three assays. Detailed analysis vs. in vivo data estimates redundancy between different tests and helps guide model selection. All models were tested on a set of case studies selected on their availability of multiple LLNA reference data in the OECD database. The predicted PoD were within or close to the area of the variation of the historical LLNA data for most of these cases studies, and overall, the models predict the in vivo value with a median fold-misprediction of a factor of around 2.5. The robustness of the models was characterized by comparing a comprehensive historical database vs. the curated dataset provided by the OECD working group on DA.

1 Introduction

The progress achieved in developing New Approach Methods (NAM) for assessing the skin sensitization potential of chemicals over the last two decades has been remarkable. The OECD led the development of the adverse outcome pathway for skin sensitization divided into mechanistic key events (OECD, 2014). Three OECD guidelines have been published that cover these mechanistic events (covalent binding to protein, keratinocyte activation, and dendritic cell activation). There are a total of eight non-animal test methods approved in OECD TG 442C, 442D, and 442E (OECD, 2018a,b, 2021d). However, none of the eight test methods are considered standalone replacements for complete hazard identification or potency determinations.

The current focus is to find ways to combine in vitro, in chemico and in silico assessments with read-across predictions from similar chemicals, to generate integrated approaches to testing and assessment (IATA) or DA. The OECD has published a new guideline (No. 497) that describes two simple DA for assessing skin sensitization (OECD, 2021a). This strategy of combining NAM data will enhance the vigour of skin sensitization hazard identification calls on chemicals and further address the critical need for predicting a chemical’s potency. Potency prediction is needed for the UN Global Harmonized System (GHS) subclassification of sensitizers into 1A (strong sensitizers) and 1B (other sensitizers). For the specific need of GHS subclassification into sub-category 1A, the kDPRA assay is accepted as a standalone assay (Natsch et al., 2020; OECD, 2021d). However, assessing potency is also required for conducting Next Generation Risk Assessments on new chemical entities where only non-animal information is available (Api et al., 2020; Bernauer et al., 2021; Dent et al., 2018; Gilmour et al., 2020). Potency assessment for risk assessment needs to be more granular than the GHS sub-classification, and it is expected that this granularity can only be provided by combining multiple methods. Such potency information is critical to ensure any new chemical is safe for exposed workers and consumers.

Multiple studies have addressed the application of quantitative in silico, in chemico, and in vitro data from the validated assays for potency prediction (Gilmour et al., 2020; Kleinstreuer et al., 2018; Natsch et al., 2018; OECD, 2021a). A few NAMs have been designed to predict potency, including the SENS-IS (Cottrez et al., 2015, 2016) the Genomic Allergen Rapid Detection (Zeller et al., 2017) and the kDPRA (Wareing et al., 2017). DA using data from NAMs have been shared for predicting potency, including a Bayesian network approach (Jaworska et al., 2015), regression models with KS and peptide reactivity data (Natsch et al., 2015) artificial neural network model (Hirotta et al., 2015, 2018) and integrated use of h-CLAT, DPRA and...
DEREK data (Takenouchi et al., 2015). It will be essential for risk assessors and regulators to evaluate these different approaches to identify which DA or individual methods provide an accurate PoD value for conducting sound risk assessments.

One promising approach based on generating linear regression models has been published using KS and kinetic peptide reactivity data to provide a predicted EC3 as a PoD (Natsch et al., 2015, 2018). Predicting an EC3 value offers the advantage of generating continuous potency values compared to predicting a chemical potency class (Cottrez et al., 2015; Jaworska et al., 2015; Zang et al., 2017; Zeller et al., 2017). It also provides the opportunity to manage uncertainty using statistical tools based on knowledge of the accuracy of the prediction. Such uncertainty could be factored in to refine the PoD value for conducting a skin sensitization risk assessment. The determination of potency has been primarily dependent on the use of the LLNA (Lovelless et al., 2010; OECD, 2010). The LLNA has long been considered the “gold standard” for potency assessment because it yields quantitative data suitable for a dose-response evaluation. An alternative, non-animal approach is urgently needed.

The previous work on regression models (Natsch et al., 2015) used kinetic rate constants generated with the Cor1-C420 assay (Natsch and Gfeller, 2008). In this paper, updated linear regression models based on data from OECD validated methods (kDPRA, KS, and/or h-CLAT) are described for predicting a PoD value for risk assessment purposes. A comprehensive database of 322 chemicals was assembled that contains data from previous papers (Natsch et al., 2015, 2020) merged with the database put together within the framework of the OECD project on DA (OECD, 2021c). The report addresses (1) contribution of the in vitro parameters for predicting LLNA potency, (2) comparison of prediction models based on an inclusive dataset versus the highly curated OECD dataset, (3) guidance in model selection with multiple input data, (4) comparison of kDPRA and Cor1-Cor420 reactivity data, and (5) case studies on chemicals with multiple LLNA EC3 values. This work further advances the 3R for skin sensitization testing as a standardized way to derive a PoD from validated methods is still a missing element in the application of NAM for sensitization assessment.

2 Materials and methods

2.1 Data sources of existing data

All in vivo and in vitro data are from a previous publication (Natsch et al., 2015) or the data compilation by Urbisch et al. (Urbisch et al., 2015) and Jaworska et al. (2015) and from the database compiled by the OECD group on Defined Approaches (OECD, 2021c). The compiled database (ESM1-1) contains data from the DRPA and the kDPRA in OECD TG 442C, KS assay (TG 442D), h-CLAT (TG 442E), LLNA data (GL 429) and data from the Cor1-C420 reactivity assay (Natsch and Gfeller, 2008). All the individual parameters are described in the supporting document.

