National Institute of Environmental Health Sciences
Division of Translational Toxicology

Evaluation of Skin Sensitization Classification Rules to Reflect Human Potency

M. Herzler[1], K.T. To[2], J. Abedini[3], D.G. Allen[2], A.M. Api[4], D. Germolec[5], J. Gordon[6], N. Kleinstreuer[5], H-S. Ko[7], J. Matheson[6], H-J. Thierse[1], J. Truax[2], J.T. Vanselow[1], J. Strickland[2]

[1]BfR, Berlin, Germany; [2]Inotiv, RTP, NC; [3]RTI International, RTP, NC; [4]RIFM, Woodcliff Lake, NJ; [5]NIH/NIEHS/DTT/NICEATM, RTP, NC; [6]CPSC, Rockville, MD; [7]FDA/CDER, Silver Spring, MD

Introduction

- To support the development of Guideline 497 on Defined Approaches for Skin Sensitization published by the Organisation for Economic Co-operation and Development (OECD; OECD 2021), we collected historical human predictive patch test (HPPT) data for use as reference data.
- We deemed data from 2255 HPPTs, representing 1366 different substances, as sufficiently reliable to assign skin sensitization potency classifications according to the United Nations Globally Harmonized System (GHS) of Classification and Labelling of Chemicals criteria and guidance (UN 2021) (Figure 1a).
- Approaches currently used to assign skin sensitizers to GHS potency subcategories consider only the dose inducing the skin sensitization response and not the frequency of induced sensitization in human subjects. Variations in conduct of assays may introduce uncertainty into otherwise valid data.
- To address these limitations, we developed a modified approach to GHS classification (Figure 1b) that incorporates a frequency metric into potency classification and also addresses uncertainty in assay results (Table 1).
- We also developed a strategy for using these classifications in a weight-of-evidence (WoE) approach with animal reference data, when classifications do not agree, to develop an overall classification.

Modification of the Standard GHS Classification System

Table 1. Modifications to the standard GHS classification system

Standard GHS Classification	Challenge	Modified GHS Classification
Substances classified as skin sensitizers are assigned potency subcategories using the dose per skin area (DSA) as the dose metric. • **DSA**: the amount in micrograms of chemical per cm^2 ($\mu g/cm^2$) of skin area.	Potency subcategorization does not account for the number of sensitized individuals contributing to a positive result, thereby ignoring an important measure of potency.	To incorporate this measure into classification, we examined two additional dose metrics: • **DSA1+**: the hypothetical DSA that sensitizes one test subject. • **DSA05**: the hypothetical DSA that sensitizes 5% of test subjects.
• A positive result at DSA ≤ 500 $\mu g/cm^2$ results in classification as a **1A,** strong sensitizer. • A positive result at DSA > 500 $\mu g/cm^2$ results in classification as a **1B,** weak sensitizer.	Variability and uncertainty associated with the HPPT data may lead to ambiguous 1A or 1B classifications. • A positive result at DSA > 500 $\mu g/cm^2$ would indicate a 1B sensitizer, but 1A cannot be ruled out because a lower dose could produce a positive result.	We derived a DSA1+/DSA05 borderline range of [375…625] $\mu g/cm^2$ (± 25% around the 500 $\mu g/cm^2$ cut-off between 1A and 1B). • Substances testing positive at [500 $\mu g/cm^2$ < DSA1+/DSA05 ≤ 625 $\mu g/cm^2$] are classified as **1B+**, indicating moderate sensitization potential (1B) with some likelihood of underclassification. • Substances testing positive at [375 $\mu g/cm^2$ < DSA1+/DSA05 ≤ 500 $\mu g/cm^2$] are classified as **1A-**, indicating strong sensitization potential (1A) with some likelihood of overclassification.
Substances that test negative are assigned a GHS designation of **NC**.	NC classifications may be ambiguous because a substance was tested at a concentration too low to produce a positive result.	We defined a DSA cut-off at 625 $\mu g/cm^2$ (the upper boundary of the DSA1+/DSA05 borderline range) and a test concentration cut-off of at least 25% (the 99th percentile of the top concentrations of negative tests). • Substances testing negative at concentrations < 25% and DSA ≥ 625 $\mu g/cm^2$ are assigned **NC/1B**, an ambiguous outcome that excludes strong skin sensitization potential. • Substances testing negative at concentrations < 25% and DSA < 625 $\mu g/cm^2$ are assigned **NC/1**, an ambiguous outcome that provides no information on skin sensitization potential.

