ICCS | INTERNATIONAL COLLABORATION ON COSMETICS SAFETY

# Best Practice Guidance Document

*Skin Sensitization Assessment:*
*Using New Approach Methodologies for Substances in Cosmetics and Personal Care Products*

July 2025

ICCS

# Foreword

The International Collaboration on Cosmetics Safety (ICCS), established in 2023, is a global initiative, headquartered in New York, focused on advancing the adoption of animal-free assessments of cosmetics, and their ingredients, for human health and environmental safety.

ICCS brings together scientists and experts from cosmetics manufacturers and suppliers, industry and research associations, and animal protection organizations to drive greater global awareness and confidence in animal-free science through scientific research, capacity-building, and regulatory engagement. Building on nearly four decades of progress in the development, evaluation, and use of animal-free approaches, ICCS aims to accelerate the transition to animal-free safety science through widespread adoption and use.

ICCS publishes Best Practice Guidance documents and makes these available on the ICCS website, free of charge. These can be accessed at: https://www.iccs-cosmetics.org/education/best-practice-guidance/bpg-skin-sensitization-assessment-using-new-approach-methods.

For questions about the Best Practice Guidance, please contact ICCS at info@iccs-cosmetics.org.

# About this Best Practice Guidance

This document was prepared by ToxStrategies LLC on behalf of the International Collaboration on Cosmetics Safety (ICCS) and has been extensively reviewed and informed by input from the following ICCS Working Groups:

- ICCS Skin Sensitization Working Group
- ICCS Best Practice Guidance Working Group

## Version

v1.1 (February 2026)

v1.2 (April 2026)

## What's New in Version 1.1

**Key updates in this version include:**

- Updates to the ICCS skin sensitization BPG workflow (Figure 3) to improve clarity.
- Update to Figure 8.
- Minor editorial revisions made throughout the document.

## What's New in Version 1.2

- Technical update, removed links to Figure 3 on pages 14 and 19 to expedite proper PDF export.

## How to Cite

Please cite as "International Collaboration on Cosmetics Safety. (2026). Best Practice Guidance Document: Skin Sensitization Assessment Using New Approach Methodologies for Substances in Cosmetics and Personal Care Products. v1.2. https://www.iccs-cosmetics.org/education/best-practice-guidance/bpg-skin-sensitization-assessment-using-new-approach-methods.

# Executive Summary

This Best Practice Guidance (BPG) document provides insights into using New Approach Methodologies (NAMs) for skin sensitization assessments for substances in cosmetics and personal care products. The document aims to increase the use and acceptance of NAMs by providing a structured workflow for safety assessors. The primary goal of this document is to inform hazard and safety assessments without the need for new animal testing.

This BPG explains the key principles and steps involved in skin sensitization assessments using NAMs, focusing on problem formulation, exposure assessment, targeted testing, and applying new data in decision-making contexts. The importance of documentation and transparency is emphasized, with guidance on reporting decisions, input variables, and assumptions. The proposed workflow for hazard and/or safety is presented covering problem formulation, substance characterization, data gathering, and evaluation.

The objectives of each step of the workflow process are listed below. For additional details on each step, the assessor is referred to the details presented in the full document following this Summary.

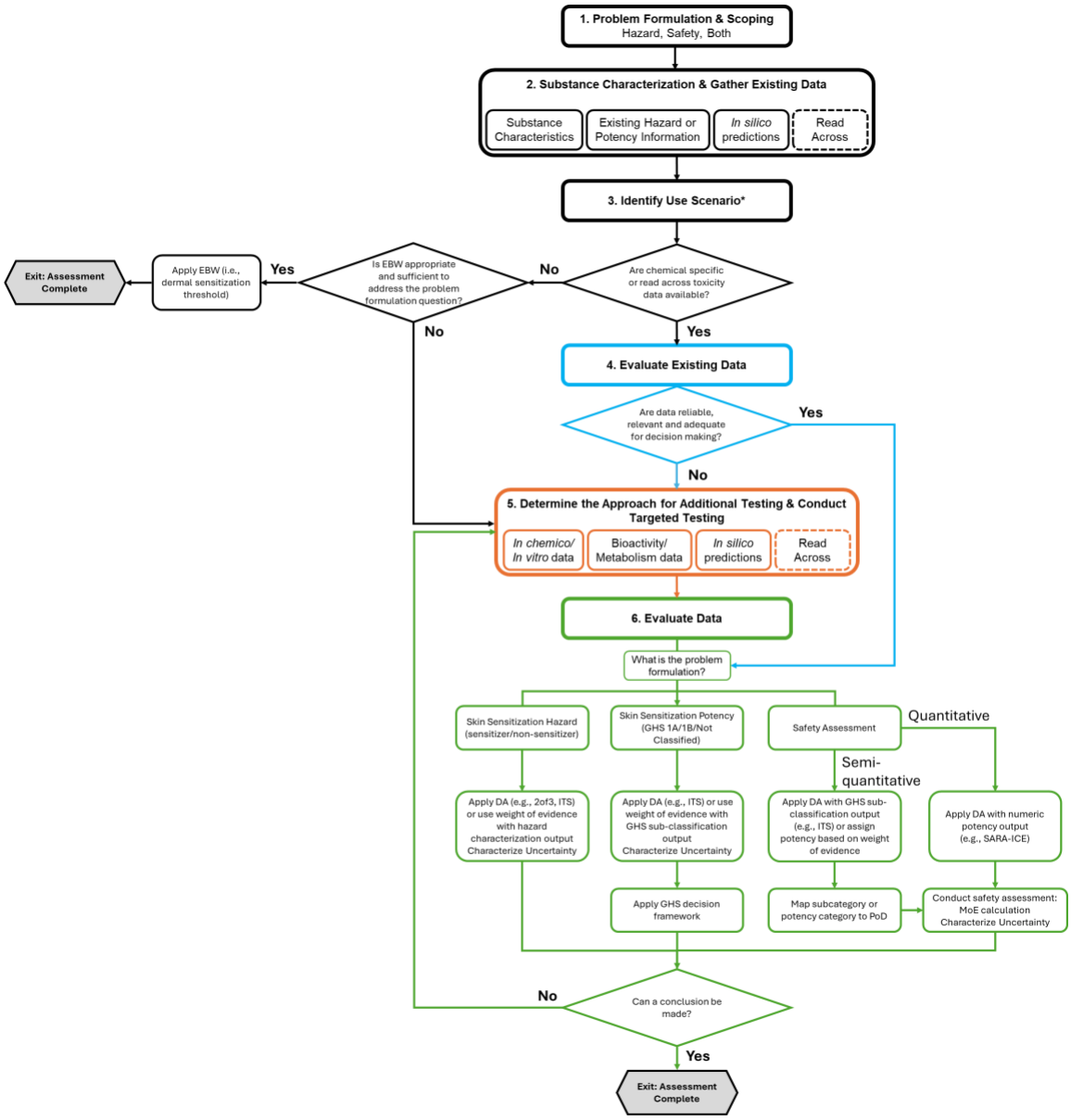
| | |
|---|---|
| **Step 1. Problem Formulation & Scoping** | Provide a clear statement of the assessment scope and the hypothesis to be evaluated. |
| **Step 2. Substance Characterization & Gather Existing Data** | Collect data on the substance that will inform the resulting skin sensitization assessment. The data collected should include substance physico-chemical properties, existing hazard and potency data, *in silico* prediction data (e.g., metabolites) and potential analogue compounds. |
| **Step 3. Identify Use Scenario** | Define the exposure scenarios for the substance and estimate the quantitative exposure level of the substance to the consumer. With this data, an assessor may, if appropriate, evaluate whether exposure-based waiving may be used to assess skin sensitization potential. |
| **Step 4. Evaluate Existing Data** | Evaluate existing data regarding reliability, relevancy, and adequacy. A data table to organize existing information for the substance of interest is recommended. |
| **Step 5. Determine the Approach for Additional Testing and** | Identify data gaps in the existing data for the substance of interest. In combination with the assessment scope and hypothesis, the test methods to fill the data gaps to assess |

| | |
|---|---|
| **Conduct Targeted Testing** | skin sensitization potential can be identified and implemented in order to inform the skin sensitization assessment. |
| **Step 6. Evaluate Data** | Identify and apply the most appropriate assessment approach based on the defined problem formulation (i.e., assessing hazard, potency, or safety). |
| **Uncertainty Characterization** | Identify potential sources of uncertainty in the safety assessment and assess the impact of the collective uncertainty on conclusion confidence. |

This guidance is designed for safety assessors experienced in traditional animal studies but who have limited experience with NAMs. It assumes a baseline understanding of skin sensitization evaluation and cosmetic safety assessment, and follows NGRA principles. This BPG was developed through an iterative process involving literature reviews and expert input. The guidance outlined herein is based on current best practices, and will be updated as scientific advancements occur.

This document provides background information on skin sensitization, including its biology, traditional animal tests, human clinical studies, and NAMs. It outlines the adverse outcome pathway (AOP) for skin sensitization and describes various *in chemico* and *in vitro* methods and *in silico* models endorsed by the Organisation for Economic Co-operation and Development (OECD), including integrated approaches to testing and assessment (IATAs) and defined approaches (DAs).

By following the approaches laid out in this BPG, safety assessors can confidently conduct a skin sensitization assessment for substances in cosmetics and personal care products and identify resources for further details, if needed.




## Overview of Process to Evaluate Skin Sensitization using Next Generation Risk Assessment

As needed, the safety assessor can iterate through the phases of this process.

Notes: *Step should be skipped if focus is only on hazard evaluation; ^additional details on hazard and safety evaluations are provided in Section 2.9; EBW = exposure-based waiving; PF = problem formulation; DA = defined approach; 2of3 = 2 out of 3 Strategy; SARA-ICE: Skin Allergy Risk Assessment – Integrated Chemical Environment Model; MoE = margin of exposure; ITS = integrated testing strategy; GHS = Globally Harmonized System classification; PoD = point of departure.

# Table of Contents

**Foreword** ..... **i**

**About this Best Practice Guidance** ..... **ii**

**Executive Summary** ..... **iii**

**Acronym List** ..... **ix**

**1 Introduction** ..... **1**

**1.1 Purpose and Scope** ..... **1**

**1.2 Intended Audience** ..... **2**

**1.3 Overview of Next Generation Risk Assessment (NGRA) Principles** ..... **3**

**1.4 Method for Document Development** ..... **3**

**1.5 Comparison with Other Guidance Documents on Skin Sensitization** ..... **6**

**1.6 Overview of the Document** ..... **6**

**2 Skin Sensitization Assessment** ..... **7**

**2.1 Background** ..... **7**

2.1.1 Skin sensitization biology and adverse outcome pathway ..... 8

2.1.2 Animal tests ..... 9

2.1.3 Human clinical studies ..... 10

2.1.4 NAMs used for skin sensitization assessment ..... 11

**2.2 Reporting** ..... **13**

**2.3 Overview of Workflow** ..... **14**

**2.4 Step 1. Problem Formulation and Scoping: Define the Context for the Decision** ..... **16**

**2.5 Step 2. Substance Characterization and Gathering Existing Data** ..... **17**

2.5.1 Substance characteristics ..... 17

2.5.2 Gather Existing Data ..... 19

**2.6 Step 3. Identify Use Scenario** ..... **23**

2.6.1 Apply Exposure-Based Waiving (i.e., dermal sensitization threshold) ..... 27

**2.7 Step 4. Evaluate Existing Data** ..... **31**

**2.8 Step 5. Determine the Approach for Additional Testing and Conduct Targeted Testing** ..... **34**

2.8.1 Determine the Approach for Additional Testing ..... 34

2.8.2 Conduct Targeted Testing ..... 38

2.8.3 Metabolism data....38
2.8.4 In chemico and in vitro data....39
**2.9 Step 6. Evaluate Data....40**
2.9.1 Assessing skin sensitization hazard (sensitizer/non-sensitizer)....41
2.9.2 Assessing skin sensitization potency (GHS 1A/GHS 1B/Not Classified)....44
2.9.3 Conducting a safety assessment using a point of departure (PoD)....46
**2.10 Uncertainty characterization of the entire skin sensitization NGRA....53**
**3 Conclusion....55**
**4 References....56**
**Appendix A. *In chemico* and *in vitro* skin sensitization methods with applicability domain information....67**
**Appendix B. Historically applied Safety Assessment Factors (SAFs)....71**

# List of Figures

Figure 1. Graphical Representation of Induction (left) and Elicitation (right) Phases of Skin Sensitization (from OECD, 2014)....7
Figure 2. Flow Diagram of the AOP and the KEs Associated with Skin Sensitization (adapted from OECD, 2014)....9
Figure 3. Overview of Process to Evaluate Skin Sensitization using Next Generation Risk Assessment....15
Figure 4. Considerations for Problem Formulation and Scoping....16
Figure 5. Overview of the Substance Characterization and Identification of Existing Information Process (dashed line indicates that read-across at this phase in the evaluation can be decided on a case-by-case basis)....17
Figure 6. Visual Depiction of Identification of Use Scenario and Exposure-Based Waiving....30
Figure 7. Evidence Integration Workflow, based on OECD (2019)....32
Figure 8. *In Chemico, In Vitro*, and *In Silico* Methods and Models for Skin Sensitization Mapped to the Adverse Outcome Pathway (adapted

from Kleinstreuer 2021; Strickland et al., 2019) to Assist in Defining the Approach for Additional Testing .....35

Figure 9. Data Interpretation Procedure for 2o3 Defined Approach for Hazard Classification (adapted from OECD 2025).....42

# List of Tables

Table 1. Key Documents Reviewed in the Process of Best Practice Guidance Document Development .....4

Table 2. *In Chemico* and *In Vitro* Skin Sensitization Methods.....11

Table 3. Summary of Data to Identify and Describe Substance of Interest Safety Assessment.....17

Table 4. Data to Collect to Support Derivation of a Consumer Exposure Level (CEL) .....24

Table 5. Summary of Example DST Values from the Published Literature .....29

Table 6. Example Data Map for Organization Existing Data for Substance* .....33

Table 7. Defined Approaches for Skin Sensitization using *In Chemico, In Vitro*, and *In Silico* Tools (and existing *in vivo* data).....37

Table 8. Overview of Results Indicating Sensitizing Potential by Assay.....42

Table 9. Scores associated with assay results for ITS, as presented in OECD GL 497 (2025).....45

Table 10. Proposed Dose Ranges Associated with Skin Sensitization Potency, from Na et al. (2022).....47

Table 11. Side by Side Comparison of the Acceptable MoE from a Traditional NESIL-Based Risk Assessment to that for $PoD_{NAM}$ from SARA-ICE .....52

# Acronym List

**ACD**: Allergic Contact Dermatitis

**ADRA**: Amino Acid Derivative Reactivity Assay

**AEL**: Acceptable Exposure Level

**AOP**: Adverse Outcome Pathway

**BPG**: Best Practice Guidance

**CASRN**: Chemical Abstracts Service Registry Numbers

**CEL**: Consumer Exposure Level

**CIR**: Cosmetics Ingredient Review

**DA**: Defined Approach

**DIP**: Data Interpretation Procedure

**DST**: Dermal Sensitization Threshold

**EBW**: Exposure-Based Waiving

**ECETOC**: European Centre for Ecotoxicology and Toxicology of Chemicals

**EINECS**: European Inventory of Existing Chemical Substances

**EPA**: United States Environmental Protection Agency

**FDA**: United States Food and Drug Administration

**GHS**: Globally Harmonized System

**GL**: Guideline

**GPMT**: Guinea Pig Maximization Test

**HRIPT**: Human Repeat Insult Patch Test

**HMT**: Human Maximization Test

**HPPT**: Human Predictive Patch Test

**ICCS**: International Collaboration on Cosmetics Safety

**IATA**: Integrated Approaches to Testing and Assessment

**INCI**: International Nomenclature of Cosmetic Ingredients

**ICCR**: International Cooperation on Cosmetics Regulation

**ITS**: Integrated Testing Strategy

**KE**: Key Event

**KERS**: Key Event Relationships

**LLNA**: Local Lymph Node Assay

**MIE**: Molecular Initiating Event

**MoE**: Margin of Exposure

**MoS**: Margin of Safety

**NAMs**: New Approach Methodologies

**NESIL**: No Expected Sensitization Induction Level

**NGRA**: Next Generation Risk Assessment

**NICEATM:** National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods

**OECD**: Organisation for Economic Co-operation and Development

**PF**: Problem Formulation

**PoD**: Point of Departure

**RIFM**: Research Institute for Fragrance Materials

**SAF**: Sensitization Assessment Factors

**SARA-ICE**: Skin Allergy Risk Assessment – Integrated Chemical Environment

**SCCS**: Scientific Committee on Consumer Safety

**TG**: Test Guideline

**UN**: United Nations

**WHO**: World Health Organization

**WNT**: OECD Working Group of the National Coordinators for the Test Guidelines Programme

**WoE**: Weight of Evidence

**WPHA**: OECD Working Party on Hazard Assessment

# 1 Introduction

## 1.1 Purpose and Scope

The purpose of this best practice guidance (BPG) document is to provide insight into using new approach methodologies (NAMs) for skin sensitization to inform hazard and safety assessment. The International Collaboration on Cosmetic Safety (ICCS) defines NAMs as any non-animal method, approach or combination thereof used to support safety assessments covering effects and exposure for humans and the environment (including fate) without new animal testing. This document explains key principles and steps involved in skin sensitization assessments using NAMs.

The goal of this document is to increase the use and acceptance of NAMs for skin sensitization by providing a workflow that safety assessors can follow. This document provides guidance associated with the workflow to facilitate the identification of critical data gaps and how to fill these gaps by considering existing data (*in vivo* or otherwise) along with NAMs for decision-making purposes, without the need for conducting additional animal testing.

This guidance covers key areas to consider in a skin sensitization hazard and/or safety assessment,[1] with a focus on problem formulation, exposure assessment, targeted testing, additional data generation, and applying new data in the context of the defined decision context (e.g., quantitative safety assessment; qualitative hazard evaluation). Guidance on interpreting and applying NAMs for decision making is also provided.

The guidance is intended to be applicable to the evaluation of the induction of skin sensitization for substances used in cosmetics or personal care products. The guidance also is intended to assess a single substance at a time. Formulation assessment is not encompassed within this guidance document. However, components within a formulation may be evaluated individually for skin sensitization potential with these individual lines of evidence coming together to inform an initial assessment of a formulation's overall sensitizing potential. This document does not provide guidance on aggregate or cumulative safety assessments.

[1] NAM-based tools for skin sensitization are designed to evaluate the risk of skin sensitization induction (i.e., the risk of a particular substance to cause an immune response), not elicitation (i.e., the reaction observed on the skin). Such design is purposeful, since if a substance does not induce an immune reaction, elicitation will not occur. See Section 2 for additional details.

This document describes how to use NAMs, alongside other information sources, such as physico-chemical properties, existing human safety data, and existing *in vivo* data in a skin sensitization hazard and/or safety assessment, hereafter referred to collectively as a skin sensitization next generation risk assessment (NGRA). Read-across evaluations to either traditional or NAM data can also be used at various points in the NGRA, and are identified as potentially supporting data. Read-across is conceptually described herein, but prescriptive approaches on when and how to perform read-across are not included in this document. A read-across BPG is under development by ICCS.[2]

Technical details of individual NAMs (e.g., how to perform assays or run models) are outside the scope of this document. When available, references to internationally adopted protocols will be provided. Similarly, methods for identifying currently available data for the substance undergoing evaluation (i.e., literature search methods); conducting an aggregate or probabilistic exposure assessment; approaches for identifying analogues; and conducting a read-across assessment are introduced herein. While the details for these methods are outside this document's scope, a range of available sources providing in-depth guidance on how to conduct each of these steps are available, and referenced throughout this guidance, where appropriate.

