

Classification of Human Reference Data for Skin Sensitization

J Strickland¹, D Allen¹, AM Api², J Gordon³, N Kleinstreuer⁴, HS Ko⁵, J Matheson³, HJ Thierse⁶, J Truax¹, M Herzler⁶

¹ILS, RTP, NC, USA; ²RIFM, Woodcliff Lake, NJ, USA; ³CPSC, Rockville, MD, USA; ⁴NIH/NIEHS/DNTP/NICEATM, RTP, NC, USA; ⁵FDA/CDER, Silver Spring, MD, USA; ⁶BfR, Berlin, Germany

Introduction

- A critical challenge for evaluating non-animal approaches for chemical safety testing is identifying appropriate reference data.
- Although humans are the species of interest, animal data are often preferred because they are more readily available, tests are standardized, and human testing may be viewed as unethical.
- Human reference data are needed to evaluate alternative test methods in a human-relevant manner. To this end, we collected historical human predictive patch test (HPPT) data used for the assessment of skin sensitization and evaluated the data's reliability.
- Data judged to be sufficiently reliable were used to assign skin sensitization potency classifications according to the Globally Harmonized System of Classification and Labeling of Chemicals (GHS) (Table 1).
- We then applied this approach to assign skin sensitization potency classifications to the 94 (out of 128) substances for which HPPT data were available in the Cosmetics Europe database. These classifications are being used to evaluate non-animal defined approaches for a guideline under development by the Organization for Economic Cooperation and Development (OECD).

Table 1. GHS Skin Sensitization Potency Categories

GHS Category	Induction Threshold for HPPT
1	Positive response with unknown dose per skin area (DSA)
1A (strong)	Positive response at DSA ≤ 500 µg/cm ² skin area
1B (other than strong)	Positive response at DSA > 500 µg/cm ² skin area
NC (not classified)	No sensitization response

Database

- The NTP Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) and the German Federal Institute for Risk Assessment (BfR) collected HPPT data, mostly from the scientific literature.
 - Most references were monographs on fragrance ingredients published by the Research Institute for Fragrance Materials (RIFM) since the early 1970s in *Food and Chemical Toxicology* (previously *Food and Cosmetics Toxicology*).
- HPPT data were from two main study types:
 - The human maximization test (HMT) includes a pre-treatment with sodium lauryl sulfate for non-irritating test substances. Five 48 h exposures are applied to a small group of subjects, usually 25.
 - The human repeat insult patch test (HR IPT) applies approximately ten 24-72 h exposures to approximately 100 subjects. Note that HR IPTs conducted by RIFM, after approval by an ethical review board, are not hazard assessments, but confirmatory tests for no-observed-effect levels obtained from other non-human tests.
- The current database of 2514 individual tests from approximately 1800 references was evaluated for reliability (Table 2) to mitigate the shortcomings of the existing data.

Difficulties Using Human Predictive Patch Test Data

- No test guidelines were available; older tests were not standardized.
- Documentation of test protocol and test outcomes is often poor.
- Quality control measures are sometimes not reported.
- Primary test reports are often unavailable; tests are reported in secondary sources.
- Because tests used only one concentration, there is no dose-response information with which to select appropriate no-observed-effect and lowest-observed-effect levels appropriate for risk assessment.
 - Negative results may be uncertain because the concentration tested may have been too low to produce positive results.
 - Positive test results may provide uncertain estimates of potency because the concentration tested may have been higher than the concentration needed for a threshold positive response.

