Evaluation of the GARD™skin Sensitization Test Method Using Substances of Regulatory Interest

National Institutes of Health
U.S. Department of Health and Human Services

Evaluation of the Genomic Allergen Rapid Detection (GARD™)skin Test Method Using Substances of Regulatory Interest

National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM)

National Institute of Environmental Health Sciences
National Institutes of Health
Department of Health and Human Services

June 2025

FOREWORD

The National Toxicology Program (NTP), established in 1978, is an interagency program within the Public Health Service of the U.S. Department of Health and Human Services. Its activities are executed through a partnership of the National Institute for Occupational Safety and Health (part of the Centers for Disease Control and Prevention), the U.S. Food and Drug Administration (primarily at the National Center for Toxicological Research), and the National Institute of Environmental Health Sciences (NIEHS; part of the National Institutes of Health), where the program is administratively located. NTP offers a unique venue for the testing, research, and analysis of agents of concern to identify toxic and biological effects, provide information that strengthens the science base, and inform decisions by health regulatory and research agencies to safeguard public health. The Division of Translational Toxicology (DTT) at NIEHS supports NTP by developing and applying new and improved methods and approaches that advance toxicology and better assess health effects from environmental exposures.

The NTP Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) is a DTT office focused on the development and evaluation of alternatives to animal use for chemical safety testing. NICEATM was established by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) Authorization Act of 2000 (42 U.S.C. 285 *l*-3) to provide support to ICCVAM. NICEATM and ICCVAM work collaboratively to evaluate new and improved testing approaches applicable to the needs of U.S. federal agencies.

NICEATM publishes reports of its test method development and evaluation activities in the scientific literature. NICEATM also issues reports of ICCVAM test method evaluations and other communications and makes these available on the NTP website, where they are available free of charge. Data from these studies are included in NTP's Chemical Effects in Biological Systems database.

For questions about the reports and studies, please contact NICEATM.

TABLE OF CONTENTS

TABLES

FIGURES

APPENDICES

EXECUTIVE SUMMARY

Integrated approaches to assess skin sensitization potential and predict the potency category combine multiple types of methods to overcome the limitations of individual tests. These integrated approaches can be used to derive a hazard or potency characterization. One type of integrated approach is the defined approach (DA), which uses predetermined data sources and defined interpretation approaches to arrive at an outcome (e.g. hazard or potency) without using expert judgement. Skin sensitization DAs combine non-animal tests that align with multiple key events in the adverse outcome pathway for skin sensitization to inform on chemical hazard and potency.

This report summarizes an evaluation of the performance of the Genomic Allergen Rapid Detection (GARD)™skin assay for predicting skin sensitization hazard and potency. It can be used to evaluate dendritic cell activation, which is the third key event in the adverse outcome pathway. The evaluation compared the results of the GARDskin assay alone and in applicable DAs with reference results derived from the murine local lymph node assay (LLNA; OECD Test Guideline 429, 2010), human skin sensitization reference data, and three other non-animal skin sensitization test methods. These three non-animal test methods assay the first three key events in the adverse outcome pathway for skin sensitization (covalent interaction with skin proteins, activation of inflammatory cytokines, and dendritic cell activation). All non-animal methods were also applied to the appropriate DAs to evaluate and compare their results against DAs that included the GARDskin assay. In total, 10 in chemico/in vitro methods and DAs were evaluated and compared to each other and to the LLNA.

This evaluation compared the GARDskin assay against tests of nominated substances using methods described in internationally harmonized test guidelines (TGs) or other technical guidelines.

Test Method/Defined Approach (DA)	**Test Guideline**	**Reference**
Direct peptide reactivity assay (DPRA)	TG 442C	OECD, 2024a
KeratinoSens test	TG 442D	OECD, 2024b
Human cell line activation test (h-CLAT)	TG 442E	OECD, 2024c
GARDskin assay	TG 442E	OECD 2023c
2 out of 3 (2o3) DA (DPRA, KeratinoSens, h-CLAT)	TG 497	OECD, 2025
Integrated Testing Strategy (ITS) DA (DPRA, h-CLAT, OECD Toolbox)	TG 497	OECD, 2025
2o3 DA GARDskin (DPRA, KeratinoSens, GARDskin)	TG 497	OECD, 2025
ITS DA GARDskin (DPRA, GARDskin, OECD Toolbox)	TG 497	OECD, 2025
Key Event 3/1 Sequential Testing Strategy (KE 3/1 STS) DA (h-CLAT, DPRA)	N/A	EPA, 2018
KE 3/1 STS DA GARDskin (GARDskin, h-CLAT)	N/A	EPA, 2018

Recognizing a need to better characterize or expand upon the types of substances that can be assessed with these non-animal methods, the National Toxicology Program Interagency Center for the Evaluation of Alternative Methods (NICEATM) solicited nominations of substances subject to regulatory assessments for skin sensitization from member agencies of the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM). Initially, 181 substances were nominated. Subsequently, NICEATM asked ICCVAM member agencies to identify a subset of the nominated substances for an initial evaluation of the GARDskin assay's robustness. Thirty-one substances were identified, and NICEATM coordinated their testing in the GARDskin assay. As part of the larger effort, these substances had already undergone testing in the DPRA, h-CLAT, and KeratinoSens, and had predictions of skin sensitization hazard made using the in silico read-across algorithms provided by the OECD QSAR Toolbox. This evaluation also included compiling reference data from the published literature, publicly available databases, or directly from the nominating agencies to allow for comparison with the non-animal results. Of the 31 substances assessed as part of the GARDskin project, reference LLNA data were available for 30 substances and human reference data were available for seven substances.

Overall, the GARDskin assay performed well against the other assays for predicting skin sensitization hazard and potency classifications.

Predictions of skin sensitization potential derived from both individual methods and the DAs were evaluated for concordance with one another and with human and LLNA reference data. Additionally, the predictive performance of the GARDskin assay was evaluated by assessing the accuracy of its predictions against the human and LLNA reference data. These assessments evaluated sensitivity (positive predictivity) and specificity (negative predictivity). The GARDskin assay as a standalone method had the highest concordance (63%), accuracy (63%), balanced accuracy (62%), and sensitivity (81%) when compared to LLNA reference data. Among the in chemico/in vitro methods evaluated, the GARDskin assay had the second-highest specificity (43%), only lower than the DPRA (54%). Performance statistics for the DAs that included the GARDskin assay suggested that its addition improved those DAs, although performance statistics for the DAs including the GARDskin assay were no better than GARDskin alone. The false positive and false negative rates among the test methods were generally high. The GARDskin test had the lowest false negative rate (19%) and the second-lowest false positive rate (57%), again only below DPRA (46%). When evaluating for potency, the ITSv2 containing the GARDskin assay was more accurate (53%) for potency and less likely to overpredict potency (5%) as compared to the standard ITSv2 (37% accuracy) and (18% potency). Concordance of the ITSv2 GARDskin with LLNA reference data was 53%.

Only seven substances had human hazard classification data; therefore, these results are considered preliminary. Performance of the GARDskin assay in predicting hazard classifications based on human results was second to the KeratinoSens for accuracy and balanced accuracy, although overall, the methods were comparable. The GARDskin assay was one of eight methods with 0% false negative rates. Potency performance using human data as a reference for ITSv2 both with and without GARDskin was similar.

Although the use of GARDskin assay alone had higher balanced accuracy than all the DAs for this set of substances, use of a DA over a single method confers more confidence in the outcomes because the DAs cover multiple key events of an adverse outcome pathway. Importantly, since only 31 substances were evaluated, these results cannot be used to infer that

the GARDskin assay will always have higher performance than DAs when applied to other sets of substances.

LIST OF ACRONYMS AND ABBREVIATIONS

2o3	2 out of 3
AOP	Adverse outcome pathway
ARE	Nrf2-dependent antioxidant response element
BRT	Burleson Research Technologies, Inc.
CAS RN	Chemical Abstracts Service Registry Number
CCTE	Center for Computational Toxicology and Exposure (U.S. Environmental Protection Agency)
CPSC	U.S. Consumer Product Safety Commission
CV75	Concentration needed to produce viability of 75%
DAs	Defined approaches
DASS	Defined approaches for skin sensitization
DPRA	Direct peptide reactivity assay
DTT	Division of Translational Toxicology (National Institute of Environmental Health Sciences)
EC150	Effective concentration of a test substance that produces a 150% increase in the expression of the CD86 cell surface marker
EC200	Effective concentration of a test substance that produces a 200% increase in expression of the CD54 cell surface marker
EPA	U.S. Environmental Protection Agency
FDA	U.S. Food and Drug Administration
FN	False negative
FP	False positive
GARD	Genomic Allergen Rapid Detection
GDAA	GARD™ Data Analysis Application
GHS	United Nations Globally Harmonized System of Classification and Labeling of Chemicals
h-CLAT	Human cell line activation test
ICCVAM	Interagency Coordinating Committee on the Validation of Alternative Methods
ICE	Integrated Chemical Environment
ITSv2	Integrated Testing Strategy version 2
KE	Key event
KE 3/1 STS	Key events 3 and 1 sequential testing strategy
LLNA	Murine local lymph node assay
NAMs	New approach methodologies
NA	Not applicable

NC	Not Classified (GHS hazard classification)
NICEATM	National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods
NTP	National Toxicology Program
OECD	Organisation for Economic Co-operation and Development
OPERA	Open Structure-activity/property Relationship App
OPP	Office of Pesticide Products (U.S. Environmental Protection Agency)
OPPT	Office of Pollution Prevention and Toxics (U.S. Environmental Protection Agency)
QSAR	Quantitative structure-activity relationships
RCC	Report Code Count
RLF	Reporter Library File
SMILES	Simplified Molecular Input Line Entry System
TG	Test guideline
TN	True negative
TP	True positive
UN	United Nations

ABOUT THIS REPORT

Authors

Judy A. Strickland[1], Heather Maldonado[1], Tripp LaPratt[1], James F. Truax[1], Kimberly T. To[1], Elizabeth Farley-Dawson[1], Emily Reinke[1], David G. Allen[1], Dori Germolec[2], Nicole Kleinstreuer[3]

[1]Inotiv, Morrisville, NC USA

[2]National Institute of Environmental Health Sciences, Division of Translational Toxicology, Systems Toxicology Branch, P.O. Box 12233, Research Triangle Park, NC USA

[3]National Institute of Environmental Health Sciences, Division of Translational Toxicology, Predictive Toxicology Branch, National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods, P.O. Box 12233, Research Triangle Park, NC USA

Integrated Laboratory Systems, LLC, an Inotiv Company, Research Triangle Park, North Carolina, USA

Managed project, developed study design, evaluated and interpreted results, and prepared report.

Judy A. Strickland, Ph.D., D.A.B.T., Study Director

James F. Truax, M.S.

Kimberly To., Ph.D.

Tripp LaPratt, M.S.

Elizabeth A. Farley-Dawson, Ph.D.

Heather Maldonado, Ph.D.

Emily Reinke., Ph.D., D.A.B.T., Principal Investigator (December 2023 – present)

David G. Allen, Ph.D., Principal Investigator (until December 2023)

National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods, Division of Translational Toxicology, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA

Project oversight, evaluated and interpreted results, and reviewed report

Nicole C. Kleinstreuer, Ph.D., NICEATM Director

Dori Germolec, Ph.D.

Contributors

Burleson Research Technologies, Morrisville, North Carolina, USA

Performed all testing and provided results

Vic Johnson, Ph.D.