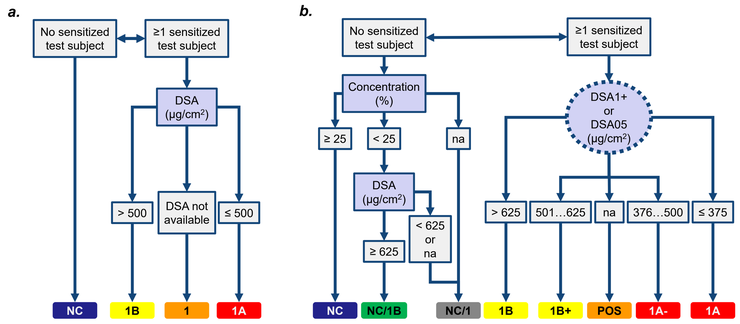

Figure 1. (a) represents the standard GHS classification approach. The modified approach we developed, shown in (b), incorporates sensitization incidence as well as ambiguous/borderline cases. Two dose metrics were applied to this approach: DSA1+ or DSA05. Derivation of the dose metrics is explained in Table 1. DSA = dose per skin area.

Classification of the HPPT Database

- We applied the modified GHS classification approach to the 2255 HPPT results to derive **extrapolated classifications (ECs)** for the 1366 substances using both the DSA1+ and DSA05 dose metrics.
- Test results for each substance were evaluated to assign overall classifications using three different modes based on GHS categories:
  - GHS_{BIN}: substance classified in a binary manner as Category 1 (sensitizer) or NC.
  - GHS_{SUB}: substance assigned to one of three classes: 1A sensitizer, 1B sensitizer, or NC.
  - GHS_{BORDER}: substance assigned to one of five classes: 1A sensitizer, 1* (sensitizer, but subclassification not possible), 1B sensitizer, NC/1B ambiguous (substance may or may not be a sensitizer, but 1A can be ruled out), or NC.
- For substances with discordant EC outcomes, overall classifications were assigned by combining the multiple results using three weight-of-evidence (WoE) approaches:
  - WoE score: average of individual scored test outcomes (Figure 2).
  - Median-like location parameter (MLLP): value at the median position of individual test outcomes, sorted by potency (adapted from Hoffmann et al. 2018) (Table 2).
  - Median sensitization potency estimate (MSPE): a slightly modified version of the MLLP (Table 3).
- Results from the three approaches were evaluated for concordance. If a WoE approach did not return a result for a substance, results from the remaining one or two approaches were evaluated.
  - If results from the WoE approaches agreed or there was only one result, then the concordant outcome was used as the overall classification.
  - If results from the WoE approaches disagreed, results from the WoE Score, MLLP, and MSPE approaches were evaluated using a consensus classification scheme or expert judgement.

Weight-of-Evidence Approaches Applied

1. Weight-of-Evidence Score

- The WoE Score approach scores each EC outcome and uses the average of the individual scores to classify a substance (Figure 2).
- NC/1 results are excluded from the combined chemical classification.

Individual Test Data

EC	Score
1A	2
1A-	1.75
POS	1.5
1B+	1.25
1B	1
NC/1B	0.5
NC/1	NA
NC	0

Combined Chemical Classification

WoE Score	Classification Mode		
	GHS_{BIN}	GHS_{SUB}	GHS_{BORDER}
1.76-2	1	1A	1A
1.51-1.75			
1.50		NA	1*
1.26-1.49		1B	
0.76-1.25			1B
0.26-0.75	NA	NA	NC/1B
0-0.25	NC	NC	NC

Calculate Weight-of-Evidence Score

Figure 2. WoE Score approach for classifying substances with multiple discordant tests. NA = not applicable, not assigned