2.2 Data transformation and normalization for statistical analysis

All data were Log-transformed and normalized as described before (Natsch et al., 2015). In case multiple LLNA data were available, we took the geometric mean of the LLNA EC3 values. Data are expressed as pEC3, a logarithmic expression taking molecular weight into account since most in vitro data are also expressed in molar concentrations and not on a per weight basis as in the LLNA.

\[ pEC3 = \log\left(\frac{MW}{EC3}\right) \]

For negative LLNA results, the pEC3 was set to zero, which is the special case for e.g., molecules with MW of 100 and an EC3 of 100%. This approach treats all negatives equally, as not all negatives were tested up to 100%. Based on this normalization, the pEC3 spans a range from 0 for non-sensitizers to 4.86 for the potent sensitizer oxazoline. In the KS assay, the IC50 value (concentration for 50% reduction in cellular viability), and the EC1.5 or EC3 values, i.e., concentration for 1.5- or 3-fold luciferase induction are determined. Chemicals are tested up to a maximal concentration of 2000 µM. For chemicals with no cytotoxicity at this concentration, the numerical IC50 was set to the arbitrary value of 4000 µM. Similarly, if the luciferase gene is not induced above a given threshold, the EC1.5 or EC3 values were set to 4000 µM. In addition, in the KS prediction model only gene induction observed at non-cytotoxic levels (> 70% viability) is considered relevant. Thus if 1.5-fold gene induction is only observed at cytotoxic levels, then EC1.5 and EC3 were set to 4000 µM too. The three parameters were linearized by log transformation and normalized by multiplication by -1 and the addition of the constant Log(4000).

Thus, as an example, the normalized IC50 is:

\[ \log(C50_{norm}) = -1 \times \log((C50_{KS}) + \log(4000)) \]

These three normalized KS parameters span a linear range from 0 (no effect at 2000 µM) to 3.61 (effect at lowest tested concentration of 0.98 µM).

In the h-CLAT, four quantitative parameters are determined: The CV75 indicates the concentration for 25% reduction in cellular viability, the EC150 indicates the concentration for 1.5-fold induction of CD86 surface expression and the EC200 indicates the concentration for 2-fold induction of CD54 surface expression. The MIT reports the smaller of the two values from the EC150 and the EC200. Most chemicals are tested up to 5000 µg/ml, but in some cases only up to 1000 or 2000 (Nukada et al., 2012). H-CLAT data are reported in µg/ml and they were thus transformed to µM, negative values were set to the arbitrary value of 25000 µM (i.e., 5000 µg/ml for a chemical with a default molecular weight of 200) and normalization was done in the same way. As an example, the normalized MIT value is:

\[ \log(MIT_{norm}) = -1 \times \log((MIT_{µM}) + \log(25000)) \]

Abbreviations: DA, Defined Approach; DPRA, Direct Peptide Reactivity Assay; GHS, Global Harmonized System; h-CLAT, Human Cell Line Activation Test; IATA, Integrated Approaches to Testing and Assessment; kDPRA, kinetic Direct Peptide Reactivity Assay; KS, KeratinoSens tests; LLNA, Local Lymph Node Assay; NAM, New Approach Methodologies; PoD, Point of Departure; OECD, The Organisation for Economic Co-operation and Development; OECD DB, OECD reference database on defined approaches; TG, Test Guideline; VP, Vapor Pressure

DOI: 10.14573/altex.2201141s1
In the kDPRA, the reaction rates are measured at different time points, and the logarithm of the maximal rate in s\(^{-1}\)M\(^{-1}\) is reported (\(\log k_{\text{max}}\)). The minimal value for chemicals negative in the assay, i.e., reaching less than 13.86% depletion of the Cysteine peptide after 24 h is < - 3.5. These values were thus normalized according to the following formula:

\[
\log k_{\text{max, norm}} = \log k_{\text{max}} + 3.5
\]

With these linear transformations, all \textit{in vitro} parameters are set to 0 in case a molecule is non-reactive in kDPRA, non-toxic beyond the selected threshold in the cell-based assays or it does not induce the cellular markers (luciferase or CD86/CD54 induction), while all \textit{in vitro} evidence according to the prediction models of the individual assays has a positive prefix. These linear transformations of logarithmic data do not distort the data but make reading the regression equations easier.

Vapor pressure was calculated with the TIMES SS software (Laboratory of Mathematical Chemistry). Vapor pressure is considered important as chemicals are tested in the LLNA with open application, while evaporation is limited in the \textit{in vitro} assays. Relationship of calculated vapor pressure to evaporation from the LLNA vehicle acetone / olive oil 4:1 (AOO) at typical skin surface temperature of 32°C was tested experimentally for 37 molecules and a linear relationship between \(\log (VP_{\text{lab}})\) and \(\log \text{half-life of the chemical in AOO}\) was demonstrated as shown in Figure S2 in Natsch et al. (Natsch et al., 2015). Chemicals with a \(\log (VP_{\text{lab}}) < 10\) Pa were found not to evaporate from AOO within 60 min significantly. Therefore, vapor pressure data were normalized as follows:

\[
\log VP_{\text{norm}} = \log (VP) - 1
\]

and values below zero are set to 0. With this approach, chemicals predicted to evaporate significantly from mouse ears within 60 min have a positive coefficient, whereas the coefficient for chemicals with low volatility (i.e., not significantly evaporating within 60 min) is set to 0.