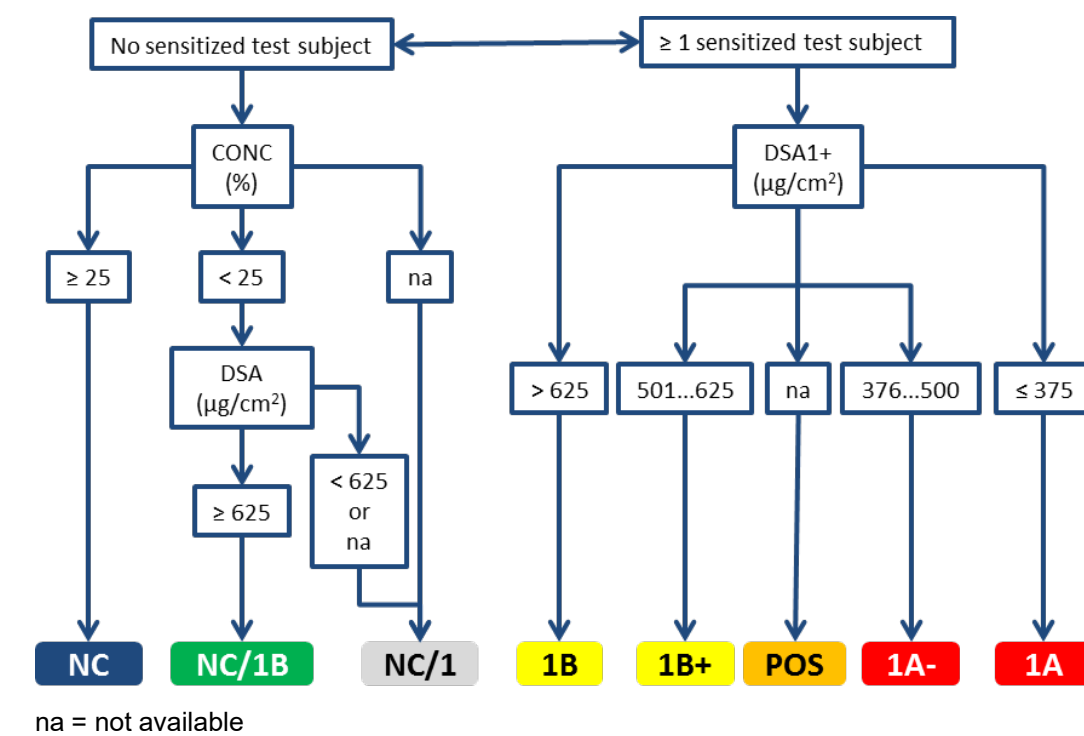
Table 2. Relative Reliability Scores for Individual HPPTs

Relative Reliability Score (RRS)	Description	Number of Tests
RRS = 1: Highly reliable	All of the following information is available from the primary report: <ul style="list-style-type: none"> Substance identity Test type Test concentration, volume and patch size, or a direct statement regarding the dose per skin area Number of subjects tested Number of subjects reacting 	380 (15.1%)
RRS = 2: Sufficiently reliable	All information listed for RRS = 1 is available, but at least some information comes from secondary references.	939 (37.4%)
RRS = 3: Reliable with some uncertainty	All information listed for RRS = 1 is available, but at least some information has to be inferred, for example, from publications by authors with access to the primary reference.	497 (19.8%)
RRS = 4: Partially reliable	Not all relevant details are available (nor can they be inferred) to calculate dose per skin area and incidence.	112 (4.5%)
RRS = 5: Not reliable	One of the following conditions applies to the data: <ul style="list-style-type: none"> No information is available to enable the calculation of incidence or dose per skin area. The identity or composition of the test substance is not clear (as for natural oils or extracts). The test substance contains structurally diverse constituents and/or its composition is variable. 	586 (23.3%)

Classification of HPPT Using Extrapolated GHS Classifications to Incorporate Uncertainty