2. Median-like Location Parameter

- Hoffmann et al. (2018) described a "median-like location parameter" (MLLP) approach to establish a representative value (RV) for describing skin sensitizer potency of substances with multiple test results.
- Test results with NC/1 or NC/1B EC outcomes were excluded if the DSA was less than the median DSA1+ of the positive tests.
- Test results with POS EC outcomes were excluded from GHS_{SUB} and GHS_{BORDER} classifications.
- An RV was assigned to each test result. For positive outcomes, the RV is the DSA1+/DSA05. For negative outcomes, the RV is the EC.
- RVs were ordered from low to high potency in the following order, with DSA1+/DSA05 values in descending order:

  NC → NC/1B → NC/1 → DSA1+/DSA05
- The value at the median position of the ordered RVs is designated the MLLP and used to classify substances as summarized in Table 2.

3. Median Sensitization Potency Estimate

- The median sensitization potency estimate (MSPE) approach was developed due to concern that the MLLP approach was insufficiently conservative in some cases.
  - Test results with NC/1 EC outcomes were excluded. Test results with POS EC outcomes were included.
- The MSPE was calculated by sorting all values from low to high potency in the following order, with DSA1+/DSA05 values in descending order:

  NC → NC/1B → DSA1+/DSA05 for 1B and 1B+ test results → POS → DSA1+/DSA05 for 1A- and 1A test results
- The value at the median position of the ordered RVs is designated the MSPE and used to classify substances as summarized in Table 3.

Table 2. MLLP approach for classifying substances with multiple discordant tests.

MLLP ($\mu g/cm^2$)	Classification Mode		
	GHS_{BIN}	GHS_{SUB}	GHS_{BORDER}
≤ 375	1	1A	1A
375 < MLLP ≤ 500			1*
500 < MLLP ≤ 625		1B	
> 625			1B
NC/1B	NA	NA	NC/1B
NC	NC	NC	NC

Table 3. MSPE approach for classifying substances with multiple discordant tests.

MSPE ($\mu g/cm^2$)	Classification Mode		
	GHS_{BIN}	GHS_{SUB}	GHS_{BORDER}
≤ 375	1	1A	1A
375 < MSPE ≤ 500			1*
POS		NA	
500 < MSPE ≤ 625		1B	
> 625			1B
NC/1B	NA	NA	NC/1B
NC	NC	NC	NC

Evaluation of Classification Approach with DSA1+ and DSA05

- Using DSA1+ (or DSA05) as the dose metric, 287 (288), 274 (277), and 1309 (1309) out of 1366 substances could be assigned GHS_{BIN}, GHS_{SUB}, and GHS_{BORDER} classifications, respectively (Table 4).
  - Among these substances, 143 (141), 134 (135), and 183 (180) substances had discordant HPPT test results and were evaluated with the WoE approaches for GHS_{BIN}, GHS_{SUB}, and GHS_{BORDER} classification (Table 4).

Table 4. Summary of substances classified with the modified GHS classification approach and the weight-of-evidence approaches

Number of Substances (N = 1366)…	DSA1+			DSA05		
	GHS_{BIN}	GHS_{SUB}	GHS_{BORDER}	GHS_{BIN}	GHS_{SUB}	GHS_{BORDER}
↳ Without overall classification.	1079	1092	57	1078	1089	57
↳ With overall classification…	287	274	1309	288	277	1309
↳ From 1 HPPT test result.	**122**	**119**	**1009**	**122**	**119**	**1009**
↳ From > 1 HPPT test results…	165	155	300	166	158	300
↳ That are concordant.	**22**	**21**	**117**	**25**	**23**	**120**
↳ That are discordant…	143	134	183	141	135	180
↳ And evaluated with only 1 WoE approach.	**15**	**17**	**37**	**10**	**12**	**37**
↳ And evaluated with > 1 WoE approaches…	128	117	146	131	123	143
↳ That are concordant.	**128**	**110**	**55**	**131**	**119**	**53**
↳ That are discordant and classified by expert judgement.	**na**	**7**	**2**	**na**	**4**	**2**
↳ That are discordant and classified by consensus.	**na**	**na**	**89**	**na**	**na**	**88**

Bold values indicate the final decision step used to classify a given chemical.