2.3 Statistical analysis

Multiple linear regression against the pEC3 was performed with the MiniTab 15 software. The R\(^2\)-values and F- and p-value of the correlations are given throughout the manuscript as measures of correlation strengths. The regression equations were used in a method similar to the kDPRA but with the more reactive peptide Cor1-C420, as reported earlier. This database is merged from our previous database (Natsch et al., 2015) with the database used to validate the kDPRA (Natsch et al., 2020) and the database put together within the frame of the OECD project on DA for skin sensitization in the Series on Testing and Assessment No. 336, Annex 2 (OECD, 2021c). It also contains additional h-CLAT and DPRA data from Urbisch et al. (2015) and Jaworska et al. (2015). Table 1 indicates the different subsets of this database used.

In the analysis below, we present regression analysis on these different datasets, and we provide a Prediction Spread of models using h-CLAT instead of (or in combination with) KS

### 3 Results

#### 3.1 A comprehensive database of \textit{in vitro} and \textit{in vivo} data

We compiled a comprehensive database on 322 chemicals (Table ESM1\(^2\)) with data from the OECD validated LLNA, KS (KS), h-CLAT, DPRA and kDPRA. In addition, this database contains physicochemical information and reaction rates measured in a method similar to the kDPRA but with the more reactive peptide Cor1-C420, as reported earlier. This database is merged from our previous database (Natsch et al., 2015) with the database used to validate the kDPRA (Natsch et al., 2020) and the database put together within the frame of the OECD project on DA for skin sensitization in the Series on Testing and Assessment No. 336, Annex 2 (OECD, 2021c). It also contains additional h-CLAT and DPRA data from Urbisch et al. (2015) and Jaworska et al. (2015). Table 1 indicates the different subsets of this database used.

In the analysis below, we present regression analysis on these different datasets, and we provide a Prediction Spread with kDPRA on subset of previous publication

### Tab. 1: Subsets of chemicals in the dataset used for the different evaluations on the robustness of predictive models

<table>
<thead>
<tr>
<th>Data subset</th>
<th>n</th>
<th>Data available for all chemicals in a subset</th>
<th>Used for</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full dataset</td>
<td>322</td>
<td>LLNA, KS, Cor1-C420 assay</td>
<td>Overall analysis in ESM4(^5)</td>
</tr>
<tr>
<td>Full dataset with kDPRA data</td>
<td>203</td>
<td>LLNA, KS, kDPRA, Cor1-C420 assay</td>
<td>Most comprehensive model integrating kDPRA and KS</td>
</tr>
<tr>
<td>Full dataset with kDPRA and h-CLAT data</td>
<td>188</td>
<td>LLNA, KS, kDPRA, Cor1-C420 assay, h-CLAT</td>
<td>Models using h-CLAT instead of (or in combination with) KS</td>
</tr>
<tr>
<td>OECD LLNA dataset</td>
<td>149</td>
<td>LLNA data from previous database, LLNA data curated by OECD DA group, KS, kDPRA, Cor1-C420 assay, h-CLAT</td>
<td>Comparison of models derived from OECD dataset with models on more comprehensive datasets; evaluation of the impact of data curation</td>
</tr>
<tr>
<td>Core set 2015</td>
<td>244</td>
<td>LLNA, KS, Cor1-C420 assay</td>
<td>For direct comparison with the 2015 analysis in ESM4(^5)</td>
</tr>
<tr>
<td>Core set 2015; subset with kDPRA data</td>
<td>158</td>
<td>LLNA, KS, Cor1-C420 assay, kDPRA</td>
<td>Predictivity with kDPRA on subset of previous publication</td>
</tr>
</tbody>
</table>

#### 3.2 A quantitative global regression model integrating kDPRA data with KS data

For a set of 203 chemicals, the database contains kDPRA and KS data. For this dataset, a regression model based on the normalized EC1.5 and IC50 from KS, \(\log k_{\text{max}}\) from the kDPRA and normalized vapor pressure was derived, similar to the global model in the 2015 study (EQ1), but just replacing reaction rates with the Cor1-C420 peptide with data from the validated kDPRA.

\[
\text{EQ1: } p\text{EC}3 = 0.42 + 0.40 \times \log k_{\text{max, norm}} + 0.15 \times \log \text{EC1.5, norm} + 0.36 \times \log \text{IC50, norm} - 0.21 \times \log VP_{\text{norm}}
\]

\(^2\) doi:10.14573/altex.2201141s2
Constant T = 4.61, p < 0.0005
Log Kmax,norm T = 9.1, p < 0.0005
Log EC1.5,norm T = 2.5, p = 0.013
Log IC50,norm T = 4.47, p < 0.0005
Log VP,norm T = -3.44, p = 0.001
n = 203; S = 0.73; R² = 62.0%; R² (adj) = 61.3%

In this equation, kmax from kDPRA and IC50 from KS have the highest statistical weight. This equation is derived from the most comprehensive dataset with available kDPRA data. EQ1 is therefore proposed as the key model in the Prediction Spreadsheet (ESM2) for chemicals with available kDPRA and KS data only.

For comparison, the same regression analysis was performed on the subset of data from the core set used in the 2015 study with kDPRA data (n = 158; EQ2). Similarly, the same model was also calculated on the subset of data with available hCLAT data. The latter can then be directly (i.e., based on the same chemicals) compared to the models below integrating hCLAT data (n = 188; EQ3).