The decision-making guidance provided herein is based on current best practices based on the state-of-the-science available at the date of publication (July, 2025, later updated April, 2026). However, this area is one of rapid scientific advancements, and, as such, this guidance will be updated as deemed necessary.

Finally, although a safety assessor may use this document to inform assessments for regulatory purposes, agency-specific considerations are beyond the scope of this document.

## 1.2 Intended Audience

This guidance is intended to be used globally by safety assessors with experience conducting skin sensitization assessments using traditional animal studies (e.g., murine local lymph node assay [LLNA]; guinea pig maximization test [GPMT]), but limited experience evaluating and applying NAMs for evaluating skin sensitization potential. This guidance assumes that the safety assessor has a baseline understanding of skin sensitization evaluation fundamentals and cosmetic safety assessment. Foundational information on skin sensitization biology is provided in Section 2.1. For background on conducting cosmetic

[2] As of February 2026

safety assessments, including for skin sensitization, the reader is directed to the Scientific Committee on Consumer Safety (SCCS) Notes of Guidance (2023).[3]

## 1.3 Overview of Next Generation Risk Assessment (NGRA) Principles

This document covers both hazard and safety assessments, using a safety assessment approach developed following NGRA principles. In 2017, nine NGRA principles were published by a working group convened under the auspices of the International Cooperation on Cosmetics Regulation (ICCR) (Dent et al., 2018; ICCR, 2017). These principles, outlined below, were identified to assist in developing integrated safety assessments without generating additional animal data, and are applicable to both skin sensitization-specific and systemic safety assessments.

1. Overall goal is human safety assessment.
2. Assessment is exposure-led.
3. Assessment is hypothesis driven.
4. Assessment is designed to prevent harm.
5. Assessment follows an appropriate appraisal of all existing information.
6. Assessment uses a tiered and iterative approach.
7. Assessment uses robust and relevant methods and strategies.
8. Sources of uncertainty should be characterized and documented.
9. Logic of the approach should be transparently and explicitly detailed.

The subsequent chapters of this document and the workflow proposed (see Executive Summary or Section 2.3) are based on these nine principles. Details for each of these principles (i.e., how each relates to the goal of the risk assessment, how it should be conducted, and how it should be documented) are further described in Dent et al. (2018). Additional sources of information that provide background on using NGRA for skin sensitization include Gilmour et al. (2020; 2023).

## 1.4 Method for Document Development

This BPG document was developed using an iterative process involving both a literature review of available guidance and best practices in skin sensitization safety and safety

[3] The reader should be advised that although the SCCS Notes of Guidance provide some perspective on NAMs, it is referenced here as a resource for understanding the more traditional approaches used in conducting safety assessments.

assessment of cosmetics and input from experts experienced in skin sensitization hazard and safety assessment using both traditional animal models and NAMs.

The literature review resulted in an inventory of available skin sensitization references on adverse outcome pathways (AOPs), defined approaches (DAs), integrated approaches to testing and assessment (IATAs), frameworks, guidance documents, and best practices incorporating NAMs. General approach methodology was also inventoried, including the use of systematic review methods, as well as various risk and exposure assessment methods (e.g., quantitative margin of exposure/margin of safety). Documents and reference materials from the Organisation for Economic Co-operation and Development (OECD), European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC), U.S. Environmental Protection Agency (EPA), Research Institute for Fragrance Materials, Inc. (RIFM), the NTP Interagency Center for the Evaluation of Alternative Methods (NICEATM), and the SCCS were reviewed and considered herein. Additional literature reviews were also conducted throughout document development to fill gaps in the initial inventory. Key citations considered when developing this BPG document are presented in Table 1.

**Table 1. Key Documents Reviewed in the Process of Best Practice Guidance Document Development**

| Organization/ Author | Title | Year |
|---|---|---|
| OECD | The adverse outcome pathway for skin sensitisation initiated by covalent binding to proteins. | 2014 |
| OECD | Integrated Approaches to Testing and Assessment (IATA) Reporting for Defined Approaches for Skin Sensitisation | 2016 |
| Pare and Kitsou | Chapter 9 Methods for Literature Reviews. | 2017 |
| ICCR | Integrated strategies for safety assessments of cosmetic ingredients - Part I | 2017 |
| OECD | Guiding principles and key elements for establishing a weight of evidence for chemical assessment | 2019 |
| Api et al. | Updating exposure assessment for skin sensitization quantitative risk assessment for fragrance materials. | 2020 |
| Gilmour et al. | Development of a next generation risk assessment framework for the evaluation of skin sensitisation of cosmetic ingredients. | 2020 |
| OECD | Users' Handbook Supplement to the Guidance Document for Developing and Assessing AOPs | 2022 |
| Gilmour et al. | Applying a next generation risk assessment framework for skin sensitization to inconsistent new approach methodology information. | 2023 |

| Organization/ Author | Title | Year |
|---|---|---|
| SCCS | The SCCS Notes of Guidance for the Testing of Cosmetic Ingredients and Their Safety Evaluation 12th Revision | 2023 |
| UN | Globally Harmonized System of Classification and Labelling of Chemicals (GHS) (10th Rev. Ed.). | 2023 |
| OECD | 497: Defined Approaches for Skin Sensitization | 2025 |
| OECD | Test No. 442C: *In Chemico* Skin Sensitisation - Assays addressing the Adverse Outcome Pathway Key Event on Covalent Binding to Proteins | 2024 |
| OECD | Test No. 442D: *In Vitro* Skin Sensitisation - Assays Addressing the Adverse Outcome Pathway Key Event on Keratinocyte Activation | 2024 |
| OECD | Test No. 442E: *In Vitro* Skin Sensitisation - Assays Addressing the Key Event on Activation of Dendritic Cells on the Adverse Outcome Pathway for Skin Sensitisation | 2024 |

Abbreviations: ICCR: International Cooperation on Cosmetics Regulation; OECD: Organisation for Economic Co-operation and Development; SCCS: Scientific Committee on Consumer Safety; UN: United Nations; WoE: weight of evidence.

This BPG will continue to be updated iteratively as the scope is expanded (e.g., to include newly developed NAMs). Topic areas such as NGRA, NAMs, aggregate and mixtures exposure, and risk assessment are rapidly evolving, and will be continually revisited in future updates.

## 1.5 Comparison with Other Guidance Documents on Skin Sensitization

As noted in Section 1.4, there are already IATAs (OECD 2016a), DAs (OECD, 2025), and primary publications covering the topic of skin sensitization (see Box 1 for definitions). What sets this BPG apart from these pre-existing documents is that it collates the information from these resources into a cohesive explanation of methods and considerations, such that it provides the safety assessor with a comprehensive summary of best practices to approaching a skin sensitization NGRA.

**Box 1. Select Definitions for Decision-Making Approaches**

**Integrated Approach to Testing and Assessment (IATA):** An approach based on multiple information sources used for hazard identification, hazard characterisation, and/or chemical safety assessment. An IATA integrates and weights all relevant existing evidence and guides the targeted generation of new data, where required, to inform regulatory decision-making regarding potential hazard and/or risk (OECD 2016a; OECD 2020).

**Data Interpretation Procedure (DIP):** A rule-based approach or mathematical model used with a defined set of information sources to derive a predicted response (OECD, 2025).

**Defined Approach (DA):** Utilizes a fixed data interpretation procedure (DIP) that is applied to a defined set of data or information sources to yield a prediction, in the absence of expert judgement (OECD, 2025).

**Weight of Evidence (WoE):** An approach that includes evaluating all available data (inclusive of *in vivo*, *in chemico*/*in vitro* and *in silico* data) to determine a chemical's toxicity.

## 1.6 Overview of the Document

This document provides details on how to approach a skin sensitization hazard and/or safety assessment (hereafter referred to collectively as a skin sensitization NGRA). It begins with background information that a safety assessor should be familiar with when conducting a skin sensitization NGRA. It next discusses the importance of reporting and transparency in the assessment, followed by a workflow. The remainder of the document follows the steps outlined in the workflow (see Section 2.3) to conduct a skin sensitization NGRA.

# 2 Skin Sensitization Assessment

## 2.1 Background

Skin sensitization, which can cause allergic contact dermatitis (ACD), is a type IV hypersensitivity reaction of the immune system in response to repeated dermal exposure to a substance. ACD is characterized by pruritus, erythema, and scaling of the skin. Skin sensitization is comprised of two phases: induction and elicitation. During the induction phase, allergens (i.e., haptens) penetrate the skin and bind to skin proteins by forming a stable conjugate, hapten-protein complex and initiating the immune response cascade, which ultimately leads to production of protein-specific memory T-cells. Following subsequent exposure to the same allergen, the elicitation phase may occur, in which an immune response is observed (e.g., release of inflammatory cytokines). The inflammatory cells migrate to the skin and induce local inflammatory reactions (e.g., rash; blisters) (OECD, 2014; Figure 1).

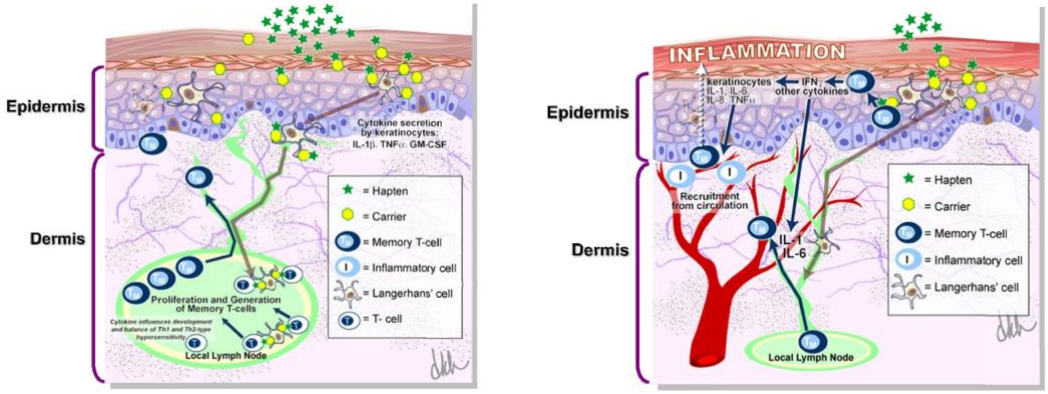



**Figure 1. Graphical Representation of Induction (left) and Elicitation (right) Phases of Skin Sensitization (from OECD, 2014)**

Allergens that induce sensitization may not be sensitizing alone, but can be transformed into a hapten to become sensitizing. These allergens are referred to as pre- or pro-haptens. A pre-hapten refers to a substance that is activated through abiotic transformation, meaning it is activated outside the skin (e.g., oxidation with air). A pro-hapten is activated in the skin (Casati et al., 2016).

Skin sensitization NGRA focuses on the induction phase and looks to ensure that the level of a substance with sensitization potential to which consumers are exposed is far below that which induces sensitization. To achieve this goal, an understanding of both the amount of exposure to a substance and the estimated dose per unit area that results in sensitization

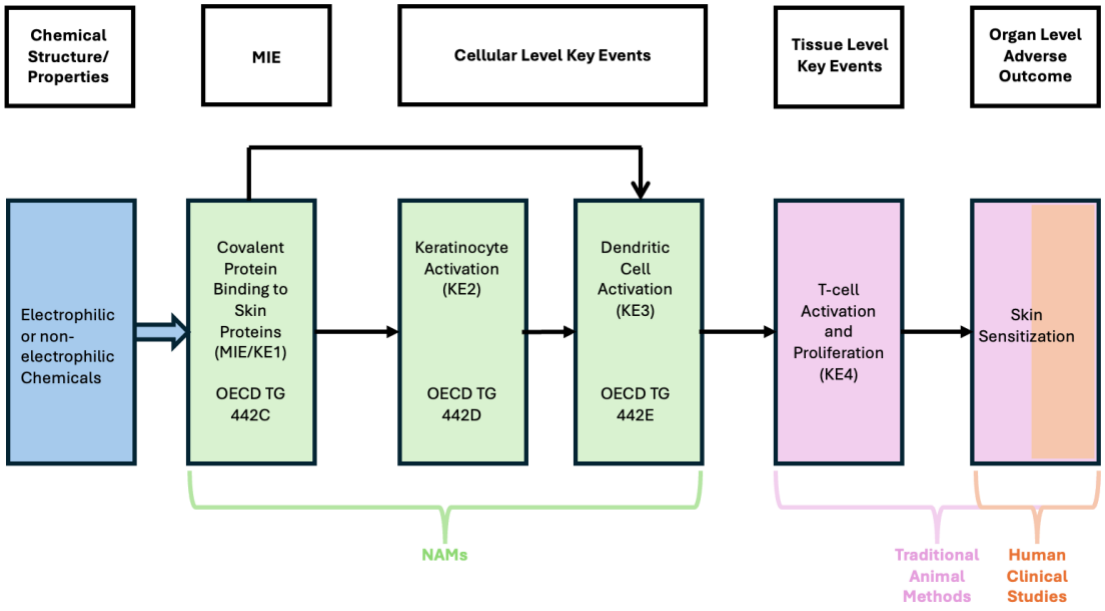
induction is needed. If the exposure is lower than the sensitization induction threshold, including consideration of any uncertainties, a low risk of sensitization induction can be assumed.

### *2.1.1 Skin sensitization biology and adverse outcome pathway*

**Box 2. Overview of the Adverse Outcome Pathway (AOP) Concept**

AOPs are frameworks for organizing scientific data and knowledge linking a molecular perturbation to an adverse outcome (AO) – in this case skin sensitization (SAAOP, 2025; OECD, 2022). AOPs are composed of key events (KEs) and are visualized as a sequence of biological and cellular events that are initiated through an initial perturbation, referred to as a molecular initiating event (MIE) or KE1. Key event relationships (KERs) identify the links between individual KEs, and represent the causal and predictive relationship between the upstream and downstream KE. AOPs, by default, provide a structure by which assess the KERs (SAAOP, 2025; OECD, 2022a).

Historically, evaluating the potential for skin sensitization has been carried out through either animal studies (see Section 2.1.2) or human studies (see Section 2.1.3). These tests were designed to measure either the skin sensitization reaction at the organism level or T-cell activation. However, as the science for skin sensitization has progressed, a well-defined pathway for the induction phase of skin sensitization, initiated by covalent binding has led to developing an adverse outcome pathway (AOP). For each key event (KE) present in the AOP, NAMs have been designed (see Section 2.1.4). These NAMs specifically measure test substance ability to induce KE activity in the AOP. Figure 2 summarizes the skin sensitization AOP, and shows which KEs in this AOP are covered by traditional animal tests, human tests, and NAMs. By focusing on upstream KEs, the advancement of NAMs have enabled safety assessments to target potential risk assessment interventions based on molecular or cellular signals from which key activation events cascade, and therefore should provide for greater confidence in the degree of protection that can be inferred from skin sensitization NGRAs.




Abbreviations: MIE: molecular initiating event; KE: key event

**Figure 2. Flow Diagram of the AOP and the KEs Associated with Skin Sensitization (adapted from OECD, 2014)**

### *2.1.2 Animal tests*

*In vivo* methods historically used to evaluate skin sensitization include the Buehler test, the GPMT, and the murine LLNA. These tests were designed to provide data on the induction phase at the organism (for Buehler and GPMT) or organ (for LLNA) level. The Buehler test and GPMT provide qualitative data on the skin sensitization potential of a substance. Additional information is available in OECD TG 406 (OECD, 2022b).

The LLNA evaluates the skin sensitization induction phase, but does so by measuring the proliferation of activated T-cell lymphocytes. Additional information is available in OECD TG 442A, 422B or 429 (OECD, 2010a; 2010b; 2024d).

Although the GPMT and Buehler test data provide a qualitative assessment of skin sensitization potential (i.e., hazard data), approaches have been applied that allow for using data from these studies in safety assessments. Specifically, the qualitative hazard classification from these tests have been mapped to approximate No Expected Sensitization

Induction Levels (NESIL)[4] by ECETOC (2003) and others (e.g., Api et al. 2017), which can then be used in quantitative risk assessments (see Api et al. 2020).

Compared to the Buehler test and GPMT, the LLNA provides quantitative data that may be used in a dose-response evaluation. Specifically, the Stimulation Index (SI) is used to derive the EC3, which is defined as the amount of substance required to induce a three-fold increase in lymph node cell proliferation compared with vehicle control values (ECETOC, 2003). The unit of EC3 can be expressed as the percentage concentration of test substance required or as dose per unit area of skin (i.e., $\mu g/cm^2$). If expressed directly as a dose per unit area, this value can be used as a NESIL. If expressed as a percentage, the percentage can be multiplied by a factor of 250 to achieve units of $\mu g/cm^2$. This value is based on the standard LLNA protocol where 25 µL test solutions are distributed over a surface of 1 $cm^2$ per mouse ear (Robinson 2000). The converted value can then be assumed to represent the NESIL. Additionally, the EC3 value may be used for skin sensitization hazard classification according to the United Nationals Globally Harmonized System of Classification (UN, 2023).

### *2.1.3 Human clinical studies*

Human studies are not to be conducted for hazard identification for ethical reasons but may be conducted to confirm lack of sensitization. Evaluating skin sensitization in humans via clinical studies can be done with a variety of methods, such as the human repeat insult patch test (HRIPT),[5] human maximization test (HMT), and diagnostic patch tests. A standardized protocol for the HRIPT used by RIFM is available for review by Politano and Api (2008). Overall, the protocol has two phases: induction and challenge. In the induction phase, patches treated with fragrance ingredients are applied to backs of volunteers for 24 hours, followed by a 24-hour rest period, and then applied again to achieve nine applications over three weeks. After a 2-week rest, a single patch is applied, and reactions are observed. An equivocal or positive response in these tests is considered indicative of a skin sensitization potential (Lee et al., 2022), and that the pre-determined NESIL is not protective of the human population. An HMT is similar to HRIPT, although the major difference between the two tests is that the HMT (a) evaluates a smaller number of individuals; and b) includes pretreatment with an irritant to enhance skin penetration. Comparatively, a diagnostic patch test is conducted on individuals to determine the source of observed ACD (Strickland et al., 2023).

---

[4] A NESIL is analogous to a point of departure for a systemic endpoint. A guide on the steps, utilized by the Research Institute for Fragrance Materials (RIFM), for NESIL derivation are outlined in Lee et al. (2022).

[5] Referred to as the "confirmation of no induction in humans" by RIFM (Lee et al., 2022).

Human studies, such as HRIPTs, may be used to confirm the animal or *in vitro*-derived NESIL for inducing a skin sensitization response in a normal human population (Lee et al., 2022; Na et al., 2020).

### *2.1.4 NAMs used for skin sensitization assessment*

As noted above, the science for skin sensitization has progressed, leading to an internationally accepted AOP. Consequently, the OECD Working Party on Hazard Assessment (WPHA) and Working Group of the National Coordinators for the Test Guidelines Programme (WNT) endorsed an AOP for skin sensitization, shown in Figure 2. It is this AOP upon which NAMs have been developed to evaluate skin sensitization. These NAMs are designed to provide more human-relevant information while negating the ethical and scientific concerns with whole-animal studies. In addition, NAMs also have shorter timelines for studies (Doke and Dhawale, 2015).

The skin sensitization NAMs described in this BPG can be considered in three main categories (EPA, 2024):

- *In chemico*: evaluate how a substance interacts or reacts with assay substrates (without animal or human cells) to produce an effect
- *In vitro*: include the use of human or animal cells
- *In silico*: computer or computationally based.

Table 2 provides a summary of the available *in chemico* and *in vitro* NAMs that are endorsed by OECD as they pertain to skin sensitization. NAMs are constantly evolving, and, as such, Table 2 does not encompass all available methods at the time of writing, since there are many more assays beyond those that OECD endorses. Literature searches and review of the adopted OECD Test Guidelines (TG) may be conducted to identify other methods and/or models that may provide relevant mechanistic skin sensitization data. Additional information about the individual *in chemico* and *in vitro* models are provided in Section 2.8.2 and Appendix A.