- The majority of substances did not have sufficient data to produce an unambiguous GHS_{BIN} and GHS_{SUB} outcome (Table 4).
  - GHS_{BORDER} provides information on the uncertainty of GHS_{BIN} and GHS_{SUB}. Using DSA1+ (or DSA05), 1021 out of 1366 (1022/1366) substances received the ambiguous GHS_{BORDER}, classification of NC/1B (Table 5c).
  - Further evaluation of test data indicated that the majority of test results in the HPPT database were negative but obtained at such low test concentrations/DSA values that a positive result at a higher concentration/DSA could not be ruled out with sufficient certainty.
- Comparison between DSA1+- and DSA05-based classifications demonstrated high concordance:
  - 287 substances could be assigned GHS_{BIN} classifications with both DSA1+ and DSA05. The 287/287 (100%) GHS_{BIN} outcomes for DSA1+ and DSA05 were concordant (Table 5a).
  - 274 substances could be assigned GHS_{SUB} classifications with both DSA1+ and DSA05. The 258/274 (94.16%) GHS_{SUB} outcomes for DSA1+ and DSA05 were concordant (Table 5b).

Table 5. Confusion matrices comparing classifications derived from DSA1+ and DSA05 for the three classification modes (a) GHS_{BIN}, (b) GHS_{SUB}, and (c) GHS_{BORDER}. Values indicate substance counts.

a. GHS_{BIN}

DSA1+ \ DSA05	1	NC	na	Total
1	234	0	0	234
NC	0	53	0	53
na	1	0	1078	1079
Total	235	53	1078	1366

b. GHS_{SUB}

DSA1+ \ DSA05	1A	1B	NC	na	Total
1A	55	9	0	0	64
1B	7	150	0	0	157
NC	0	0	53	0	53
na	0	3	0	1089	1092
Total	62	162	53	1089	1366

c. GHS_{BORDER}

DSA1+ \ DSA05	1A	1	1B	NC/1B	NC	na	Total
1A	41	7	4	0	0	0	52
1	3	15	9	0	0	0	27
1B	5	1	149	0	0	0	155
NC/1B	0	0	1	1021	0	0	1022
NC	0	0	0	0	53	0	53
na	0	0	0	0	0	57	57
Total	49	23	163	1021	53	57	1366

Reproducibility of HPPT-based WoE Classifications

- For substances with more than two unambiguous test results, reproducibility was estimated by comparing the individual ECs to the GHS_{BIN} and GHS_{SUB} classifications (Table 6).
- GHS_{BIN} and GHS_{SUB} are used as the "true" reference classification.
- Reproducibility for a substance is estimated as the fraction of unambiguous EC outcomes equal to the WoE-based overall classification.
- The mean reproducibility of the GHS_{BIN} classification was on the order of 99%, indicating that very few of the available test results disagreed with the overall classification outcome (Table 6).
- For GHS_{SUB}, the mean reproducibility was on the order of 80%, ranging from 76 to 84%.

Table 6. Reproducibility estimates for the WoE-based overall classifications of substances with at least two test results relevant to the respective classification mode.

Classification Mode	Number of test results available	Number of substances		Reproducibility (%) Mean (SD)	
		DSA1+	DSA05	DSA1+	DSA05
GHS_{BIN}	> 1	97	98	99.4 (3.6)	99.1 (4.9)
	> 2	53	54	98.9 (4.9)	98.3 (6.5)
	> 3	37	37	98.5 (5.8)	98.5 (5.8)
	> 4	27	27	99.8 (1.1)	99.8 (1.1)
GHS_{SUB}	> 1	96	97	82.5 (22.3)	84.2 (22.6)
	> 2	53	57	79.7 (21.7)	79.2 (23.6)
	> 3	40	39	77.2 (21.5)	77.3 (22.9)
	> 4	28	28	76.4 (21.0)	77.3 (23.0)

