Subcategorization of Skin Sensitizers into UN GHS Categories Using GARD®skin Dose-Response

Andy Forreryd[1], Robin Gradin[1], Tim Lindberg[1], Henrik Johansson[1] | [1] SenzaGen AB, Lund, Sweden

- *In Vitro* Skin Sensitization Testing
- Potency Classification
- Meta-Analysis of Peer-Reviewed GARD® Data

New Approach Methodology

SENZA GEN

Abstract ID #3242

1. Introduction

Current trends in predictive toxicology advocate a transition from *in vivo* methods to New Approach Methodologies (NAMs). For the purpose of hazard identification of skin sensitizers, numerous successful innovations have led to the development and validation of several NAMs. Such methods can further be combined into Defined Approaches for Skin Sensitization (DASS), which facilitate also potency-associated UN GHS subcategorizations.

While this has constituted a significant milestone for potency-assessment in a regulatory context, opportunities exist to further improve or complement strategies for potency-associated subclassifications, in terms of performance, applicability and the number of data sources required to obtain robust and accurate results.

The herein presented meta-analysis of GARD®skin Dose-Response demonstrates that the method can, on its own, provide UN GHS subcategorizations at performance and reproducibility levels comparable to, or surpassing, those of the DASS methods, as defined in OECD GD 497.

2. Methods

The GARD®skin method (OECD TG 442E) is an *in vitro* assay for the assessment of chemical skin sensitizers (Figure 1). The assay readout is a Decision Value (DV), where a positive value indicates a skin sensitizer, while a negative one indicates a Non-sensitizer.

The GARD®skin Dose-Response (Figure 2) is an adaptation of GARD®skin and is currently included in the OECD Test Guideline Program (TGP 4.106). In this assay, the test chemicals are evaluated in an extended range of concentrations and provides a quantitative estimation of sensitizing potency, referred to as cDV_0, which corresponds to the lowest required dose able to generate a positive mean DV. To facilitate interpretation, cDV_0 can be further converted into a Predicted Potency Value (PPV) in the unit of ug/cm^2, which enables intuitive understanding in a parameter space identical to those utilized by, e.g., the LLNA and human HRIPT methods. As such, the UN GHS classification threshold of 500 ug/cm^2 (corresponding to an LLNA EC3 of 2%) can be applied to subcategorize test chemicals as 1A or non-1A skin sensitizers.

Lastly, a Borderline Range (BR) can be applied to the classification threshold, thereby recognizing the uncertainty of measurements, by considering results within the BR as inconclusive.

3. Results

A meta-analysis of peer-reviewed GARD®skin Dose-Response data (comprising >150 unique test chemicals) was conducted to evaluate the method's reproducibility and predictive performance.

The reproducibility of the method was evaluated by considering all chemicals with repeated assessments in at least two independent experiments, across the published data sources (n=42). Results indicate that the method was 92.9% reproducible when applying a "hard" classification threshold of 500 ug/cm^2, with concordant subcategorizations as cat. 1A or non-cat. 1A obtained for 39 out of the 42 chemicals. When implementing the proposed BR, 7 test chemicals were considered inconclusive. All remaining 35 test chemicals had concordant results across experiments, indicating an estimated reproducibility of 100%. A graphical representation of generated results for the 42 test chemicals with repeated test results is presented in Figure 3.

The predictive performance of the method was evaluated by comparing obtained results with corresponding reference values (human and/or LLNA) available in the Annex 2 of the Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitization (n=87). Results indicated that GARD®skin Dose-Response was able to subcategorize test chemicals as cat. 1A or non-cat. 1A with a predictive accuracy of 88.6% and 91.3%, when considering human (n=38) and LLNA (n=69) references, respectively, when utilizing a "hard" classification threshold of 500 ug/cm^2. When implementing the proposed BR, the predictive accuracy increased to 90.9% and 92.2%, when considering human (n=33) and LLNA (n=64) references, respectively.

To put obtained estimations of predictive performance into context, the above summarized results were compared with the current in vitro gold standard for UN GHS subcategorization, i.e., the ITS v.1 and v.2 described in OECD GL 497. A comparative analysis of the different methods' predictive performances on an identical dataset is presented in Table 1, including various performance statistics. Here, it was concluded that for 1A/non-1A subcategorization, GARD®skin Dose-Response is on its own predictive at levels similar to those of ITS v.1 and v.2, or slightly improved. When implementing the proposed borderline range, the GARD®skin Dose-Response performance was additionally improved further beyond the performance observed for ITS v.1 and v.2.