**Table 2. *In Chemico* and *In Vitro* Skin Sensitization Methods**

| Test Method/OECD Test Guideline | Key Event Assessed (based on AOP shown in Figure 2) |
|---|---|
| *In chemico* | |
| Direct Peptide Reactivity Assay (DPRA)/OECD TG 442C | MIE/KE1: covalent binding of substances to nucleophilic centers in proteins |
| Amino acid Derivative Reactivity Assay (ADRA)/OECD TG 442C | MIE/KE1: covalent binding of substances to nucleophilic centers in proteins |
| Kinetic DPRA (kDPRA)/OECD TG 442C | MIE/KE1: covalent binding of substances to nucleophilic centers in proteins |

| Test Method/OECD Test Guideline | Key Event Assessed (based on AOP shown in Figure 2) |
|---|---|
| *In vitro* | |
| KeratinoSens™/OECD TG 442D | KE2: changes in induction of Nrf2-mediated activation of antioxidant response element (measured by luciferase) |
| LuSens/OECD TG 442D | KE2: changes in induction of Nrf2-mediated activation of antioxidant response element (measured by luciferase) |
| EpiSensA/OECD TG 442D | KE2: changes in the expression of four marker genes associated with keratinocyte activation in a reconstructed human epidermis model |
| human cell line activation test (h-CLAT)/OECD TG 442E | KE3: activation of dendritic cells |
| U937 cell line activation test (U-SENS™)/OECD TG 442E | KE3: activation of dendritic cells |
| Interleukin-8 reporter gene assay (IL-8 Luc assay)/OECD TG 442E | KE3: activation of dendritic cells |
| Genomic Allergen Rapid Detection (GARD™) for assessment of skin sensitizers (GARD™skin)/OECD TG 442E | KE3: activation of dendritic cells |

Abbreviations: KE: key event; MIE: molecular initiating event; OECD: Organisation of Economic Co-operation and Development

As shown in Table 2, the *in chemico* or *in vitro* OECD TG methods address the first three KEs of the skin sensitization AOP (Figure 2).

In addition to these tools, a range of *in silico* models have also been developed to address the endpoint of skin sensitization (e.g., OECD QSAR Toolbox; Derek Nexus; TIMES-SS; LeadScope Model Applier; StopTox; iSafeRat). These models provide a range of outputs, including LLNA potency, potency classification and presence of reactivity domains. The quality and confidence in the predictions is dependent upon a range of parameters (including quality of the data used in model development and chemical space encompassed by these models) (Wilm et al., 2018; Gilmour et al., 2020; Selvestrel et al., 2022). Additional details about available models and associated uncertainties are provided in 2.5.2.2.

While results from most assays or *in silico* models are not usually considered as direct replacements of animal data for assessing skin sensitization potential, these data can be used in a DA or in a WoE evaluation, which will be further described in the remainder of this document.

## 2.2 Reporting

In conducting any safety assessment, such as an NGRA, documentation and transparency throughout the process is critical for a variety of reasons, including being a requirement for regulatory approval in most jurisdictions. Therefore, reporting should:

- Include adequate information in order replicate findings
- Allow for critical evaluation of methods and results

Additional information on adequate reporting for skin sensitization DAs and individual information sources to be used within IATA can be reviewed in OECD guideline 255 and 256 (OECD, 2017a, 2017b) and a general template for IATA case studies has also been published (OECD, 2022c). Further, any expert judgement should be explained and justified to the fullest extent such that any conclusions are well supported.

Though a reporting template is not provided in this guidance document, indications for recording decisions, input variables, and assumptions throughout the process are called out. The safety assessor is referred to Box 3 for OECD resources that propose templates as they pertain to reporting when using IATAs and DAs, along with associated case studies using these templates.

**Box 3. References for Templates and Examples as They Relate to Reporting**

- Guidance on the reporting of defined approaches to be used within integrated approaches to testing and assessment (OECD GD 255): Provides templates for reporting defined approaches to testing and assessment, and for reporting individual information sources. Accessible at: https://doi.org/10.1787/9789264274822-en.

- Guidance document on the reporting of defined approaches and individual information sources to be used within integrated approaches to testing and assessment for skin sensitization (OECD GD 256): Provides case studies for skin sensitization which include demonstration on the use of reporting templates. Accessible at: https://doi.org/10.1787/9789264279285-en.

- Report on considerations from case studies integrated approaches for testing and assessment (IATA): Series on Testing and Assessment No. 369 (Annex 6). Accessible at: https://www.oecd.org/en/topics/sub-issues/testing-of-chemicals/publications-on-testing-and-assessment-of-chemicals.html.

## 2.3 Overview of Workflow

The proposed workflow with which to conduct a skin sensitization NGRA is presented in Figure 3. This workflow incorporates NGRA concepts and the available approaches for skin sensitization hazard or safety assessment. The workflow is designed to be flexible, and to address a range of questions related to skin sensitization, including:

- Does the substance have a potential for skin sensitization: yes or no?
- What is the potency of the substance (e.g., GHS classification)?
- Is there an adequate margin of exposure for the defined use scenario for a safety assessment?

The remainder of this guidance document will walk through each of the steps proposed in the workflow for both hazard and safety.

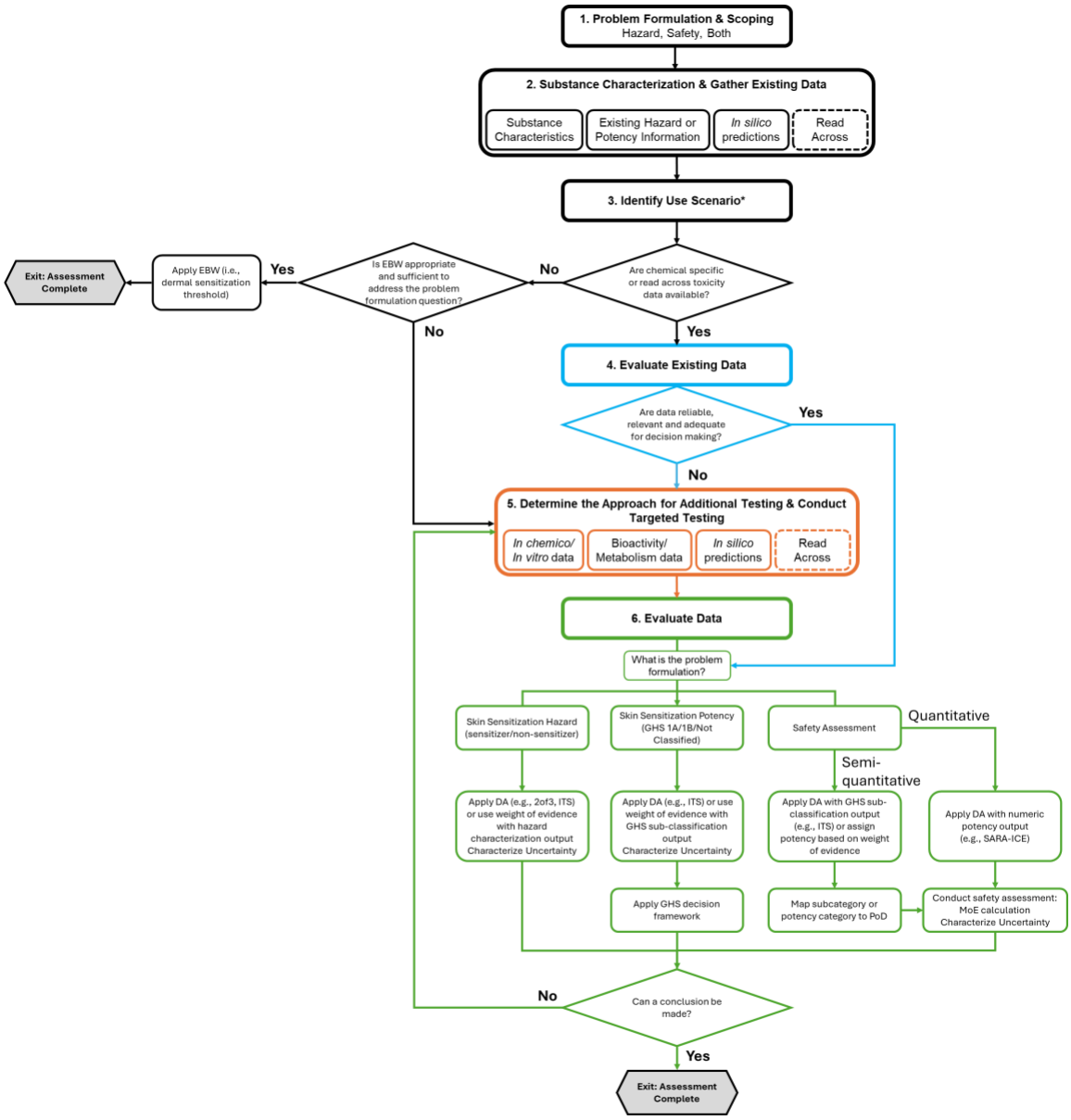



**Figure 3. Overview of Process to Evaluate Skin Sensitization using Next Generation Risk Assessment**

Notes: As needed, the safety assessor can iterate through the phases of this process.
*Step should be skipped if focus is only on hazard evaluation; ^additional details on hazard and safety evaluations are provided in Section 2.9; EBW = exposure based waiving; PF = problem formulation; DA = defined approach; 2of3 = 2 out of 3 Strategy; SARA-ICE: Skin Allergy Risk Assessment – Integrated Chemical Environment Model; MoE = margin of exposure; ITS = integrated testing strategy; GHS = Globally Harmonized System classification; PoD = point of departure

## 2.4 Step 1. Problem Formulation and Scoping: Define the Context for the Decision

The initial step for any cosmetic NGRA is problem formulation and scoping. During this step, the context of the assessment will be defined and documented. The assessment objectives will also be defined. The scoping process for each assessment will be different, and will be an iterative process carried out by assessors as appropriate. The general considerations for this step are presented in Figure 4.

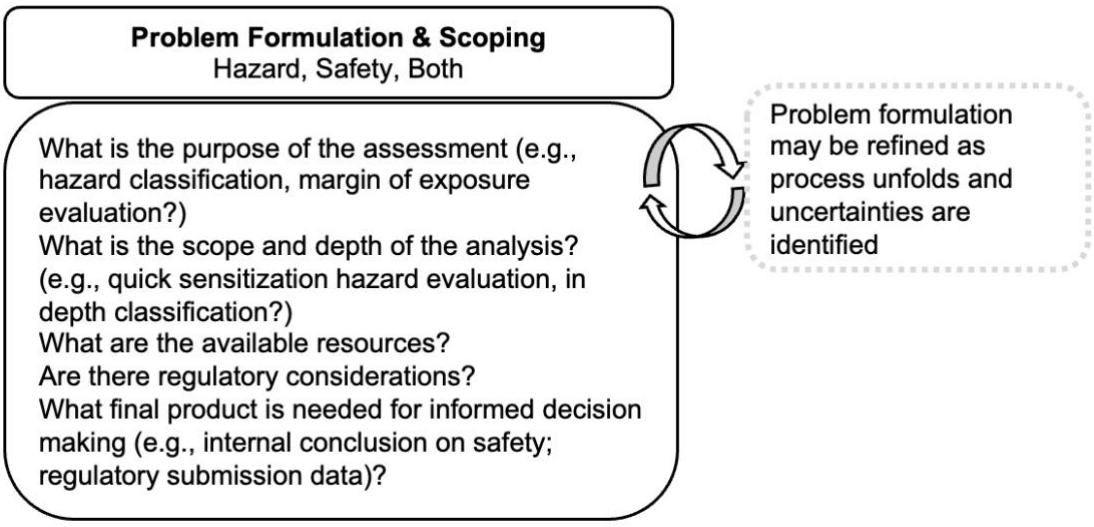



**Figure 4. Considerations for Problem Formulation and Scoping**

Critical to setting the scope, information should be provided to define the context of the assessment. This step could include:

- A description of why there is interest in evaluating the substance. This explanation could range from a research & development department identifying a new substance of interest and needing to determine if it is worth pursuing, to a company developing a safety assessment prior to a full product launch, to full regulatory use.
- Identification of any specific skin sensitization concerns regarding the substance that prompted the evaluation (e.g., increase in adverse events being reported related to skin rashes).
- Reference to stakeholder interest, questions, or concerns.
- Timing or resource considerations (e.g., the need for a rapid review).

**Step 1 Objective**: Provide a clear statement of the assessment scope and the hypothesis to be evaluated.

## 2.5 Step 2. Substance Characterization and Gathering Existing Data

Once the scope has been defined, the next step is to characterize the substance of interest and to collect the available existing hazard data available on the substance. An overview of this step is depicted in Figure 5, with additional description following in the narrative.

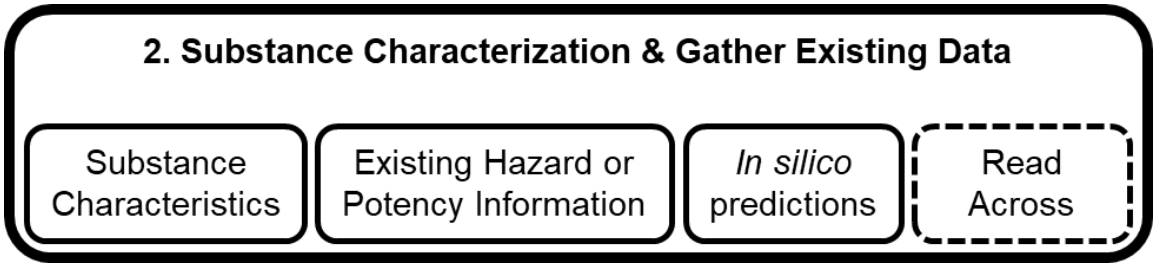



**Figure 5. Overview of the Substance Characterization and Identification of Existing Information Process (dashed line indicates that read-across at this phase in the evaluation can be decided on a case-by-case basis)**

### *2.5.1 Substance characteristics*

Characterizing the substance is critical to understanding its safety, given that substance properties play an integral role in a substance's skin sensitization potential. A basic data matrix required to characterize the substance of interest is provided in Table 3. These data should be collected and recorded. These data points in the table should not be deemed to be exhaustive, and physico-chemical properties proposed to play a role in activity should be captured for consideration (e.g., pH). Conditions in which experimental values for properties (e.g., temperature used in assessing solubility) exist should also be captured for completeness. If computational models are used to predict these properties, the platform and version of the model(s) must also be captured.

**Table 3. Summary of Data to Identify and Describe Substance of Interest**

| Element | Description* |
|---|---|
| Common name | [Name which the substance will be referred to] |
| INCI name | [International Nomenclature of Cosmetic Ingredients name] |
| Synonyms | [The most commonly used synonyms.] |
| CASRN | [If there are multiple Chemical Abstracts Service Registry Numbers (CASRNs), note the distinction between them. (If European Inventory of Existing Commercial Chemical Substances (EINECS) number is also important, an additional field can be added)] |
| Structure | [Insert molecular structure] |
| Molecular formula | [Insert substance molecular formula] |
| Molecular weight (or MW distribution, if applicable) | [Identify substance molecular weight] |

| Element | Description* |
|---|---|
| Particulate size or size range (if applicable) | [If relevant, identify the size of substance particulates. If range of size, enter the total range of particulate sizes encompassing the minimum and maximum sizes] |
| Purity | [Identify the substance purity] |
| Known impurities | [Identify known impurities within the substance. If concentrations are available, include information] |
| Potential impurities | [Identify potential impurities within the substance] |
| Physical form | [Identify the physical form of the pure substance (e.g., solid)] |
| Water solubility | [Identify the predicted or experimental water solubility of the substance] |
| Partition coefficient | [Identify the partition coefficient between two immiscible solvents] |
| Vapor pressure | [Identify the predicted or experimental vapor pressure] |
| Additional physico-chemical properties that impact skin sensitization | (e.g., presence of electrophilic functional groups, pH; density) |
| Non-cosmetic uses or exposures | [Insert any known uses of the substance that are non-cosmetic] Examples: antiseptic; solvent; food additive |

Substance physico-chemical properties that may be of key relevance to evaluating skin sensitization potential include those that would inform skin exposure and penetration, include partition coefficient, vapor pressure, water solubility, and molecular weight (Strickland et al., 2017). These data points may be collected from a wide range of sources, including:

- PubChem: https://pubchem.ncbi.nlm.nih.gov/
- ECHA CHEM Database: https://chem.echa.europa.eu/
- ChemSpider: https://www.chemspider.com/
- NIST Chemistry WebBook: https://webbook.nist.gov/chemistry/
- Merck Index: https://merckindex.rsc.org/
- Chemical supplier catalogs (e.g., Sigma Aldrich): https://www.sigmaaldrich.com/

When collecting these data, experimental properties are sometimes preferred over calculated properties. Reliability of experimental data should also be considered and differing calculated properties should be evaluated further. As shown in Figure 2, the initiation of a skin sensitization response is the interaction of a substance with skin proteins to produce a covalent bond. The formation of a covalent bond may occur through electrophilic or non-electrophilic reactions. Substances may undergo transformation to an electrophile (i.e., pro-hapten). Many of the amino acids present in proteins contain structural features (i.e., nucleophilic centers) that are reactive. Reaction mechanisms that

may occur are: Michael acceptors and pro-Michael type acceptors, SNAr reaction, SN2 reaction, Schiff Base, acylating agents, and no-evidence electrophile or pro-electrophilic features (Aptula et al., 2005). Therefore, the presence of electrophilic or pro-electrophilic functional groups in the substance structure should be recorded by the assessor.

### *2.5.2 Gather Existing Data*

The process of gathering existing information is critical to framing a complete assessment, as seen in Figure 3. If enough data are available to conclude an adequate WoE to inform decision making,[6] additional data generation or application of non-testing methods such as exposure-based waiving may not be needed. The available data also will provide direction as to what data are needed to best inform the decision-making process. As such, when entering this phase of the assessment, the safety assessor should think critically about the approach for reviewing the existing data, including literature, information from *in silico* tools, and potentially data from analogue substances (read-across). Additionally, this step is not an isolated phase in the assessment process, but instead encompasses collating and distilling data that may be used and integrated through all steps presented in this guidance document.

Ultimately, the objective of the substance characterization and gathering existing data step is to collect data on the substance that will inform the resulting skin sensitization assessment. The data collected should include substance physico-chemical properties, existing hazard and potency data, *in silico* prediction data (e.g., metabolites) and potential analogue compounds.

#### *2.5.2.1 Existing hazard or potency information*

Existing hazard, potency, or safety data for the test substance should be gathered using a transparent protocol. Because each assessment will be different, this section will provide general methods to gather and assess data efficiently to inform substance characterization, hazard identification, and safety assessment. Those not familiar with generally agreed upon literature review processes and steps should refer to Pare and Kitsiou (2017).

Before embarking on a *de novo* literature review for primary publications, existing collated information should be identified. This task can be done by reviewing existing authoritative assessments by government agencies and scientific advisory groups (e.g., U.S. Food and Drug Administration (FDA), EPA, SCCS, World Health Organization (WHO), Cosmetics

[6] This document assumes that the safety assessor is aware of the minimum data requirements needed to do a traditional skin sensitization safety assessments, such that that is not explained herein. If additional information is needed to inform a skin sensitization safety assessment based on animal or human data, the reader is directed to the SCCS Notes of Guidance (2023).