- In HPPTs, a substance is considered a skin sensitizer if at least one test subject exhibits a sensitization reaction. Table 1 provides the GHS criteria for classifying sensitizers as 1A (strong) and 1B (other than strong) sensitizers, based on the dose per skin area (DSA) applied during induction. If the DSA is unknown for a positive response, a subcategory cannot be assigned, and the substance is classified as Category 1.
- Test results in our database were classified using "extrapolated" GHS classifications.
 - The applied DSA for a positive test was converted into a "DSA1+" value, the hypothetical DSA producing exactly one sensitized test subject, which was calculated by DSA/number of sensitized subjects.
 - Aside from a normalization of the test results, extrapolation compensates for the discrepancy between applied DSA and GHS cut-offs, where applicable. In addition, "uncertainty/borderline zones" were introduced to account for the inherent uncertainty of classifications based on test results close to the 1A/1B and 1B/NC borders.
 - Figure 1 provides a decision tree for classification.
- Criteria for negative tests:
 - Negatives tested at concentrations ≥25% were "not classified" (NC). This concentration cutoff was selected because 99% of the positive results in the database used concentrations less than this. Thus, substances were not likely to be positive if tested at higher concentrations.
 - Negatives tested at concentrations <25% and producing a DSA ≥625 µg/cm² were classified as NC/1B, an ambiguous outcome that enables exclusion of a strong skin sensitization potential.
 - Negatives tested at <25% and producing a DSA <625 µg/cm² were considered NC/1, an ambiguous classification that provided no information on the skin sensitization potential of the chemical.
- Criteria for positive tests:
 - Positives with DSA1+ ≤375 µg/cm² or >625 µg/cm² were classified as 1A or 1B sensitizers, respectively.
 - Positives with DSA1+ between 375 µg/cm² and 500 µg/cm² were assigned the borderline classification of 1A- because they suggested a strong sensitization potential (1A) with some (non-quantifiable) likelihood of overclassification (i.e., classification as a stronger sensitizer than it actually is).
 - Positives with DSA1+ between 501 µg/cm² and 625 µg/cm² were assigned the borderline classification of 1B+ because they suggested moderate sensitization potential (1B) with some (non-quantifiable) likelihood of underclassification (i.e., classification as a weaker sensitizer than it actually is).
 - Positives for which a DSA1+ was not available were classified as positive (POS) because subcategorization was not possible.

Figure 1. Decision Tree for Extrapolated GHS Classifications



Evaluation of Cosmetics Europe Database

- The Cosmetics Europe reference database of 128 substances (Hoffmann et al. 2018) is currently being used to evaluate defined approaches for skin sensitization assessments for a guideline project under development at the OECD.
- Our current HPPT database included 555 HPPTs among 94 substances in the Cosmetics Europe reference database. We classified results from tests with RRS = 1-4.
- Substances with multiple non-concordant results were classified by combining the multiple results using a weight-of-evidence approach (Figure 3) and two median-based methods: the median-like location parameter (MLLP) (Hoffmann et al. 2018) (Table 3) and the median sensitization potency estimate (MSPE), a slightly modified version of the MLLP (Table 4). The workflow is shown in Figure 2.
- Substances were classified in three different modes based on GHS categories:
 - GHS_{BIN}: binary classification as a sensitizer or nonsensitizer.
 - GHS_{SUB}: three classes, including 1A sensitizers, 1B sensitizers, and nonsensitizers.
 - GHS_{BORDER}: five classes, including the three classes used in GHS_{SUB} with different criteria, 1 (subclassification not possible), and NC/1B (substance may or may not be a sensitizer, but 1A can be ruled out).