EQ2 pEC3 = 0.34 + 0.38 × Log kmax,norm + 0.16 × Log EC1.5,norm + 0.42 × Log IC50,norm - 0.18 × Log VP,norm
n = 158; S = 0.72; R² = 63.0%; R² (adj) = 61.2%
EQ3 pEC3 = 0.40 + 0.40 × Log kmax,norm + 0.14 × Log EC1.5,norm + 0.38 × Log IC50,norm - 0.21 × Log VP,norm
n = 188; S = 0.75; R² = 58.4%; R² (adj) = 57.5%

The resulting equations are very similar to EQ1, indicating that the chemicals added to the core set used previously have no major impact on the overall predictive model and that the model is relatively stable when used with different subsets of the input data.

3.3 A quantitative global regression model integrating kDPRA data with hCLAT data
Next, a similar model was calculated using only data from the kDPRA and hCLAT to predict the potency of chemicals with these two sources of data available, thus normalized EC1.5 and IC50 were replaced with the normalized MIT and the CV75 from hCLAT, both calculated in µM (n = 188; EQ4).

EQ4 pEC3 = 0.18 + 0.36 × Log kmax,norm + 0.21 × Log MIT,norm + 0.35 × Log CV75,norm - 0.19 × Log VP,norm
Constant T = 1.65, p = 0.102
Log kmax,norm T = 7.59, p < 0.0005
Log MIT,norm T = 3.06, p = 0.004
Log CV75,norm T = 3.69, p < 0.0005
Log VP,norm T = -3.0, p = 0.003
n = 188; S = 0.72; R² = 61.1%; R² (adj) = 60.3%

Compared to EQ3, on the same dataset, kmax has a similar weight in this equation, and MIT and CV75 can replace the KS data with a very similar overall statistical weight. EQ4 is thus added to the Prediction Spreadsheet to be used for those chemicals for which only kDPRA and hCLAT data are available.

3.4 A quantitative global regression model integrating kDPRA data with KS and hCLAT data
For some chemicals, comprehensive data from all three OECD TG will be available. Thus, the most comprehensive model integrates kDPRA, KS and hCLAT data (n = 188; EQ5).

EQ5 pEC3 = 0.20 + 0.34 × Log kmax,norm + 0.20 × Log MIT,norm + 0.09 × Log EC1.5,norm + 0.21 × Log CV75,norm + 0.11 × Log IC50,norm - 0.19 × Log VP,norm
Constant T = 1.98, p = 0.080
Log kmax,norm T = 6.88, p < 0.0005
Log EC1.5,norm T = 1.41, p = 0.159
Log IC50,norm T = 0.94, p = 0.347
Log MIT,norm T = 2.74, p = 0.004
Log CV75,norm T = 1.63, p = 0.119
Log VP,norm T = -2.84, p = 0.005
n = 188; S = 0.72; R² = 61.6%; R² (adj) = 60.6%

The predictive power of EQ5 is only very marginally improved vs. EQ3 and EQ4, indicating that with both KS and hCLAT data, the overall predictivity does only slightly improve as compared to the situation when only one cellular assay is available. This has already been observed in our previous comprehensive analysis (Natsch et al., 2020), indicating significant data redundancy between these two cell-based assays. This is especially the case for the cytotoxicity readout. When both IC50 and CV75 are used in EQ5, they contribute additively, but the coefficient and statistical parameters indicate a weak contribution of each parameter. Alternative models using either CV75 or IC50 alone have the same R² value (61%), but in both cases, the contribution of the cytotoxicity parameter has more statistical weight (p = 0.014 for IC50 and p = 0.005 for CV75; regression equations not shown), which confirms data redundancy especially for the two cytotoxicity indicators, which is not a surprising finding.
3.5 A quantitative global regression model integrating KS and h-CLAT data in the absence of kDPRA data

Finally, in some cases, chemicals are rated as positive in a DA based on a positive result in the cell-based assays only despite a negative result in the kDPRA and the DPRA. For these and complex extracts or multi-constituent substances with data gaps in the kDPRA due to incompatibility with the assay, a model based on KS and h-CLAT data only may be of interest. This model is presented in Equation 6.

EQ6 \[ p_{EC3} = 0.09 + 0.276 \times \log \text{MIT}_{\text{norm}} + 0.22 \times \log \text{EC1.5}_{\text{norm}} + 0.34 \times \log \text{CV75}_{\text{norm}} + 0.06 \times \log \text{IC50}_{\text{norm}} - 0.12 \times \log \text{VP}_{\text{norm}} \]

Constant \( T = 0.67, p = 0.50 \)

Log \( \text{EC1.5}_{\text{norm}} \) \( T = 3.22, p = 0.002 \)

Log \( \text{IC50}_{\text{norm}} \) \( T = 0.43, p = 0.671 \)

Log \( \text{MIT}_{\text{norm}} \) \( T = 3.75, p < 0.0005 \)

Log \( \text{CV75}_{\text{norm}} \) \( T = 2.29, p = 0.025 \)

Log \( \text{VP}_{\text{norm}} \) \( T = -1.81, p = 0.072 \)

\( n = 188; S = 0.81; R^2 = 51.8\%; R^2 (adj) = 50.5\% \)

All the above models (EQ1, 4 and 5) were also calculated with KS EC3 derived from KS, i.e., the concentration for threefold stimulation of the luciferase signal in KS, instead of EC 1.5 and models with very similar \( R^2 \) were obtained (data not shown). However, in the situation without reactivity data, EQ7 using KS EC3 has improved statistical power, with KS EC3 having the highest statistical weight.