Integrating Data from HPPT and Local Lymph Node Assay

- To further explore the utility of our proposed classification approach, we evaluated the concordance of HPPT-based classifications with classifications using LLNA data for the set of reference chemicals in OECD Guideline 497 (OECD 2021).
  - For GHS_{BIN}, 56/196 OECD reference chemicals had classifications based on both HPPT and LLNA data. Concordance of HPPT with LLNA was 82% for both DSA1+ or DSA05 outcomes.
  - For GHS_{SUB}, 47/196 OECD reference chemicals had classifications based on both HPPT and LLNA data. Concordance of HPPT with LLNA was 60% for DSA1+ and 58% for DSA05 outcomes.
- We then developed a strategy to integrate HPPT-based reference classifications using DSA1+ or DSA05 with those obtained using LLNA data to develop an overall WoE classification (Figure 3). The concordance of LLNA, DSA1+, and DSA05 classifications with overall WoE classifications is shown in Table 7.

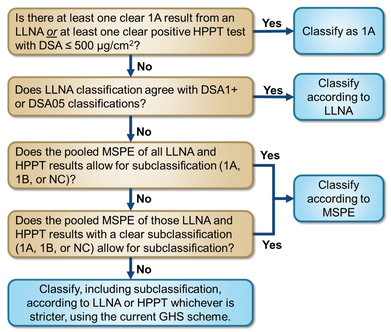

Figure 3. Decision scheme for obtaining an overall classification based on all available LLNA and HPPT data, in cases where the classifications based on LLNA, DSA1+, and DSA05 do not fully agree.

Table 7. Comparison of LLNA, DSA1+, DSA05, and overall WoE outcomes for the (a) GHS_{BIN} and (b) GHS_{SUB} classification modes for OECD reference substances with both LLNA- and HPPT-based classifications.

a. GHS_{BIN}

		Overall WoE	
		1	NC
LLNA	1	49	2
	NC	3	2
DSA1+	1	47	0
	NC	5	4
DSA05	1	47	0
	NC	5	4

b. GHS_{SUB}

		Overall WoE		
		1A	1B	NC
LLNA	1A	12	0	0
	1B	3	25	2
	NC	0	3	2
DSA1+	1A	12	4	0
	1B	3	20	0
	NC	0	4	4
DSA05	1A	10	3	0
	1B	5	21	0
	NC	0	4	4

Summary

- We collected a large data set of historical HPPT studies from the scientific literature to use as reference data for development of OECD Guideline 497.
- We developed a new approach for hazard and potency classification of these tests based on GHS categories (Figure 1b).
  - The modified GHS classification approach addresses uncertain or borderline results, incorporates the number of sensitized subjects to better inform on potency, and considers the validity of negative test results tested at low concentrations (Table 1).
  - WoE approaches were applied to resolve multiple discordant results for single substances. The WoE approaches provided reproducible results using either DSA1+ or DSA05 as the dose metric (Table 6).
  - Overall, substance classifications based on HPPT results were consistent with LLNA classifications.
- We developed a stepwise strategy that integrates LLNA results with HPPT results to address cases where there is higher uncertainty in the HPPT results (Figure 3).
- We conclude that using a modified GHS approach to classifying HPPT data provided good reproducibility and concordance with animal reference data while considering potency and uncertainty.

References

- Hoffmann et al. 2018. Crit Rev Toxicol 48(5):344-358. https://doi.org/10.1080/10408444.2018.1429385
- OECD. 2021. Test Guideline No. 497. https://doi.org/10.1787/b92879a4-en.
- UN. 2021. Globally Harmonized System of Classification and Labelling of Chemicals. https://unece.org/transport/standards/transport/dangerous-goods/ghs-rev9-2021

Acknowledgments and More Information

The Intramural Research Program of the National Institute of Environmental Health Sciences (NIEHS) supported this poster. Technical support was provided by Inotiv under NIEHS contract HHSN273201500010C. The views expressed above do not necessarily represent the official positions of any Federal agency. Since the poster was written as part of the official duties of the authors, it can be freely copied.

To get announcements of NICEATM activities, visit the NIH mailing list page for NICEATM News at https://list.nih.gov/cgi-bin/wa.exe?SUBED1=niceatm-l&A=1 and click "Subscribe."