The current landscape of NAM Methods for Skin Sensitization Hazard Identification:

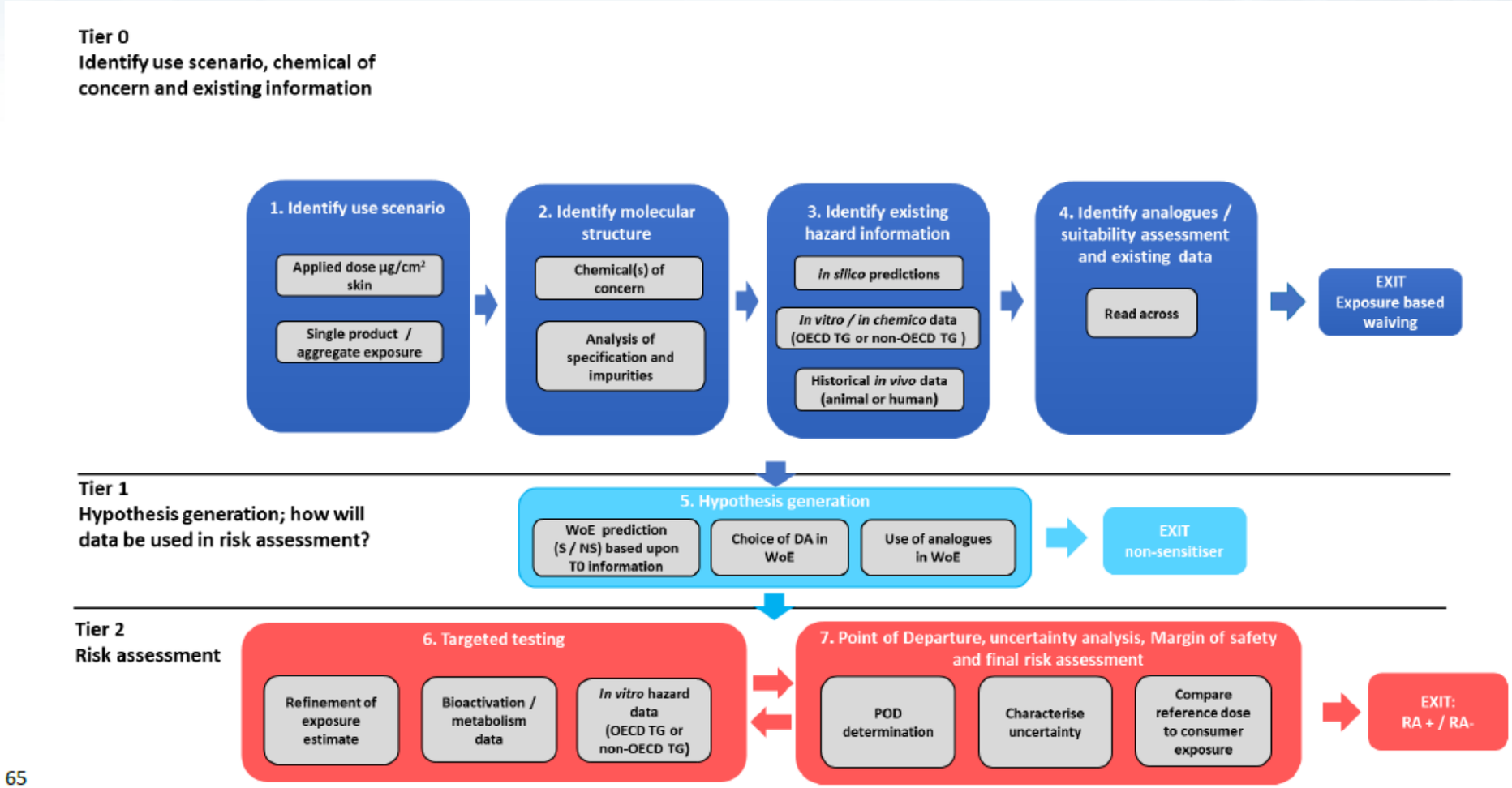
Progress towards Potency Assessment

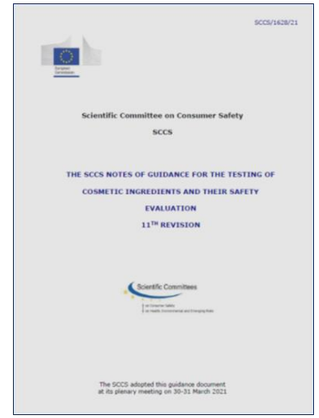
Petra Kern

October 7th, 2022

Human Safety
Ensuring Safe Products

NGRA Framework for Skin Sensitisation




SCCS/1628/21

Scientific Committee on Consumer Safety

SCCS

THE SCCS NOTES OF GUIDANCE FOR THE TESTING OF COSMETIC INGREDIENTS AND THEIR SAFETY EVALUATION 11TH REVISION

Scientific Committee

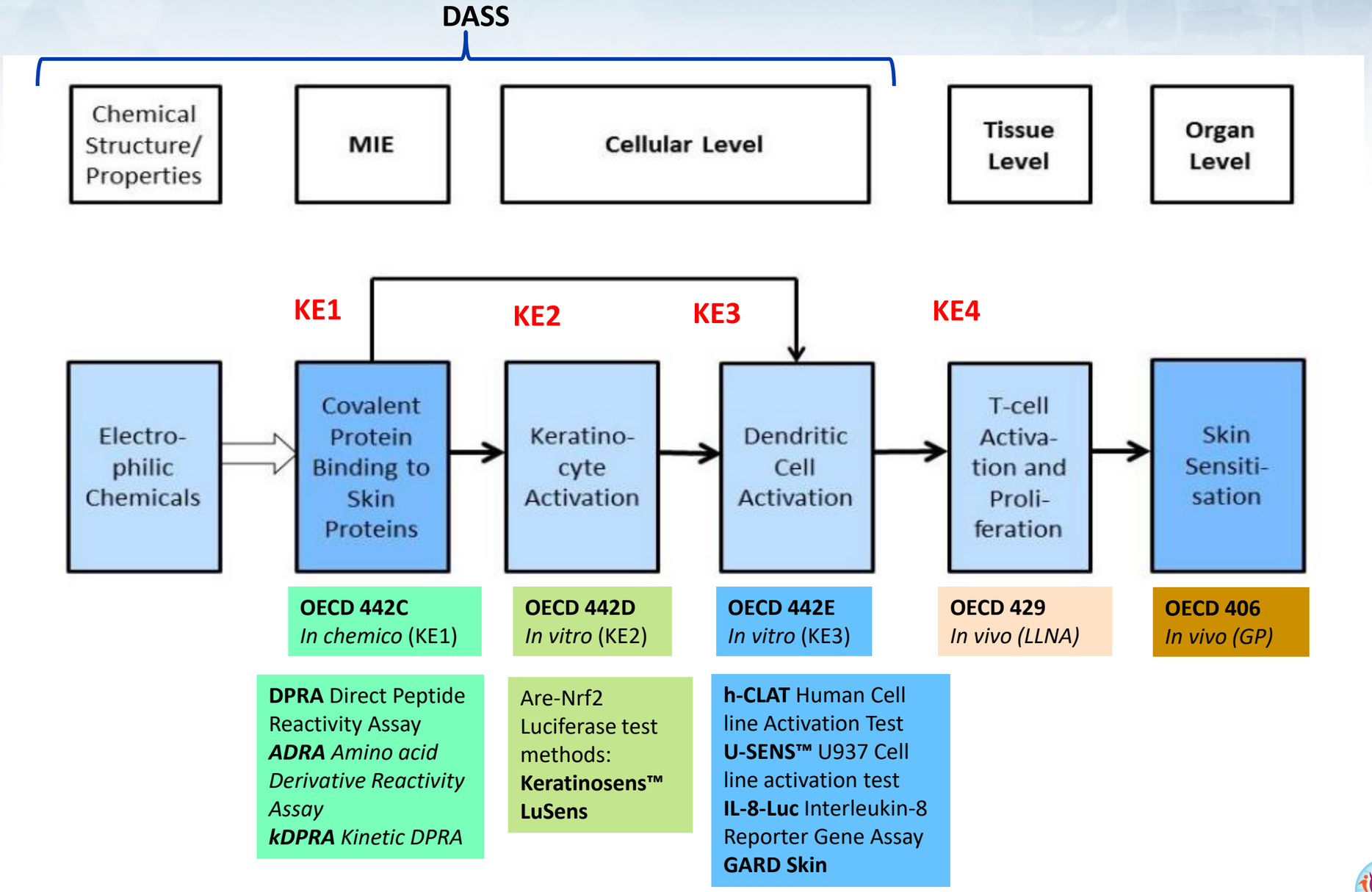
The SCCS adopted this guidance document at its plenary meeting on 30-31 March 2021