EQ7 \[ p_{EC3} = 0.202 + 0.222 \times \log \text{MIT}_{\text{norm}} + 0.40 \times \log \text{EC3}_{\text{norm}} + 0.313 \times \log \text{CV75}_{\text{norm}} + 0.023 \times \log \text{IC50}_{\text{norm}} - 0.151 \times \log \text{VP}_{\text{norm}} \]

Constant \( T = 1.67, p = 0.098 \)

Log \( \text{EC3}_{\text{norm}} \) \( T = 5.29, p < 0.0005 \)

Log \( \text{IC50}_{\text{norm}} \) \( T = 0.18, p = 0.855 \)

Log \( \text{MIT}_{\text{norm}} \) \( T = 3.02, p = 0.003 \)

Log \( \text{CV75}_{\text{norm}} \) \( T = 2.28, p = 0.024 \)

Log \( \text{VP}_{\text{norm}} \) \( T = -2.28, p = 0.023 \)

\( S = 0.77; R^2 = 55.9\%; R^2 (adj) = 54.7\% \)

These models without kDPRA data have less statistical power, but they would primarily be used either in the case of data gaps for the kDPRA, for chemicals that are negative in the kDPRA, or chemicals outside of the applicability domain of the kDPRA (See ESM3). These latter chemicals, in most cases, are weak or moderate sensitizers, the majority of strong sensitizers having a direct peptide reactivity in the kDPRA and DPRA (Natsch et al., 2020).

3.6 Overall predictive capacity of the different models

Equations 1, 4, 5 and 6 (with 7 as an alternative) are implemented in the Prediction Spreadsheet (ESM2) to predict potency of new chemicals, depending on the set of available in vitro data. The overall predictive capacity of these models is summarized in Table 2. For all models, the median of the fold-misprediction is around 2.5-fold, while the geometric mean is around 3.3-fold.

Tab. 2: Predictivity on the dataset (\( n = 188 \)) with KS, h-CLAT and kDPRA data

<table>
<thead>
<tr>
<th>Model</th>
<th>Input parameters</th>
<th>Fold-misprediction* (geomean)</th>
<th>Fold-misprediction (median)</th>
<th>Chemicals &gt; 5-fold under-predicted* n, (%)</th>
<th>Chemicals &gt; 10-fold under-predicted n, (%)</th>
<th>Chemicals &gt; 5-fold over-predicted n, (%)</th>
<th>Chemicals &gt; 10-fold over-predicted n, (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQ1</td>
<td>kDPRA, KS</td>
<td>3.3</td>
<td>2.5</td>
<td>33 (18%)</td>
<td>20 (11%)</td>
<td>16 (9%)</td>
<td>7 (4%)</td>
</tr>
<tr>
<td>EQ4</td>
<td>kDPRA, h-CLAT</td>
<td>3.2</td>
<td>2.4</td>
<td>30 (16%)</td>
<td>17 (9%)</td>
<td>16 (9%)</td>
<td>7 (4%)</td>
</tr>
<tr>
<td>EQ5</td>
<td>kDPRA, KS, h-CLAT</td>
<td>3.1</td>
<td>2.3</td>
<td>35 (19%)</td>
<td>17 (9%)</td>
<td>18 (10%)</td>
<td>6 (3%)</td>
</tr>
<tr>
<td>EQ6</td>
<td>KS, h-CLAT</td>
<td>3.5</td>
<td>2.6</td>
<td>33 (18%)</td>
<td>19 (10%)</td>
<td>19 (11%)</td>
<td>8 (4%)</td>
</tr>
<tr>
<td>EQ7</td>
<td>KS, h-CLAT</td>
<td>3.4</td>
<td>2.7</td>
<td>31 (16%)</td>
<td>19 (10%)</td>
<td>18 (10%)</td>
<td>6 (3%)</td>
</tr>
<tr>
<td>Global model 2015</td>
<td>Cort-C420 assay, KS</td>
<td>3.6</td>
<td>2.7</td>
<td>39 (21%)</td>
<td>24 (13%)</td>
<td>19 (10%)</td>
<td>4 (2%)</td>
</tr>
</tbody>
</table>

* The ratio between the higher and the lower values of the measured and predicted EC3 value. Predicted EC3 > 100% were set to 100%.

* Under-predicted chemicals: those for which the measured LLNA EC3 is < than the predicted EC3; over-predicted chemicals: Those with measured LLNA EC3 > than the predicted value.

4 doi:10.14573/altex.2201141s3
3.7 Models trained on the subset of data in the OECD database

Very recently, the data curated by the OECD expert group on DA has been made publicly available (OECD, 2021c), and this database will probably be used as a gold-standard in the future. This database contains data on a total of 196 chemicals, but only for 154 a discrete LLNA EC3 value or an unambiguous rating as non-sensitizer by the LLNA is available, as for some chemicals, no LLNA data are included, or the LLNA data were not considered sufficient to conclude on an EC3 value by the data curation team. For 149 of these 154 chemicals, data are available in our database on kDPRA, KS and h-CLAT. This dataset is interesting to interrogate two questions:

a) Does the smaller data subset collected by the OECD group lead to significantly different equations compared to those presented above on a more comprehensive set?

b) Do the LLNA EC3 values after the data curation lead to significantly different predictive equations compared to the LLNA data in the previously published databases?