U.S. Food and Drug Administration

Nominated substances for testing

Office of Pesticide Programs, Office of Pollution Prevention and Toxics, U.S. Environmental Protection Agency, Washington, District of Columbia, USA

Nominated substances for testing

Office of Pollution Prevention and Toxics, Office of Chemical Safety and Pollution Prevention, U.S. Environmental Protection Agency, Washington, District of Columbia, USA

Provided nominations for chemical testing

Center for Computational Toxicology and Exposure, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina, USA

Nominated substances for testing

U.S. Consumer Protection Safety Commission, Rockville, Maryland, USA

Nominated substances for testing

Division of Translational Toxicology, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA

Provided nominations for chemical testing and provided contract oversight

Pei-Li Yao, Ph.D.

Proofread report

Mary S. Wolfe, Ph.D.

Review

This document has been reviewed by, and reflects comments received from, the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) Skin Sensitization Expert Group. The Skin Sensitization Expert Group is comprised of experts from seven different ICCVAM member agencies, who are listed below.

- Joanna Matheson, Ph.D., U.S. Consumer Product Safety Commission
- Renee Beardslee, PhD, U.S. Environmental Protection Agency, Office of Pollution Prevention and Toxics
- Simona Bancos, Ph.D., U.S. Food and Drug Administration (FDA), Center for Devices and Radiological Health
- Hon-Sum Ko, Ph.D., FDA, Center for Drug Evaluation and Research

Acknowledgements

This project was supported by the Intramural Research Program (ES# 103386) at the National Institute of Environmental Health Sciences, National Institutes of Health under contract HHSN273201500010C.

PUBLICATION DETAILS

Publisher: Division of Translational Toxicology

Publishing Location: Research Triangle Park, NC

DOI: https://doi.org/10.22427/NICEATM-06

Report Series: NICEATM Report

Report Series Number: 06

Official citation: Strickland JA, Maldonado H, LaPratt T, Truax JF, To KT, Farley-Dawson E, Reinke E, Allen DG, Germolec D, Kleinstreuer N. 2025. Evaluation of the Genomic Allergen Rapid Detection (GARD™)skin test method using substances of regulatory interest. Research Triangle Park, NC: Division of Translational Toxicology. NICEATM Report 06.

1. INTRODUCTION

Recent advances in the development of alternative or new approach methodologies (NAMs) for skin sensitization testing have made it possible to combine the results from several test methods as replacements for traditional in vivo test methods such as the guinea pig maximization test, the Buehler test, or the murine local lymph node assay (LLNA). In vitro/in chemico methods such as the direct peptide reactivity assay (DPRA), the KeratinoSens™ assay, and the human cell line activation test (h-CLAT) can now be used as inputs in defined approaches (DAs) to evaluate substances for skin sensitization hazard and potency classification. DAs use predetermined data sources to arrive at an outcome without the use of expert judgement. This is accomplished through the application of the results of the methods used in the DA in an explicitly defined data interpretation protocol.

The Organisation for Economic Co-operation and Development (OECD) issued the Guideline on Defined Approaches for Skin Sensitisation (DASS), Guideline No. 497, in 2021, with an update in 2025 (OECD 2025). OECD Guideline 497 was the first internationally harmonized guideline to describe a non-animal approach that can be used to fully replace an animal test to identify skin sensitizers. The guideline describes two validated DAs to classify substances for skin sensitization hazard and/or potency: the 2 out of 3 (2o3) and the Integrated Testing Strategy (ITS). The 2025 update expanded the in chemico and in vitro information sources for application to the DAs. A third DA, Key Event 3/1 Sequential Testing Strategy (KE 3/1 STS), has been accepted by the U.S. Environmental Protection Agency (EPA) for skin sensitization hazard and potency (EPA 2018).

1.1. Background

The National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) is an office within the National Institute of Environmental Health Sciences Division of Translational Toxicology (DTT). NICEATM is focused on the development and evaluation of alternatives to animal use for chemical safety testing. In response to a recognized need to expand the number and types of substances with available in vitro skin sensitization testing data, NICEATM embarked on a project to evaluate NAMs for skin sensitization potential. DTT-NICEATM developed a partnership with SenzaGen, the developer of the Genomic Allergen Rapid Detection (GARD™) system, to evaluate the GARD™skin assay for federal agency use by testing substances of interest to multiple agencies. The GARDskin project is a satellite project to a larger NICEATM led effort to evaluate skin sensitization in vitro/in chemico methods with substances outside the OECD evaluated substance types. NICEATM requested a subset of nominations from the larger substance list from Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) member agencies. ICCVAM is composed of 18 U.S. federal regulatory and research agencies that require, use, generate, or disseminate toxicological and safety testing information. DTT, the EPA Office of Pesticide Products (OPP), the EPA Office of Pollution Prevention and Toxics (OPPT), the EPA Center for Computational Toxicology and Exposure (CCTE), the U.S. Consumer Product Safety Commission (CPSC), and the U.S. Food and Drug Administration (FDA) nominated substances for testing. The National Toxicology Program Interagency Center for the Evaluation of Alternative Test Methods (NICEATM) coordinated testing, data analysis, and report completion for this project.

The GARDskin assay predicts skin sensitization hazard classification (i.e., sensitizer or nonsensitizer) using gene signatures from 196 biomarkers. The method uses nanostring RNA technology developed by SenzaGen AB that provides a genomic prediction signature from the proprietary myeloid leukemia SenzaCell™ cell line. GARDskin testing on nominated substances was performed by the DTT contract testing laboratory, Burleson Research Technologies, Inc. (BRT), in Morrisville, NC, using technology provided by SenzaGen.

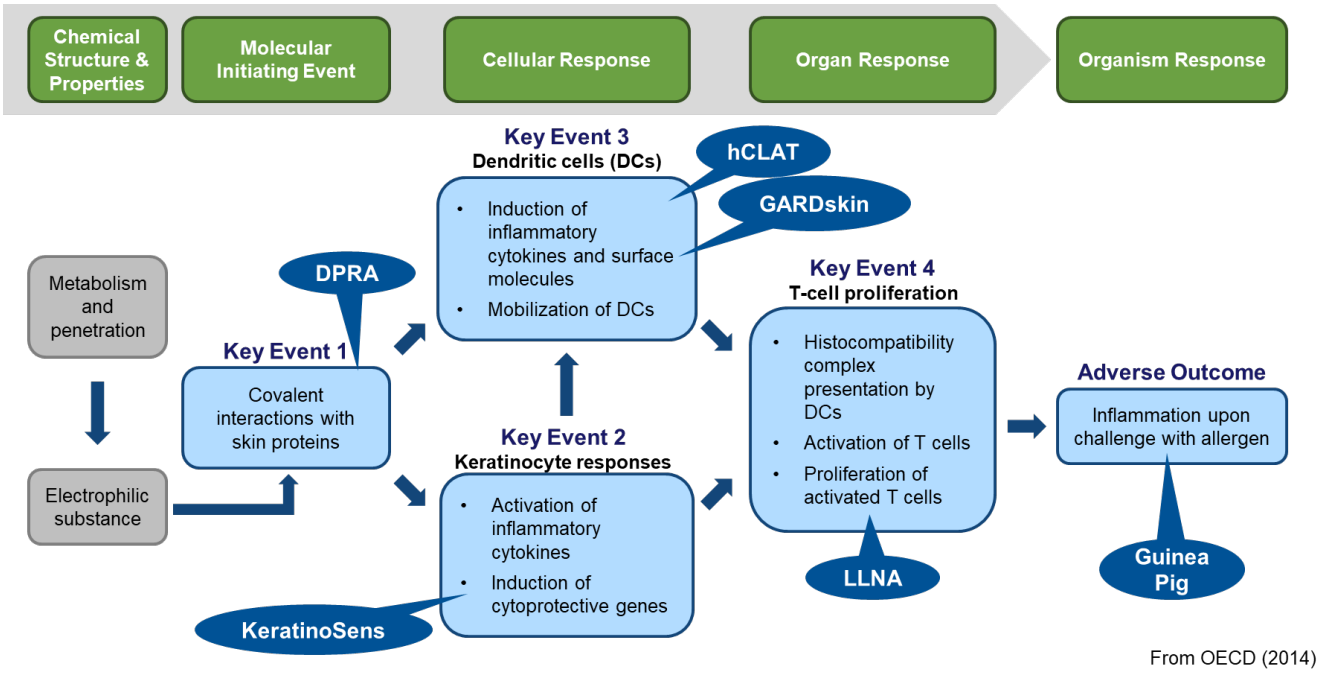
1.2. AOP for Skin Sensitization with Key Events as Targets of Alternative Method Development

Skin sensitization is an adverse outcome that occurs in two phases: induction of sensitization followed by elicitation of an immune reaction (Kimber et al., 2002). The adverse outcome pathway (AOP) for skin sensitization initiated by covalent binding to proteins (Figure 1) has been published by OECD (OECD, 2014) and described by others (MacKay et al., 2013; Maxwell et al., 2014; Strickland et al., 2019). The molecular initiating event, Key Event (KE)1, occurs when a chemical that is either naturally electrophilic or has been made electrophilic via auto-oxidation or metabolism penetrates the skin and binds covalently to lysine or cysteine residues in epidermal proteins, triggering an immune response. The molecular initiating event in KE1 is assessed by DPRA, an in chemico test method that measures depletion of synthetic peptides containing lysine or cysteine as test substances covalently bind to the synthetic peptides (OECD, 2023b).

KE2 is the initiation of an inflammatory response with induction of inflammatory cytokines and cytoprotective genes including the Nrf2-dependent antioxidant response element (ARE) in keratinocytes. KE2 is addressed by the KeratinoSens™ assay (OECD, 2022b). This assay measures luciferase gene induction using the KeratinoSens cell line, which has a stably inserted luciferase reporter gene under the control of the ARE element.

KE3 is the activation of dendritic cells with induction of inflammatory cytokines and surface molecules and mobilization of dendritic cells. KE3 is addressed by the h-CLAT (OECD, 2023c) and the GARDskin assay. The h-CLAT measures the cell surface marker expression of CD86 and CD54 on human monocytic leukemia THP-1 cells, which are surrogate dendritic cells. The GARDskin assay measures KE3 using the GARDskin Genomic Prediction Signature that results when the myeloid leukemia SenzaCell cell line, also surrogate dendritic cells, is exposed to test substances.

KE4 is T-cell activation with histocompatibility complexes presented by dendritic cells leading to T-cell proliferation, which is typically measured by the in vivo LLNA (OECD, 2010). The adverse outcome is an inflammatory response upon challenge with an allergen and can be assessed using the guinea pig maximization test or the Buehler test (OECD, 2022a). No single in chemico or in vitro assay can be used to derive a skin sensitization hazard or potency classification for regulatory purposes, but such assays can be used within the DAs.



From OECD (2014)

Figure 1. Adverse Outcome Pathway for Skin Sensitization

1.3. Objectives

The objectives of this study were to: 1) use the GARDskin assay to characterize the skin sensitization hazard and potency of substances nominated by ICCVAM agency members, 2) compare the GARDskin assay hazard outcomes to those produced by other NAMs and to in vivo outcomes, and 3) apply the NAM results to DAs for hazard and potency classification and compare the DA outcomes to in vivo outcomes.

2. METHODS

2.1. Substances Nominated by Agency Partners for NAM Evaluation

In all, 31 substances were nominated for GARDskin testing: two from FDA, 13 from DTT, three from CPSC, two from EPA CCTE, seven from EPA OPP, and four from EPA OPPT. For each substance, name, other identifiers, lot number, and nominating agency are provided in Table 1.