SCCS 11th NoG 2021

Gilmour & Kern et al, RTP, 116, 2020 https://authors.elsevier.com/a/1bRLv%7E81IJLAV

Human Safety
Ensuring Safe Products

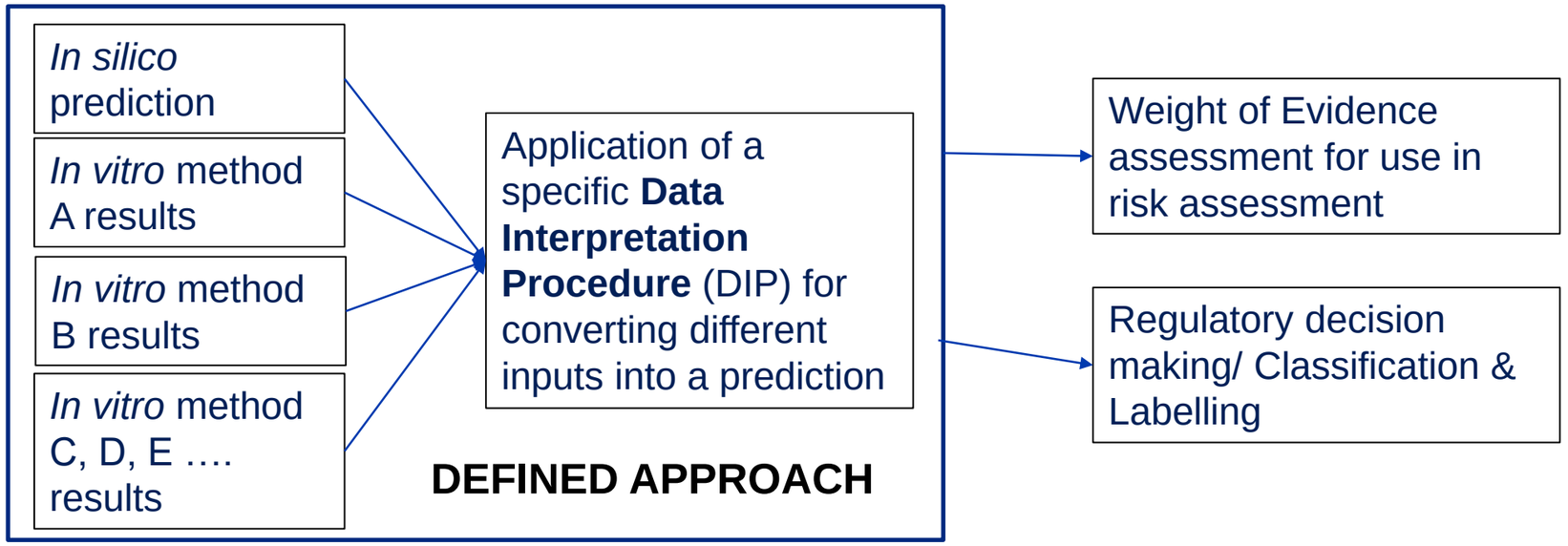
Skin Sensitisation AOP

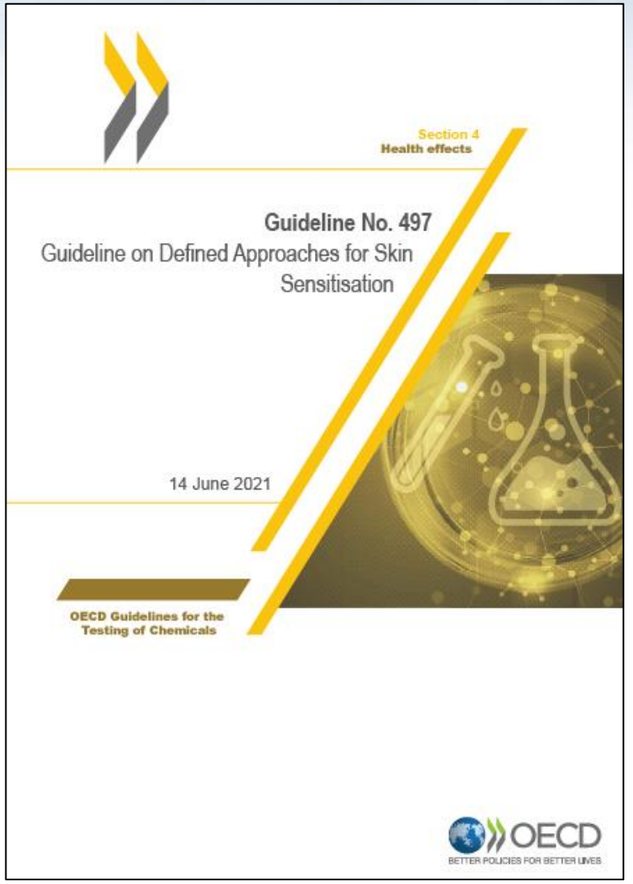


OECD, 2012. The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins. Series on Testing and Assessment No. 168.

Human Safety
Ensuring Safe Products

Defined Approaches





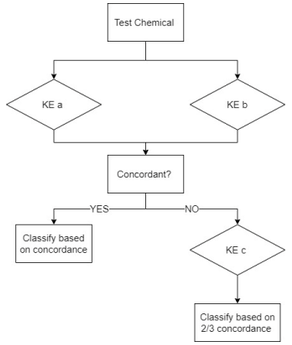
OECD 497
Defined Approach (DASS)

Human Safety
Ensuring Safe Products

NAM - Defined Approaches

Hazard | **Potency (GHS 1A/ 1B)** | **Potency grouping** | **Continuous PoD values**

OECD (2021), Guideline No. 497

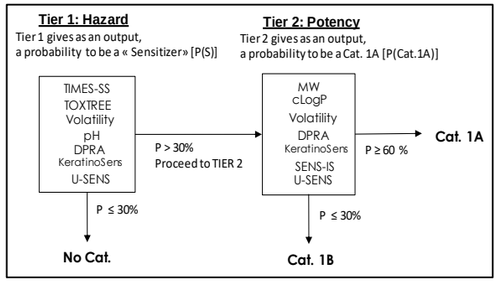


2 out of 3

Score	h-CLAT MIT µg/mL	DPRA mean Cysteine and Lysine% depletion	DPRA Cysteine % depletion*	*In silico* (ITSv1: DEREK; ITSv2: OECD TB)
3	≤10	≥42.47	≥98.24	
2	>10, ≤150	≥22.62, <42.47	≥23.09, <98.24	
1	>150, ≤5000	≥6.38, <22.62	≥13.89, <23.09	Positive
0	not calculated	<6.38	<13.89	Negative

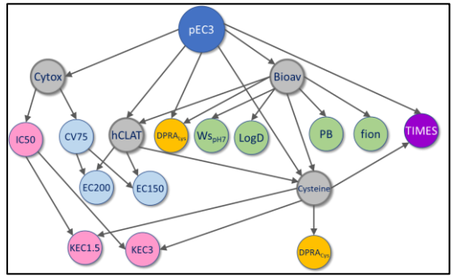
Potency	Total Battery Score
UN GHS 1A	6-7
UN GHS 1B	2-5
Not classified	0-1