Thus, in Table 3, comparisons of the model parameters are provided as calculated for the chemicals in the OECD database only (n=149) vs. (i) published LLNA data and (ii) curated LLNA data and these models are compared to the models established on the more comprehensive sets of chemicals (EQ1, EQ4 and EQ5).

Overall, the additional models (EQ 8 – EQ13) derived from the subset in the OECD database are similar to EQ1, EQ4 and EQ5, integrating the larger datasets of all available data for a given data combination. Especially, only minor differences are observed between the models on the historical and the curated LLNA data (EQ 8 vs. 9 /EQ10 vs. 11/ EQ12 vs. 13).

### Tab. 3: Regression coefficients and statistics for predictive models implemented in the Prediction Spreadsheet compared to models derived from the OECD dataset only

<table>
<thead>
<tr>
<th>Model</th>
<th>Con-</th>
<th>kmax</th>
<th>EC1.5</th>
<th>IC50</th>
<th>MIT</th>
<th>CV75</th>
<th>Vp</th>
<th>R²</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQ1 (KS+kDPRA)</td>
<td>0.42</td>
<td>0.40</td>
<td>0.15</td>
<td>0.36</td>
<td>-0.21</td>
<td>62%</td>
<td>203</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KS+kDPRA, OECD set, published LLNA values (EQ 8)</td>
<td>0.45</td>
<td>0.32</td>
<td>0.26</td>
<td>0.25</td>
<td>-0.20</td>
<td>55%</td>
<td>149</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KS+kDPRA, OECD set, curated LLNA value (EQ 9)</td>
<td>0.50</td>
<td>0.33</td>
<td>0.21</td>
<td>0.29</td>
<td>-0.25</td>
<td>54%</td>
<td>149</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EQ4 (h-CLAT+kDPRA)</td>
<td>0.18</td>
<td>0.36</td>
<td>0.21</td>
<td>0.35</td>
<td>-0.19</td>
<td>61%</td>
<td>188</td>
<td></td>
<td></td>
</tr>
<tr>
<td>h-CLAT+kDPRA, OECD set, published LLNA values (EQ10)</td>
<td>0.24</td>
<td>0.31</td>
<td>0.22</td>
<td>0.34</td>
<td>-0.16</td>
<td>57%</td>
<td>149</td>
<td></td>
<td></td>
</tr>
<tr>
<td>h-CLAT+kDPRA, OECD set, curated LLNA values (EQ11)</td>
<td>0.28</td>
<td>0.30</td>
<td>0.19</td>
<td>0.37</td>
<td>-0.22</td>
<td>57%</td>
<td>149</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EQ5 (KS+h-CLAT+kDPRA)</td>
<td>0.20</td>
<td>0.34</td>
<td>0.09</td>
<td>0.11</td>
<td>0.20</td>
<td>-0.19</td>
<td>62%</td>
<td>188</td>
<td></td>
</tr>
<tr>
<td>KS+h-CLAT+kDPRA, OECD set, published LLNA values (EQ12)</td>
<td>0.23</td>
<td>0.26</td>
<td>0.21</td>
<td>0.03</td>
<td>0.22</td>
<td>-0.17</td>
<td>59%</td>
<td>149</td>
<td></td>
</tr>
<tr>
<td>KS+h-CLAT+kDPRA, OECD set, curated LLNA values (EQ13)</td>
<td>0.27</td>
<td>0.26</td>
<td>0.15</td>
<td>0.00</td>
<td>0.19</td>
<td>-0.22</td>
<td>58%</td>
<td>149</td>
<td></td>
</tr>
</tbody>
</table>

3.8 Practical application: An open spreadsheet for calculating a PoD and case studies

To further facilitate application, here we provide a Prediction Spreadsheet as ESM2. Users can simply enter the experimental data from the different in vitro assays along with the molecular weight and a calculated vapor pressure to calculate a predicted EC3, which can be used as a PoD and, if applying adequate safety factors and taking into account all available information and applicability domain limitations, for a quantitative risk assessment. To illustrate its use, the Prediction Spreadsheet also contains the input data for two examples, DNBC and cinnamic aldehyde.

For all chemicals, additional spreadsheets in ESM1 (Tables ESM1-2 – ESM2-4) give the individual predictions by the different equations on the different data subsets shown in Table 1. Here we show individual data on key case studies (Table 4). We selected all chemicals for which the OECD LLNA database provides n ≥ 5 individual EC3 values. Selecting these chemicals as case studies has two advantages: (i) The overall weight of evidence for the LLNA EC3 is strong as it comes from multiple independent tests and (ii) for these chemicals, we have a good understanding of the variability of the target, i.e., the range of LLNA EC3 measured, and thus the predictions can also be compared to this intrinsic variability of the prediction target.

For most of the 16 case studies, the predicted EC3 values fall within the experimental variability of the LLNA or are close to this range. In general, the predictions based on kDPRA and/or cell-based assay are quite close for these chemicals.

3.9 Model selection with multiple input data

Here we show how different combinations of source data from OECD TG 442C, 442D and 442E can be used in different regression equations. Depending on the available data, a user is faced with two critical questions:

a) Once data from two positive tests are available, is it worth collecting data from the third test for a more accurate potency assessment?

b) If the available data allows calculating a predicted PoD with different equations, which result should be taken forward?