Table 1. Selected Compounds for GARDskin Testing

CAS RN(s)	BRT Number	Chemical Name	SMILES	Lot Number	Nominator
501-98-4	BRTG-1	trans-p-Hydroxycinnamic acid	OC(=O)\C=C\C1=CC=C(O)C=C1	BCBS8872	EPA OPPT
138261-41-3	BRTG-2	IMA-jet 10	NA	WTN-20420	EPA OPP
15625-89-5	BRTG-3	Trimethylolpropane triacrylate	CCC(COC(=O)C=C)(COC(=O)C=C)COC(=O)C=C	X10E027	DTT
611-06-3	BRTG-4	2,4-Dichloronitrobenzene	[O-][N+](=O)C1=CC=C(Cl)C=C1Cl	06822BD	EPA CCTE
81103-11-9	BRTG-5	Clarithromycin	CC[C@H]1OC(=O)[C@H](C)[C@@H](O[C@H]2C[C@@](C)(OC)[C@@H](O)[C@H](C)O2)[C@H](C)[C@@H](O[C@@H]2O[C@H](C)C[C@@H]([C@H]2O)N(C)C)[C@@](C)(C[C@@H](C)C(=O)[C@H](C)[C@@H](O)[C@]1(C)O)OC	UYQXL	DTT
97-74-5	BRTG-6	Tetramethylthiuram monosulfide	CN(C)C(=S)SC(=S)N(C)C	03816EJ	CPSC
479500-35-1	BRTG-7	1-Butyl-1-methylpyrrolidinium chloride	[Cl-].CCCC[N+]1(C)CCCC1	20100610	DTT
1912-24-9 1912-24-9	BRTG-8	Atrazine (container A) Atrazine (container B)	CCNC1=NC(NC(C)C)=NC(Cl)=N1	77P7D	DTT
81-48-1	BRTG-9	1-Hydroxy-4-(p-toluidino) anthraquinone	CC1=CC=C(NC2=C3C(=O)C4=C(C=CC=C4)C(=O)C3=C(O)C=C2)C=C1	77P7D	EPA OPPT
119-36-8	BRTG-10	Methyl salicylate	COC(=O)C1=C(O)C=CC=C1	2AH0634	DTT
120-32-1	BRTG-11	o-Benzyl-p-chlorophenol	OC1=C(CC2=CC=CC=C2)C=C(Cl)C=C1	KM11195	DTT
7783-18-8	BRTG-12	Ammonium thiosulfate	[NH4+].[NH4+].[O-]S([S-])(=O)=O	12128JE	EPA OPPT
2893-78-9	BRTG-13	Aquatabs	NA	BN M804	EPA OPP
142-31-4	BRTG-14	Sodium octyl sulfate	[Na+].CCCCCCCCOS([O-])(=O)=O	SLBR5188V	EPA OPPT
2892-51-5	BRTG-15	Squaric acid (phosphate buffer)	OC1=C(O)C(=O)C1=O	R7JKB	CPSC
14324-55-1	BRTG-16	Zinc diethyldithiocarbamate	[Zn++].CCN(CC)C([S-])=S.CCN(CC)C([S-])=S	08319ME	CPSC
102-71-6	BRTG-17	Triethanolamine	OCCN(CCO)CCO	03421DJ	FDA
62924-70-3	BRTG-18	Flumetralin	CCN(CC1=C(Cl)C=CC=C1F)C1=C(C=C(C=C1[N+]([O-])=O)C(F)(F)F)[N+]([O-])=O	5-LXM-47-1	EPA OPP
6834-92-0	BRTG-19	Sodium metasilicate	[Na+].[Na+].[O-][Si]([O-])=O	MKCH6862	DTT
71751-41-2*	BRTG-20	Abamectin	CNC(=O)C1=C(NC(=O)C2=CC(Br)=NN2C2=C(Cl)C=CC=N2)C(C)=CC(=C1)C#N	6-ABY-69-1	EPA OPP
736994-63-1	BRTG-21	Cyantraniliprole	CC1(C)CC(CC(C)(CN=C=O)C1)N=C=O	1-MLM-17-1	EPA OPP
4098-71-9	BRTG-22	Isophorone diisocyanate	CCCCCCCC\C=C/CCCCCCCC(O)=O	STBH3457	DTT
112-80-1	BRTG-23	Oleic Acid	CCOC(=O)CC(SP(=S)(OC)OC)C(=O)OCC	SLBQ3165V	EPA CCTE

Table 1. Selected Compounds for GARDskin Testing (Continued)

CAS #(s)	BRT Number	Chemical Name	SMILES	Lot Number	Nominator
121-75-5	BRTG-24	Malathion	[Cl-].CC(C)(C)CC(C)(C)C1=CC=C(OCCOCC[N+](C)(C)CC2=CC=CC=C2)C=C1	4-ABY-188-1	EPA OPP
121-54-0	BRTG-25	Benzethonium chloride	Cl.COC(=O)CCC(=O)CN	W0061	DTT
79416-27-6	BRTG-26	Methyl aminolevulinate hydrochloride	CCC(=O)NC1=CC=C(Cl)C(Cl)=C1	67865	FDA
709-98-8	BRTG-27	Propanil	CC1CC2=C(CC1(C)C(C)=O)C(C)(C)CCC2	19-ABY-118-1	EPA OPP
54464-57-2	BRTG-28	Iso E Super™	CC(\C=C\C=C(C)\C=C\C(O)=O)=C\C=C/C=C(C)\C=C/C=C(C)\C=C\C(O)=O	1-LWJ-148-1	DTT
1393-63-1	BRTG-29	Annatto	CC(\C=C\C=C(C)\C=C\C(O)=O)=C\C=C/C=C(C)\C=C/C=C(C)\C=C\C(O)=O	202005040021	DTT
87-66-1	BRTG-30	Pyrogallol	OC1=CC=CC(O)=C1O	010326	DTT
100-39-0	BRTG-31	Benzyl bromide	BrCC1=CC=CC=C1	34796HJ	DTT

*For abamectin, two different containers with the same lot number were tested.

2.2. In Vitro Data Generated for this Project

BRT, the DTT contract laboratory for immunotoxicity testing, tested 31 substances nominated by ICCVAM agencies in the GARDskin assay. BRT tested 29 of these substances in the DPRA, KeratinoSens, and h-CLAT (NICEATM, 2023). DPRA, KeratinoSens, and h-CLAT data for squaric acid (BRTG-15) and benzyl bromide (BRTG-31) were obtained from the Integrated Chemical Environment (ICE; Abedini et al., 2021; https://ice.ntp.niehs.nih.gov/).

Sections 2.2.1 to 2.2.4 review the tests performed by BRT. The comprehensive test report, which includes detailed protocols for the methods and results, is provided in Appendix B.

2.2.1. Genomic Allergen Rapid Detection for Assessment of Skin Sensitizers (GARD™skin)

For testing with GARDskin, test substances were dissolved in dimethyl sulfoxide or sterile water to a stock concentration of 500 mM or the maximum soluble concentration in accordance with OECD Test Guideline 442E (OECD, 2023c). A preliminary input finder assay was conducted to determine a concentration range that ensured at least 90% viability. In this assay, SenzaCells were exposed to each test substance at a range of concentrations for 24 hours and cytotoxicity was measured by flow cytometry with propidium iodide. This concentration range was then used for three independent main experiments, performed in parallel or sequentially, in which, SenzaCells were exposed to the appropriately adjusted test chemical concentrations for 24 hours and RNA was isolated with TRIzol reagent®.

The level of RNA transcription was measured to determine changes in gene expression using nanostring technology. RNA samples were blinded by BRT and shipped to SenzaGen AB, where the nanostring analysis was conducted. These analyses generate two outputs files, the Report Code Count (RCC) gene expression file and the Reporter Library File (RLF) nanostring code set mapping file. The RCC files record the expression levels for each probe in the code set for all RNA samples analyzed in a specific assay. For each lot of code set, an RLF file provides specific mapping coordinates for each probe in the code set to ensure that expression values are assigned to the correct genes. The nanostring gene expression data (RCC and associated RLF files) were provided to BRT by SenzaGen AB in order to predict skin sensitization hazard using the GARD™ Data Analysis Application (GDAA).

Nanostring results were analyzed using GDAA v.2.2.1, which was accessed through a cloud interface (https://senzagen.shinyapps.io/GDAA_v2_2_1/). The nanostring RNA expression files (RCC and RLF files) and sample annotation files were uploaded into the GDAA. These files were automatically analyzed for quality followed by normalization of expression values based on reference genes to facilitate prediction of sensitization potential based on the GARD™skin support vector machine prediction algorithm. The GDAA software generated a decision value for each replicate and then a mean decision value representing the average of the decision value for the three replicates. Substances with a mean decision value greater than or equal to zero were classified as sensitizers, while substances with mean decision values less than zero were classified as nonsensitizers.

2.2.2. DPRA

DPRA testing was carried out in accordance with OECD Test Guideline 442C (OECD, 2023b). Test substances were evaluated for their reactivity with synthetic peptides containing cysteine or

lysine. The concentrations of each peptide in the reaction solution were measured by high pressure liquid chromatography to determine percent depletion over the 24-hour incubation period. Assay acceptance criteria for assay controls and test substance results were applied as described in OECD 442C (OECD, 2023b). A test substance was classified as positive if the average lysine or cysteine peptide depletion was higher than 6.38%. If the test substance co-eluted with the lysine peptide, substance classification was based only on cysteine peptide depletion. In this case, cysteine peptide depletion greater than 13.89% was used to classify substances as positive. Depletion of the peptides was also used to classify each test substance as unreactive or having minimal, low, moderate, or high reactivity. Substances classified as unreactive or having minimal reactivity are negative in the DPRA and substances assigned to any other classes are positive in the DPRA.

2.2.3. KeratinoSens

Test substances were evaluated for activation of the ARE-dependent pathway using the immortalized, human-derived keratinocyte cell line KeratinoSens™ as described in OECD Test Guideline 442D (OECD, 2022b). These cells have been transfected with a plasmid containing the luciferase gene with expression under the control of the AKR1C2 gene ARE sequence upstream of the SV40 promoter. Increase in luciferase expression is associated with keratinocyte activation and is used to classify test substances as nonsensitizers or sensitizers.

Appropriate dose-ranges were determined in a dose range-finding assay for all test substances. This dose range was used for the full KeratinoSens assay. After cell lysing, activation of the ARE-dependent pathway was determined by measuring luminescence with a luminometer (Molecular Devices SpectraMax® i3 or i3x; data analyzed using SoftMax® Pro GxP v 6.5.1 or 7.03, respectively). Cell viability was measured using the MTT cytotoxicity assay. Acceptance criteria for the assay controls and test substance results were applied as described in OECD 442D (OECD, 2022b). A test substance was considered to be a skin sensitizer when all of the following conditions were met:

- Average maximum fold induction of luciferase activity was at least 1.5-fold over the solvent control value.
- Cell viability was greater than 70% at the lowest concentration that induced luciferase activity at least 1.5-fold over the solvent control value.
- The effective concentration at 1.5-fold induction was less than 1000 μM.
- There was a dose-dependent increase in luciferase induction.

2.2.4. The Human Cell Line Activation Test (h-CLAT)

The h-CLAT measures dendritic cell activation in response to test substance exposure using the immortalized human monocytic leukemia cell line THP-1 as a dendritic cell surrogate. Activation of dendritic cells is assessed by measuring cell surface expression of the costimulatory molecules CD86 and CD54 that parallel production of pro-inflammatory cytokines that induce inflammation. THP-1 cells were initially treated with a range of concentrations of each test substance in a dose range-finding assay in accordance with OECD Test Guideline 442E (OECD, 2023c). The concentration needed to produce viability of 75% (CV75) was determined from these results and used to calculate the starting concentration for the main assay. Expression of CD86 or CD54 was determined by flow cytometry (BD Accuri™ C6; data analysis performed

with CFlowPlus v1.0.264.21). Propidium iodide staining was used to assess cell viability concurrently in the same cell population. Acceptance criteria for assay controls and test substance results were applied as described in OECD Test Guideline 442E (OECD, 2023c). An increase in the relative fluorescence intensity greater than or equal to 150% for CD86 and/or greater than or equal to 200% for CD54 expression was indicative of DC activation. This was considered a positive response if cell viability at those concentrations was at least 50%. The minimum induction threshold, which is the lowest value of the CD54 EC200 and the CD86 EC150, has been derived from the results reported.