ITSv1/v2



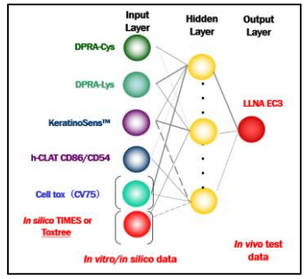
Sequential Testing Strategy

(Tourneix et al. 2019)

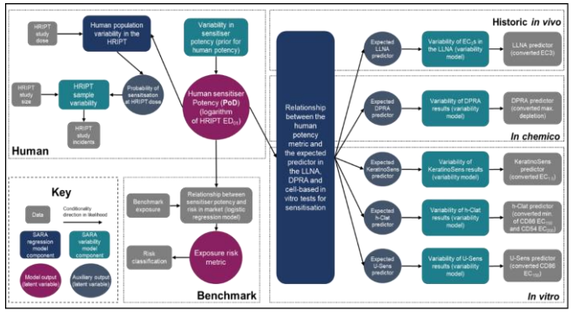


BN-ITS

(Jaworska et al. 2015, Kern et al. 2022 in prep)



ANN EC3 *(Hirota et al 2015, 2018)*



(Reynolds et al. 2019, Gilmour et al 2022) **SARA**

Human Safety
Ensuring Safe Products

NAM - Defined Approaches - More Methods

Hazard | **Potency (GHS 1A/ 1B)** | **Potency grouping** | **Continuous PoD values**

GARD" skin
200 genes

Binary hazard identification of skin sensitizing chemicals.

GARD" potency
51 genes

An add-on *in vitro* test to GARDskin for potency classification according to GHS/CLP (1A or 1 B)

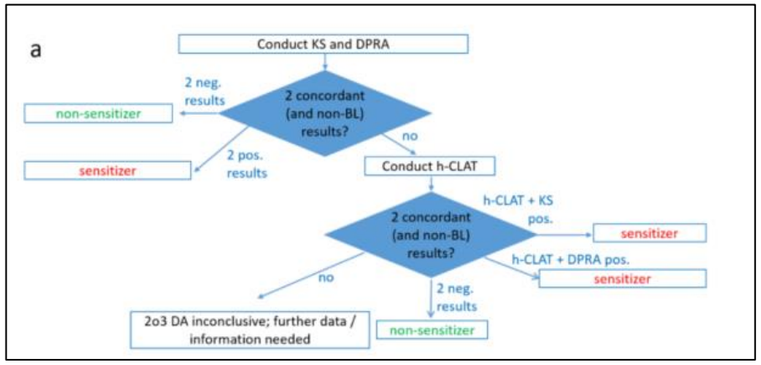
SENS-IS

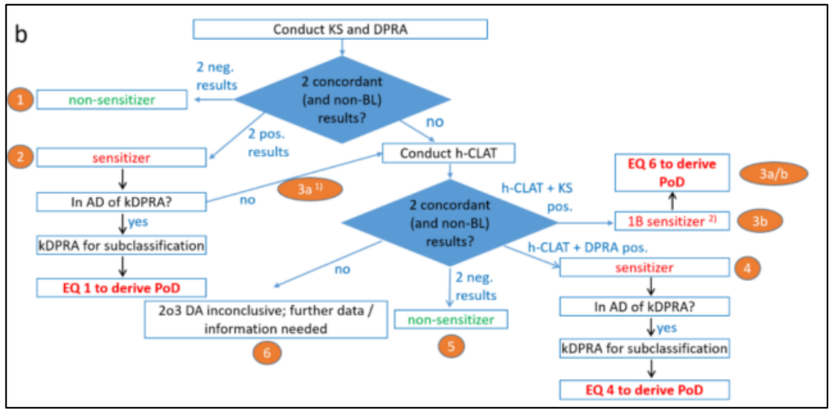
Derek nexus

In Silico QSAR Models

GARD" skin Dose-Response
200 genes

Quantitative assessment of skin sensitization potency on a continuous scale.