Figure 2. Workflow for GHS Classification of Substances Using HPPT Data

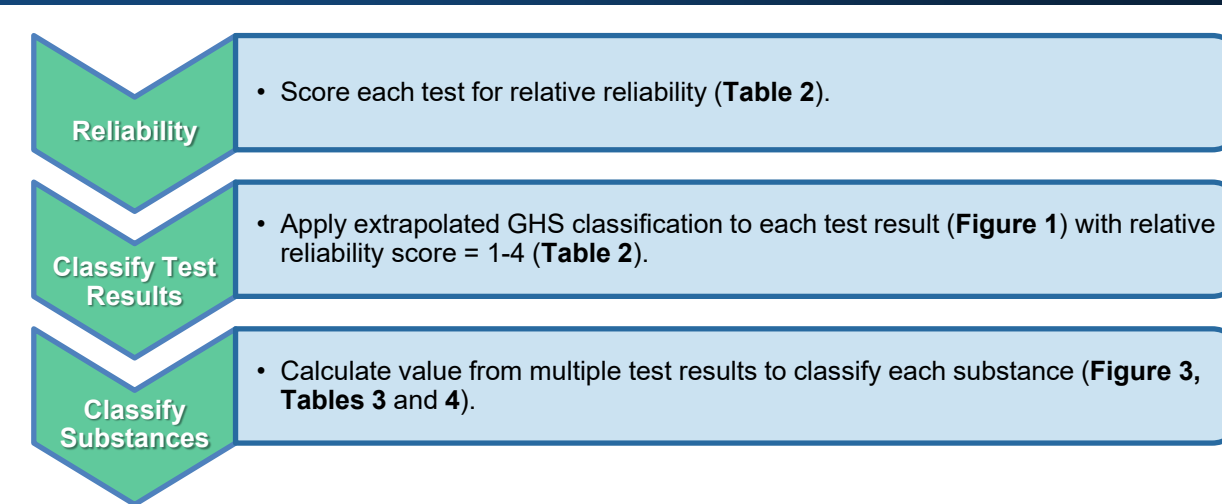


Figure 3. Weight-of-Evidence Approach to Classification of Each Substance with Multiple Nonconcordant Tests

Individual Test Data		Combined Chemical Classification		
Extrapolated Classification (Figure 1)	Score	WoE Score	Classification Mode	
1A	2	1.76-2	GHS _{BIN}	1A
1A-	1.75	1.51-1.75	GHS _{SUB}	1A*
POS	1.5	1.50	GHS _{BORDER}	1**
1B+	1.25	1.26-1.49	GHS _{SUB}	1B*
1B	1	0.76-1.25	GHS _{BORDER}	1B
NC/1	NA	0.26-0.75	GHS _{BORDER}	NC/1B
NC/1B	0.5	0-0.25	GHS _{BORDER}	NC
NC	0		GHS _{BORDER}	NC

NA = not applicable

* For GHS_{SUB}, individual test results with 1A and 1A- are considered as 1A, and both 1B and 1B+ are considered as 1B.

** For GHS_{BORDER}, 1A- and 1B+ are considered as Category 1.

NC/1 tests were excluded from the combined chemical classification.

Median-like Location Parameter Approach

- The MLLP approach (Hoffmann et al. 2018) is summarized in Table 3.
 - Test results with NC/1 or NC/1B outcomes were included as negatives if tests with a positive outcome (1A, 1B, or 1) were available and the concentration applied in the ambiguous test was higher than or equal to the median DSA1+ of the positive tests.
 - The substance was considered a sensitizer under GHS_{BIN} if the majority of the unambiguous tests were positive; the GHS_{BIN} classification was NC if the majority of the unambiguous tests were negative. If the number of positives and negatives was equal, the overall reference classification was 1.
 - For GHS subcategorization (GHS_{SUB}), ambiguous test results for subclassification (1) were excluded (in addition to the NC/1 and NC/1B tests with concentrations that were too low).
 - Classification was performed using the approach shown in Table 3 after calculating the MLLP of the remaining tests.

Table 3. MLLP Approach to Classification of Each Substance

Median-like Location Parameter (µg/cm ²)	Classification Mode		
	GHS _{BIN}	GHS _{SUB}	GHS _{BORDER}
≤ 375		1A	1A
375 < MLLP ≤ 500	1	1A	1
500 < MLLP ≤ 625		1B	1B
> 625			
NC/1B	NA	NA	NC/1B
NC	NC	NC	NC

NA = not applicable

Median Sensitization Potency Estimate Approach

- The MSPE approach, a slightly modified version of the MLLP approach, is summarized in Table 4.
 - NC/1 test results were excluded, but positive test results with no DSA1+ values were included in the median calculation.
 - The MSPE was calculated by sorting all values from low to high potency in the following order:
 - NC → NC/1B → Numerical DSA1+ values > 500 µg/cm² in descending order → POS → Numerical concentration values ≤ 2.0% in descending order.
 - When only positive tests were available and the number of 1A results equaled that of 1B, the MSPE was "POS".