2.3. Generation of In Silico Read-Across Hazard Predictions for Skin Sensitization Hazard

In silico read-across hazard predictions for skin sensitization hazard for the nominated test substances were generated using OECD QSAR Toolbox v4.5, which is freely available software (OECD, 2021). The Simplified Molecular Input Line Entry System (SMILES) specifications of chemical structure and CAS RNs for each substance were used as inputs to QSAR Toolbox. SMILES matching the CAS RNs were obtained from EPA CompTox Chemicals Dashboard (Williams et al., 2017, https://comptox.epa.gov/dashboard/ version 2.1). SMILES that were unavailable in the CompTox Chemicals Dashboard were found in QSAR Toolbox. Skin sensitization hazard predictions were made for organic substances using the QSAR Toolbox automated workflow for "EC3 from LLNA or Skin sensitization from GPMT assays for defined approaches (SS AW for DASS)." The workflow provides a read-across prediction of skin sensitization hazard (positive or negative) as well as an assessment of whether each substance evaluated is covered by the applicability domain of the automated workflow. The applicability domain is based on the training set of 2268 substances used to develop the automated workflow that have LLNA and/or guinea pig maximization test experimental data. If the automated workflow could not make a prediction because an ingredient was a salt, the salt was dissociated and the automated workflow was applied to the organic portion of the substance to make a prediction. QSAR Toolbox does not make skin sensitization hazard predictions for inorganic structures or for ingredients with undefined structures (e.g., substances of unknown or variable composition, complex reaction products, or biological materials).

2.4. Physicochemical Properties

Physicochemical properties for each tested substance were determined using the Open (Quantitative) Structure-activity/property relationship App (OPERA) v2.7 (Mansouri et al., 2018; https://github.com/NIEHS/OPERA). OPERA is a free and open-source/open-data suite of QSAR models providing predictions of physicochemical properties, environmental fate parameters, and toxicity endpoints.

2.5. In Vivo Reference Data

Historical human and animal data were used as reference data; no new animal tests were conducted for this project.

LLNA data were obtained from multiple literature sources, a 2013 NICEATM LLNA database (https://ntp.niehs.nih.gov/whatwestudy/niceatm/test-method-evaluations/skin-sens/llna/index.html), the National Toxicology Program's Chemical Effects in Biological

Systems database (https://cebs.niehs.nih.gov/cebs/), and OPP and OPPT. Historical LLNA data were generated using the traditional radiolabeled LLNA described in TG 429 (OECD, 2010a) as well as modified LLNAs such as ex vivo LLNA, cell count LLNA, nonradiolabeled guideline versions (OECD 2010b, OECD 2024) and other nonguideline methods. LLNA data were used for skin sensitization hazard classification according to the Globally Harmonized System of Classification and Labeling of Chemicals (GHS; (UN, 2021). Under this system, skin sensitizers are described as 1A (high frequency of occurrence or potency), 1B (moderate to low frequency of occurrence or low potency), and Not Classified (NC).

Human reference data came primarily from historical human predictive patch tests, the human maximization test and the human repeat insult patch test, which are performed using normal human volunteers (Strickland et al., 2023). These data were sourced primarily from ICE or from the European Chemicals Agency website for Information on Chemicals (https://echa.europa.eu/information-on-chemicals). Human data were available for seven substances.

In vivo reference data and source information are provided in Appendix A.

2.6. Defined Approaches Used

Three accepted DAs were evaluated: the 2o3 method, the ITS, and the KE 3/1 STS. In addition to using the guideline methods for assigning classifications for the OECD methods, the KE 3/1 was used in a modified form with GARDskin assay results replacing h-CLAT results. The DAs were developed based on the KEs of the skin sensitization AOP. Per the 2025 update to Guideline 497, h-CLAT and GARDskin are interchangeable, only requiring the application of the relevant scoring system for each method within the ITS. As the usage of the GARDskin for KE 3/1 STS is exploratory, no scoring system has yet been developed for potency assessment, so only hazard was assigned for this DA, regardless of test methods used.

The DASS App (To et al., 2024; https://ntp.niehs.nih.gov/go/952311) is a freely available web tool via which users can obtain a hazard or potency classification from each of the three DAs derived from their own data. For this project, the DASS App was used to obtain classifications of all three DAs.

2.6.1. 2 out of 3

The 2o3 DA (Bauch et al., 2012; Urbisch et al., 2015) has been incorporated into OECD Guideline 497 (OECD, 2025). The 2o3 provides a skin sensitization hazard classification. This DA uses the outcomes from the in chemico DPRA (KE1), and the in vitro KeratinoSens (KE2), and h-CLAT or GARDskin assays (KE3) (Figure 2). A hazard classification from the 2o3 DA is based on concordant hazard classifications from at least two assays, regardless of the testing order. If a concordant classification is not obtained with two assays, a third assay is conducted. To differentiate the two different 2o3 DAs, we refer to them as the "2o3" and "2o3 GARDskin DA." This terminology is in recognition that the h-CLAT containing 2o3 was the original combination, the 2o3 GARDskin is a permutation of the original.

OECD Guideline 497 has a workflow to identify and remove borderline results for DPRA, KeratinoSens, and h-CLAT (OECD, 2025). The removal of borderline results improved the performance of the 2o3 during the validation assessment for Guideline 497 but also increased the number of substances with inconclusive results.

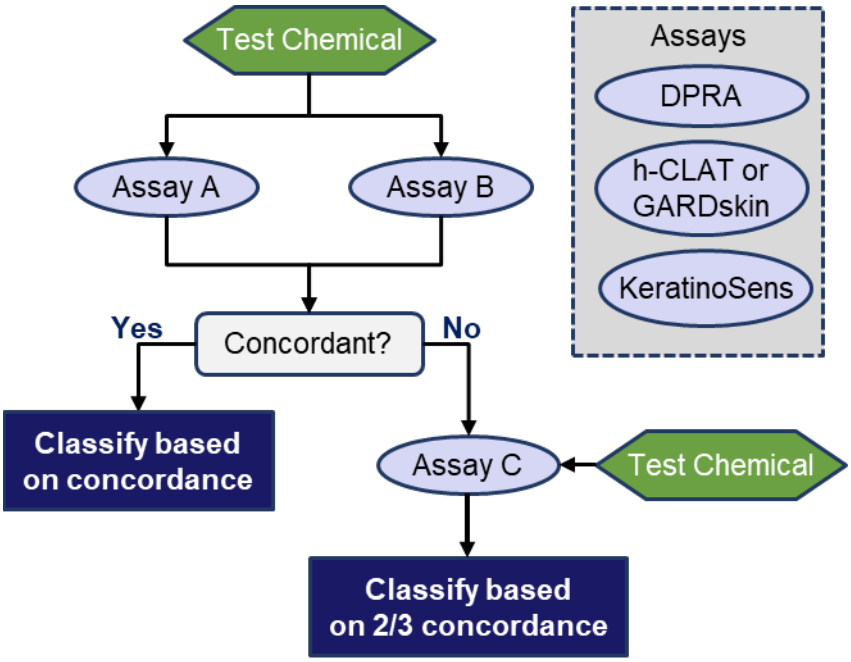


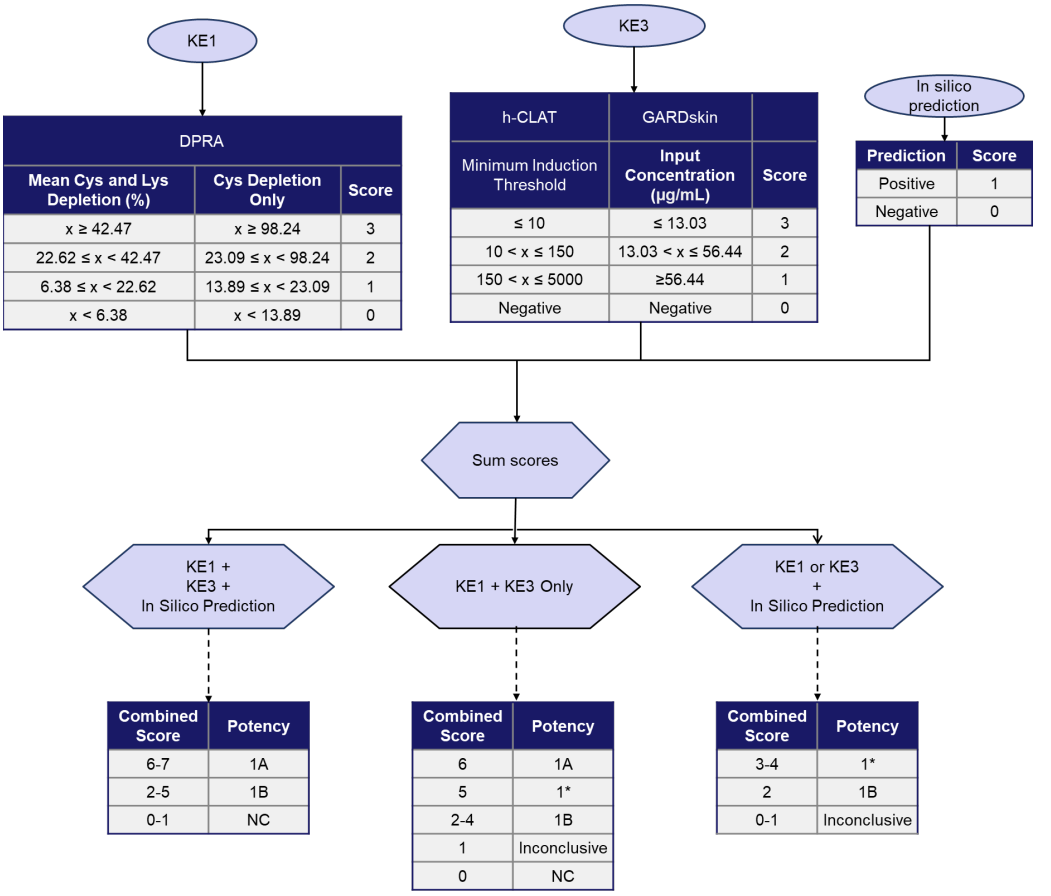
Figure 2. 2o3 DA

2.6.2. Integrated Testing Strategy

The ITS DA was first described by Nukada et al. (2013). Takenouchi et al. (2015) modified the original ITS using an expanded data set, and this version was included in OECD Guideline 497 (OECD, 2025). The ITS provides both skin sensitization hazard and GHS potency classification (i.e., 1A, 1B, or Not Classified)

There are two in silico tools defined within Guideline 497, OECD QSAR Toolbox v4.5 (ITSv2) and Derek Nexus v6.1.0 (ITSv1). We selected to use ITSv2 because OECD QSAR Toolbox is a freely available software, while the alternative is proprietary software. The ITS addresses KE3 of the AOP using h-CLAT or GARDskin and KE1 using DPRA (OECD, 2025). To differentiate the two different ITS DAs, we refer to them as the "ITSv2" and "ITSv2 GARDskin." This terminology is in recognition that the h-CLAT containing ITS was the original combination, the ITS GARDskin is a permutation of the original.

The ITS generates a prediction of skin sensitization hazard and GHS potency by calculating a sum of classification scores from the individual inputs. The ITS uses a scoring system of 0 to 3 for h-CLAT minimum induction threshold or GARDskin input concentration and DPRA peptide depletion results, with a score of 0 to 1 for OECD Toolbox hazard (Figure 3). OECD Guideline 497 includes a workflow for interpreting the total score to consider partial information (e.g., situations in which one input is unavailable) or out-of-domain results for the in silico hazard prediction. In some cases, potency category may not be assigned.