The detailed data analysis to test these questions is presented in ESM3, and a summary is given here. If two positive results are available, including kDPRA data, studying the third key event, in general, is not needed because (i) if the third outcome would be negative, the output of the two positive tests can directly be used with the corresponding equation EQ1 or EQ4 and if (ii) the third outcome would be positive, either equation EQ1, EQ4 or EQ5 could then be used, but these would lead to very similar predictions. Thus, for the 73 chemicals with three positive tests, the correlation between the predicted EC3 from EQ1 (KS + kDPRA) and EQ5 (KS, h-CLAT and kDPRA) has an R² value of 0.95 with a slope close to 1 and y-intercept close to zero. Similarly, the correlation between EQ4 (h-CLAT and kDPRA) and EQ5 has an R² value of 0.99. Thus adding the third test and moving from EQ1 or EQ4 to EQ5 has little effect on the final prediction.
Thus, based on the measured Cor1-C420 data, we calculated predicted kDPRA values based on EQ14 and used these data to train the same regression model on the same chemicals as in EQ1 (EQ15).

---

EQ15 \[ pEC3 = 0.40 + 0.49 \times \log k_{\text{max}} \text{norm} (\text{calculated from Corl-420 result}) + 0.12 \times \log \text{EC1.5norm} + 0.28 \times \log \text{IC50norm} - 0.22 \times \log VP_{\text{norm}} \]
\[ S = 0.73; R^2 = 61.5\%; R^2 (\text{adj}) = 60.7\% \]

This equation has a similar predictive capacity as EQ1 (calculated on the same dataset) and overall a similar contribution of the individual parameters, although contribution of the reactivity parameter is slightly higher and the contribution of KS slightly weaker in EQ15 vs EQ1. The reactivity rates for the Corl-420 assay presented in the full database are all presented as measured values and also transformed to predicted kDPRA \( k_{\text{max}} \) values according to EQ14. In another data column, data for measured kDPRA \( k_{\text{max}} \) are shown and data gaps were filled with the predicted kDPRA \( k_{\text{max}} \) values according to EQ14. Due to data redundancy between the two tests, these data can be used in the absence of a true kDPRA experimental value for any further modeling on the larger database, e.g., to perform domain-based assessments or quantitative read-across assessment based on data from the target and the read–across substance using all the chemicals in the full database (For details see ESM4).

4 Discussion

This paper presents global regression models, generated from different NAM datasets, for producing a PoD value for risk assessment purposes. The models are based on the use of OECD accepted NAMs that include the recently accepted kDPRA (OECD, 2021d) as well as the KS and h-CLAT (OECD, 2018a,b). A comprehensive database of 322 chemicals was assembled for this analysis (Natsch et al., 2015; 2020; OECD, 2021c) which may also be useful in further refinements or the development of alternative models.

4.1 Robustness of the models when compared to previous models

Previous work describing the development of regression models using NAM data used a more reactive peptide Corl-C420 (Natsch and Gfeller, 2008) for assessing reactivity (Natsch et al., 2015, 2018). With the update of the OECD guideline 442C (OECD, 2021d) that now includes the kDPRA, it made sense to evaluate its use for input into these regression models. A linear correlation is observed when comparing reaction constants from the kDPRA and Corl-C420 assays (see ESM4). Thus, the two reactivity assays can be used interchangeably. To illustrate this, using EQ14, measured Corl-C420 rate constants were used to calculate predicted kDPRA values for use in generating the same regression model as in EQ1. Not surprisingly, the equation generated (EQ15) has the same predicted capacity as EQ1 and overall, a similar contribution of the individual parameters, indicating similar information content of the two reactivity assays and a robustness of the overall approach.

The regression model using a comprehensive dataset with available kDPRA and KS data (n = 203) was based on the normalized EC1.5 and IC50 from KS, Log \( k_{\text{max}} \) from the kDPRA and normalized vapor pressure. This model (EQ1) is based on the highest number of chemicals compared to other models and is very similar to the global model published previously (Natsch et al., 2015), but replacing Corl-C420 reactivity data with kDPRA data.

4.2 Complementary and redundancy of the cell-based assays

A vital feature described in the paper is the flexibility in what input NAM data is used to make a PoD prediction. A model integrating kDPRA data with h-CLAT instead of KS showed a similar statistical weight using normalized MIT and CV75 data (n = 188; EQ4). This shows that either KS or h-CLAT data can be used to obtain comparable predictions. It is well known that having data from all three assays is not always possible due to assay compatibility factors (Kolle et al., 2019). However, in cases where data from all three OECD TG data is available, a comprehensive model can be obtained that integrates kDPRA, KS and h-CLAT data (n = 188; EQ5).

Interestingly, having data from both KS and h-CLAT along with kDPRA data only slightly improves the predictive power compared to one cellular assay. This redundancy has been observed previously (Natsch et al., 2020) and indicates that there is significant redundancy between the two cell-based assays. For situations where kDPRA data is absent, possibly due to an incompatibility issue using complex mixtures or technical issues (e.g., thiols), a regression model integrating KS and h-CLAT data is available (n = 188; EQ6), too. In the absence of reactivity data, the parameters for luciferase induction in KS or surface marker induction in h-CLAT receive more statistical weight. The best statistical power is obtained using KS EC3 instead of EC1.5 suggesting that this parameter for strong Nrf2-dependent luciferase induction can partially compensate for the lack of reactivity data (EQ7).