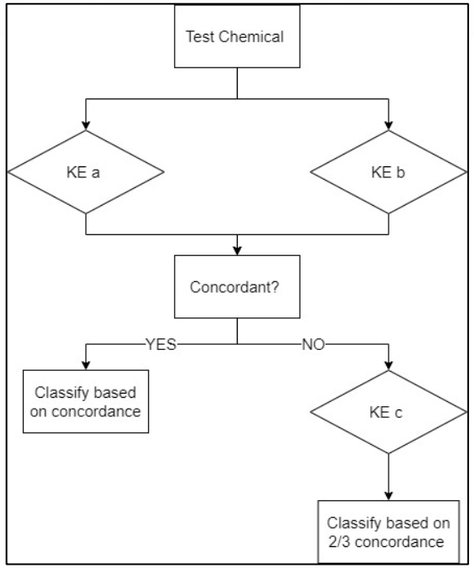


Natsch & Gerberick 2022

DISCLAIMER: not a complete list

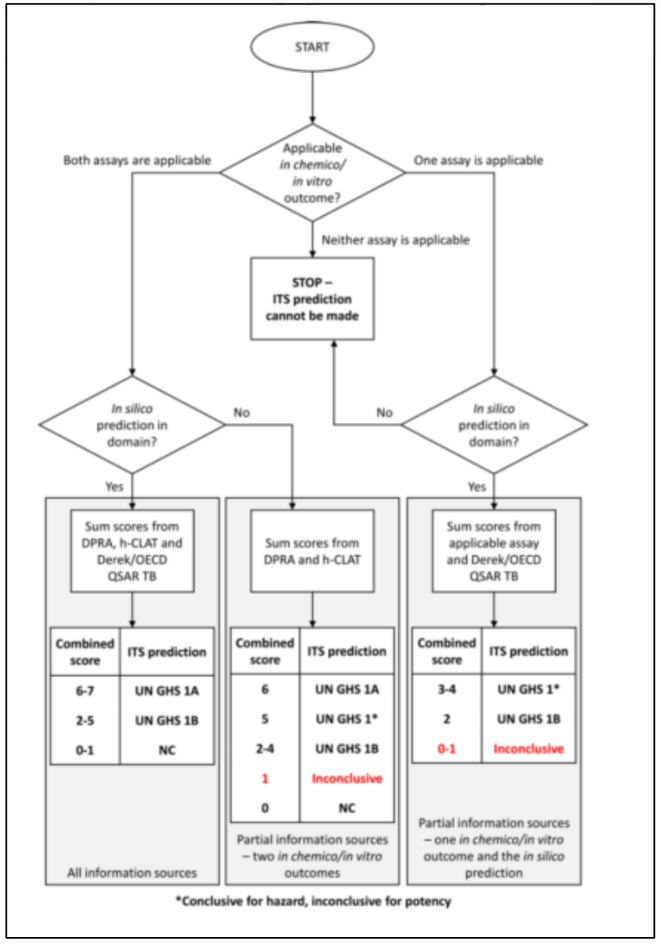
Human Safety
Ensuring Safe Products

GL497 Defined Approaches



2 out of 3
Hazard prediction only

ITSv1/v2
GHS 1A, 1B, NC, inconclusive

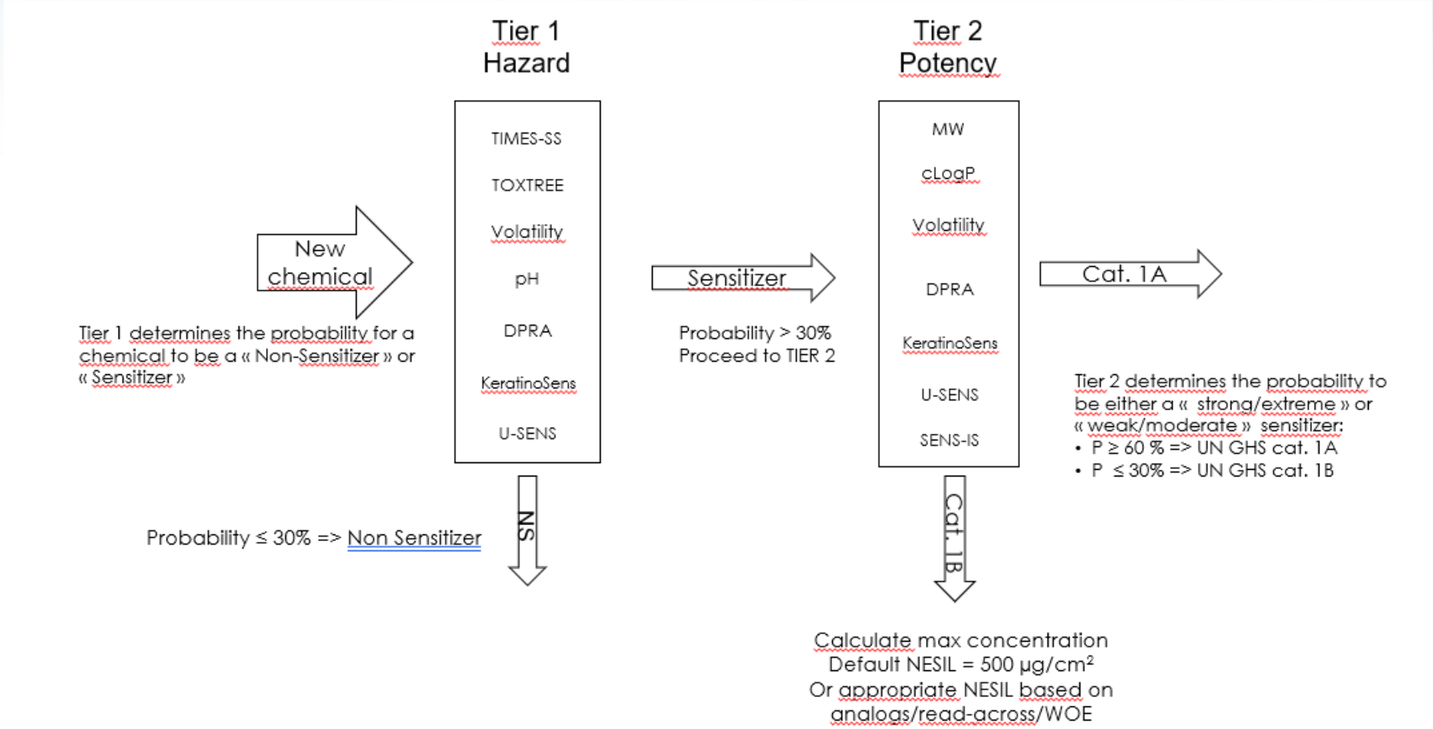


DA/Method	Information Sources	Capability (Hazard and/or Potency)	Hazard Performance vs. LLNA	Hazard Performance vs. Human	Potency Performance vs. LLNA (Accuracy)	Potency Performance vs. Human (Accuracy)
2o3 DA	DPRA, KeratinoSens™, h-CLAT	Hazard	84% BA, 82% Sens, 85% Spec	88% BA, 89% Sens, 88% Spec	-	-
ITSv1 DA	DPRA, h-CLAT, DEREK Nexus v6.1.0	Hazard, Potency	81% BA, 92% Sens, 70% Spec	69% BA, 93% Sens, 44% Spec	70% NC, 71% 1B, 74% 1A	44% NC, 77% 1B, 65% 1A
ITSv2 DA	DPRA, h-CLAT, OECD QSAR Toolbox v4.5	Hazard, Potency	80% BA, 93% Sens, 67% Spec	69% BA, 94% Sens, 44% Spec	67% NC, 72% 1B, 72% 1A	44% NC, 80% 1B, 67% 1A
LLNA (provided for comparison)	*in vivo*	Hazard, Potency	-	58% BA, 94% Sens, 22% Spec	-	25% NC, 74% 1B, 56% 1A

ECHA
EUROPEAN CHEMICALS AGENCY
October 2021
Skin sensitisation

GLOBALLY HARMONIZED SYSTEM OF CLASSIFICATION AND LABELLING OF CHEMICALS (GHS)
Ninth revised edition
UNITED NATIONS
New York and Geneva, 2021

Sequential Stacking Meta-Model



- 2 Tier model: hazard ID followed by potency categorization
- GHS subcategories (1A, 1B, NC)
- Bayesian approach: probability information – confidence in prediction
- LLNA training set

Table 2
Predictive performance of the defined approach using a stacking meta-model for human and LLNA hazard data with ≤30 and ≥70% probability cut-offs.

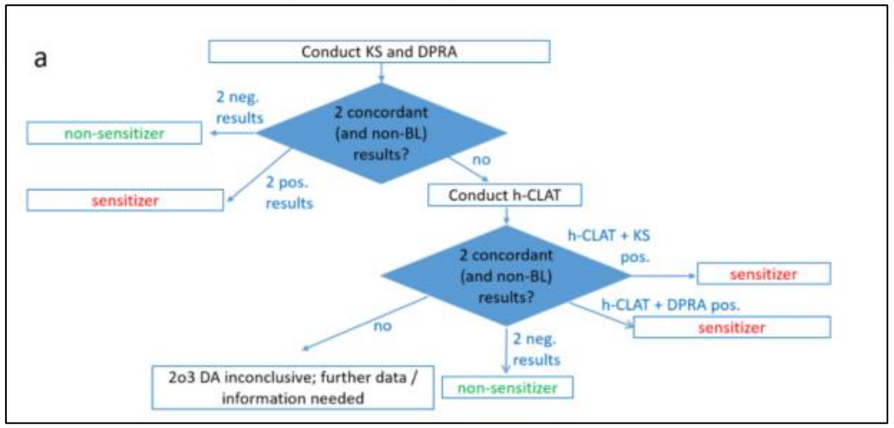
Probability cut-offs: ≤30 and ≥70% applied to the stacking meta-model	Chemicals with hazard prediction	Sensitivity (%)	Specificity (%)	Accuracy (%)	Kappa
DA predictions *vs.* human	97	91 (58/64)	76 (25/33)	86 (83/97)	0.67
DA predictions *vs.* LLNA	97	85 (64/75)	91 (20/22)	87 (84/97)	0.67
DA predictions *vs.* composite reference[a]	97	87 (65/75)	95 (21/22)	89 (86/97)	0.72
LLNA *vs.* human	97	92 (59/64)	51 (17/33)	78 (76/97)	0.48
LLNA *vs.* human	105	92 (62/67)	53 (20/38)	78 (82/105)	0.49
DA predictions *vs.* human	66[b]	88 (38/43)	65 (15/23)	80 (53/66)	0.55
LLNA *vs.* human	66[b]	91 (39/43)	39 (9/23)	73 (48/66)	0.33

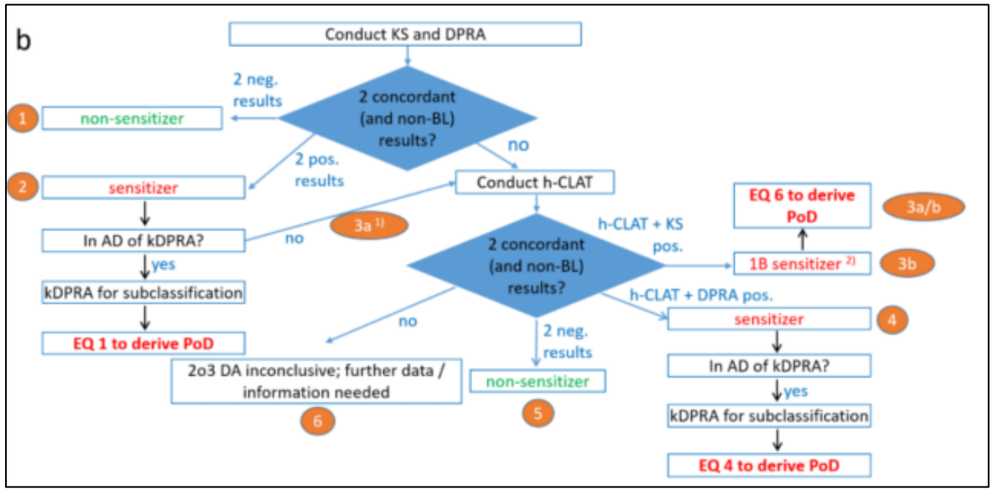
[a] Composite ref. for Human potency cat.5 + LLNA neg. = NS and cat.5 + LLNA pos. = S.
[b] Test set of naïve chemicals not used to develop nor challenge the defined approach for which a hazard prediction was generated using the 30/70% probability cut-offs (Del Bufalo et al., 2018).