DPRA		
Mean Cys and Lys Depletion (%)	Cys Depletion Only	Score
x ≥ 42.47	x ≥ 98.24	3
22.62 ≤ x < 42.47	23.09 ≤ x < 98.24	2
6.38 ≤ x < 22.62	13.89 ≤ x < 23.09	1
x < 6.38	x < 13.89	0

h-CLAT	GARDskin	
Minimum Induction Threshold	Input Concentration (µg/mL)	Score
≤ 10	≤ 13.03	3
10 < x ≤ 150	13.03 < x ≤ 56.44	2
150 < x ≤ 5000	≥56.44	1
Negative	Negative	0

Prediction	Score
Positive	1
Negative	0

Combined Score	Potency
6-7	1A
2-5	1B
0-1	NC

Combined Score	Potency
6	1A
5	1*
2-4	1B
1	Inconclusive
0	NC

Combined Score	Potency
3-4	1*
2	1B
0-1	Inconclusive

1* indicates conclusive sensitizer hazard prediction and inconclusive potency

Figure 3. ITSv2 DA

2.6.3. KE 3/1 Sequential Testing Strategy

The KE 3/1 STS is accepted by the EPA but is not included in OECD Guideline 497 (OECD, 2025). The STS was developed by Nukada et al. (2013) and addresses KEs 1 and 3 in the AOP for skin sensitization using the DPRA and h-CLAT (or GARDskin), respectively. To differentiate the two different KE 3/1 STS DAs, we refer to them as the "KE 3/1 STS" and "KE 3/1 STS GARDskin," in recognition that the h-CLAT containing KE 3/1 STS was the original combination. The STS can provide both skin sensitization hazard and GHS potency classification (i.e., 1A, 1B, or Not Classified). For potency classification a cutoff value for the measurement endpoint of the KE3 assay to separate 1A and 1B sensitizers is required and no such cutoff has been established for GARDskin. Therefore, consistent with EPA use, we used the STS to provide hazard classification only.

This DA is conducted using sequential testing beginning with the h-CLAT (or GARDskin) (Figure 4). A test substance is classified as positive if the KE3 assay result is positive. If the test substance is negative, it is tested using DPRA. A substance testing positive in the DPRA is classified as a sensitizer, and a substance testing negative is a nonsensitizer ("Not Classified").

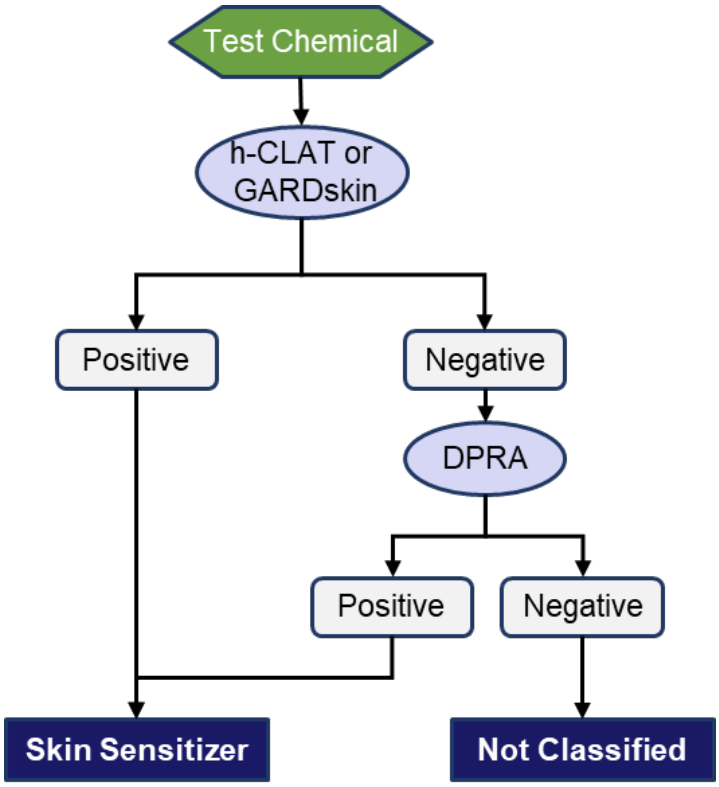


Figure 4. KE 3/1 STS

2.7. Data Analyses

2.7.1. Concordance Analyses

Concordance of the hazard classifications among the individual NAMs and DAs was evaluated, as well as concordance of these methods with classifications based on LLNA and human data. The individual in chemico, in vitro, and in silico read-across predictions are not used for potency classification. As noted above, GARDskin has not been assigned scores or cutoffs so that it could be substituted into the KE 3/1 STS that provides potency categorization (Nukada et al., 2013), however it is included in ITSv1 and ITSv2.

2.7.2. Performance Analyses

The performance of the individual test methods and DAs for hazard classification was calculated by counting the number of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) outcomes relative to the LLNA or human data. Accuracy, sensitivity, specificity, and balanced accuracy were calculated as follows:

$$\textbf{Accuracy}\ (\%) = \left[\frac{\textbf{TP} + \textbf{TN}}{\textbf{TP} + \textbf{TN} + \textbf{FP} + \textbf{FN}}\right] * \textbf{100}$$

$$\textbf{Sensitivity}\ (\%) = [\textbf{TP}/(\textbf{TP} + \textbf{FN})] * \textbf{100}$$

$$\textbf{Specificity}\ (\%) = [\textbf{TN}/(\textbf{TN} + \textbf{FP})] * \textbf{100}$$

$$\textbf{Balanced Accuracy}\ (\%) = [\textbf{Sensitivity}\ (\%) + \textbf{Specificity}\ (\%)]/\textbf{2}$$

3. RESULTS

A total of 31 substances were tested by BRT for skin sensitization potential using the GARDskin assay. The comprehensive test report, which includes detailed protocols and results, is provided in Appendix B. BRT previously tested 29 of these substances in the DPRA, KeratinoSens, and h-CLAT. The comprehensive results for these tests can be found in NICEATM (2025). DPRA, KeratinoSens, and h-CLAT data for squaric acid (BRTG-15) and benzyl bromide (BRTG-31,) were obtained from ICE. The hazard classifications based on DPRA, KeratinoSens, h-CLAT, and GARDskin results are presented in Table 2. Also shown are hazard classifications based on OECD Toolbox predictions, LLNA, the 2o3 DA, the ITSv2 DA, and the KE 3/1 STS DA. Table 3 shows potency classification predictions based on the LLNA, ITSv2 using h-CLAT, and ITSv2 using GARDskin.

The GARDskin assay produced conclusive classifications for all 31 substances. Two substances were missing DPRA or h-CLAT data, and these classifications are listed as "not tested (NT)" in Table 2. Five substances produced inconclusive results in the KeratinoSens assay because the substances were nontoxic (viability >70%), gene induction was less than 1.5-fold greater than controls, and the substances were tested at maximum concentrations less than 1000 µM due to solubility limitations. These classifications are listed as "inconclusive (INC)" in Table 2. Four of these substances are pesticide active ingredients.

DPRA classified 16/30 (53%) substances as sensitizers, KeratinoSens classified 14/26 (54%) substances as sensitizers, h-CLAT classified 21/30 (70%) substances as sensitizers, and GARDskin classified 22/31 (71%) substances as sensitizers. Trimethylolpropane triacrylate (BRTG-3), Aquatabs (BRTG-13), malathion (BRTG-24), and benzyl bromide (BRTG-31) were classified as sensitizers by all four individual in chemico/in vitro methods. IMA-jet 10 (BRTG-2) and methyl salicylate (BRTG-10) were classified as nonsensitizers by all four individual in chemico/in vitro methods.

OECD QSAR Toobox v4.5 was not able to provide hazard predictions for seven substances. These are marked "NA" in Table 2. Two were pesticide products with unknown composition (IMA-jet 10, BRTG-2 and Aquatabs, BRTG-13). The Toolbox does not make skin sensitization hazard predictions for mixtures unless the composition is known. Predictions could not be made for two substances because they were inorganic (sodium metasilicate, BRTG-19 and ammonium thiosulfate, BRTG-12). Predictions could not be made for the remaining three substances presumably because a sufficient number of similar substances with skin sensitization data were not available in the QSAR Toolbox data base for comparison (1-butyl-1-methylpyrrolidinium, BRTG-7; benzethonium chloride, BRTG-25; and methyl aminolevulinate hydrochloride, BRTG-26). The Toolbox predicted 18 substances to be sensitizers, but one of these substances was outside the applicability domain (trans-p-hydroxycinnamic acid, BRTG-1). Six substances were

predicted to be nonsensitizers, three of which were outside the applicability domain (tetramethylthiuram monosulfide, BRTG-6; zinc diethyldithiocarbamate, BRTG-16; and cyantraniliprole, BRTG-21). Predictions that were outside the applicability domain of the Toolbox automated read-across workflow for skin sensitization are marked with "INC" in Table 2.

For reference hazard and potency classifications, 30 substances had LLNA data. The substance with no LLNA data, methyl aminolevulinate hydrochloride (BRTG-26), is listed as "ND" in Table 2 and Table 3. However, this substance tested positive in a human study.

Table 2. Skin Sensitization Hazard Classification Results

BRTG Number	DPRA Hazard	KS Hazard	h-CLAT Hazard	GARDskin Hazard	QSAR TBv4.5 Hazard	LLNA Hazard	2o3 Hazard (NoBL/BL)**	2o3 GARDskin Hazard (NoBL/BL)	ITSv2 Hazard	ITSv2 GARDskin Hazard	STS Hazard	STS GARDskin Hazard
BRTG-1	0	0	1	1	INC	1	0/0	0/0	INC	INC	1	1
BRTG-2	0	0	0	0	NA	1	0/INC	0/INC	0	0	0	0
BRTG-3[a]	1	1	1	1	1	1	1/1	1/1	1	1	1	1
BRTG-4	0	1	1	1	1	1	1/INC	1/INC	1	1	1	1
BRTG-5	NT	1	1	1	1	0	1/1	1/1	1	1	1	1
BRTG-6	1	1	0	1	INC	1	1/1	1/1	1	1	1	1
BRTG-7	0	1	1	0	NA	0	1/INC	0/0	INC	INC	1	0
BRTG-8	0	INC	1	1	1	0	INC/INC	INC/INC	1	1	1	1
BRTG-9	0	1	0	1	1	1	0/0	1/INC	0	1	0	1
BRTG-10	0	0	0	0	0	0	0/0	0/0	0	0	0	0
BRTG-11	1	0	1	1	1	1	1/1	1/1	1	1	1	1
BRTG-12	0	0	1	0	NA	0	0/INC	0/0	INC	INC	1	0
BRTG-13	1	1	1	1	NA	0	1/1	1/1	1	1	1	1
BRTG-14	1	1	1	0	0	0	1/1	1/1	1	1	1	1
BRTG-15	1	0	NT	0	1	1	INC/INC	0/INC	1	1	INC	0
BRTG-16	1	0	1	1	INC	1	1/INC	1/INC	1	1	1	1
BRTG-17	0	0	1	0	0	0	0/INC	0/0	0	0	1	0
BRTG-18	0	1	1	1	1	0	1/INC	1/INC	1	1	1	1
BRTG-19	1	0	1	0	NA	0	1/INC	0/INC	1	1	1	1
BRTG-20	1	INC	1	1	1	0	1/1	1/1	1	1	1	1
BRTG-21	1	INC	0	1	INC	0	INC/INC	1/1	1	1	1	1
BRTG-22	1	0	0	1	1	1	0/INC	1/INC	1	1	1	1
BRTG-23	0	INC	0	0	1	1	0/INC	0/INC	0	1	0	0
BRTG-24	1	1	1	1	1	0	1/1	1/1	1	1	1	1
BRTG-25	0	1	1	1	NA	0	1/INC	1/1	1	1	1	1
BRTG-26	1	1	0	1	NA	ND	1/1	1/1	1	1	1	1
BRTG-27[a]	0	INC	1	1	1	1	INC/INC	INC/INC	1	1	1	1
BRTG-28	0	0	1	1	1	1	0/INC	0/INC	1	1	1	1
BRTG-29	1	1	0	1	1	1	1/INC	1/INC	1	1	1	1
BRTG-30	1	0	1	1	1	1	1/INC	1/INC	1	1	1	1
BRTG-31	1	1	1	1	1	1	1/INC	1/1	1	1	1	1

Abbreviations: BL = borderline call, INC= inconclusive; NA= available; outside of Toolbox applicability domain; ND = no data; historic test data not available; NT= not tested for technical reasons, 1 = positive; 0 = negative.