Table 4. MSPE Approach to Classification of Each Substance

Median Sensitization Potency Estimate (µg/cm ²)	Classification Mode		
	GHS _{BIN}	GHS _{SUB}	GHS _{BORDER}
≤ 375		1A	1A
375 < MSPE ≤ 500		1A	1A
POS	1	NA	1
500 < MSPE ≤ 625		1B	1B
> 625			
NC/1B	NA	NA	NC/1B
NC	NC	NC	NC

NA = not applicable

Acknowledgements

We thank Catherine Sprankle, ILS, for editing the poster text. This project was funded in part with federal funds from the National Institute of Environmental Health Sciences, National Institutes of Health under Contract No. HHSN273201500010C to ILS in support of NICEATM.

The views expressed above do not necessarily represent the official positions of any federal agency.

Subscribe to the NICEATM News Email List



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."

Results

- For the very few substances for which the classifications from the weight-of-evidence score, the MLLP approach, and/or MSPE approach disagreed, the classifications were assigned, after discussion, by the OECD Expert Group on Defined Approaches for Skin Sensitization.
- We excluded 48/128 substances in the Cosmetics Europe database:
 - 14 with unclear or variable composition (e.g., jasmine absolute, oakmos, etc.).
 - 34 with no reliable HPPT data.
- The 80 remaining substances were classified in three different modes based on GHS categories (Table 5).

Table 5. HPPT Classification Results for 80 Substances from the Cosmetics Europe Database Using Three Classification Modes for the GHS

GHS _{BIN} : Binary Classification as Sensitizer or Not Classified	
Category 1 Sensitizers	38
Not Classified	8
Classification Ambiguous (but substance is not Category 1A)	34
GHS _{SUB} : Classification as 1A Sensitizer, 1B Sensitizer, or Not Classified*	
Category 1A (strong) Sensitizers	13
Category 1B (other than strong) Sensitizers	25
Not Classified	8
Classification Ambiguous (but substance is not Category 1A)	34
GHS _{BORDER} : Classification as 1A Sensitizer, 1B Sensitizer, Sensitizer with Unclear Potency, Not Classified, Ambiguous**	
Category 1A (strong) Sensitizers	9
Category 1 Sensitizers (unclear potency)	7
Category 1B (other than strong) Sensitizers	22
Not Classified	8
Classification Ambiguous (but substance is not Category 1A)	34

* For GHS_{SUB}, individual test results with 1A and 1A- are considered as 1A, and both 1B and 1B+ are considered as 1B.

** For GHS_{BORDER}, 1A- and 1B+ are considered as Category 1.

Summary and Discussion

- We collated a large dataset of historical HPPT studies from the scientific literature and analyzed the data for reliability.
- New methodology was developed to explore the potential of these data to classify chemicals using the GHS, including methodology to better account for uncertain/borderline cases.
- Following final processing, the entire NICEATM/BfR HPPT database will be made available to the public in the Integrated Chemical Environment (ICE) database (<https://ice.niehs.nih.gov/>).
 - See Poster 2976 (Poster Board P607): NICEATM Computational Tools and Resources Supporting Alternative Test Method Development and Evaluation.
 - Wednesday, March 18, in Computational Toxicology II.
 - See ICE demonstrations at the National Toxicology Program Booth 1517
 - 10:00 a.m. – 12:00 noon, Monday, March 16, through Wednesday, March 18.
- The database will also be donated to the OECD QSAR Toolbox (<https://qsartoolbox.org/>).
- Future publications about the entire HPPT database will:
 - Describe collection and curation of the HPPT data.
 - Characterize sources of uncertainty and variability in HPPT results.
 - Consider practical uses, including classification and potency assessment and how to use these data for dose-response analysis.

References

Hoffmann S, et al. 2018. Non-animal methods to predict skin sensitization (I): the Cosmetics Europe database. *Crit Rev Toxicol* 48(5):344-358.
 UN. 2019. Globally Harmonized System of Classification and Labeling of Chemicals (GHS). 8th Revised Edition. United Nations: New York.