For each model based on a dataset of 188 chemicals with data from KS, h-CLAT or kDPRA, their predictive capacity is similar. For all models, the median of the fold-misprediction is around 2.5-fold while the geometric mean is around 3.3-fold (Table 2). This finding demonstrates the flexibility in the utility of using these models based on what test data is available. A description of how to select an appropriate model is provided in ESM3 and a parallel paper describes how the models can be combined with the “2 out of 3” DA (Natsch and Gerberick, 2022). The analysis shows that for chemicals with three positive tests, adding the third test, had little effect on the final prediction. Thus, a third key event study is not needed if two positive tests, including kDPRA data, are available. There is also not a significant difference whether starting with the kDPRA and either of the two cell-based assays. When multiple PoD values are available, the most predictive choice would be to use EQ5, integrating all three assays, although a different choice might be made when the chemical is a putative pro- or pre-hapten, in which case EQ6 might be optimal.

4.3 Assessment of using the comprehensive database vs. the curated OECD DA database

Very similar models to EQ 1 were obtained using the core set of data used in 2015 for which kDPRA data was available (n = 158; EQ2) and the set of chemicals with available h-CLAT data (n = 188; EQ3) indicating the model is robust and not strongly dependent on the training set used. A curated database developed by an OECD expert group has been made publicly available (OECD, 2021c). The database contains 196 chemicals, of which 154 contain LLNA data and 149 of them are available with
NAM data in our database. The regression coefficients for the models described (EQ1, EQ4, EQ5) are very similar to equivalent models derived from the OECD curated subset (Table 3). However, the contribution of reactivity from kDPRA tends to be lower. This may be attributed to the fact that the OECD group had stringent criteria to allow for extrapolation of LLNA results (OECD, 2021b). This led to the exclusion of LLNA data for several strong and extreme sensitizers, which are highly reactive chemicals and for which the reaction rate has a high weight in the assessment. On the other hand, the KS EC1.5 has a higher weight on this selected subset. When, for the same chemical set, the LLNA EC3 from previously published databases is used as the target as compared to the curated EC3 values (e.g., EQ 8 vs EQ 9), there is only a very small difference in the resulting regression equation, indicating that the data curation had no major impact on the overall picture. In addition, models established on a large number of chemicals are not inferior to the models built upon the OECD curated data (see results for EQ13 and EQ5 in Table 4), although the data curation certainly makes a significant difference for some individual chemicals. Going forward, it therefore appears appropriate to use models incorporating the maximal weight of evidence from the largest databases publicly available rather than only base models on the curated OECD data.

4.4 Transparency of the regression models
As compared to complex tools like neural networks or Bayesian nets, linear regression models may appear as relatively simple statistical tools. However, they have the advantage of high transparency. A Prediction Spreadsheet (ESM2) is provided that is designed to allow one to enter experimental data from the different NAMs along with molecular weight and a calculated vapor pressure to calculate a PoD value. This should facilitate practical application.

4.5 Key misprediction
For most of the 16 case studies, it is evident that the PoD values, based on using kDPRA and one or both cellular assays, are quite similar to the LLNA values based on ≥ 5 individual EC3 values. Clear outliers under-predicted by the models are 3-dimethylaminopropylamine and p-phenylenediamine (see discussion on applicability domain in ESM3). The extreme sensitizer oxazolone is predicted as a strong sensitizer by the models, but the predicted PoD value is clearly underestimated. The unique reactivity with lysine residues of oxazolone explaining its unsurpassed sensitization potential has been investigated before and is not captured by the kDPRA (Natsch et al., 2010). A more detailed discussion on mispredicted chemicals will be provided in a parallel paper.

4.6 Outlook
For risk assessors responsible for assessing the skin sensitization risk of new chemical entities or chemicals lacking sufficient data, it is critical to have tools available that are predominantly dependent on using NAM data. The PoD value obtained from these regression models can thus be used to assist in the conduct of skin sensitization risk assessments. The predictive regression models (EQ1, 4, 5, 6 and 7 as an alternate) have been built from a comprehensive database of 188 - 203 chemicals along with a comparative analysis using the curated OECD database (OECD, 2021c). They show similar statistical strength and prediction accuracy. Performance is similar when using different data input parameters or when comparing models generated from different datasets. The PoD derived from these models may, using appropriate assessment factors to account for uncertainty and taking all information into account, be used as a starting point to determine safe use levels in products. The application and guidance of how to use these regression models when using the “2 out 3” DA is covered in a separate paper.

Electronic supplementary material
ESM1 gives the full database in Sheet 1 and the predictions by the different models for the individual chemicals for different data subsets and vs. different LLNA datasets in Sheet 2 – 4.
ESM2 gives the predictive Spreadsheet to calculate the PoD from in vitro data.
ESM3 discusses choice of the different regression models based on test availability.
ESM4 compares predictivity of the different reactivity assays.

References
Cottrez, F., Boitel, E., Ourlin, J. C., Peiffer, J. L. et al. (2016). SENS-IS, a 3D reconstructed epidermis based model for quantifying chemical sensitization potency: Reproducibility and predictivity results from an inter-laboratory study. Toxicol In Vitro 32, 248-60. doi:10.1016/j.tiv.2016.01.007


OECD (2021a). Guideline No. 497: Defined Approaches on Skin Sensitisation. https://www.oecd-ilibrary.org/content/publication/b92879a4-en


Conflict of interest statement
The authors declare no competing interests.
Data availability statement
All data of this publication are made publicly available, and all models and tests used are freely available.