[a] Due to conflicting RLF files for at least one replicate, decision values were produced using GDAA for each replicate and the mean decision value was calculated manually. **Please refer to Section 2.6.1 for how borderline calls are assessed.

Table 3. Skin Sensitization Potency Results

BRTG Number	LLNA Potency	ITSv2 Potency	ITSv2 GARDskin Potency
BRTG-1	1B	INC	INC
BRTG-2	1	NC	NC
BRTG-3[a]	1A	1A	1A
BRTG-4	1	1B	1B
BRTG-5	NC	INC	INC
BRTG-6	1	1B	1A
BRTG-7	NC	INC	INC
BRTG-8	NC	1B	1B
BRTG-9	1B	NC	1B
BRTG-10	NC	NC	NC
BRTG-11	1	1B	1B
BRTG-12	NC	INC	INC
BRTG-13	NC	1B	1B
BRTG-14	NC	1B	1B
BRTG-15	1	INC	1B
BRTG-16	1	1B	1B
BRTG-17	NC	NC	NC
BRTG-18	NC	1B	1B
BRTG-19	NC	1B	1B
BRTG-20	NC	1A	1A
BRTG-21	NC	1B	1B
BRTG-22	1A	1B	1A
BRTG-23	1B	NC	1B
BRTG-24	NC	1B	1B
BRTG-25	NC	1B	1B
BRTG-26	ND	1B	1B
BRTG-27[a]	1B	1B	1B
BRTG-28	1B	1B	1B
BRTG-29	1B	1B	1B
BRTG-30	1A	1B	1B
BRTG-31	1A	1A	1A

Abbreviations: INC= inconclusive; NC = Not Classified; ND = no data; 1 = GHS skin sensitizer; 1A = GHS skin sensitizer with likelihood of high frequency or potency; 1B = GHS skin sensitizer with likelihood of low to moderate frequency or potency.
[a] Due to conflicting RLF files for at least one replicate, decision values were produced using GDAA for each replicate and the mean decision value was calculated manually.

For the traditional 2o3 DA hazard classifications that use DPRA, KeratinoSens, and h-CLAT, four substances had inconclusive results because two concordant tests were not available. Three of these (atrazine, BRTG-8; cyantraniliprole, BRTG-21; and propanil, BRTG-27) had inconclusive KeratinoSens results with discordant DPRA and h-CLAT results. The fourth substance, squaric acid (BRTG-15), had no h-CLAT data with discordant DPRA and KeratinoSens results. Exclusion of borderline results for the traditional 2o3 resulted in an additional 15 substances that were inconclusive, which are specified in Table 2.

For the GARDskin 2o3 DA that replaced h-CLAT with GARDskin, two substances had inconclusive results: atrazine (BRTG-8) and propanil (BRTG-27). These substances had inconclusive KeratinoSens results with discordant DPRA and GARDskin results. Exclusion of

borderline results in the GARDskin 2o3 DA resulted in an additional 12 substances with inconclusive results, which are specified in Table 2.

For the traditional KE 3/1 STS hazard classifications that use DPRA, KeratinoSens, and h-CLAT, all substances produced conclusive results except for squaric acid (BRTG-15), which had no h-CLAT data. The GARDskin KE 3/1 STS yielded conclusive hazard classifications for all 31 substances.

Potency predictions were made for all substances by both the ITSv2 and ITSv2 GARDskin. These predictions used the scoring system developed as part of an OECD test guideline evaluation project to include GARDskin as part of Guideline 497. ITSv2 produced 26 conclusive calls, while ITSv2 GARDskin produced 27 conclusive calls (Table 3).

3.1. Concordance of Hazard and Potency Classifications from GARDskin, Other NAMs, and LLNA

Concordances of skin sensitization hazard and potency classifications are summarized in Figure 5 and Figure 6 as heatmaps. Darker colors indicate higher concordance, and lighter or brighter colors indicate lower concordance values. Concordance was higher among the NAMs than between NAMs and LLNA data. The highest concordance was among the DAs.

Without application of borderline assessment, the concordance of the four in chemico/in vitro methods with one another ranged from 30% to 69% (Figure 5A). The highest concordance was between KeratinoSens and GARDskin (69%) and the lowest concordance was between DPRA and h-CLAT (48%). The concordance of GARDskin and h-CLAT, the two KE3 methods, was 63%. The concordance of the four in chemico/in vitro methods with the LLNA ranged from 38% (h-CLAT) to 63% (GARDskin). When compared to human data, concordance ranged from 43% (h-CLAT) to 86% (KeratinoSens). Exclusion of borderline outcomes resulted in a range of 30% to 67% for the same methods (Figure 5B). The highest concordances were between DPRA and GARDskin and between KeratinoSens and GARDskin (67% each) and the lowest was between KeratinoSens and h-CLAT (30%). The concordance between GARDskin and h-CLAT decreased to 57%. Concordances with the LLNA were similar to those observed for methods without borderline results, ranging from 33% (KeratinoSens) to 58% (GARDskin). Concordances of individual method results with classifications based on human data after borderlines results were excluded ranged from 17% (h-CLAT) to 86% (KeratinoSens). Similar results were found with the 181-chemical study, with a range of 58% to 78% concordance across individual methods (NICEATM, 2025). Given that only a limited number of substances had human data available for comparison, these data should be interpreted with caution.

The concordance of the DAs for hazard prediction ranged from 75% to 96% without borderline exclusions for the 2o3/2o3 GARDskin and 75 to 100% with borderline assessments applied for the 2o3/2o3 GARDskin (Figure 5). The 2o3 (borderline), 2o3 GARDskin (borderline), and ITSv2 all had 100% concordance with each other. Concordances between the ITSv2 GARDskin and each of the other DAs were at least 89%. The KE 3/1 STS DA and the 2o3 GARDskin had the lowest concordance with one another (75%), which increased to 76% with application of borderline to the 2o3 GARDskin. The concordance of the DAs with the LLNA ranged from 36% (traditional 2o3 DA with borderline) to 59% (ITSv2 GARDskin). The DAs with the highest concordance with the LLNA had GARDskin as the KE3 assay (2o3 GARDskin, no borderline, at 54%, ITSv2 GARDskin at 59%, and KE3/1 STS GARDskin at 57%). However, the concordance

of GARDskin alone with LLNA results (63%) was higher than the concordance of any of the DAs with the LLNA (42% to 59%).

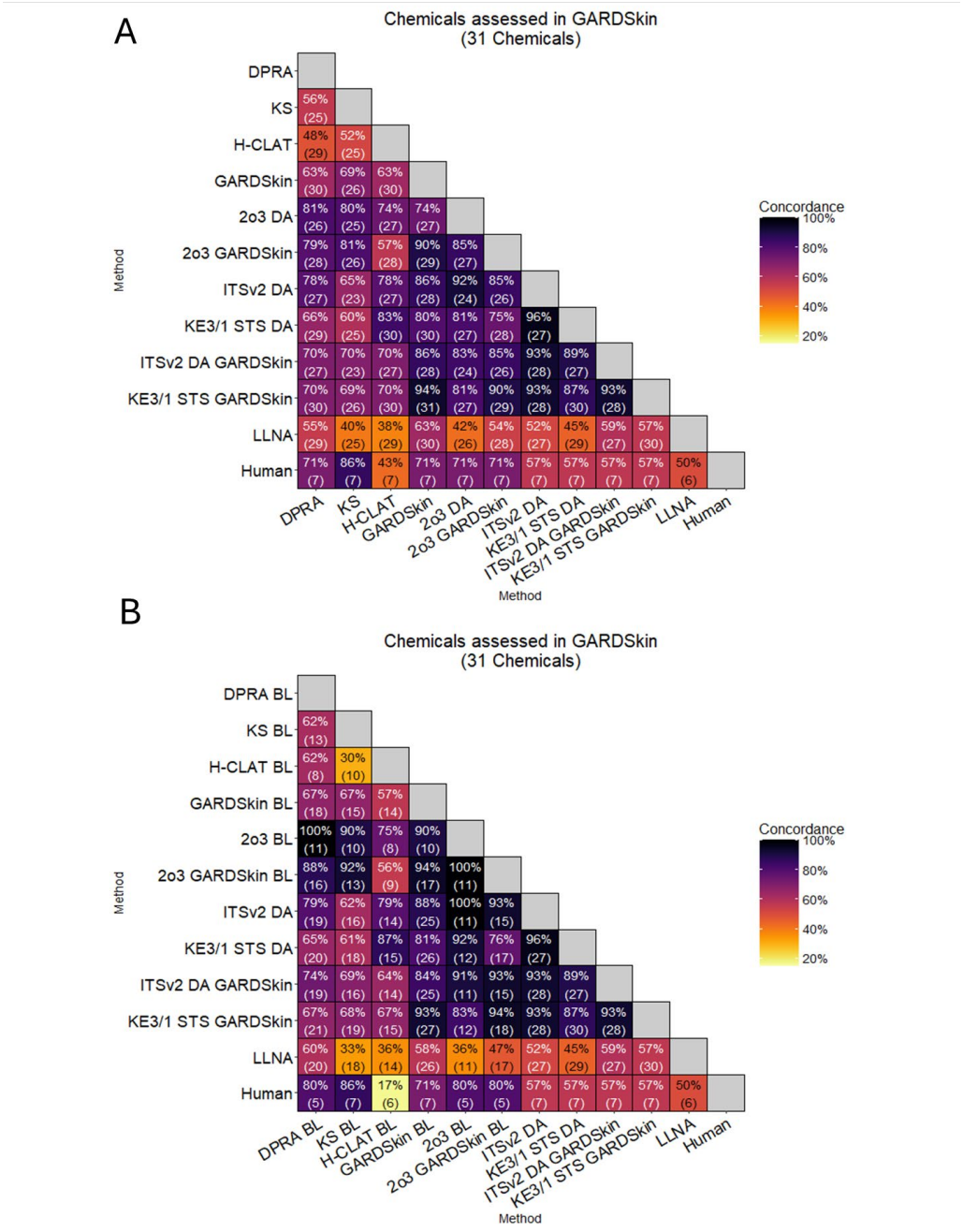


Figure 5. Concordance of Skin Sensitization Hazard Classifications for NAMs and LLNA. A) Without borderline assessments applied to the 2o3 and 2o3 GARDskin. B) With borderline assessments applied to the 2o3 and the 2o3 GARDskin. Darker colors indicate higher concordance, and lighter/brighter colors indicate lower concordance.

With the development of scoring criteria for the GARDskin, it is possible to predict the potency of substances tested with the GARDskin in the ITSv2 (ITSv2 GARDskin). The ITSv2 and ITSv2 GARDskin were 85% concordant, which is the same concordance as the ITSv2 and KE3/1 STS (Figure 6). As noted above, a measurement cutoff that would be required to use GARDskin in the KE 3/1 STS for potency has not yet been developed. The ITSv2 GARDskin had the highest concordance with LLNA potency assessments, at 53%.