Ingredient Review (CIR)). If an existing data collation from a trusted scientific body is available, the reviewer should critically evaluate the information in comparison to the context of the current assessment, and determine if any scope refinements are warranted. In addition, the assessor should determine how previous assessments will (or will not) be used. Based on this evaluation, the reviewer should determine if the problem formulation question needs to be refined. Prompting questions to consider in this evaluation include, but are not limited to:

- Are there assessments of skin sensitization potential in existing SCCS or CIR reviews? If yes, how does the scope compare? What is the year of assessment? How can these be used pragmatically for the current assessments?
- In relevant assessments, was a NESIL identified? Are there hazard conclusions by endpoint that could be considered?
- How can the current assessment be made more efficient and pragmatic based on existing knowledge?
- Do the existing data demonstrate a particularly sensitive subpopulation that needs to be considered in the assessment (e.g., genetically susceptible groups; children)?
- Are there any key issues identified across the available assessments that this assessment needs to specifically address?

Next, a search should be conducted to evaluate the availability of literature that may address outstanding data gaps or new questions raised after evaluating existing authoritative assessments. The systematic review process is considered a gold standard when evaluating a body of literature (Cochrane, 2023). The reviewer will need to balance the needs of conducting an independent assessment with considering the volume of literature and assessment scope when making such determinations. The level of reliance on previous assessments will vary on a case-by-case basis, and could range from no reliance at all to heavily relying on a previous assessment.

As part of developing the search strategy, the substance name and synonyms, INCI name, and CAS registry number should be identified. Searches should be conducted to identify data relevant to the problem formulation and data gaps identified after reviewing authoritative assessments. Data reviewed may include *in vivo* experimental animal data (e.g., LLNA; GPMT), and *in chemico* and *in vitro* studies (described in Section 2.1). Human exposure studies (e.g., HRIPT, HMT) and epidemiological data that include skin sensitization should be evaluated, if available.

The overall search strategy, including searches for authoritative documents and searches for primary literature, should be recorded by the safety assessor. This recording should be done in a manner allowing for full transparency and reproducibility. As such, the

documentation should include: databases or websites searched; dates searched; and specific search terms and/or filters used for each search. For literature citation databases (e.g., PubMed), the search syntax used should be reported, as well as the number of hits identified. Citations that were reviewed, but not identified through literature citation database searches, should also be captured.

The data gathering and evaluation approach can align with systematic review methods, on a case-by-case basis, as it fits with the problem formulation and scope. The systematic review process includes developing a protocol, critically appraising all included studies, and formally assessing confidence in the body of evidence and conclusions (Wikoff et al., 2020). As such, the data gathering and evaluation approach described herein aligns with systematic review methods, but does not describe a formal systematic review.

#### *2.5.2.2 In silico data*

In addition to literature sources, *in silico* tools may be used as part of the data gathering process to assist with assessing potential skin sensitization hazard and potency. While not

**Box 4. Examples of *in silico* Platforms and Models for Skin Sensitization Hazard and Potency Evaluations**

**Hazard assessments**

- Leadscope Model Profiler (https://www.instem.com/solutions/discovery/leadscope-model-applier/)
- Toxtree: Skin Sensitization Reactivity Domains (https://toxtree.sourceforge.net/)
- OECD QSAR Toolbox: Protein Binding Alerts for Skin sensitization by OASIS (https://qsartoolbox.org/)
- Derek Nexus Skin Sensitization (https://www.lhasalimited.org/)
- Danish QSAR Database (https://qsar.food.dtu.dk/)
- TIMES-SS (https://oasis-lmc.org/products/software/times.aspx)
- StopTox (https://stoptox.mml.unc.edu/)
- iSafeRat (https://isaferat.kreatis.eu/)

**Potency predictions**

Leadscope Model Profiler (https://www.instem.com/solutions/discovery/leadscope-model-applier/)

- Derek Nexus: Skin Sensitization EC3 (https://www.lhasalimited.org/)
- SARA-ICE (https://ntp.niehs.nih.gov/go/n465041)

**Mechanistic endpoints**

- OECD QSAR Toolbox (https://qsartoolbox.org/)
- Leadscope Model Applier (https://www.instem.com/solutions/discovery/leadscope-model-applier/)

intended to be all inclusive, particularly considering the continuous developments around *in silico* tools, a list of several models is included in Box 4 for reference.

Selecting the appropriate *in silico* model will depend on the assessment purpose (hazard identification versus potency or point of departure prediction). For hazard assessments (i.e., sensitizer vs. non-sensitizer), models that assess the presence of structural features that may covalently bind to proteins in the skin and/or result in a binary prediction of skin sensitization (sensitizer/non-sensitizer) may be used. Models that provide potency or point of departure predictions (e.g., predicted LLNA EC3 values) may be useful for safety assessments. Models assessing specific mechanistic endpoints (e.g., keratinocyte activation) relevant to skin sensitization also are available, and may provide supporting evidence for skin sensitization. Additional information about the individual modeled mechanistic endpoints (e.g., mechanism evaluated; KE information) is provided in Section 2.8. Expert guidance on identifying appropriate models and interpreting predictions is detailed in the OECD QSAR Assessment Framework (OECD, 2023a).

As detailed in Gilmour et al. (2020), the quality and usefulness of the predictions from computational models is related to the size, quality, and chemical space of the training set used to develop the model. These three factors identify a model's applicability domain. The *in silico* model applicability domain refers to the chemical space (as defined by model inputs) for which the model is considered scientifically valid. The model inputs used to define the chemical space are the parameters used to develop the model. These parameters may include structural features and/or physico-chemical properties (Sahigara et al., 2012). If the model and results predicted fulfill OECD Validation Principles and the QSAR Assessment Framework, the results may be considered acceptable. However, expert evaluation should be incorporated when assessing *in silico* predictions and confidence in the results based on the applicability domain. Additional information on assessing the applicability domain of models and impact on confidence in the prediction results are provided in the QSAR Assessment Framework (OECD, 2023a).

For all *in silico* assessments, documentation should be maintained for reporting transparency. The *in silico* platform, models, and version of both the platform and model should be retained. Information regarding any limitations or constraints placed on the model (e.g., not including an evaluation of tautomeric forms of the substance) should be noted, and the rationale for these selections maintained. Rationale for selecting and using selected platforms and models also should be maintained for transparency. Finally, assessing confidence in the results should be maintained with the output results. Additional information on documentation is also reviewed in OECD (2023b; 2023c).

#### *2.5.2.3 Read-across*

The scientific community uses read-across at different points in the workflow presented. Read-across assessments can also be used to support results for endpoints that have limited or conflicting data. Using this type of assessment is based on the needs as assessed by problem formulation and the available evidence base identified in Sections 2.4 and 2.5.

For all read-across evaluations, the rationale for analogue selection and groupings should be clearly provided. If groupings are based on the presence of structural features, rationale and supporting evidence should be provided (e.g., presence of structural feature that will produce a covalent bond with skin proteins). Structural similarity rationales should be accompanied by the calculation method and calculated similarity scores. Inclusion and/or exclusion of substances for read-across should also be provided for transparency. Publications such as the OECD Guidance on Grouping of Chemicals (OECD, 2017c), and the upcoming ICCS Best Practice Guidance on read-across may be useful resources on how to conduct a read-across assessment. There are many other resources available for read-across assessment (e.g., Helman et al., 2019; Patlewicz et al., 2019; EFSA, 2024), but a comprehensive list is outside the scope of this document.

**Step 2 Objective:** Collect data on the substance that will inform the resulting skin sensitization assessment. The data collected should include substance physico-chemical properties, existing hazard and potency data, *in silico* prediction data (e.g., structural alerts, metabolites) and potential analogue compounds.

## 2.6 Step 3. Identify Use Scenario

Identifying the use scenario and exposure amount is an initial step in conducting a substance safety assessment for any endpoint.[7]

For skin sensitization, the consumer exposure level (CEL) or the amount of the substance for which there is local exposure at the site-of-contact, is needed. This dose metric is in units of $\mu g/cm^2$. To determine an accurate CEL, the safety assessor should gather the data outlined in Table 4. Evaluating the exposure can be done using a deterministic (first tier) or probabilistic (higher tier) approach.

[7] This step is not needed for a hazard assessment.

**Table 4. Data to Collect to Support Derivation of a Consumer Exposure Level (CEL)**

| Element | Description |
|---|---|
| Product types of interest | [Insert product types to be assessed for the substance]<br>Example: Body lotion and shower gel |
| Concentration of substance in each product | [Insert the substance amount that will be in each product of interest]<br>Example: 1% in all products (provide data source) |
| Method and location of application for each product type | [Insert description of application method and intention for leave on/rinse off]<br>Example: Rubbed on body lotion, applied to full body, and then rinsed off |
| Population, including targeted or special consumer groups (if applicable) | [Insert description of targeted or special consumer group considerations]<br>Examples: substance or product known to be used by children or people with sensitive skin |

In addition, the assessor should consider the problem formulation and whether the use scenario is most appropriate for a single product, or if an exposure aggregation across multiple products is needed. In this section, a deterministic, single product exposure approach is described. For additional details on more complex exposure assessments that integrate probabilistic evaluations, the reader is referred to useful resources referenced in Box 5. The SCCS (2023) Notes of Guidance also provide some insight regarding aggregate exposure considerations.

**Box 5. Recommendations for Further Reading for Aggregate and Probabilistic Exposure Evaluations**

- ECETOC. 2016. Guidance for Effective Use of Human Exposure Data in Risk Assessment of Chemicals. Available online at: https://www.ecetoc.org/publication/technical-report-no-126-guidance-for-effective-use-of-human-exposure-data-in-risk-assessment-of-chemicals/
- EPA. 2021. Exposure Assessment Tools by Tiers and Types – Deterministic and Probabilistic Assessments. Available online at: https://www.epa.gov/expobox/exposure-assessment-tools-tiers-and-types-deterministic-and-probabilistic-assessments
- Delmaar CJE, et al. 2022. PACEMweb: A tool for aggregate consumer exposure assessment. https://doi.org/10.1038/s41370-022-00509-7.
- Dudzina T, et al. 2015. The probabilistic aggregate consumer exposure model (PACEM): Validation and comparison to a lower-tier assessment for the cyclic siloxane D5. https://doi.org/10.1016/j.envint.2015.03.006
- Safford B, et al. 2015. Use of an aggregate exposure model to estimate consumer exposure to fragrance ingredients in personal care and cosmetic products. https://doi.org/10.1016/j.yrtph.2015.05.017
- Safford B, et al. 2017. Application of the expanded Creme RIFM consumer exposure model to fragrance ingredients in cosmetic, personal care and air care products. https://doi.org/10.1016/j.yrtph.2017.02.021

Based on the substance's proposed use, several sources of data that can assist in identifying pre-established CELs, including the SCCS (2023) Notes of Guidance. In addition, Ficheux et al. (2019) presents a comprehensive review of consumption data (e.g., percentage of users; frequency of use; amount used, etc.) of cosmetic products across the globe. As described by the authors, "vehicle or formulation effects and frequency results obtained in the different studies, these data are not presented [in the publication]. Readers interested in one or more studies are invited to consult the results available in the original publication" (Ficheux et al. 2019, p. 281). This data source can be used to locate alternate studies that could potentially inform refinements to the CEL or variables for its calculation.

If an assessor needs to calculate a *de novo* CEL, this process begins by understanding $E_{dermal,}$ the product amount to which a user is dermally exposed. It is calculated as the product of the substance concentration in the personal care product ($C$), the amount of the product applied on a daily basis ($q$), and how much of the product stays on the skin (or in the mouth) after application, also known as the retention factor ($f_{ret}$). For leave-on products

(e.g., body lotion), the retention factor is usually one, whereas for rinse-off products (e.g., shampoo) the retention factor is less. Typical retention values for specific products can be found in the published literature, and also within other guidance documents (SCCS, 2023).

Regarding $q$, if a published value is not readily available, data on the amount of a product used per application, multiplied by the number of times in a day it is used, will result in $q$. That is:

$$q = AA \times f \qquad \text{(Equation 1)}$$

Where:

- $q$ (µg/day) = total amount of cosmetic product that is applied per day
- $AA$ (µg) = Amount of cosmetic product applied in a typical application. If the data allow, using the 90$^{th}$ percentile from a distribution for this parameter is recommended[8]
- $f$ (per day) = Frequency of application of the finished cosmetic product, based on product use information available in the literature.

Using data on $q$ combined with the retention factor for a product category $E_{product}$ is calculated as follows:

$$E_{product} = q \times f_{ret} \qquad \text{(Equation 2)}$$

Where:

- $E_{product}$ (µg/day) = the daily amount of the cosmetic product to which a user is externally exposed
- $q$ (µg/day) = total amount of product that is applied per day
- $f_{ret}$ = Product-specific retention factor (ranges from 0.01 to 1).

To translate $E_{product}$ to the amount of the substance a user would be exposed to ($E_{substance}$), the assessor needs to combine the data on the concentration ($C$) of the substance in the product with $E_{product}$ for the product category to calculate the substance's estimated external dose ($E_{dermal}$). When conducting the safety assessment, to be representative of 'worst case'

[8] 90th percentile of the Monte-Carlo distributions or derived distributions is recommended in accordance with SCCS practice. 95th percentile will be used in cases in which the 90th percentile is not reported.

the concentration used in these equations should be informed by the maximum amount potentially intended for consumer exposure.

$$E_{dermal} = \left(\frac{C}{100}\right) \times E_{product} \quad \text{(Equation 3)}$$

Where:

- $E_{dermal}$ (µg/day) = the daily amount of the substance to which the skin is exposed
- $C$ (%) = concentration of the substance in the cosmetic product
- $E_{product}$ (µg/day) = the daily amount of the cosmetic product to which a user is externally exposed.

For a skin sensitization assessment, the $E_{dermal}$ value in the unit of µg/day needs to be normalized by the area of the skin applied. That is, the local dose should be divided by the skin surface area (SSA) to derive a CEL. Equation 4 below presents the method to calculate the CEL:

$$CEL = \frac{E_{dermal}}{SSA} \quad \text{(Equation 4)}$$

Where:

- CEL (µg/cm$^2$) = consumer exposure level
- $E_{dermal}$ (µg/day) = the daily substance amount to which the skin is exposed
- SSA (cm$^2$) = Skin surface area expected to be exposed to the cosmetic product, based on product use information available in the literature. If the data allow, the 90$^{th}$ percentile from a distribution for this parameter should be used.

All inputs selected for deriving the CEL should be recorded by the safety assessor.

### *2.6.1 Apply Exposure-Based Waiving (i.e., dermal sensitization threshold)*

At this phase in a safety assessment, in the absence of substance-specific information or as part of the WoE, the assessor should evaluate if the exposure estimated is such that exposure-based waiving could be implemented,[9] as is standard practice when implementing a NGRA (Berggren et al. 2017). Exposure-based waiving is when a full quantitative safety assessment is determined unnecessary because the estimated exposure

[9] Exposure-based waiving assessment is not needed when conducting a skin sensitization evaluation specifically for hazard evaluation purposes.

to a chemical is anticipated to be negligible or below a pre-defined safety threshold. For sensitization, exposure-based waiving occurs by comparing the estimated exposure to a pre-determined threshold based on the reactivity potential of the substance, as shown in Figure 6.

The dermal sensitization threshold (DST) approach, first proposed by Safford (2008), proposes dermal exposure levels that can be determined for a substance for which there is no appreciable risk of skin sensitization induction; the DST is based on the principles of the threshold of toxicological concern (TTC) approach. A fundamental difference in the two approaches is that application of a DST requires risk assessors to apply sensitization assessment factors (SAFs) to determine the acceptable exposure level (AEL) to be compared with the consumer exposure level (CEL). Commonly applied SAFs are found within Appendix B. The DST approach was expanded by Safford et al. (2011), who increased the underlying dataset from 167 to 271 skin sensitizing substances and better defined the approach's chemical applicability domain. Most recently, Chilton et al. (2022) expanded the data set to update the DST values using more than 1,100 data points and using an *in silico* system to assign reactivity domains. In addition, Nishijo et al. (2022) estimated the threshold for high potency substances. Proposed DST values based on reactivity domain are presented in Table 5.

The safety assessor is referred to a recent publication by Nishijo et al. (2022), in which a narrative explanation (see also Table 5 footnote), accompanied by a flow chart, is provided to describe the workflow for classifying a substance of interest into one of the above listed categories. Determining the substance's reactivity can also be supported with *in silico* tools, such as ToxTree, Derek Nexus, or OECD Toolbox.

**Table 5. Summary of Example DST Values from the Published Literature**

| | Dermal Sensitization Threshold Value (µg/cm$^2$) | | |
|---|---|---|---|
| **Publication** | **Non-Reactive Substances** | **Reactive Substances[a]** | **High Potency Substances[b]** |
| Chilton et al. 2022 | 710 | 73 | 1.0 |
| Nishijo et al. 2020; 2022 | 900[c] | 64[c] | 1.5 |
| Safford et al. 2011; 2015 | 900 | 64 | Not provided |

[a]Nishijo et al. (2022) defines reactive substances as "those that bind covalently to skin proteins via Michael addition, Schiff base formation, bimolecular nucleophilic substitution (SN2), nucleophilic aromatic substitution (SNAr), or acyl transfer. Further, chemicals can also be classified as reactive in special cases, including when they do not fit in any of the above five domains but are electrophilic (e.g. SN1 electrophiles)" (p. 53).

[b] Nishijo et al. (2022) defines high potency substances as those that "include protein derivatization agents, organic peroxides, and structurally complex compounds. The third rule of "complex structure" is to cover [high potency] chemicals that are not assigned any further examination of the reactive domain, but cannot be confidently classified as non-reactive, for example, substances that have an unfamiliar substituent with uncertain reactivity or those with several substituents, the combination of which could lead to significant reactivity" (p. 53).

[c] based on values provided by Safford et al 2011 and 2015.

If existing data is available and/or exposure-based waiving is not feasible, the safety assessor should continue through the workflow process presented in Figure 3.

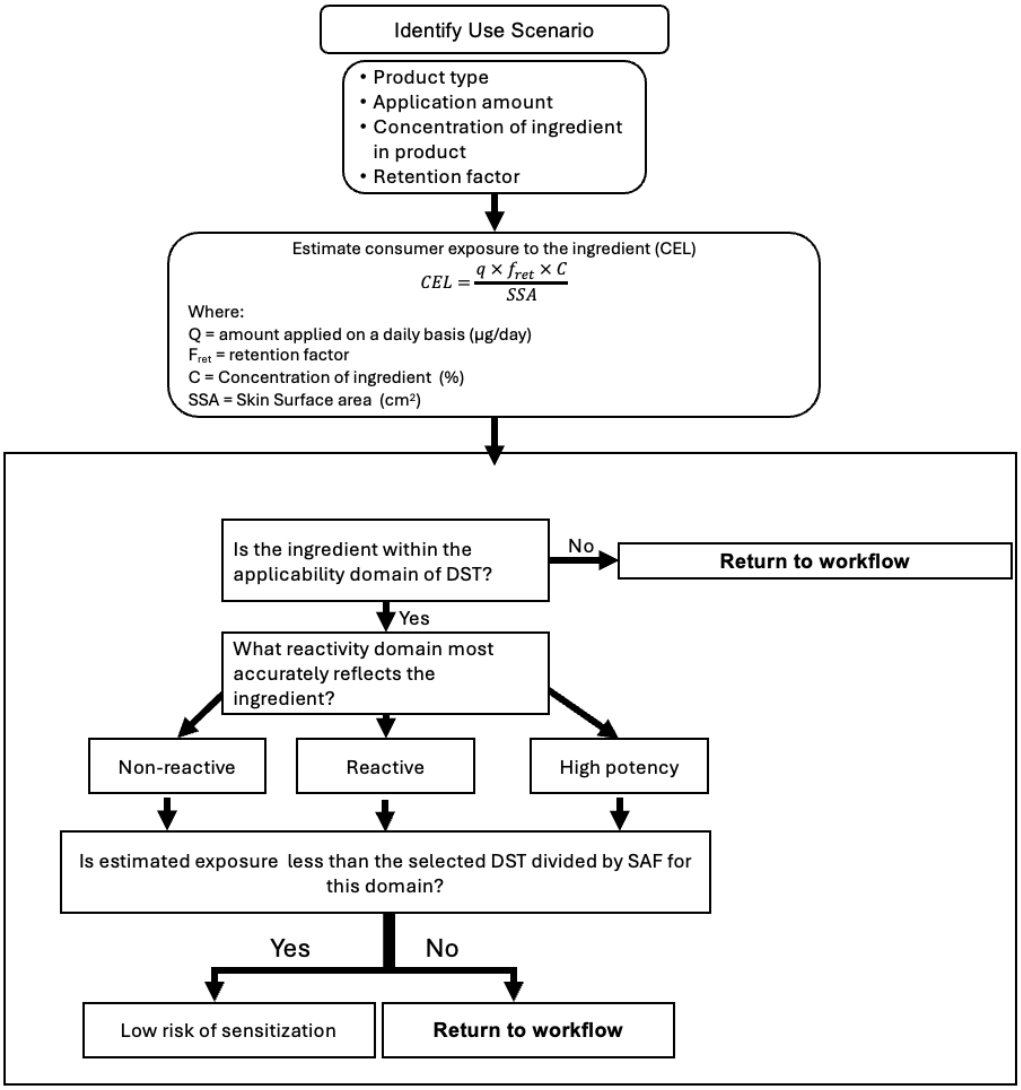



**Figure 6. Visual Depiction of Identification of Use Scenario and Exposure-Based Waiving**

Note: DST = dermal sensitization threshold; SAF = safety assessment factors

**Step 3 Objective:** Define the exposure scenarios for the substance and estimate the quantitative exposure level of the substance to the consumer. With this data, an assessor may, if appropriate, evaluate whether exposure-based waiving may be used to assess skin sensitization potential.