L'ORÉAL
Research & Innovation

P&G Human Safety
Ensuring Safe Products

kDPRA and “2 out of 3”





- 2 Tier model: hazard ID followed by potency prediction
- Testing sequence including GHS categories and POD determination combined with the “2 out of 3” approach
- Several multiple linear regression models
- Use of kDPRa data in combination with others NAM
- Use of an uncertainty factor to be applied
- LLNA training set

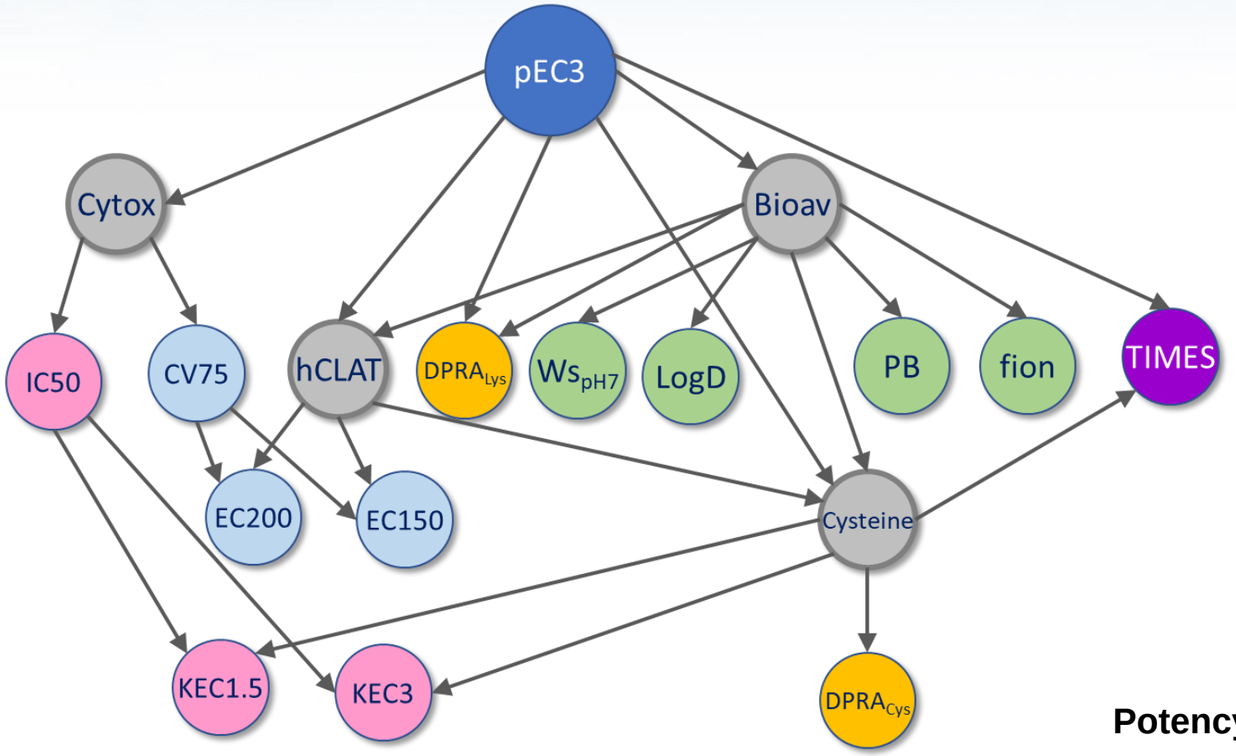
Tab. 2: GHS sub-classification of the chemicals in the OECD database by the 2o3 DA combined with kDPRA

	LLNA Result		
Prediction 2o3 DA with kDPRA	**NC (n =26)**	**1B (n = 85)**	**1A (n = 38)**
NC	21	16	0
1B	3	34	7 (4)[a]
1A	1	14	26 (29)
Correct	84%	53%	79% (88%)
Under	NA	25%	21% (12%)
Over	16%	22%	NA
Inconclusive	n = 8	n =21	n = 5

[a] The values are based on applying only the prediction model of the kDPRA and 2o3 DA. In brackets are given the values if taking the AD of the kDPRA into account and applying Scenario 3a in Figure 1 (Using EQ6 for chemicals outside of AD of kDPRA).

Human Safety
Ensuring Safe Products

BN-ITS Defined Approach



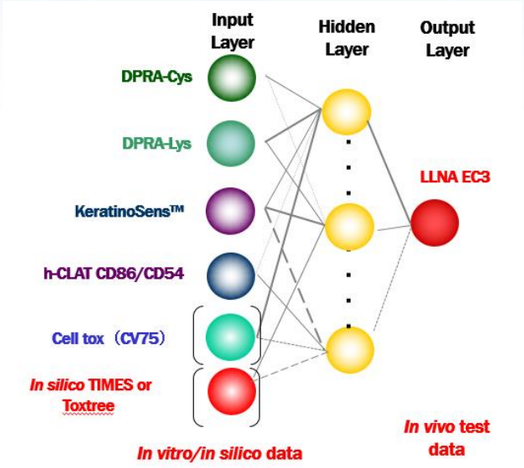
- Bayesian network model: captures conditional dependencies between variables
- Hazard ID and Potency evaluation
- Target data: LLNA potency categories based on EC3 values (NS, weak, moderate, strong)
- Prediction of potency categories and uncertainty evaluation using probability distribution and Bayes factors
- Conversion to EC3 values (continuum POD) possible

Potency		pEC3 Category (Experimental)			
		1	**2**	**3**	**4**
pEC3 Category (Predicted)	**1**	62	6	2	2
	2	3	31	7	0
	3	5	8	48	9
	4	3	4	13	34

Accuracy = 74%
Accuracy +1/-1 = 93%
Over-predicted = 15%
Under-predicted = 11%

Human Safety
Ensuring Safe Products

ANN



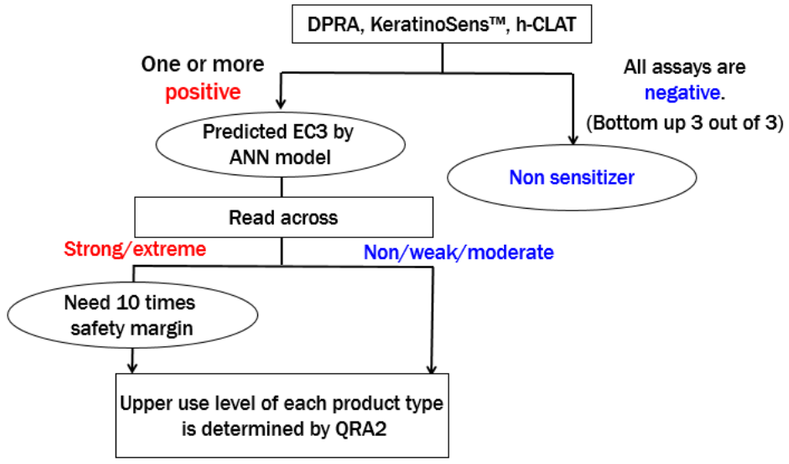


TABLE 5 Correlations between LLNA EC3 classified into four categories and predicted LLNA classification obtained with ANN models 1, 2, 4 and 6 (n = 134)