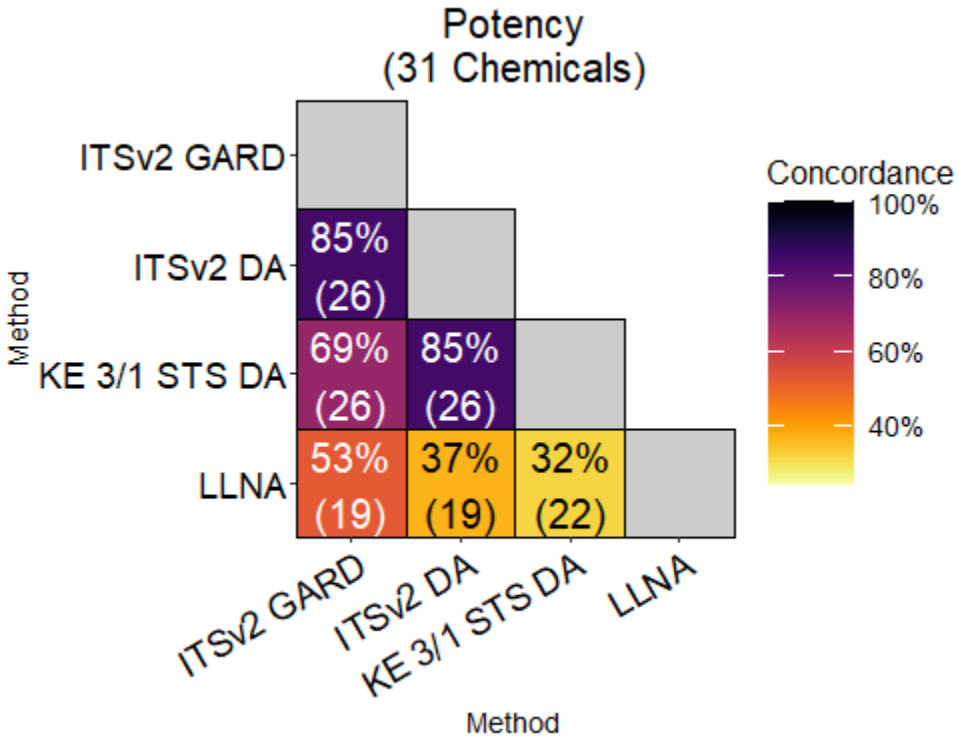


Figure 6. Concordance of Skin Sensitization Potency Classifications for NAMs and LLNA. Darker colors indicate higher concordance, and lighter/brighter colors indicate lower concordance.

3.2. Performance of the GARDskin and Other NAMs for Hazard and Potency Classification with Respect to LLNA

We calculated performance statistics for hazard classification as described in Section 2.7.2 using LLNA data as the reference classification (Table 4). Consistent with its higher concordance to the LLNA (Figure 5), Table 4 also shows that the GARDskin assay had the highest accuracy, 63.3%, of the individual test methods. GARDskin also had the highest balanced accuracy, 62.1%. The accuracies of the remaining individual test methods ranged from 37.9% (h-CLAT) to 55.2% (DPRA) and their balanced accuracies ranged from 37.1% (h-CLAT) to 55.0% (DPRA). GARDskin also had the highest sensitivity of the individual test methods at 81.3%. DPRA had the highest specificity at 53.8% and h-CLAT had the lowest specificity at 14.3%. The FP and FN rates were quite high for the individual test methods. DPRA had the lowest FP rate at 46.2% and GARDskin had the lowest FN rate at 18.8%.

The DAs that used GARDskin assay had higher accuracies, 53.6% to 59.3%, and balanced accuracies, 52.6% to 55%, than the traditional DAs (with accuracies ranging from 42.3% to 51.9% and balanced accuracies ranging from 41.1% to 48.3%). The FP and FN rates were quite high for the DAs, both those that included the GARDskin assay and those that did not. The FP rates were 61.5% (2o3 GARDskin DA) to 92.9% (KE3/1 STS DA) and the FN rates were at 6.7% (ITSv2 GARDskin) to 42.9% (2o3 DA). The 2o3 GARDskin had the lowest FP rate at 61.5% and the 2o3 GARDskin had the lowest FN rate at 6.7%.

Table 4. Performance of GARDskin and Other NAMs Compared to the LLNA for Skin Sensitization Hazard Classification

Test Method	n	Accuracy (%)	Sensitivity (%)	Specificity (%)	Balanced Accuracy (%)	False Positives (%)	False Negatives (%)
DPRA	29	55.2 (16/29)	56.3 (9/16)	53.8 (7/13)	55.0	46.2 (6/13)	43.7 (7/16)
DPRA BL	20	60 (12/20)	66.7 (6/9)	54.6 (6/11)	60.6	45.5 (5/11)	33.3 (3/9)
KeratinoSens	25	40.0 (10/25)	42.9 (6/14)	36.4 (4/11)	39.6	63.6 (7/11)	57.1 (8/14)
KeratinoSens BL	18	33.3 (6/18)	40 (4/10)	25 (2/8)	32.5%	75 (6/8)	60 (6/10)
h-CLAT	29	37.9 (11/29)	60.0 (9/15)	14.3 (2/14)	37.1	85.7 (12/14)	40.0 (6/15)
h-CLAT BL	14	35.7 (5/14)	71.4 (5/7)	0 (0/7)	35.7	100 (7/7)	28.6 (2/7)
GARDskin	30	63.3 (19/30)	81.3 (13/16)	42.9 (6/14)	62.1	57.1 (8/14)	18.8 (3/16)
GARDskin BL	26	57.7 (15/26)	75 (9/12)	42.9 (6/14)	58.9	57.1 (8/14)	25 (3/12)
2o3 DA	26	42.3 (11/26)	57.1 (8/14)	25.0 (3/12)	41.1	75.0 (9/12)	42.9 (6/14)
2o3 BL	11	36.4 (4/11)	60 (3/5)	16.7 (1/6)	38.2	83 (5/6)	40 (2/5)
2o3 GARDskin	28	53.6 (15/28)	66.7 (10/15)	38.5 (5/13)	52.6	61.5 (8/13)	33.3 (5/15)
2o3 GARDskin BL	17	47 (8/17)	66.7 (4/6)	36.4 (4/11)	51.5	63.6 (7/11)	33.3 (2/6)
ITSv2 DA	27	51.9 (14/27)	80 (12/15)	16.7 (2/12)	48.3	83.3 (10/12)	20 (3/15)
KE3/1 STS DA	29	44.8 (13/29)	80.0 (12/15)	7.1 (1/14)	43.6	92.9 (13/14)	20.0 (3/15)
ITSv2 GARDskin	27	59.3 (16/27)	93.3 (14/15)	16.7 (2/12)	55	83.3 (10/12)	6.7 (1/15)
KE3/1 GARDskin	30	56.7 (17/30)	81.3 (13/16)	28.6 (4/14)	54.9	71.4 (10/14)	18.8 (3/16)

BL = borderline procedures applied

Potency predictions were made with the ITSv2 DA, KE3/1, and ITSv2 GARDskin (Table 5). For concurrence with LLNA-based classifications, the ITSv2 GARDskin was the most accurate at 53%, the KE3/1 was the least accurate with 37% accuracy. The ITSv2 GARDskin underpredicted potency 42% of the time (same as ITSv2), and overpredicted potency only 5% (versus 21% for the ITSv2). Underprediction results in application of a lower potency classification, such as a 1A being classified as a 1B or NC or a 1B being classified as an NC. Overprediction results in a lower potency substance being classified as a higher potency substance, such as a 1B being classified as 1A or an NC being classified as 1A or 1B.

Table 5. Performance of DAs with or without GARDskin compared to the LLNA for Skin Sensitization Potency Classification

Test Method	n	Accuracy (%)	Underpredicted (%)	Overpredicted (%)
ITSv2	19	36.8	42.1	21.1
KE3/1 STS DA	22	31.82	50	18.2
ITSv2 GARDskin	19	52.6	42.1	5.3

3.3. Performance of GARDskin and Other NAMs for Hazard and Potency Classification with Respect to Human Data

We calculated performance statistics for hazard classification as described in Section 2.7.2 using human data as the reference classifications (Table 6). As reported in Section 2.6, human data were available for only seven substances. Six of those substances also had LLNA hazard classifications. Due to the very small number of substances, the statistics here must be interpreted with caution.

For this limited number of test substances, all the individual test methods except for h-CLAT and all the DAs performed better than the LLNA with respect to accuracy or balanced accuracy for predicting human hazard classifications. KeratinoSens had the highest accuracy and balanced accuracy at 86% and 88%, respectively. Eight of the nine individual methods or DAs accurately predicted all three sensitizers (i.e., sensitivity = 100%). However, the methods performed poorly at predicting nonsensitizers. Specificity for these was typically 25 or 50%, although the specificity of KeratinoSens was 75%. Likewise, FP rates were typically high at 50% or 75%, although again KeratinoSens performed better than the other methods, with an FP of 25%. Most of the methods and DAs (8 of 9) had no false negatives. The highest FN rates were for LLNA (50%) and h-CLAT (33%). Application of borderline exclusions to the 2o3 methods resulted in equivocal performance between the 2o3 and 2o3 GARDskin.

Table 6. Performance of Hazard Classifications from GARDskin and Other NAMs with Respect to Human Hazard Classifications

Test Method	n	Accuracy (%)	Sensitivity (%)	Specificity (%)	Balanced Accuracy (%)	False Positive (%)	False Negative (%)
LLNA	6	50.0 (3/6)	50.0 (1/2)	50.0 (2/4)	50.0	50.0 (2/4)	50.0 (1/2)
DPRA	7	71.4 (5/7)	100.0 (3/3)	50.0 (2/4)	75.0	50.0 (2/4)	0.0 (0/3)
DPRA (BL)	5	80.0 (4/5)	100.0 (3/3)	50 (1/2)	75.0	50.0 (1/2)	0.0 (0/3)
KeratinoSens	7	85.7 (6/7)	100.0 (3/3)	75.0 (3/4)	87.5	25.0 (1/4)	0.0 (0/3)
KeratinoSens (BL)	7	85.7 (6/7)	100.0 (3/3)	75.0 (3/4)	87.5	25.0 (1/4)	0.0 (0/3)
h-CLAT	7	42.9 (3/7)	66.7 (2/3)	25.0 (1/4)	45.8	75.0 (3/4)	33.3 (1/3)
h-CLAT (BL)	6	16.7 (1/6)	50.0 (1/2)	0.0 (0/4)	25.0	100.0 (4/4)	50.0 (1/2)
GARDskin	7	71.4 (5/7)	100.0 (3/3)	50.0 (2/4)	75.0	50.0 (2/4)	0.0 (0/3)
GARDskin (BL)	7	71.43 (5/7)	100.0 (3/3)	50.0 (2/4)	75.0	50.0 (2/4)	0.0 (0/3)
2o3 DA	7	71.4 (5/7)	100.0 (3/3)	50.0 (2/4)	75.0	50.0 (2/4)	0.0 (0/3)
2o3 (BL)	5	80.0 (4/5)	100.0 (3/3)	50.0 (1/2)	75.0	50.0 (1/2)	0.0 (0/3)
2o3 GARDskin	7	71.4 (5/7)	100.0 (3/3)	50.0 (2/4)	75.0	50.0 (2/4)	0.0 (0/3)
2o3 GARDskin (BL)	5	80.0 (4/5)	100.0 (3/3)	50.0 (1/2)	75.0	50.0 (1/2)	0.0 (0/3)
ITSv2 DA	7	57 (4/7)	100 (3/3)	25 (1/4)	62.5	75 (3/4)	0 (0/3)
KE3/1 STS DA	7	57.1 (4/7)	100.0 (3/3)	25.0 (1/4)	62.5	75.0 (3/4)	0.0 (0/3)
ITSv2 GARDskin DA	7	57 (4/7)	100 (3/3)	25.0 (2/4)	75.0	75.0 (3/4)	0.0 (0/3)
KE3/1 GARDskin	7	57.1 (4/7)	100.0 (3/3)	25.0 (1/4)	62.5	75.0 (3/4)	0.0 (0/3)

BL = borderline procedures applied

Potency predictions were made with the ITSv2 DA, KE3/1, and ITSv2 GARDskin (Table 7). With this very limited data set, all methods performed equally, with very low accuracy of 20% for this chemical set.