## 2.7 Step 4. Evaluate Existing Data

At this phase in the process, the assessor must now evaluate whether the available data provide adequate information to inform the question defined in the problem formulation. This process includes evaluating all the available historical data (inclusive of *in vivo, in chemico/in vitro* and *in silico* data) to determine what the data might conclude about the substance's sensitization potential, and whether the existing information is sufficient for hazard or safety decision-making purposes. This process also guides higher tier refinements and any new data generation decisions.

Given that each assessment is unique to a specific use and decision-making scenario, a prescriptive workflow to conduct such an assessment is not feasible. However, OECD (2019) describes guiding principles (e.g., transparency; assessment of potential bias) and key elements (e.g., evidence evaluation; evidence weighting; evidence integration) for consideration when conducting a WoE, IATA, or DA assessment. These key elements can help determine the strength of the evidence to be used to answer the hypothesis under evaluation. Some of these key elements have been integrated into the workflow process in Figure 7 in a step-wise manner, and are summarized in Box 6.

**Box 6. Considerations for Evidence Evaluation**

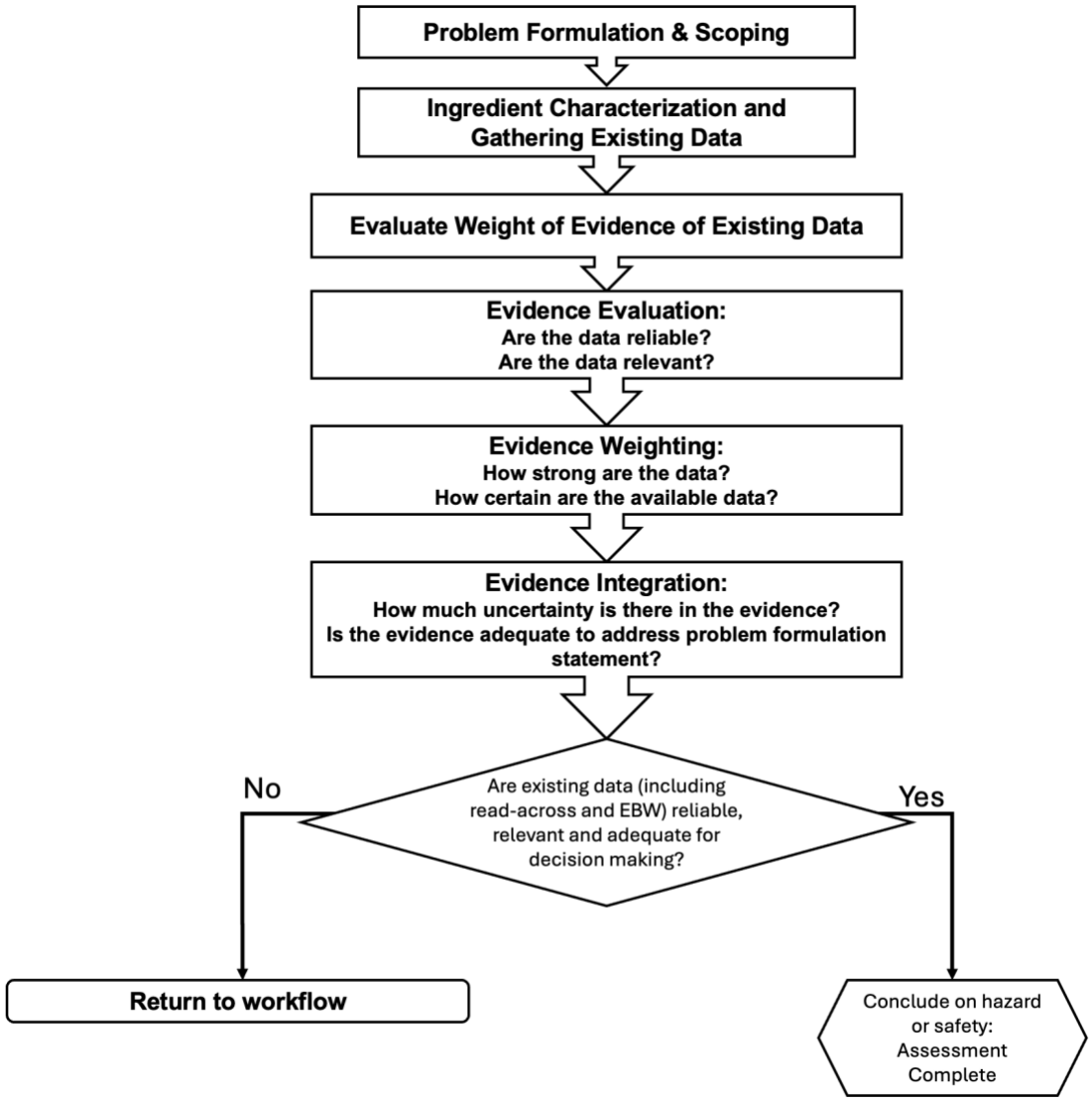
**Reliability** - evaluating the inherent quality of a test report or publication relating to preferably standardized methodology, and the way that the experimental procedure and results are described to give evidence of the clarity and plausibility of the findings.

**Relevance** - the extent to which data and/or tests are appropriate for a particular hazard identification or safety assessment.

**Adequacy** - the usefulness of data for hazard and/or safety assessment purposes.

**Uncertainty** - combination of all limitations in available knowledge that could affect confirming a hypothesis.

A range of methods may be used for each key element. For example, data reliability and uncertainty assessments that are encompassed in the evidence evaluation step may be assessed using Klimisch scoring or the OHAT NTP Risk of Bias Tool (Klimisch et al., 1997, NTP, 2015). The output of these assessments would then inform the evidence weighting (e.g., if an assessor has high or low confidence in a study). The specific methods that are used and the rationale for selecting each key element should be documented and maintained for transparency.




Note: EBW = exposure-based waiving

**Figure 7. Evidence Integration Workflow, based on OECD (2019)**

A data map or matrix can be used to organize data and inform WoE evaluations. The data map can be used to identify data gaps and what the data indicates regarding sensitization, when sufficient data are available. Additionally, the data map can be used to determine if the available data are sufficient to address the hypothesis developed in Section 2.4.

Considering a hazard-focused assessment, the summarized data should indicate whether the substance is a sensitizer or non-sensitizer, or if enough data are available to conduct hazard classification. This assessment can be based either on animal studies or a combination of *in chemico*, *in vitro*, and/or *in silico* data (see Table 7 for information on

defined approaches that provide a hazard classification). If adequate data exist, the workflow can be exited, and a hazard determination may be made based on the currently available data.

If a safety assessment is needed, the data should be evaluated to determine whether a NESIL (or PoD) can be determined from currently available historical animal data.[10] If adequate data are identified, the workflow can be exited, and the relevant data can be used to develop a NESIL.

If this phase determines that data are insufficient for addressing the question posed in problem formulation, the assessor should continue through the workflow. The workflow's next phase will help define the evidence generation approach needed to address the identified gaps so a final safety or hazard assessment conclusion can be made.

**Table 6. Example Data Map for Organization of Existing Data for Substance***

| Evidence Stream | KE | Data availability | Data reliability | Result | Comment |
|---|---|---|---|---|---|
| Physico-chemical properties | NA | | | | |
| Animal | | | | | |
| Buehler | AO | | | | |
| GPMT | AO | | | | |
| LLNA | AO | | | | |
| Other | NA | | | | |
| Human/Clinical | | | | | |
| HRIPT | AO | | | | |
| HMT | AO | | | | |
| Diagnostic Patch Testing | AO | | | | |
| Other | NA | | | | |
| *In Chemico* | | | | | |
| DPRA | MIE/KE1 | | | | |
| ADRA | MIE/KE1 | | | | |
| kDPRA | MIE/KE1 | | | | |
| Other | NA | | | | |
| *In Vitro* | | | | | |
| KeratinoSens™ | KE2 | | | | |
| LuSens | KE2 | | | | |

[10] The 7th amendment to EU Directive 76/768/EEC provides a regulatory framework for phasing out animal testing for cosmetic ingredients and finished products that are marketed or sold in the EU (European Economic Community, 1992). Therefore, any animal studies used for a risk assessment evaluation should have been conducted prior to 2013.

| Evidence Stream | KE | Data availability | Data reliability | Result | Comment |
|---|---|---|---|---|---|
| EpiSensA | KE2 | | | | |
| h-CLAT | KE3 | | | | |
| U-SENS™ | KE3 | | | | |
| IL-8 Luc | KE3 | | | | |
| GARD™skin | KE3 | | | | |
| Metabolism | NA | | | | |
| Other Assay | NA | | | | |
| *In Silico* | | | | | |
| Tool name & version | NA | | | | |
| Metabolism | NA | | | | |
| Other | NA | | | | |
| Analogue Data | | | | | |
| Read-across data** | NA | | | | |
| **Sufficient data for concluding hazard/safety?** | | | | | |

Abbreviations: AO: adverse outcome; KE: key event; MIE: molecular initiating event; NA: not applicable

*The evidence streams and methods or models noted in the example table are not exhaustive. The models and methods used for an evaluation should be updated on a case-by-case basis.

**For either a hazard or safety assessment, read-across data may be used to increase confidence in the conclusions developed using existing data. As discussed in Section 2.5.2.3, empirical data for structurally or biologically similar compounds to the substance may be used as supporting evidence for hazard characterization. These read-across data can be incorporated into the evidence evaluation step shown in Figure 7.

**Step 4 Objective:** The objective of this step is to evaluate existing data regarding reliability, relevancy, and adequacy. A data table to organize existing information for the substance of interest is recommended.

## 2.8 Step 5. Determine the Approach for Additional Testing and Conduct Targeted Testing

Under circumstances where it is determined that there are insufficient data available for decision making, the next step is to understand what data gaps need to be filled, and what assays or tools are available to fill them.

### *2.8.1 Determine the Approach for Additional Testing*

This evaluation should be done in the context of the problem formulation and scope that were defined at the start of the assessment, along with considering the identified available data.

One way to approach this step is to expand upon the data matrix developed when evaluating the WoE (Table 6) and align the available data to the skin sensitization AOP. Figure 8 demonstrates where available non-animal assay data align with individual KEs and the MIE of the skin sensitization AOP. Using this figure, the safety assessor can visualize which AOP KEs are not covered by the existing evidence base for the substance.

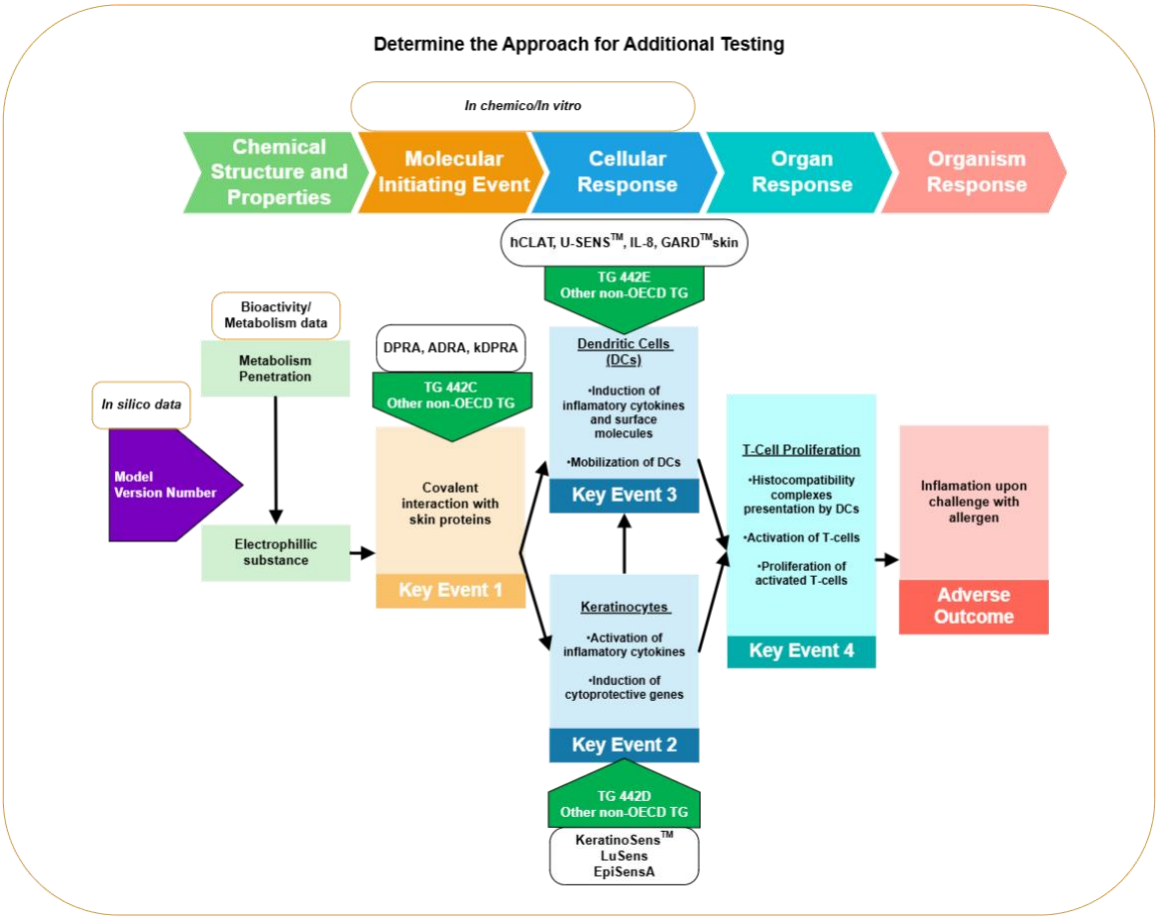



**Figure 8. *In Chemico, In Vitro*, and *In Silico* Methods and Models for Skin Sensitization Mapped to the Adverse Outcome Pathway (adapted from Kleinstreuer 2021; Strickland et al., 2019) to Assist in Determining the Approach for Additional Testing**

In addition to determining which data gaps need to be filled, the scope of the evaluation should be considered (e.g., hazard assessment; safety assessment). Questions for consideration to help identify targeted testing needs include:

- Is qualitative hazard, hazard category, or quantitative point of departure information needed?

- Do relevant metabolic pathways in the skin need to be probed for more in-depth understanding?
- Which *in chemico* and *in vitro* methods, or *in silico* tools, will inform the identified data gaps?

To minimize the need for expert judgement in skin sensitization assessments, OECD GL 497 (2025) describes three DAs that can be used to assess skin sensitization potential. Two of the DAs, the 2 out of 3 (2o3) and Integrated Testing Strategy (ITS), use validated OECD test methods for hazard identification and UN GHS potency characterization, respectively. These DAs utilize a combination of *in chemico* and *in vitro* data to predict skin sensitization (and an *in silico* prediction for the ITS DA).

The third DA uses the Skin Allergy Risk Assessment – Integrated Chemical Environment (SARA-ICE) DA. SARA-ICE uses a Bayesian statistical model for deriving a quantitative PoD using *in chemico*, *in vitro*, and/or existing *in vivo* data. The model estimates the dose associated with a 1% chance of skin sensitization in a patch test. Inputs to this model may include any combination of data from the following sources: DPRA, kDPRA, h-CLAT, KeratinoSens™, U-SENS™ and historical LLNA or human predictive patch test (HPPT) data.

These defined approaches dictate the information requirements needed to assess skin sensitization potential, and are used as described in OECD GL 497 (OECD, 2025). The three DAs are described with respect to the problem formulation being addressed, and their inputs and outputs in Table 7. Irrespective of the DA used, detailed information regarding the data inputs, results, and conclusions should be documented for transparency.

**Table 7. Defined Approaches for Skin Sensitization using *In Chemico, In Vitro*, and *In Silico* Tools (and existing *in vivo* data)**

| Defined Approach | Problem Formulation Addressed | Input | | | |
|---|---|---|---|---|---|
| | | **KE1: ADRA or DPRA** | **KE2: KeratinoSens™, EpiSensA, or LuSens** | **KE3: GARD™skin, h-CLAT, IL-8 Luc, or U-SENS™** | ***In Silico:* Derek Nexus or OECD QSAR Toolbox** |
| 2o3 | Hazard (GHS 1 vs GHS NC) | X | X | X | - |
| ITS | Hazard (GHS 1 vs GHS NC) Potency Category (GHS 1A vs. GHS 1B) | X | - | X | X |
| **Defined Approach** | **Problem Formulation Addressed** | **Input** | | | |
| | | **KE1: DPRA and/or kDPRA** | **KE2: KeratinoSens™** | **KE3: h-CLAT and/or U-SENS™** | **KE4/AO: LLNA/HPPT°** |
| SARA-ICE | Point of Departure ($ED_{01}$*) | X | X | X | X |

°Existing *in vivo* data only.

*$ED_{01}$ is the estimate of a dermal dose at which there is a 1% chance of inducing sensitization using a human predictive patch test. Further information may be reviewed in OECD (2025).

### *2.8.2 Conduct Targeted Testing*

Once the data needs have been identified and the approach used to evaluate the skin sensitization hazard or safety profile of the substance has been defined, the targeted testing should be implemented. Targeted testing will be focused on filling data gaps and reducing uncertainties present in the existing data matrix, with a focus on NAMs. For example, metabolism targeted testing may be conducted to clarify whether metabolic activation impacts activity. Targeted testing also may be conducted to complete the inputs needed to conduct one of the DAs described in Section 2.8.

This section details the types of NAMs that are available and accepted by OECD in a manner that follows the AOP for skin sensitization, from the MIE through to KE3. Non-guideline test methods are also available but are not yet detailed in this BPG. Testing at the organ or organism response are *in vivo* models that are outside the scope of this document. Additional details, including further descriptions of the test methods and their applicability domain, are presented in Appendix A. Before implementing any assay, the safety assessor should discuss the substance of interest, any of its unique attributes, or any concerns with the study director who will be conducting the evaluation to confirm the analysis will result in useful data (OECD 2018). In addition, the assessor should confirm study method documentation, outline any variances from protocol, as appropriate and communicate results in a manner that can be repeated, understood and interpreted further to sound scientific principles.