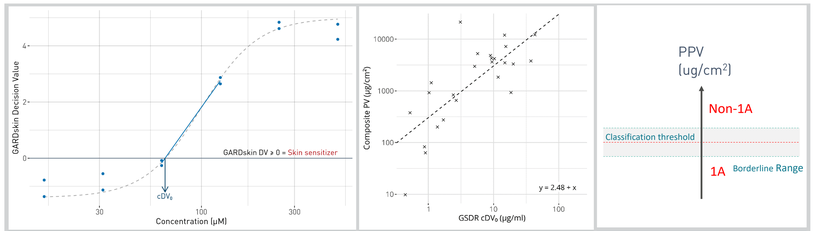
Expose cells to the test sample at determined concentration.

Measure the gene expression levels of the 196 biomarkers.

Analyse gene expression patterns using the GARD® Data Analysis Application.

Classify the test sample based on the Decision Value (DV).

Figure 1. Overview of the GARD®skin method (OECD TG 442E). The assay readout is a Decision Value (DV), where DV ≥ 0 indicates a skin sensitizer and DV < 0 indicates a Non-skin sensitizer.



cDV$_0$ estimation from GARD®skin Dose-Response | Conversion of cDV$_0$ to PPV | UN GHS potency classification

Figure 2. Schematic overview of GARD®skin Dose-Response for UN GHS classification. Step 1: Test chemical-specific cDV_0-values are estimated from dose-response testing. Step 2: Individual cDV_0-values are converted to PPV (composite PV). Step 3: The individual PPV:s are evaluated with respect to the UN GHS classification threshold of 500 ug/cm^2.

Table 1. Comparison of results generated with GARD®skin Dose-Response and ITS v.1 and v.2 for an overlapping dataset comprising 87 test chemicals.

Predictor/Assay	Reference[3]	Accuracy	Sensitivity	Specificity	Balanced Accuracy	N Classifications	N BL
GSDR[1]	LLNA	91.3	77.8	93.3	85.6	69	0
GSDR BR[2]	LLNA	92.2	85.7	93.0	89.3	64	5
ITS v1/v2[3, 4]	LLNA	89.9	77.8	91.7	84.7	69	0
GSDR[1]	Human	88.6	77.8	92.0	84.9	38	0
GSDR BR[2]	Human	90.9	75.0	96.0	85.5	33	5
ITS v1/v2[3, 4]	Human	86.8	66.7	93.1	79.9	38	0

[1]) GARD®skin Dose-Response with a hard classification threshold.
[2]) GARD®skin Dose-Response with implemented borderline range.
[3]) Extracted from (Annex 2 of the supporting document to OECD GD 497).
[4]) The ITS v1 and v2 perform identically in the herein investigated dataset.

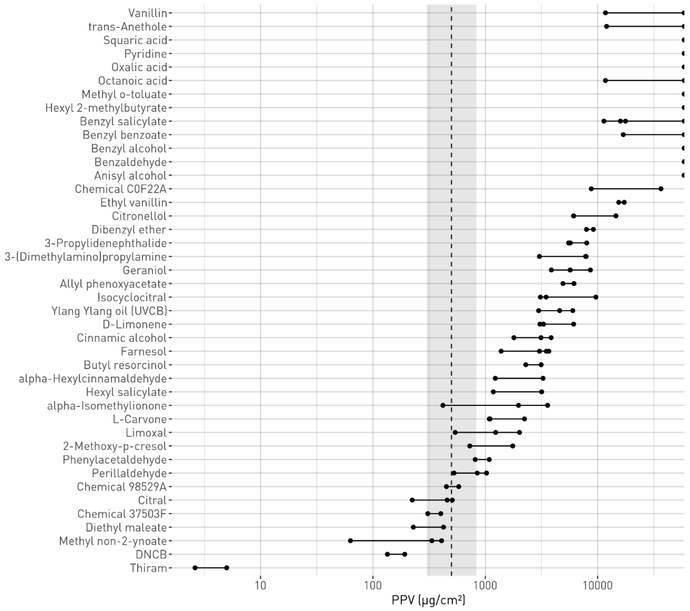


Figure 3. Graphical representation of GARD®skin Dose-Response used to estimate reproducibility. Data points indicate obtained PPV:s from repeated assessments of each test chemical. The classification threshold (500 µg/cm^2) is indicated with a dashed line. The borderline range is indicated with a grey area.

4. Conclusion

GARD®skin Dose-Response provides accurate and robust UN GHS subclassifications, with performance and reproducibility levels comparable to, or surpassing, those of current *in vitro* counterparts.

Contact: Henrik Johansson, henrik.johansson@senzagen.com