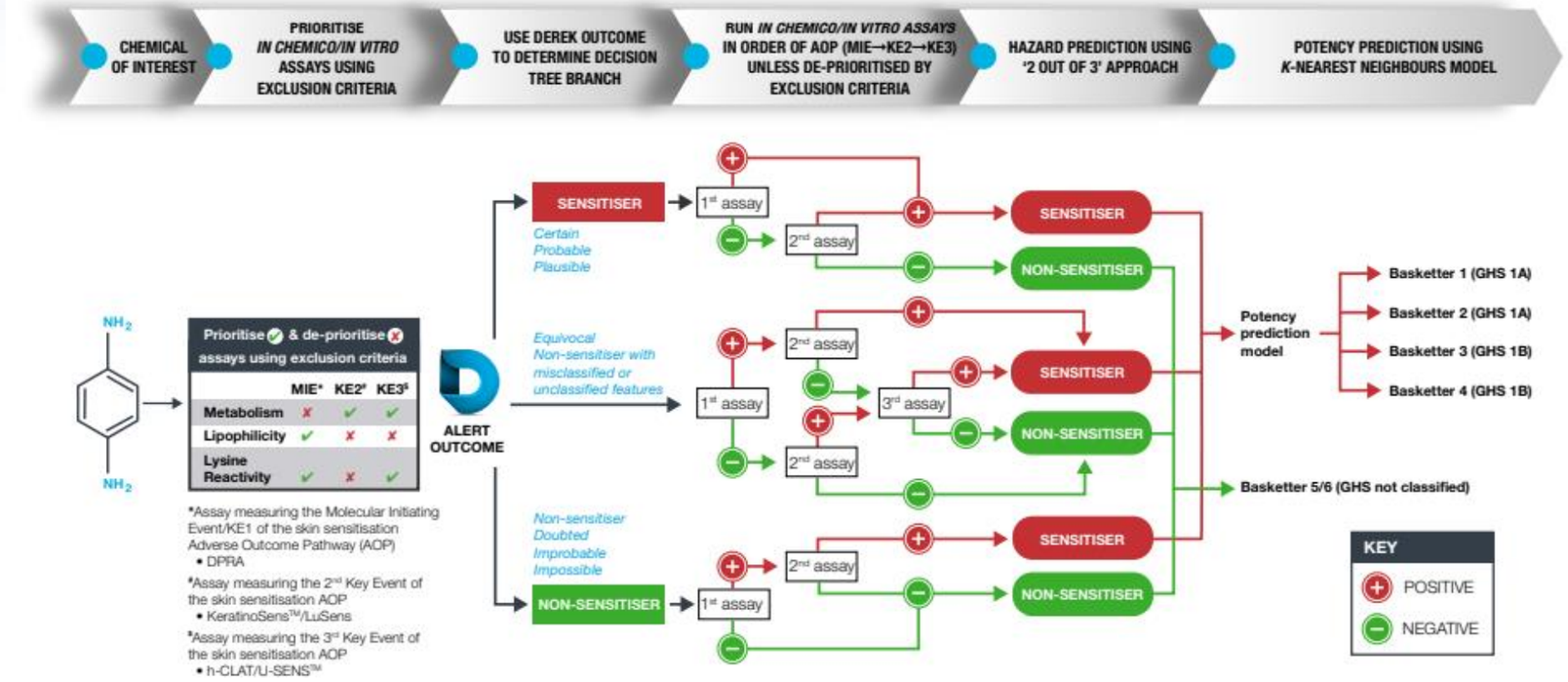
		LLNA classification (in vivo)			
	Potency category	Extreme or strong (26)	Moderate (37)	Weak (31)	NS (40)
(A) ANN model 1 (Min(h-CLAT&cell toxicity)/ DPRA/ KeratinoSens™)	Extreme or strong	19	3	0	0
	Moderate	6	17	7	0
	Weak	1	15	23	24
	Negative, predicted EC3 > 100%	0	2	1	16
	Accuracy (%)	56.0			
	Overpredicted (%)	25.4			
	Underpredicted (%)	18.7			
(B) ANN model 2 (h-CLAT/ DPRA/ KeratinoSens™/ cell toxicity)	Extreme or strong	21	3	0	0
	Moderate	4	15	10	0
	Weak	1	17	19	24
	Negative, predicted EC3 > 100%	0	2	2	16
	Accuracy (%)	53.0			
	Overpredicted (%)	27.6			
	Underpredicted (%)	19.4			
(C) ANN model 4 (h-CLAT/ DPRA/ KeratinoSens/cell toxicity™/TIMES-M)	Extreme or strong	20	4	0	0
	Moderate	6	25	9	1
	Weak	0	6	21	14
	Negative, predicted EC3 > 100%	0	2	1	25
	Accuracy (%)	67.9			
	Overpredicted (%)	20.9			
	Underpredicted (%)	11.2			
(D) ANN model 6 (h-CLAT/DPRA/ KeratinoSens/cell toxicity™/ Toxtree)	Extreme or strong	22	4	0	0
	Moderate	3	18	10	1
	Weak	1	13	20	20
	Negative, predicted EC3 > 100%	0	2	1	19
	Accuracy (%)	59.0			
	Overpredicted (%)	26.1			
	Underpredicted (%)	14.9			

- 2 Tier model: hazard ID followed by potency prediction
- LLNA EC3% predicted – continuous value
- No categorization
- No probability information
- LLNA training set

- Applied at EPA for isothiazolinone RA

SHISEIDO P&G Human Safety Ensuring Safe Products

DEREK NEXUS (LHASA)



- Tiered approach: Hazard ID and potency categorization
- 6 categories
- Combination of DEREK alerts with “2 out of 3” and a nearest neighbour prediction model
- Updated with EC3 prediction

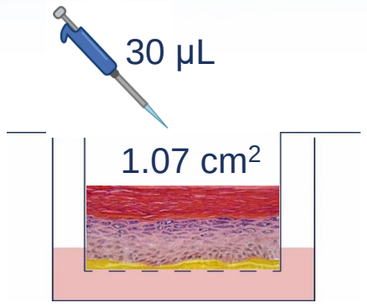
Table 7
The skin sensitisation potency predictivity of the DA (and LLNA) compared to both LLNA and human reference data.

Benchmark data	Potency category				GHS classification			
	n	Over-predicted	Correctly predicted	Under-predicted	*n*	Over-predicted	Correctly predicted	Under-predicted
DA vs LLNA	174[a]	20%	59%	21%	174[a]	12%	73%	15%
DA vs Human	79[b]	14%	68%	18%	79[b]	10%	76%	14%
LLNA vs Human	89	25%	54%	21%	79	20%	65%	15%

[a] 20 chemicals with a hazard prediction not given a potency prediction by the DA.
[b] 23 chemicals with a hazard prediction not given a potency prediction by the DA.

(Macmillan & Chilton 2019)

SENS-IS assay versus Weight of Evidence Skin Sensitisation Potency Categories



62 SENS-IS biomarkers
(ARE, SENS-IS, Irritation genes)

Concentration leading to Pos.	Potency category	Dose per unit area
0.1%	Extreme	28 µg/cm²
1%	Strong	280 µg/cm²
10%	Moderate	2803 µg/cm²
50%	Weak	14000 µg/cm²
100%	Very Weak	28000 µg/cm²

174

Category Name	Human Cat. (µg/cm²)
Extreme	<25
Strong	25 - 500
Moderate	500 - 2,500
Weak	2,500 - 10,000
Very Weak	> 10,000
Non Sensitizer	Negative

Na et al. reg. Tox pharm. 2022

Cottrez et al. (2016)

Human Safety
Ensuring Safe Products

SENS-IS assay versus Weight of Evidence Skin Sensitisation Potency Categories

SENS-IS / WoE	Extreme	Strong	Moderate	Weak	Very Weak	NS
Extreme	**2**	**2**	**0**	**1**	**0**	**0**
Strong	**1**	**12**	**5**	**1**	**0**	**0**
Moderate	**0**	**5**	**18**	**6**	**3**	**2**
Weak	**1**	**1**	**14**	**27**	**5**	**6**
Very Weak	**0**	**0**	**7**	**17**	**2**	**12**
NS	**0**	**0**	**0**	**4**	**1**	**19**