### *2.8.3 Metabolism data*

A structural review may not clearly indicate whether a substance requires metabolism to become a hapten. A pre- or pro-hapten may not be sufficiently metabolized in to its electrophilic metabolite and thus the potential to produce a sensitization response will be decreased. In this case, *in silico* models and *in vitro* methods can be used to assess the potential activation or deactivation of a substance within the skin environment. There are *in silico* tools available, such as OECD QSAR Toolbox, has a skin metabolism simulator that may be used to assess skin metabolism potential (OECD QSAR Toolbox, 2025). The simulator requires input of the substance structure, which is then evaluated for potential metabolite formation in the skin. Several other *in silico* tools have similar models, including Meteor Nexus (https://www.lhasalimited.org/solutions/metabolite-identification-and-analysis/) and ADMET Predictor (https://www.simulations-plus.com/software/admetpredictor/).

*In vitro* assays (e.g., skin homogenates, reconstructed human epithelial models and skin explants) can be used to experimentally assess potential skin metabolism. If non-human

skin sources are utilized for studies, differences in metabolic pathways should be considered when assessing results (Genies et al., 2019; Kazem et al., 2019; SCCS, 2010).

### *2.8.4* In chemico *and* in vitro *data*

As shown in Figure 8, *in chemico* and *in vitro* data can be used to fill specific mechanistic data gaps (e.g., to complete inputs required for a specific DA) that can be used in support of available data to assess the skin sensitization potential of a substance.

The *in chemico* methods DPRA, ADRA, and kDPRA provide information for the MIE of the skin sensitization AOP. These three methods quantify the reactivity of a test substance towards a synthetic model peptide containing cysteine and/or lysine (OECD, 2024a). DPRA and ADRA measure the depletion of the peptide and correlate the depletion levels to identify the substance as a skin sensitizer or non-sensitizer. The kDPRA measures reactivity towards the model peptide in a time- and concentration-dependent manner, which can be used to identify a GHS category 1A skin sensitizer. Detailed information on how to conduct each of these methods, including methodology, prediction model application, acceptance criteria, and applicability domain, can be obtained from OECD TG 442C (OECD, 2024a).

The *in vitro* test methods evaluating KE2 of the skin sensitization AOP, KeratinoSens™ and LuSens, quantify activation of NRf2-mediated antioxidant response element-dependent genes in immortalized cell lines derived from human keratinocytes (OECD, 2024b). EpiSensA quantifies changes in the expression of marker genes associated with keratinocyte activation in a reconstructed human epidermis model (OECD, 2024b). Detailed information on how to conduct each of these methods, including methodology, prediction model application, acceptance criteria, and applicability domain can be obtained from OECD TG 442D (OECD, 2024b).

The *in vitro* test methods h-CLAT, U-SENS™, IL-8 Luc assay, and GARD™skin provide information for KE3 of the skin sensitization AOP. h-CLAT, U-SENS™, and IL-8 Luc assays quantify changes in cell surface markers associated with dendritic cells or monocyte activation (OECD, 2024c). GARD™skin evaluates transcriptional patterns of genomic biomarker signatures (OECD, 2024c). These methods are typically used to assess skin sensitization hazard. Detailed information on how to conduct each of these methods, including methodology, prediction model application, acceptance criteria, and applicability domain, can be obtained from OECD TG 442E (OECD, 2024c).

*In chemico* and *in vitro* methods noted in this section have limited metabolic capacity. Please refer to the individual OECD TG for additional information on limitations and

consider *in vitro* or *in silico* metabolism predictions if metabolic activation is likely to play a role in substance activity.

**Step 5 Objective:** Identify data gaps in the existing data for the substance of interest. In combination with the assessment scope and hypothesis, the test methods to fill the data gaps to assess skin sensitization potential can be identified and implemented in order to inform the skin sensitization assessment.

## 2.9 Step 6. Evaluate Data

The new data generated using NAMs can be used (a) in combination with historical animal data, *in silico* predictions, and read-across assessment to conduct a WoE evaluation or utilize an IATA; or (b) in a specific DIP, as defined in a DA.

If a WoE assessment is utilized, the process described in Section 2.7 should be followed.

If, using a WoE approach, IATA, or outcome of a DA, the substance under assessment is determined to have no skin sensitizing potential (and the determination is deemed appropriate based on the stated problem formulation), the assessor may conclude that no further evaluation is needed. The assessor must document all the inputs that lead to this conclusion, along with any assumptions and uncertainties (see Section 2.10) associated with the conclusion. If it is not appropriate to draw a conclusion, based on the problem formulation or the given data, the assessor should continue through the workflow to address the question.

The subsequent sections focus on applying and interpreting available data using DAs (as opposed to WoE, which has been previously discussed). Specifically, in these sections, the assessor will be guided through the DIP for each DA. A DIP is a rule-based approach or mathematical model used with a defined set of information sources to derive a predicted response. A DA does not require expert judgement to make a hazard conclusion, but can also be used in a larger WoE assessment or IATA, which may require expert judgement. In this section, the focus is on DAs published in OECD GL 497 as of June 2025, the 2o3, the ITS and the SARA-ICE DA. The web-based DASS-App also provides support for using the 2o3 and ITS DAs and their DIPs. The DASS-App allows for skin sensitization predictions according to the defined approaches published in OECD GL 497 and noted in Table 7. More information about the interface and its use can be found at: https://ntp.niehs.nih.gov/go/952311 (NICEATM, 2024a). SARA-ICE and its User Guide can be accessed here https://ntp.niehs.nih.gov/sites/default/files/sara-ice/?NTPW-4081-DEPLOY.

### *2.9.1 Assessing skin sensitization hazard (sensitizer/non-sensitizer)*

To evaluate whether a substance is a sensitizer or a non-sensitizer, either the 2o3 or ITS DA, published in OECD GL 497, can be used. Additional DAs are available in the literature, and can also provide information on skin sensitization. Selected approaches will be included in a future iteration of this BPG such as Gilmour et al. (2023), Jaworska et al. (2015); Kleinstreuer et al. (2018), Natsch (2023); Natsch and Gerberick (2022a, 2022b), Reinke et al. (2025), Tourneix et al. (2020, 2024).

#### *2.9.1.1 The 2 of 3 Defined Approach (2o3 DA)*

The 2o3 DA is summarized in OECD GL 497 using a decision tree, which has been replicated in Figure 9. To apply this DA, an assessor starts by evaluating the results for assays that address two of the three key events (see Table 7 or Figure 9). If the results from these two assays are concordant in terms of predicting if the substance is a sensitizer or not, the assessor may make a conclusion regarding the substance's sensitization potential. If the data from these two assays are not concordant, then data from an assay representing the third KE (which has not yet been evaluated) is reviewed. The result (sensitizer: yes or no) is concluded based on the two concordant assays of the three conducted. For each assay, borderline ranges have been statistically derived (see OECD GL 497). If the results fall into these borderline ranges, there is lower confidence in the result, and in some cases, a prediction of sensitizer/non-sensitizer may not be possible. As such, and as depicted in Figure 10, results in the borderline range are to be interpreted considering the decision-making context, and as addressed further in Annex I of OECD GL 497.

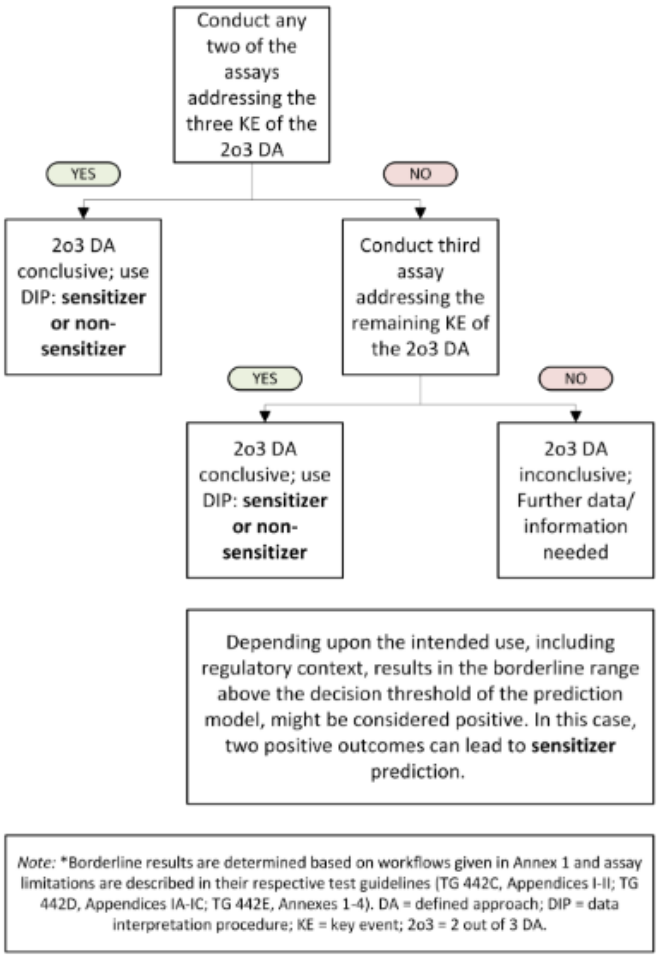



**Figure 9. Data Interpretation Procedure for 2o3 Defined Approach for Hazard Classification (adapted from OECD 2025)**

The assays, and the output resulting in a conclusion of 'sensitizer' for the 2o3 DA are summarized in Table 8 below.

**Table 8. Overview of Results Indicating Sensitizing Potential by Assay**

| Assay | Description of Result Indicating Sensitization Potential (as described in OECD 2025) |
|---|---|
| DPRA (OECD TG 442C) | Test substance induces a mean peptide depletion of cysteine-and lysine-containing peptide above 6.38% (or in the case of co-elution, cysteine-only depletion above 13.89%). |
| ADRA (OECD TG 442C) | Test substance induces mean percent depletion of N-(2-(1-naphthyl)acetyl)-L-cysteine (NAC) and α-N-(2-(1-naphthyl)acetyl)-L-lysine (NAL) >4.9% (or in the case of co-elution, NAC depletion above 5.6%). |
| KeratinoSens™ (OECD TG 442D) | Test substance causes >1.5-fold luciferase induction, at viabilities >70% when compared to vehicle control. |

| Assay | Description of Result Indicating Sensitization Potential (as described in OECD 2025) |
|---|---|
| LuSens (OECD TG 442D) | Test substance induces luciferase induction ≥ 1.5-fold, and is statistically significant compared to the solvent control in at least two consecutive non-cytotoxic tested concentrations (i.e., cellular viability ≥ 70%), where at least three tested concentrations should be non-cytotoxic. |
| EpiSensA (OECD TG 442D) | At least one of the following results occurs for at least one tested concentration::<br>*Imax for ATF3 is > 15<br>*Imax for GCLM is > 2.<br>*Imax for DNAJB4 is > 2.<br>* Imax for IL-8 is > 4. |
| h-CLAT (OECD TG 442E) | CD86 induction exceeds 1.5-fold and/or CD54 exceeds 2-fold at viabilities >50% when compared to the vehicle control. |
| U-SENS™ (OECD TG 442E) | CD86 stimulation index exceeds 1.5-fold observed when compared to the vehicle control for 2 independent runs. |
| IL-8 Luc (OECD TG 442E) | Test substance has an Ind-IL8LA > 1.4 and the lower limit of the 95% confidence interval of Ind-IL8LA < 1.0. |
| GARD™skin (OECD TG 442E) | Test substance with a calculated mean Decision Value ≥ 0. |

#### *2.9.1.2 The Integrated Testing Strategy Defined Approach (ITS DA)*

Another approach for answering the question of whether a substance is or is not a sensitizer is using the ITS DA. The DIP for the ITS uses scores derived from quantitative results from a KE1 assay, a KE3 assay, and a score from an *in silico* tool, either Derek Nexus or OECD QSAR Toolbox. The score from each information source is then summed. If the sum is above or equal to two, the substance is considered a sensitizer (Table 9). Please refer to OECD GL 497 for information on the protocol and version of the *in silico* tool permitted for use in the ITS DA.

### *2.9.2 Assessing skin sensitization potency (GHS 1A/GHS 1B/Not Classified)*

#### *2.9.2.1 The Integrated Testing Strategy Defined Approach (ITS DA)*

In GHS revision 10 (2023) the UN notes that "skin sensitizers shall be classified in Category 1 where subcategorization is not required… or where data are not sufficient for sub categorization" (GHS, 2023, p. 171). The skin sensitization classification system categorizes a substance as a Category 1 skin sensitizer if there is evidence that the substance (1) can induce sensitization in a substantial number of people; (2) achieves positive results from an appropriate animal test; or (3) is validated by the DIP from a DA according to international procedures (i.e., OECD GL 497).

Using NAMs, the assessor can utilize the output from the 2o3 DA to classify the substance as GHS 1 (sensitizer) or Not Classified (non-sensitizer), or use the ITS DA scores to classify a substance as subcategory 1A, 1B, or Not Classified. For additional details on using animal and human data in the classification process, the assessor is referred to GHS (2023).

To evaluate a substance's GHS sub-classification, the ITS DA, published in OECD GL 497, can be used. The DIP for the ITS uses scores derived from quantitative results from a KE1 assay, a KE3 assay, and a score from an *in silico* tool, either Derek Nexus or OECD QSAR Toolbox. The score from each information source is then summed. If the sum is above or equal to six, the substance is considered a GHS 1A sensitizer. If the sum is between two and five then the substance is considered a GHS 1B sensitizer. If the sum is equal to zero or one then the substance is considered a non-sensitizer (Table 9). Please refer to OECD GL 497 for information on the protocol and version of the *in silico* tool permitted for use in the ITS DA.

**Table 9. Scores associated with assay results for ITS, as presented in OECD GL 497 (2025)**

| **Information Source Type** | **KE1 (OECD TG 442C)** | | | | **KE3 (OECD TG 442E)** | | | ***In silico*** |
|---|---|---|---|---|---|---|---|---|
| **Score** | ADRA Mean NAC & NAL % depletion | ADRA NAC % depletion* | DPRA Mean Cys & Lys % depletion | DPRA Cys % depletion† | GARD™skin (µM) | h-CLAT MIT (µg/mL) | U-SENS™ EC150 (µg/mL) | Derek Nexus, OECD QSAR Toolbox |
| **3** | ≥46.4 | ≥67.4 | ≥42.47 | ≥98.24 | ≤13.03 | ≤10 | ≤3 | |
| **2** | ≥15.5, <46.4 | ≥17.5, <67.4 | ≥22.62, <42.47 | ≥23.09, <98.24 | >13.03, ≤56.44 | >10, ≤150 | >3, ≤35 | |
| **1** | ≥4.9, <15.5 | ≥5.6, <17.5 | ≥6.38, <22.62 | ≥13.89, <23.09 | >56.44 | >150, ≤5000 | >35, <200 | Positive |
| **0** | <4.9 | <5.6 | <6.38 | <13.89 | - | not calculated | not calculated /≥200 | Negative |
| **Hazard** | **Potency** | **Total Score from ITS DA** | | | | | | |
| Sensitizer/GHS 1 | GHS 1A | 6-7 | | | | | | |
| Sensitizer/GHS 1 | GHS 1B | 2-5 | | | | | | |
| Non-sensitizer | NC | 0-1 | | | | | | |

Abbreviations: ADRA: Amino acid Derivative Reactivity Assay; NAC: N-(2-(1-naphthyl) acetyl)-L-cysteine, NAL: α-N-(2-(1-naphthyl) acetyl)-L-lysine; DPRA: Direct Peptide Reactivity Assay; Cys: cysteine, Lys: Lysine; GARD: Genomic Allergen Rapid Detection; GHS: Globally Harmonized System; h-CLAT: Human Cell Line Activation Test; MIT: minimum induction threshold; NC: Not Classified; U-SENS: U-937 cell line activation test; EC: effective concentration.
For KE1, KE3, and *in silico*, only one assay/model selected from each is used for determining a score. *NAC-only depletion thresholds are used in the case of co-elution with the NAL peptide. †Cysteine-only depletion thresholds are only used in the case of co-elution with the lysine peptide.

### *2.9.3 Conducting a safety assessment using a point of departure (PoD)*

To conduct a skin sensitization safety assessment, a comparison between the estimated exposure amount (in $\mu g/cm^2$) and a defined point of departure (PoD) is needed.

#### *2.9.3.1 Derivation of a PoD*

A PoD is the point on a dose–response curve, established from experimental data, from which low-dose extrapolation can be done to determine the level of exposure below which there would be no appreciable risk. For a skin sensitization safety assessment, a NESIL (usually given in $\mu g/cm^2$) is considered a human relevant PoD. Traditionally, NESILs have been derived using all available data (e.g., *in vivo*; read-across, human data) in a WoE approach (Gauthier et al., 2023). If sufficient information is available, it can also be established directly from human data. However, PoDs can now also be derived from NAM data ($PoD_{NAM}$).

When a $PoD_{NAM}$ reflects a "no effect level", it could be considered equivalent to a NESIL when appropriately adjusted to be human relevant. However, in some cases, a $PoD_{NAM}$ can be derived to reflect a low-effect level, as is the case for the SARA-ICE DA. SARA-ICE provides the dermal dose at which there is a 1% chance of inducing sensitization in a human predictive patch test [$ED_{01}$]) from which the geometric mean, conditional on it being a sensitizer, can be used as a $POD_{NAM}$ in lieu of a NESIL.

#### *2.9.3.2 Derivation of a quantitative $PoD_{NAM}$, based on a semi-quantitative (classification) DA*

As some DAs do not provide a quantitative $PoD_{NAM}$, the assessor can adapt a semi-quantitative output, such as a GHS classification grouping, and convert it into a quantitative value for safety assessment. If this approach is undertaken, it must be done in a conservative way to ensure the resulting $PoD_{NAM}$ is protective.

The GHS subcategorizations obtained from the ITS DA can be translated from a qualitative categorical assignment of GHS 1A or GHS 1B to a quantitative NESIL. As formally agreed upon NESIL values for these subcategories are not available, a safety assessor may look to previously published case studies for proposed NESIL values based on these subcategorizations.

For example, in Gilmour et al. (2023), when a substance was classified as GHS 1B based on the ITS DA, a PoD of > 500 $\mu g/cm^2$ was selected for use in a quantitative safety assessment, with a note that the exact PoD value is undetermined. The justification for selecting this NESIL is that, according to GHS 2023, 1B substances are classified as such

when the results from an LLNA model indicate the EC3 value is greater than 2%[11] or positive results in HRIPTs occurs at concentrations > 500 µg/cm$^2$. As such, the NESIL can be set to the lower conservative bound value of 500 µg/cm$^2$ and remain protective of human health.

As for GHS 1A substances, GHS notes these are classified as such if an HRIPT demonstrates a positive result at concentration ≤ 500 µg/cm$^2$ or an LLNA EC3 value ≤ 2%. Given the values provided in GHS for 1A are upper bounds (as opposed to lower bounds, as for GHS 1B). The NESIL for 1A substances is less clear. One option to select an appropriate NESIL in these instances involves a WoE approach. This approach results in the safety assessor assigning a potency category based on the available data, which can then be converted to a NESIL. This concept is presented in Na et al. (2022), who provide estimated dose ranges based on potency categories from two different sources, one from Api et al. (2017) and the other from ECETOC (2003). Once an assessor assigns the potency category, the NESIL can be selected as the lower end of the given dose range. As indicated in Table 10, the lower bound of the proposed moderate potency dose range from Api et al. (2017) is concordant with that used when considering a GHS 1B classification in Gilmour et al. (2023).