Table 7. Performance of DAs with or without GARDskin Compared to Human for Skin Sensitization Potency Classification

Test Method	n	Accuracy (%)	Underpredicted (%)	Overpredicted (%)
ITSv2	5	20	60	20
KE3/1 STS DA	5	20	60	20
ITSv2 GARDskin	5	20	60	20

4. CONCLUSIONS

This study evaluated how the GARDskin assay predicted skin sensitization hazard, both individually and as part of accepted DAs, for a group of 31 substances of interest to ICCVAM agencies. Compared to other assays measuring skin sensitization key events, GARDskin performed well. For predicting classifications based on LLNA data, GARDskin had the highest balanced accuracy, 62%, of any individual test method or DA. The specificity of the GARDskin assay was low, 43%, but it was higher than any other method or DA except for DPRA. GARDskin had a high FP rate, but it was nearly the lowest of the methods/DAs evaluated at 57% and it had the lowest FN rate at 19%. Application of borderline exclusions to the individual methods results in a slight decrease in the balanced accuracy for the GARDskin assay, to 59%. In this regard, the GARDskin assay resembled most of the other individual methods in exhibiting decreased performance. The exception was DPRA, for which performance improved when borderline exclusions were applied. The individual assays had similar concordance rates with one another and with GARDskin (57% to 69%).

For hazard prediction, the DAs that used GARDskin had higher balanced accuracy, 51% to 55%, than the traditional DAs (38% to 48%). Concordance among all the DAs was similar, typically above 80% for all combinations. When potency was considered, as compared to LLNA reference data, the ITSv2 GARDskin was the most accurate (53%) and the least likely to overpredict potency (5%) compared to the ITSv2 (37% accurate, 21% overprediction) and KE3/1 STS (32% accurate, 18% overprediction). It also had the highest concordance with LLNA data at 53%.

Because only seven of the tested substances had human data, the evaluation of GARDskin against human data should be considered preliminary. Hazard classifications based on GARDskin data generally agreed with those based on human data. The GARDskin had a higher balanced accuracy in predicting human hazard (75%) than the LLNA (50%), however, it yielded the same balanced accuracy as DPRA, the 2o3 DA, and the 2o3 GARDskin DA; the only NAM with a higher balanced accuracy was KeratinoSens at 88%. The ITSv2 and KE 3/1 STS DAs with and without GARDskin all performed similarly to each other, both for hazard and potency.

Although the GARDskin assay alone had higher overall performance than the DAs for this set of substances, application of the DAs, which cover multiple key events of the AOP, confers more confidence in the outcomes. The results for this small set of substances cannot be used to infer that the GARDskin assay will always have higher performance than the DAs when applied to other sets of substances. However, overall, the GARDskin assay performs as well as or better than the other individual methods and significantly outperforms the h-CLAT as a solo method for this set of substances. The GARDskin also has equivocal or better performance as a KE3

replacement in the DAs, highlighting that GARDskin is a suitable substitute for the h-CLAT in situations with more difficult-to-assess substances.

5. REFERENCES

Abedini, J., Cook, B., Bell, S., Chang, X., Choksi, N., Daniel, A. B., Hines, D., Karmaus, A. L., Mansouri, K., McAfee, E., Phillips, J., Rooney, J., Sprankle, C., Allen, D., Casey, W., & Kleinstreuer, N. (2021). Application of new approach methodologies: ICE tools to support chemical evaluations. *Computational Toxicology, 20*, 100184. https://doi.org/10.1016/j.comtox.2021.100184

Bauch, C., Kolle, S. N., Ramirez, T., Eltze, T., Fabian, E., Mehling, A., Teubner, W., van Ravenzwaay, B., & Landsiedel, R. (2012). Putting the parts together: Combining in vitro methods to test for skin sensitizing potentials. *Regulatory Toxicology And Pharmacology, 63*(3), 489-504. https://doi.org/10.1016/j.yrtph.2012.05.013

EPA. (2018). Draft Interim Science Policy: Use of Alternative Approaches for Skin Sensitization as a Replacement for Laboratory Animal Testing.10 April 2018. https://www.regulations.gov/document/EPA-HQ-OPP-2016-0093-0090

Kimber, I., Basketter, D. A., Gerberick, G. F., & Dearman, R. J. (2002, Feb). Allergic contact dermatitis. *International Immunopharmacology, 2*(2-3), 201-211. https://doi.org/10.1016/s1567-5769(01)00173-4

MacKay, C., Davies, M., Summerfield, V., & Maxwell, G. (2013). From pathways to people: Applying the adverse outcome pathway (AOP) for skin sensitization to risk assessment. *ALTEX - Alternatives to Animal Experimentation, 30*(4), 473-486. https://doi.org/10.14573/altex.2013.4.473

Mansouri, K., Grulke, C. M., Judson, R. S., & Williams, A. J. (2018). OPERA models for predicting physicochemical properties and environmental fate endpoints. *Journal of Cheminformatics, 10*(1), 10. https://doi.org/10.1186/s13321-018-0263-1

Maxwell, G., MacKay, C., Cubberley, R., Davies, M., Gellatly, N., Glavin, S., Gouin, T., Jacquoilleot, S., Moore, C., Pendlington, R., Saib, O., Sheffield, D., Stark, R., & Summerfield, V. (2014). Applying the skin sensitisation adverse outcome pathway (AOP) to quantitative risk assessment. *Toxicology in Vitro, 28*(1), 8-12. https://doi.org/10.1016/j.tiv.2013.10.013

NICEATM. (2025). Evaluation of substances of regulatory interest using non-animal skin sensitization test methods. Research Triangle Park, NC: Division of Translational Toxicology. NICEATM Report 05. https://doi.org/10.22427/NICEATM-05

Nukada, Y., Miyazawa, M., Kazutoshi, S., Sakaguchi, H., & Nishiyama, N. (2013). Data integration of non-animal tests for the development of a test battery to predict the skin sensitizing potential and potency of chemicals. *Toxicol In Vitro, 27*(2), 609-618. https://doi.org/10.1016/j.tiv.2012.11.006

OECD. (2010). Test No. 429. Skin Sensitisation: Local Lymph Node Assay. In *OECD Guidelines for the Testing of Chemicals, Section 4: Health Effects*. OECD Publishing. https://doi.org/10.1787/9789264071100-en

OECD. (2014). The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins. In *OECD Series on Testing and Assessment, No. 168*. OECD Publishing. https://doi.org/10.1787/9789264221444-en

OECD. (2021). *The OECD QSAR Toolbox v4.5*. https://www.oecd.org/en/data/tools/oecd-qsar-toolbox.html

OECD. (2022a). Test No. 406: Skin Sensitisation. In *OECD Guidelines for the Testing of Chemicals, Section 4*. OECD Publishing. https://doi.org/10.1787/9789264070660-en

OECD. (2022b). Test No. 442D: In Vitro Skin Sensitisation: ARE-Nrf2 Luciferase Test Method. In *OECD Guidelines for the Testing of Chemicals, Section 4*. OECD Publishing. https://doi.org/10.1787/9789264229822-en

OECD. (2023b). Test No. 442C: In Chemico Skin Sensitisation Assays addressing the Adverse Outcome Pathway key event on covalent binding to proteins. In *OECD Guidelines for the Testing of Chemicals, Section 4*. OECD Publishing. https://doi.org/10.1787/9789264229709-en

OECD. (2023c). Test No. 442E: In Vitro Skin Sensitisation: In Vitro Skin Sensitisation assays addressing the Key Event on activation of dendritic cells on the Adverse Outcome Pathway for Skin Sensitisation. In *OECD Guidelines for the Testing of Chemicals, Section 4*. OECD Publishing. https://doi.org/10.1787/9789264264359-en

OECD. (2025). Guideline No. 497: Defined Approaches on Skin Sensitisation. In *OECD Guidelines for the Testing of Chemicals, Section 4*. OECD Publishing. https://doi.org/10.1787/b92879a4-en

Strickland, J., Abedini, J., Allen, D. G., Gordon, J., Hull, V., Kleinstreuer, N., Ko, H.-S., Matheson, J., Thierse, H.-J., Truax, J., Vanselow, J., & Herzler, M. (2023). A database of human predictive patch test data for skin sensitization. *Archives of Toxicology*. https://doi.org/10.1007/s00204-023-03530-3

Strickland, J., Daniel, A. B., Allen, D., Aguila, C., Ahir, S., Bancos, S., Craig, E., Germolec, D., Ghosh, C., Hudson, N. L., Jacobs, A., Lehmann, D. M., Matheson, J., Reinke, E. N., Sadrieh, N., Vukmanovic, S., & Kleinstreuer, N. (2019). Skin sensitization testing needs and data uses by US regulatory and research agencies. *Archives of Toxicology, 93*(2), 273-291. https://doi.org/10.1007/s00204-018-2341-6

Takenouchi, O., Fukui, S., Okamoto, K., Kurotani, S., Imai, N., Fujishiro, M., Kyotani, D., Kato, Y., Kasahara, T., Fujita, M., Toyoda, A., Sekiya, D., Watanabe, S., Seto, H., Hirota, M., Ashikaga, T., & Miyazawa, M. (2015, Nov). Test battery with the human cell line activation test, direct peptide reactivity assay and DEREK based on a 139 chemical data set for predicting skin sensitizing potential and potency of chemicals. *Journal of Applied Toxicology, 35*(11), 1318-1332. https://doi.org/10.1002/jat.3127

To, K. T., Strickland, J., Reinke, E., Borrel, A., Truax, J., Maldonado, H., Allen, D., & Kleinstreuer, N. (2024). Computational application of internationally harmonized defined approaches to skin sensitization: DASS App. *BMC Bioinformatics*, 25(4). https://doi.org/10.1186/s12859-023-05617-1

UN. (2021). *Globally Harmonized System of Classification and Labelling of Chemicals (GHS)* (9 ed.). United Nations. https://unece.org/sites/default/files/2021-09/GHS_Rev9E_0.pdf

Urbisch, D., Mehling, A., Guth, K., Ramirez, T., Honarvar, N., Kolle, S., Landsiedel, R., Jaworska, J., Kern, P. S., Gerberick, F., Natsch, A., Emter, R., Ashikaga, T., Miyazawa, M., & Sakaguchi, H. (2015). Assessing skin sensitization hazard in mice and men using non-animal test

methods. *Regulatory Toxicology and Pharmacology, 71*(2), 337-351. https://doi.org/10.1016/j.yrtph.2014.12.008

Williams, A. J., Grulke, C. M., Edwards, J., McEachran, A. D., Mansouri, K., Baker, N. C., Patlewicz, G., Shah, I., Wambaugh, J. F., Judson, R. S., & Richard, A. M. (2017, 2017/11/28). The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. *Journal of Cheminformatics, 9*(1), 61. https://doi.org/10.1186/s13321-017-0247-6

Appendix A:
In Vitro Results, In Silico Data, Physiochemical Data, In Vivo Reference Data, and Defined Approach Results

Data available at
https://doi.org/10.22427/NICEATM-06

Appendix B:
BRT In Vitro Testing Results

Data available at
https://doi.org/10.22427/NICEATM-DATA-NICEATM-06