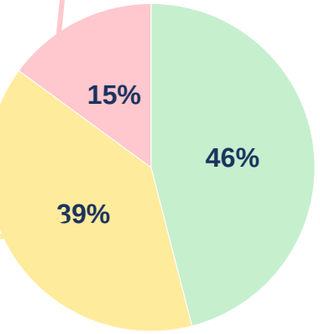
Exact match **Approximate** **Off**

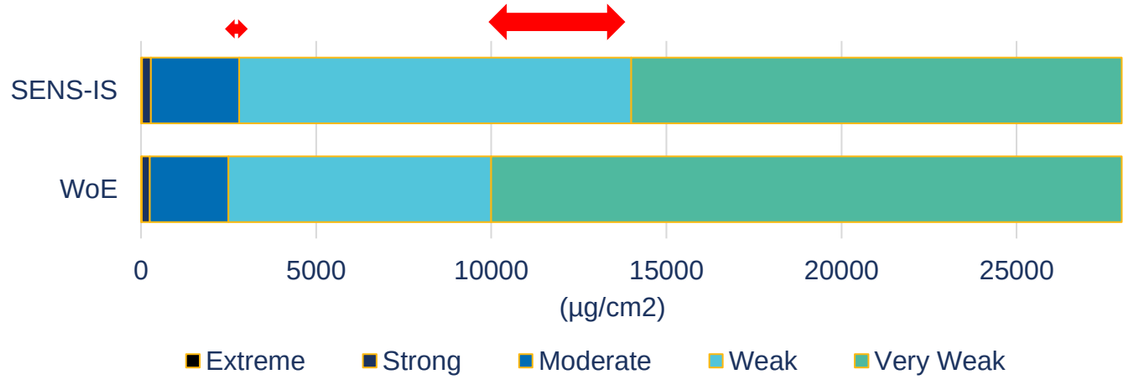
13/26 predicted to be stronger sensitizers by SENS-IS

13/26 predicted to be weaker sensitizers by SENS-IS

38/68 predicted to be stronger sensitizers by SENS-IS

30/68 predicted to be weaker sensitizers by SENS-IS





RIFM RESEARCH INSTITUTE FOR FRAGRANCE MATERIALS

Na et al. reg. Tox pharm. 2022

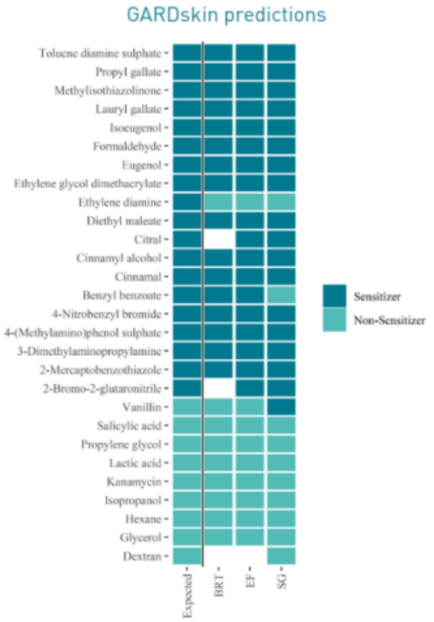
P&G Human Safety Ensuring Safe Products

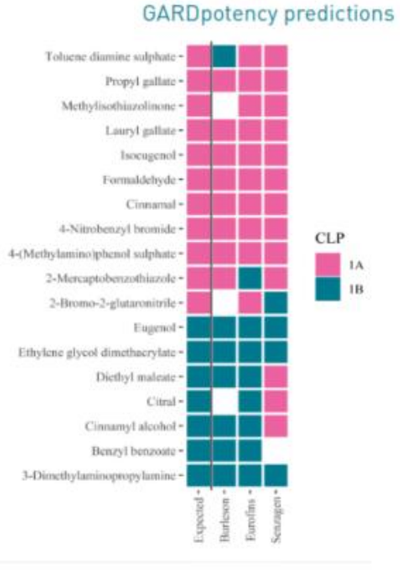
GARD Skin Sensitisation Assay Portfolio

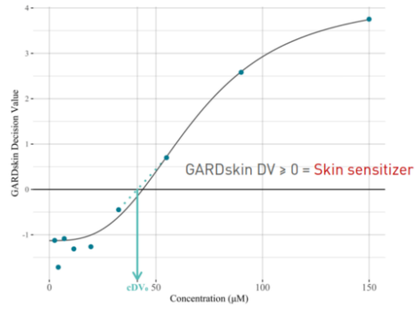
OECD TG 442e

From binary hazard identification to quantitative potency information on a continuous scale

GARD" skin 200 genes	GARD" potency 51 genes	GARD" skin Dose-Response 200 genes
Binary hazard identification of skin sensitizing chemicals.	An add-on *in vitro* test to GARDskin for potency classification according to GHS/CLP (1A or 1 B)	Quantitative assessment of skin sensitization potency on a continuous scale.







Performance statistics:

GARDskin accuracy:	94%
GARDpotency accuracy:	88%
GARD Defined Approach:	86%

	GARD	LLNA
Response value	DV	SI
Binary Threshold	DV = 0	SI = 3
Readout	cDV_0 (DV_0 Concentration)	EC3 Concentration