**Table 10. Proposed Dose Ranges Associated with Skin Sensitization Potency, from Na et al. (2022)**

| Potency Category | Dose Range (µg/cm$^2$) | |
|---|---|---|
| | As proposed in Api et al. (2017) | As proposed in ECETOC 2003 |
| Extreme | <25 | <25 |
| Strong | 25-500 | 25-<250 |
| Moderate | 500-2500 | 250-<2500 |
| Weak | >2500-10,000 | 2500-25,000 |
| Very weak | >10,000 | |
| Non sensitizer | -- | -- |

Other DAs that can be used to set PoD$_{NAM}$ have been described in Gilmour et al. (2023) (e.g., artificial neural network models (Hirota et al., 2018); Bayesian network ITS DA

[11] Using animal data, a NESIL was derived based on the EC3 values derived from local lymph node assays. If the EC3 value is a percentage, it must be converted to µg/cm$^2$ to conduct the quantitative risk assessment. To achieve the proper units, the percentage should be multiplied by a factor of 250. This value from Robinson [12] et al. (2000) is based on the standard LLNA protocol, where 25 µL test solutions are distributed over a surface of 1 cm$^2$ per mouse ear.

(Jaworska et al., 2015); sequential testing strategy DA (Takenouchi et al., 2015), in Tourneix et al. (2024) and Natsch and Gerberick (2022a, 2022b). All of these DAs provide a potency category prediction that can be converted to a protective NESIL for safety assessment, similar to what is described above.

#### *2.9.3.3 Derivation of quantitative $PoD_{NAM}$, based on a quantitative approach (i.e., a no or low effect level ($ED_{01}$))*

Another way to set a $PoD_{NAM}$ is using a DA, which provides a quantitative output. This determination is usually done with mathematical models, allowing users to input data from either a combination of NAM and historical animal data, or NAM-only data, to estimate a data-informed PoD. Depending on the model used, the output may be a no-effect level (which can be adjusted to a NESIL), or a low-effect level (e.g., an $ED_{01}$).

The only OECD-accepted mathematical model for deriving a no- or low-effect level is the SARA-ICE DA (as of June 2025; OECD, 2025). This model uses Bayesian statistical analysis to estimate a dermal dose of a substance at which there is a 1% chance of inducing sensitization in a human predictive patch test ($ED_{01}$, in units of µg/cm$^2$). Inputs to this model must include at least two information sources addressing different key events: DPRA, kDPRA, h-CLAT, KeratinoSens™, U-SENS™, and historical LLNA or HPPT data. A $PoD_{NAM}$ is computed as the geometric mean of the $ED_{01}$ distribution, conditional on it being less than 60,000 µg/cm$^2$. This mean $ED_{01}$ represents the best estimate of the substance's potency, assuming it is a sensitizer.

As SARA-ICE is a probabilistic method, the outputs have a minor degree of variation, judged to represent an acceptable level of reproducibility for regulatory use. Additional details on this model can be found in Reinke et al. (2025),[12] and the tool, and its User Guide, can be accessed at: https://ntp.niehs.nih.gov/whatwestudy/niceatm/test-method-evaluations/skin-sens/da/SARA-ICE (NICEATM, 2024b).

Finally, there are several other approaches that use various statistical approaches, such as regression analysis and artificial neural networks, which are also available to support PoD identification for safety assessment. Some additional literature on this topic can be found below in Box 7.

[12] In addition to the SARA-ICE DA, a SARA-ICE 'extended' version has also been created that provides additional statistics, such as the $ED_{01}$ distribution. Using this model version allows users to select alternative PoDs. Suitably low quantiles (e.g., 0.05) would represent a more conservative estimate of potency (Reinke et al., 2025).

**Box 7. Literature Examples of Methods to Develop Point of Departure Estimates for Safety Assessment using *in vitro* Data**

Natsch and Gerberick. 2022a. Integrated skin sensitization assessment based on OECD methods (I): Deriving a point of departure for risk assessment. ALTEX 39:636-646.

Natsch A, Gerberick GF. 2022b. Integrated skin sensitization assessment based on OECD methods (II): Hazard and potency by combining kinetic peptide reactivity and the "2 out of 3" defined approach. ALTEX 39(4):647-655.

Natsch et al. 2018. Deriving a no expected sensitization induction level for fragrance ingredients without animal testing: An integrated approach applied to specific case studies. Toxicol Sci 165:170-185.

Lee et al. 2025. Determining a point of departure for skin sensitization potency and quantitative risk assessment of fragrance ingredients using the GARDskin dose-response assay. ALTEX 42(2):263-277.

Hirota et al. 2015. Evaluation of combinations of in vitro sensitization test descriptors for the artificial neural network-based risk assessment model of skin sensitization. J Appl Toxicol 35:1333-1347.

Hirota et al. 2018. Development of an artificial neural network model for risk assessment of skin sensitization using human cell line activation test, direct peptide reactivity assay, KeratinoSens and in silico structure alert parameter. J Appl Toxicol 38:514-526.

Jaworska et al. 2015. Bayesian integrated testing strategy (ITS) for skin sensitization potency assessment: A decision support system for quantitative weight of evidence and adaptive testing strategy. Arch Toxicol 89(12):2355-2383.

Tourneix et al. 2020. Skin sensitisation testing in practice: Applying a stacking meta model to cosmetic ingredients. Toxicol. Vitro 66: 104831-104844.

Tourneix et al. 2024. Deriving a continuous point of departure for skin sensitization risk assessment using a bayesian network model. Toxics 12:536-552.

#### *2.9.3.4 MoE/MoS Determination*

For skin sensitization safety assessments, there are two approaches available to determine the magnitude between the defined PoD/$PoD_{NAM}$ and the consumer exposure level (CEL).

These are margin of *exposure* (MoE) and the margin of *safety* (MoS). The MoE (Equation 5) is defined as:

$$MoE = \frac{PoD\ (or\ PoD_{NAM})}{CEL} \qquad \text{(Equation 5)}$$

When using the MoE approach, the magnitude of the MoE needed to deem exposure levels acceptable is equivalent to the magnitude of uncertainty, which has traditionally been defined by safety assessment factors (SAF) (Felter et al., 2002; Basketter and Safford, 2016; SCCS 2023). SAFs account for "interindividual variability, skin condition, occlusion, vehicle or formulation effects and frequency/ duration of exposure" (Gilmour et al. 2020, p. 9).

The MoS (Equation 6) is defined as:

$$MoS = \frac{\left(\frac{PoD\ (or\ PoD_{NAM})}{SAF}\right)}{CEL} \qquad \text{(Equation 6)}$$

For the MoS, the SAF is incorporated into the calculation, such that if the MoS is greater than one, it is deemed acceptable. Therefore, the acceptable MoE is equivalent to the overall SAF. To avoid confusion, the term acceptable MoE will be used from here on in.

Once the $PoD_{NAM}$ and CEL have been determined, they are compared, and a safety decision can be made. Traditionally, when NESILs are used as the PoD for risk assessment, they are typically divided by predetermined SAFs, as outlined in an approach for quantitative risk assessment for fragrance materials (Api et al., 2020). A NESIL divided by a SAF results in an acceptable exposure level (AEL). Whether a certain chemical can be safely used at a given concentration depends on the magnitude of the ratio AEL versus CEL, which is equivalent to an MoS.

The acceptable MoE necessary to conclude safety should be predefined, similar to what is done when deriving a NESIL in a WoE approach. In available case studies, an illustrative MoE of 100 (equal to a SAF of 100) has been proposed (Gilmour et al., 2023; 2020; Assaf et al. 2021). This MoE has been deemed sufficiently protective in those examples.

The relative novelty of $PoD_{NAM}$s poses a question as to the acceptability/appropriateness of currently used acceptable MoEs. One approach to address this appropriateness question

uses benchmark consumer exposure scenarios. Reynolds et al. (2022) offer an example benchmark consumer exposure scenario dataset for skin sensitization risk. This dataset consists of example (current and historic) ingredient uses in consumer products, and each scenario is annotated as being either high or low risk for sensitization induction based on 'in market' experience. Datasets of this type can be used to quantify the protectiveness afforded by different choices of the acceptable MoE.

In Reynolds et al. (2025), a statistical framework was proposed to convert acceptable MoEs deemed appropriate for traditional NESILs to acceptable MoEs appropriate for $PoD_{NAM}$s. This conversion is based on the idea that an acceptable MoE for a $PoD_{NAM}$ should be at least as protective as the acceptable MoE used with a traditional NESIL.

The components of this framework are a) a Bayesian statistical model of the relationship between traditional NESILs and benchmark exposures; and b) a Bayesian statistical model of the relationship between NESILs and $PoD_{NAM}$. $PoD_{NAM}$-specific parameters characterizing the $PoD_{NAM}$ bias and error are used as a predictor of a traditional NESIL. Then, using the Bayesian statistical modeling framework, a formula that has input variables for the MoE, the estimated bias in the NAM estimate, the bias between the maximum low-risk consumer exposure level and the NESIL, and the estimated variances of the observed NAMs and NESILs can be used to output a distribution that corresponds to a traditional MoE.

The formula was applied and compared the acceptable MoE from traditional approaches to that of $PoD_{NAM}$ derived from SARA-ICE.[13] The findings are presented in Table 11 below, and demonstrate that when MoEs using traditional approaches are equal to or less than 100, the MoE is maintained for $PoD_{NAM}$s. However, if the traditional acceptable MoE (SAF) is greater than approximately 100, a larger acceptable MoE should be used for these $PoD_{NAM}$s (Table 11).

[13] Additional comparisons were also done with the GARD™skin dose-response assay and Natsch and Gerberick (2022a) proposed approach, but as these are not yet within the OECD TG, they are not presented in this BPG (Reynolds et al., 2025).

**Table 11. Side by Side Comparison of the Acceptable MoE from a Traditional NESIL-Based Risk Assessment to that for $PoD_{NAM}$ from SARA-ICE**

| Acceptable MoE for traditional, NESIL-based risk assessment | Median of the distribution for the acceptable MoE for SARA-ICE PoDs |
|---|---|
| 3 | 3 |
| 10 | 10 |
| 30 | 30 |
| 100 | 100 |
| 300 | 360 |
| 1,000 | 1,700 |
| Source: Reynolds et al. 2025 | |

This approach is applicable for any $PoD_{NAM}$, and is the first attempt to provide quantitative justification for choosing an acceptable MoE enabling application of a $PoD_{NAM}$ within a skin sensitization safety assessment. An example safety assessment using a $PoD_{NAM}$ and NAM acceptable MoE is presented in Box 8. An example safety assessment using a $PoD_{NAM}$ and MoS is presented in Box 9.

**Box 8. Example Safety Assessment Using $PoD_{NAM}$ and NAM Acceptable MoE:**

*Scenario:* 0.02% of 'Substance X' in a deodorant

*Exposure*: CEL = 5 µg/cm$^2$

$PoD_{NAM}$ : = 4800 µg/cm$^2$ (derived using SARA-ICE DA)

$MoE = \frac{4800}{5} = 960$

*Acceptable MoE* = 360 (translated from a traditional Acceptable MoE of 300 for products applied to the underarm (Appendix B. Historically applied Safety Assessment Factors (SAFs)to a NAM Acceptable MoE using Reynolds et al., 2025 as in Table 11)

***Safety Assessment Conclusion*:** No appreciable risk; MoE of 960 > Acceptable MoE of 360

**Box 9. Example Safety Assessment Using $PoD_{NAM}$ and MoS:**

*Scenario:* 0.02% of 'Substance X' in a deodorant

*Exposure*: CEL = 5 µg/cm$^2$

$PoD_{NAM}$ : = 4800 µg/cm$^2$ (derived using WoE)

*SAF* = 300 (products applied to the underarm (Appendix B. Historically applied Safety Assessment Factors (SAFs)

$MoS = \frac{4800/300}{5} = 3.2$

***Safety Assessment Conclusion*:** No appreciable risk; MoS > 1

**Step 6 Objective:** Identify and apply the most appropriate assessment approach based on the defined problem formulation (i.e., assessing hazard, potency, or safety).

## 2.10 Uncertainty characterization of the entire skin sensitization NGRA

As with any toxicological assessment (hazard or safety), an uncertainty characterization should be included. Uncertainty analysis in safety assessment is the process of identifying and evaluating the impact of knowledge or assumption limitations on the overall conclusions of a safety assessment. This type of analysis is important for characterizing the confidence in conclusions and is critical to making risk management decisions. Consistent with safety assessment best practices, an overall uncertainty characterization in the safety assessment should be provided (Beck et al., 2016; EFSA, 2018; NRC, 2014). A variety of methods, both qualitative and quantitative, are available; however, the form, nature, and depth of the assessment of uncertainty analysis should be determined by the safety assessor on a case-by-case basis and as presented in the problem formulation.

Guidance from EFSA (2018) may be helpful for determining possible approaches and aspects of uncertainty of greatest importance to each safety assessment. EFSA guidance identifies the following as the main elements of uncertainty analysis:

- Identifying uncertainties affecting the assessment
- Prioritizing uncertainties within the assessment
- Dividing the uncertainty analysis into parts
- Ensuring the questions or quantities of interest are well-defined
- Characterizing uncertainty for parts of the uncertainty analysis
- Combining uncertainty from different parts of the uncertainty analysis

- Characterizing overall uncertainty
- Prioritizing uncertainties for future investigation
- Reporting uncertainty analysis

When considering these elements of uncertainty in the context of the safety assessment process, the default approach for uncertainty assessment in each assessment is recommended to involve the following steps:

1. **Identify and document uncertainties due to limitations in the evidence base for characterizing hazard, exposure, or any other aspect of the safety assessment process.** For example, when considering NAM-based data, the applicability domain can impact data uncertainty. The applicability domain of *in chemico* and *in vitro* NAMs refers to properties (e.g., logP; molecular weight; particulate size) of a test substance for which the method is scientifically valid. The applicability domain also provides information regarding the chemical space for which the model is useful. These substance structure and physico-chemical property boundaries define the scope of where the model makes predictions with a pre-defined reliability (Sahigara et al., 2012). As such, the applicability domain of individual models should be reviewed in the relevant test guideline noted. The noted properties of relevance should be compared to the substance evaluated. If at least one property is identified as outside the model's applicability domain, an expert (e.g., a study director) should evaluate the impact this issue has on interpreting the available results.

   Another source of uncertainty that may impact the data is metabolic activation. As discussed above, the available *in chemico* and *in vitro* models do not include metabolic capabilities. If the substance evaluated is a pre- or pro-hapten, lack of metabolic activation may lead to a negative, inaccurate response. If the substance is, or may be, potentially a pre- or pro-hapten, expert evaluation should be conducted to evaluate the impact on interpreting the data. Alternatively, *in vitro* metabolism studies or models (see Section 2.8.3) should be conducted.

2. **Identify and document uncertainties associated with expert judgements** in the assessment, including judgements required for characterizing hazard, calculating exposure, selection of and PoDs, and uncertainty/safety assessment factors. In the approach described in this BPG document the uncertainty will vary if the approach to assesses hazard or safety was based on a defined approach versus WoE. In addition, inherent to a safety assessment based on NAMs, is the uncertainty currently surrounding the appropriate uncertainty factors to apply to an assessment.

3. **Collectively assess and document the impact of uncertainties** by describing potential impacts qualitatively. If appropriate and within scope, a sensitivity analyses with deterministic methods could be implemented by substituting alternative inputs for key parameters in the safety assessment (prioritizing those with the greatest uncertainty), including PoD, key exposure parameters (e.g., dermal absorption), and uncertainty factors/safety assessment factors. Ranges of values and resulting MoEs can help characterize sensitivity to particular inputs and decisions, which may help to develop conclusions and/or describe confidence in the assessment.

4. **Determine and document the overall level of confidence in the assessment conclusions** based on the collective consideration of uncertainty. Note that this step is most often done qualitatively, but quantitative approaches are also available (Reinke et al., 2025).

**Uncertainty Characterization Objective**: Identify potential sources of uncertainty in the risk assessment and assess the impact of the collective uncertainty on conclusion confidence.

# 3 Conclusion

By using the approaches presented in this document, safety assessors can feel confident in their ability to conduct a skin sensitization assessment for a substance in a cosmetic or personal care product. The presented approaches are exposure-driven, human relevant, and health protective. When using the approaches described in this document, the safety assessor should document all decisions and assumptions used to create a transparent evaluation. The field of NAMs, especially in respect to skin sensitization, is rapidly evolving. As such, this document will be updated as deemed appropriate by the International Collaboration on Cosmetics Safety (ICCS).

# 4 References


Api AM, Parakhia R, O'Brien D, Basketter DA. 2017. Fragrances categorized according to relative human skin sensitization potency. Dermatitis 28(5):299–307. https://doi.org/10.1097/DER.0000000000000304.

Api AM, Basketter D, Bridges J, Cadby P, Ellis G., Gilmour N, Greim H, Griem P, et al. 2020. Updating exposure assessment for skin sensitization quantitative risk assessment for fragrance materials. Regul Toxicol Pharmacol 118:104805. https://doi.org/10.1016/j.yrtph.2020.104805.

Aptula, A.O., Patlewicz, G., Roberts, D.W. 2005. Skin sensitization: reaction mechanistic applicability domains for structure-activity relationships. Chem Res Toxicol. 18(9):1420-6. https://doi.org/10.1021/tx050075m.

Assaf Vandecasteele H, Gautier F, Tourneix F, van Vliet E, Bury D, Alépée N. 2021. Next generation risk assessment for skin sensitisation: A case study with propyl paraben. Regul Toxicol Pharmacol 123:104936. https:/doi.org/10.1016/j.yrtph.2021.104936.

Basketter D, Safford B. 2016. Skin sensitization quantitative risk assessment: A review of underlying assumptions. Regul Toxicol Pharmacol. 74:105-16. https://doi.org/10.1016/j.yrtph.2015.11.013.

Beck NB, Becker RA, Erraguntla N., Farland, WH, Grant RL, Gray G, Kirman C, LaKind JS, et al. 2016. Approaches for describing and communicating overall uncertainty in toxicity characterizations: U.S. Environmental Protection Agency's Integrated Risk Information System (IRIS) as a case study. Environ Int 89-90:110-128. https://doi.org/10.1016/j.envint.2015.12.031.

Berggren E, White A, Ouedraogo G, Paini A, Richarz A, Bois FY, Exner T, Leite S, et al. 2017. Ab initio chemical safety assessment: A workflow based on exposure considerations and non-animal methods. Comput Toxicol 4: 31-44. https://doi.org/10.1016/j.comtox.2017.10.001.

Casati, S., Aschberger, K., Asturiol, D., Basketter, D., Dimitrov, S., Dumont, C., Karlberg, A-T., Lepoittevin, J-P., Patlewicz, G., Roberts, D.W., Worth, A. 2016. Ability of non-animal methods for skin sensitisation to detect pre- and pro-haptens: Report and recommendations of an EURL ECVAM expert meeting; EUR 27752 EN; https:/doi.org/10.2788/01803.

Chilton ML, Api AM, Foster RS, Gerberick GF, Lavelle M, Macmillan DS, Na M, O'Brien D, et al. 2022. Updating the Dermal Sensitisation Thresholds using an expanded dataset and an in silico expert system. Regul Toxicol Pharmacol 133: 105200. https://doi.org/10.1016/j.yrtph.2022.105200.

Cochrane J. 2023. What health evidence can we trust when we need it most? https://www.cochrane.org/about-us/news/what-health-evidence-can-we-trust-when-we-need-it-most.

Delmaar CJE, Schreurs R, Bakker MI, Minnema J, Bokkers BGH. 2022. PACEMweb: A tool for aggregate consumer exposure assessment. J Exp Sci Environ Epidemiol 33(6):971-979. https://doi.org/10.1038/s41370-022-00509-7.