Gradin et al. 2021

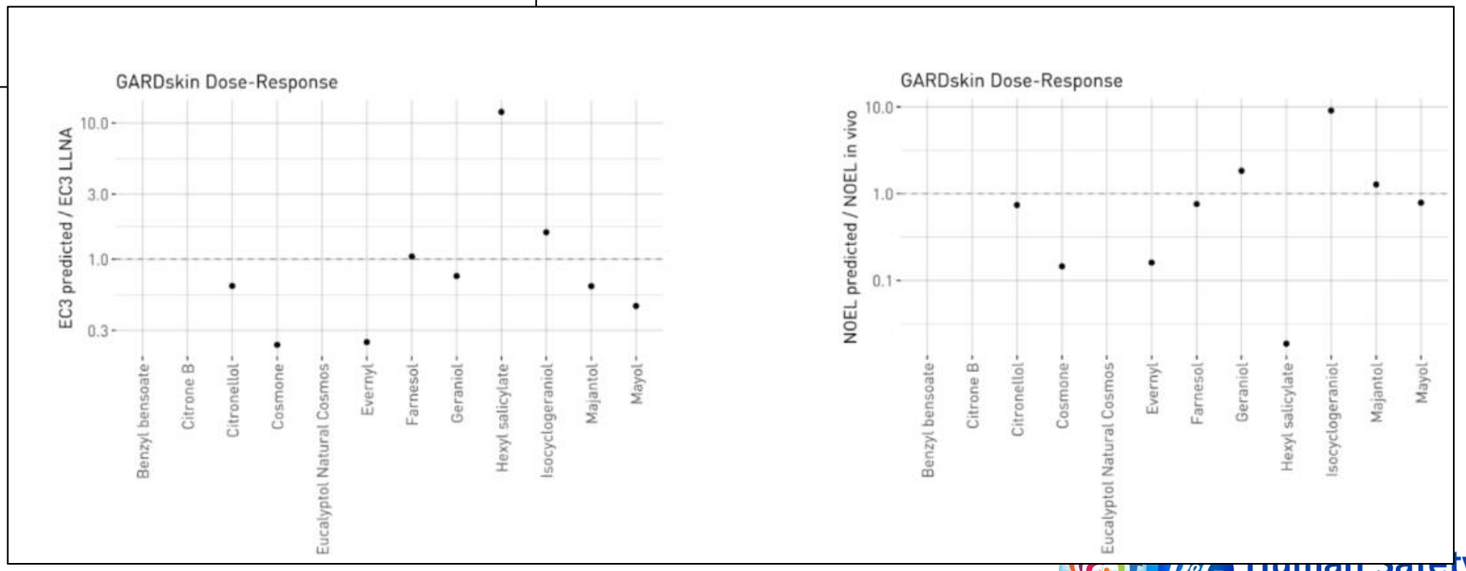
SENZA GEN

P&G Human Safety
Ensuring Safe Products

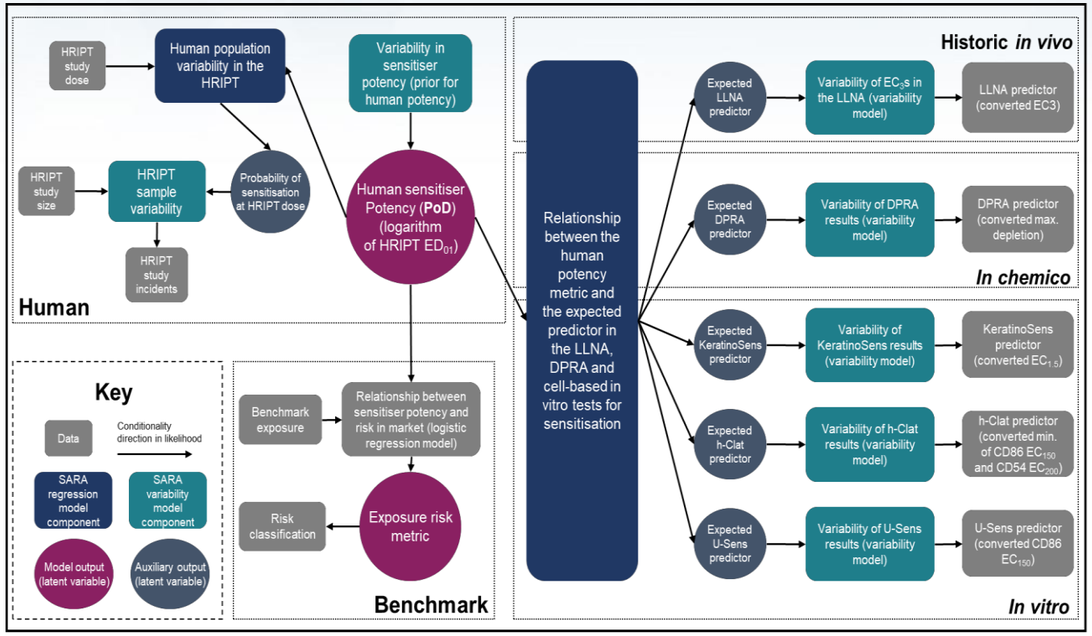
RIFM – GARD Evaluation

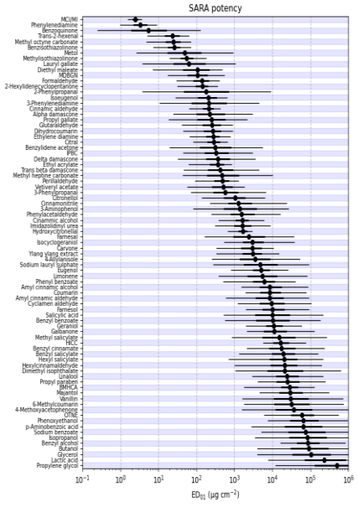
Test Item	Name	Predicted LLNA EC3 (%)[I]	Reference LLNA EC3 (%)[I]	Predicted HP Cat	Reference HP cat[II]	Predicted NOEL (µg/cm²)	Reference NOEL (µg/cm²)[II]
S1	Farnesol	5.02 (2.46, 10.3)	4.8	4 (0.357)	3	2090 (795, 5490)	2755
S2	Citronellol	27.6 (8.14, 93.9)	43.5	5 (0.481)	5	21800 (3560, 134000)	29528
S3	Iso-cyclo Geraniol	39.4 (10.2, 100)	25	5 (0.547)	4	35500 (4660, 271000)	3898
S4	Geraniol	12.2 (4.75, 31.2)	16.2	4 (0.38)	4	7070 (1840, 27100)	3875
S5	Citrone B	Non-Sens	64.98	Non-Sens	4	Non-Sens	2780
S6	Majantol	18.5 (6.29, 54.5)	29.2	5 (0.406)	4	12600 (2600, 60900)	9900
S7	Hexyl salicylate	2.17 (1.13, 4.17)	0.18	4 (0.285)	4	658 (270, 1600)	35433
S8	Cosmone	3.86 (1.96, 7.59)	16.4	4 (0.338)	5	1450 (589, 3590)	10000
S9	Evernyl	4.68 (2.32, 9.45)	19	4 (0.352)	5	1890 (736, 4880)	11810
S10	Eucalyptol Natural Cosmos	Non-Sens	65.9	Non-Sens	4	Non-Sens	590
S11	Mayol	19.9 (6.6, 60.3)	44	5 (0.42)	5	13900 (2760, 70300)	17717
S12	Benzyl benzoate	Non-Sens	17	Non-Sens	5	Non-Sens	20690

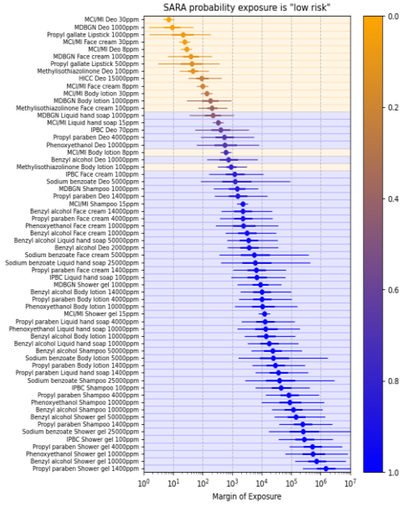
I) LLNA EC3 values were provided by RIFM.
II) Human NOEL and HP cat were obtained from Basketter (2016) and API



SARA Defined Approach







- Bayesian statistics to infer a probability that a consumer exposure to some chemical can be considered low risk (SARA risk metric) for induction
- Uses a database of public NAM data, and historic LLNA and HRIPT data
- The PoD metric is a dose with a 1% chance of human skin sensitisation (ED01)
- The model accounts for variability in the data
- Incorporates benchmark exposure
- SARA can also predict hazard, GHS categories, POD (continuous scale)

Unilever

P&G Human Safety
Ensuring Safe Products

Training sets and Performance

- NAM/ DA not trained / built with human data
- LLNA dataset used for all (most) as training set
  - Potency categories or EC3 continuous values
  - Similar data used for all DA/ models (applicability domain)
- Human benchmarks incorporated in some to determine risk

- Performances evaluated against LLNA
  - OECD LLNA reference database (168, 123 with 1A/1B) or others
  - Comparable performance (for hazard, potency difficult to compare)
  - Better for hazard than potency (drops with more categories)
- Relevance to humans / Performances against human data
  - Some NAM/ DA evaluated against human data
  - OECD human reference database (66, 55 with 1A/ 1B)

Use for QRA/ NGRA: Conversion of Prediction into POD

Potency Category	LLNA EC3 % ranges (ECETOC)	LLNA EC3 % conversion to µg/cm² (based on LLNA dosing*)	Default NESIL µg/cm² to be used as POD
Extreme (Potent)	< 0.1	< 25	1
Strong	≥ 0.1 - < 1	≥25 - <250	10
Moderate	≥ 1 - < 10	≥250 - <2500	100
Weak	≥ 10 - ≤ 100	≥2500 - <25000	1000
GHS 1A	< 2%	< 500	< 500
GHS 1B	> 2%	> 500	> 500

- Conversion of LLNA categories or EC3% to µg/cm² using a conversion factor (1% = 250 µg/cm²)
- GHS categorization models limited in deriving POD (only < or > 500 µg/cm²)

- Some methods (Bayesian) allow confidence evaluation around predicted value
- Genomic methods – unique conversion approaches
- PoD metric of SARA is a dose with a 1% chance of human skin sensitisation (ED01)

Some final Comments…

- NAM/ DA remain categorical for most, but progress was made towards better potency prediction (ie. Continuous values)
  - What is sufficient?
- Progress made towards POD setting for QRA.

- Conversion to POD not uniquely done
- Conservatism / confidence in prediction differs between NAM/ DA
- Bayesian DAs enable experimental data variability to be modelled and uncertainty in POD & risk metrics can be factured into decision making

- Shortcomings of the animal reference standard must be acknowledged
- Human biological and mechanistic relevance of the DA needs to be established

Thanks to the CE Skin Tolerance Task Force and colleagues!