Dent M, Amaral RT, Da Silva PA, Ansell J, Boisleve F, Hata, M, Hirose A, Kasai Y, et al. 2018. Principles underpinning the use of new methodologies in the risk assessment of cosmetic ingredients. Computat Toxicol 7:20-26. https://doi.org/10.1016/j.comtox.2018.06.001.

Doke SK, Dhawale SC. 2015. Alternatives to animal testing: A review. Saudi Pharml J 23(3):223. https://doi.org/10.1016/j.jsps.2013.11.002.

Dudzina T, Delmaar CJE, Bisterbos JWH, Bakker MI, Bokkers BGH, Scheepers PTJ, van Engelen JGM, Hungerbuehler K, et al. 2015. The probabilistic aggregate consumer exposure model (PACEM): Validation and comparison to a lower-tier assessment for the cyclic siloxane D5. Environ Int 79:8-16. https://doi.org/10.1016/j.envint.2015.03.006.

ECETOC (European Centre for Ecotoxicology and Toxicology of Chemicals). 2016. Guidance for Effective use of Human Exposure Data in Risk Assessment of Chemicals. Technical Report No. 126. https://www.ecetoc.org/publication/technical-report-no-126-guidance-for-effective-use-of-human-exposure-data-in-risk-assessment-of-chemicals/

ECETOC. 2003. Contact Sensitisation: Classification According to Potency. Technical Report No. 87. https://www.ecetoc.org/publication/tr-087-contact-sensitisation-classification-according-to-potency/.

EFSA (European Food Safety Authority). 2024. Read-Across Application for Food or Feed Ingredients. https://doi.org/10.2903/sp.efsa.2024.EN-8811.

EFSA. 2018. Guidance on Uncertainty Analysis in Scientific Assessments. EFSA J 16(1):5123. https://doi.org/10.2903/j.efsa.2018.5123.

EPA (U.S. Environmental Protection Agency). 2021. Exposure Assessment Tools by Tiers and Types – Deterministic and Probabilistic Assessments. Retrieved June 3, 2025 from: https://www.epa.gov/expobox/exposure-assessment-tools-tiers-and-types-deterministic-and-probabilistic-assessments.

EPA. 2024. Alternative Test Methods and Strategies to Reduce Vertebrate Animal Testing. Retrieved June 3, 2025 from: https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/alternative-test-methods-and-strategies-reduce#:~:text=Examples%20of%20NAMs%20would%20be,for%20computer%2Ddriven%20predictive%20tools.

European Economic Community. 1992. Council Directive 92/86/EEC of 21 October 1992 amending for the seventh time Directive 76/768/EEC on the approximation of the laws of the Member States relating to cosmetic products. Official Journal of the European Communities L325: 18–21. http://data.europa.eu/eli/dir/1992/86/oj.

Felter SP, Robinson MK, Basketter DA, Gerberick GF. 2002. A review of the scientific basis for uncertainty factors for use in quantitative risk assessment for the induction of allergic contact dermatitis. Contact Dermatitis. 47(5):257-66. https://doi.org/10.1034/j.1600-0536.2002.470502.x.

Ficheux AS, Gomez-Berrada MP, Roudot AC, Ferret PJ. 2019. Consumption and exposure to finished cosmetic products: A systematic review. Food Chem Toxicol. 124:280-299. https://doi.org/10.1016/j.fct.2018.11.060.

Gautier F, Assaf Vandecasteele H, Tourneix F, van Vliet E, Alépée N, Bury D. 2023. Skin sensitisation prediction using read-across, an illustrative next generation risk assessment (NGRA) case study for vanillin. Regul Toxicol and Pharmacol 143, 105458. https://doi.org/10.1016/j.yrtph.2023.105458.

Genies C, Jamin, EL, Debrauwer L, Zalko D, Person EN, Eilstein J, Gregoire S, Schepky, A, et al. 2019. Comparison of the metabolism of 10 chemicals in human and pig skin explants. J Appl Toxicol 38(2):385-397. https://doi.org/10.1002/jat.3730.

Gilmour N, Kern PS, Alépée N, Boisleve F, Bury D, Clouet E, Hirota M, Hoffmann S, et al. 2020. Development of a next generation risk assessment framework for the evaluation of skin sensitisation of cosmetic ingredients. Regul Toxicol Pharmacol 116:104721. https://doi.org/10.1016/j.yrtph.2020.104721.

Gilmour N, Alépée N, Hoffmann S, Kern PS, van Vliet E, Bury D, Miyazawa M, Nishida H, et al. 2023. Applying a next generation risk assessment framework for skin sensitization

to inconsistent new approach methodology information. ALTEX 40(3):439-451. https://doi.org/10.14573/altex.2211161.

Helman G, Shah I, Williams AJ, Edwards J, Dunne J, Patlewicz G. 2019. Generalized Read-Across (GenRA): A workflow implemented into the EPA CompTox Chemicals Dashboard. ALTEX 36(3):462-465. https://doi.org/10.14573/altex.1811292.

Hirota M, Ashikaga T, Kouzuki H. 2018. Development of an artificial neural network model for risk assessment of skin sensitization using human cell line activation test, direct peptide reactivity assay, KeratinoSens and in silico structure alert parameter. J Appl Toxicol 38:514-526. https://doi.org/10.1002/jat.3558.

Hirota M, 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, Takenouchi O, Ashikaga T, Miyazawa M. 2015. Evaluation of combinations of in vitro sensitization test descriptors for the artificial neural network-based risk assessment model of skin sensitization. J Appl Toxicol 35:1333-1347. https://doi.org/10.1002/jat.3105.

ICCR (International Cooperation on Cosmetics Regulation). 2010. Integrated strategies for safety assessments of cosmetic ingredients - Part I. Retrieved June 3, 2025 from: https://www.iccr-cosmetics.org/topics-documents/5-integrated-strategies-for-safety-assesment.

Jaworska JS, Natsch A, Ryan C, Strickland J, Ashikaga T, Miyazawa M. 2015. Bayesian integrated testing strategy (ITS) for skin sensitization potency assessment: A decision support system for quantitative weight of evidence and adaptive testing strategy. Arch Toxicol 89(12):2355-2383. https://doi.org/10.1007/s00204-015-1634-2.

Kazem S, Linssen EC, Gibbs S. 2019. Skin metabolism phase I and phase II enzymes in native and reconstructed human skin: A short review. Drug Discov Today 24(9):1898-1910. https://doi.org/10.1016/j.drudis.2019.06.002.

Kleinstreuer N. 2021. International Progress in Skin Sensitization Hazard and Risk Assessment. Current Concepts in Quantitative Risk Assessment for Skin Sensitization (Webinar Series), virtual. https://ntp.niehs.nih.gov/whatwestudy/niceatm/3rs-meetings/past-meetings/ssqra/ssqra-web.

Kleinstreuer NC, Hoffmann S, Alépée N, Allen D, Ashikaga T, Casey W, Clouet E, et al. 2018. Non-animal methods to predict skin sensitization (II): An assessment of defined approaches. Crit Rev Toxicol 48 (5):359–74. https://doi.org/10.1080/10408444.2018.1429386.

Klimisch H-J, Andreae M, Tillman U. 1997. A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. Regul Toxicol Pharmacol 25(1):1-5. https://doi.org/10.1006/rtph.1996.1076.

Lee I, Forreryd A, Na M, Schember I, Lavelle M, Gradin R, Mattson U, Johansson H, et al. 2025. Determining a point of departure for skin sensitization potency and quantitative risk assessment of fragrance ingredients using the GARDskin dose-response assay. ALTEX 42(2):263-277. https://doi.org/10.14573/altex.2405131.

Lee I, Na M., Lavelle M, Api AM. 2022. Derivation of the no expected sensitization induction level for dermal quantitative risk assessment of fragrance ingredients using a weight of evidence approach. Food Chem Toxicol 159:112705. https://doi.org/10.1016/j.fct.2021.112705.

Na M, Ritacco G, O'Brien D, Lavelle M, Api AM, Basketter D. 2020. Fragrance skin sensitization evaluation and human testing: 30-year experience. Dermatitis 32(5):339-352. https://doi.org/10.1097/der.0000000000000684.

Na M, O'Brien D, Lavelle M, Lee I., Gerberick GF, Api AM. 2022. Weight of evidence approach for skin sensitization potency categorization of fragrance ingredients. Dermatitis 33(2):161-175. https://doi.org/10.1097/der.0000000000000854.

Natsch A. 2023. Integrated skin sensitization assessment based on OECD methods (III): Adding human data to the assessment. ALTEX 40(4):571-583. https://doi.org/10.14573/altex.2302081.

Natsch A, Gerberick GF. 2022a. Integrated skin sensitization assessment based on OECD methods (I): Deriving a point of departure for risk assessment. ALTEX 39(4):636-646. https://doi.org/10.14573/altex.2201141.

Natsch A, Gerberick GF. 2022b Integrated skin sensitization assessment based on OECD methods (II): Hazard and potency by combining kinetic peptide reactivity and the "2 out of 3" defined approach. ALTEX 39(4):647-655. https://doi.org/10.14573/altex.2201142.

Natsch A, Emter R, Haupt T, Ellis G. 2018. Deriving a no expected sensitization induction level for fragrance ingredients without animal testing: An integrated approach applied to specific case studies. Toxicol Sci 165:170-185. https://doi.org/10.1093/toxsci/kfy135.

NICEATM (National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods). 2024. DASS App User Guide (Version 2.0). Retrieved June 3, 2025 from: https://ntp.niehs.nih.gov/go/952311.

NICEATM. 2024. SARA-ICE: Skin sensitization risk assessment – Integrated chemical environment model. Retrieved June 3, 2025 from: https://ntp.niehs.nih.gov/whatwestudy/niceatm/test-method-evaluations/skin-sens/da/SARA-ICE.

Nishijo T, Api AM, Gerberick GF, Miyazawa M, Na M, Sakaguchi H. 2022. Implementation of a dermal sensitization threshold (DST) concept for risk assessment: Structure-based DST and in vitro data-based DST. Crit Rev Toxicol 52(1):51-65. https://doi.org/10.1080/10408444.2022.2033162.

NRC (National Research Council). 2014. Review of EPA's Integrated Risk Information System (IRIS) Process. The National Academies Press. https://doi.org/10.17226/18764.

NTP (National Toxicology Program). 2015. OHAT Risk of Bias Rating Tool for Human and Animal Studies. https://ntp.niehs.nih.gov/research/assessments/noncancer/riskbias.

OECD (Organisation for Economic Co-Operation and Development). 2025. Guideline No. 497: Defined Approaches on Skin Sensitisation, OECD Guidelines for the Testing of Chemicals, Section 4. Paris: OECD Publishing. https://doi.org/10.1787/b92879a4-en.

OECD. 2024a. Test No. 442C: In Chemico Skin Sensitisation - Assays Addressing the Adverse Outcome Pathway Key Event on Covalent Binding to Proteins. OECD Guidelines for the Testing of Chemicals, Section 4. https://doi.org/10.1787/9789264229709-en.

OECD. 2024b. Test No. 442D: In Vitro Skin Sensitisation - Assays Addressing the Adverse Outcome Pathway Key Event on Keratinocyte Activation. OECD Guidelines for the Testing of Chemicals, Section 4. https://doi.org/10.1787/9789264229822-en.

OECD. 2024c. Test No. 442E: In Vitro Skin Sensitisation - Assays Addressing the Key Event on Activation of Dendritic Cells on the Adverse Outcome Pathway for Skin Sensitisation. OECD Guidelines for the Testing of Chemicals, Section 4. https://doi.org/10.1787/9789264264359-en.

OECD. 2024d. Test No. 442B: Skin Sensitisation - Local Lymph Node Assay: BrdU-ELISA or -FCM.OECD Guidelines for the Testing of Chemicals, Section 4. https://doi.org/10.1787/9789264090996-en.

OECD. 2023a. (Q)SAR Assessment Framework: Guidance for the Regulatory Assessment of (Quantitative) Structure Activity Relationship Models and Predictions. https://doi.org/10.1787/d96118f6-en.

OECD. 2023b. (Q)SAR Assessment Framework: Guidance for the Regulatory Assessment of (Quantitative) Structure Activity Relationship Models and Predictions. Annex I – (Q)SAR model reporting format (QMRF) v2.1. https://www.oecd.org/chemicalsafety/risk-assessment/qsar-assessment-framework-annex-1-qsar-model-reporting-format.docx.

OECD. 2023c. (Q)SAR Assessment Framework: Guidance for the Regulatory Assessment of (Quantitative) Structure Activity Relationship Models and Predictions. Annex II – (Q)SAR prediction reporting format (QPRF) v.2.0. https://www.oecd.org/chemicalsafety/risk-assessment/qsar-assessment-framework-annex-2-qsar-prediction-reporting-format.docx.

OECD. 2022a. Users' Handbook Supplement to the Guidance Document for Developing and Assessing AOPs. OECD Series on Adverse Outcome Pathways No. 1. https://doi.org/10.1787/5jlv1m9d1g32-en.

OECD. 2022b. Test No. 406: Skin Sensitisation. OECD Guidelines for the Testing of Chemicals, Section 4. https://doi.org/10.1787/9789264070660-en.

OECD. 2022c. Report on Considerations from Case Studies on Integrated Approaches for Testing and Assessment (IATA): Series on Testing and Assessment No. 369. https://www.oecd.org/en/topics/sub-issues/testing-of-chemicals/publications-on-testing-and-assessment-of-chemicals.html.
https://one.oecd.org/document/ENV/CBC/MONO(2022)36/en/pdf.

OECD. 2020. Overview of Concepts and Available Guidance Related to Integrated Approaches to Testing and Assessment (IATA), OECD Series on Testing and Assessment, No. 329. https://doi.org/10.1787/cd920ca4-en.

OECD. 2019. Guiding Principles and Key Elements for Establishing a Weight of Evidence for Chemical Assessment. Series on Testing and Assessment, No. 311. https://doi.org/10.1787/f11597f6-en.

OECD. 2018. Guidance Document on Good in vitro Method Practices (GIVIMP). OECD Series on Testing and Assessment, No. 286. https://doi.org/10.1787/9789264304796-en.

OECD. 2017a. Guidance Document on the Reporting of Defined Approaches to be Used within Integrated Approaches to Testing and Assessment. Series on Testing & Assessment, No. 255. https://doi.org/10.1787/9789264274822-en.

OECD. 2017b. Guidance Document on the Reporting of Defined Approaches and Individual Information Sources to be Used within Integrated Approaches to Testing and Assessment (IATA) for Skin Sensitization. OECD Series on Testing and Assessment, No. 256. https://doi.org/10.1787/9789264279285-en.

OECD. 2017c. Guidance on Grouping of Chemicals, Second Edition, OECD Series on Testing and Assessment, No. 194. https://doi.org/10.1787/9789264274679-en

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

OECD. 2010a. Test No. 429: Skin Sensitization: Local Lymph Node Assay. https://doi.org/10.1787/9789264071100-en.

OECD. 2010b. Test No. 442A: Skin Sensitization: Local Lymph Node Assay: DA. https://doi.org/10.1787/9789264090972-en.

OECD QSAR Toolbox. 2025. Metabolisms (Observed and simulated). Retrieved June 3, 2025 from: https://qsartoolbox.org/resources/metabolisms/.

Paré G, Kitsiou S. 2017. Chapter 9: Methods for literature reviews. In: Lau F, Kuziemsky C, eds. In: Handbook of eHealth Evaluation: An Evidence-based Approach [Internet]. Victoria (BC): University of Victoria. 2017 Feb 27. Retrieved June 3, 2025 from: https://www.ncbi.nlm.nih.gov/books/NBK481583.

Patlewicz G, Lizarraga LE, Rua D, Allen DG, Daniel AB, Fitzpatrick SC, Garcia-Reyero N, Gordon J, Hakkinen P, Howard AS, Karmaus A, Matheson J, Mumtaz M, Richarz AN, Ruiz P, Scarano L, Yamada T, Kleinstreuer N. 2019. Exploring current read-across applications and needs among selected U.S. Federal Agencies. Regul Toxicol Pharmacol 106:197-209. https://doi.org/10.1016/j.yrtph.2019.05.011.

Politano VT, Api AM. 2008. The Research Institute for Fragrance Materials' human repeated insult patch test protocol. Regul Toxicol Pharmacol 52(1):35-8. https://doi.org/10.1016/j.yrtph.2007.11.004.

Reinke EN, Reynolds J, Gilmour N, Reynolds G, Strickland J, Germolec D, Allen, DG, Maxwell G, Kleinstreuer NC. 2025. The Skin Allergy Risk Assessment-Integrated Chemical Environment (SARA-ICE) defined approach to derive points of departure for skin sensitization. Curr Res Toxicol 8:100205. https://doi.org/10.1016/j.crtox.2024.100205.

Reynolds J, Gilmour N, Baltazar MT, Reynolds G, Windebank S, Maxwell G. 2022. Decision making in next generation risk assessment for skin allergy: Using historical clinical experience to benchmark risk. Regul Toxicol Pharmacol 134:105219. https://doi.org/10.1016/j.yrtph.2022.105219.

Reynolds J, Aleksic M, Golmour N, Reynolds G, 2025. Determination of acceptable margins of exposure (MoE) for NAM-based skin sensitisation risk assessment, ALTEX (submitted).

Robinson MK, Gerberick GF, Ryan CA, McNamee P, White IR, Basketter DA 2000. The importance of exposure estimation in the assessment of skin sensitization risk. Cont Dermatitis 42(5):251-259. https://doi.org/10.1034/j.1600-0536.2000.042005251.x.

SAAOP (Society for Advancement for AOPs). 2025. AOP Wiki. Retrieved June 3, 2025 from: https://aopwiki.org/.

Safford B, Api AM, Barratt C, Comiskey D, Dal, EJ, Ellis G, McNamara C, O'Mahony C, et al. 2015. Use of an aggregate exposure model to estimate consumer exposure to fragrance ingredients in personal care and cosmetic products. Regul Toxicol Pharmacol 72(3):673-682. https://doi.org/10.1016/j.yrtph.2015.05.017.

Safford B, Api AM, Barratt C, Comiskey D, Ellis G, McNamara C, O'Mahony C, Robison S, et al. 2017. Application of the expanded Creme RIFM consumer exposure model to fragrance ingredients in cosmetic, personal care and air care products. Regul Toxicol Pharmacol 86:148-156. https://doi.org/10.1016/j.yrtph.2017.02.021.

Safford RJ. 2008. The Dermal Sensitisation Threshold— A TTC approach for allergic contact dermatitis. Regul Toxicol Pharmacol 51(2):195-200. https://doi.org/10.1016/j.yrtph.2008.02.010.

Safford RJ, Aptula AO, Gilmour, N. 2011. Refinement of the Dermal Sensitisation Threshold (DST) approach using a larger dataset and incorporating mechanistic chemistry domains. Regul Toxicol Pharmacol 60(2):218-224. https://doi.org/10.1016/j.yrtph.2011.03.009.

Sahigara F, Mansouri K, Ballabio D, Mauri A, Consonni V, Todeschini, R. 2012. Comparison of different approaches to define the applicability domain of QSAR models. Molecules 17(5):4791–4810. https://doi.org/10.3390/molecules17054791.

SCCS (Scientific Committee on Consumer Safety). 2010. Basic Criteria for the In Vitro Assessment of Dermal Absorption of Cosmetic Ingredients. Scientific Committee on

Consumer Safety (SCCS/1358/10). Retrieved June 3, 2025 from: https://ec.europa.eu/health/scientific_committees/consumer_safety/docs/sccs_s_002.pdf.

SCCS. 2023. SCCS Notes of Guidance for the Testing of Cosmetic Ingredients and Their Safety Evaluation (12th Revision). Scientific Committee on Consumer Safety (SCCS/1647/22). https://health.ec.europa.eu/publications/sccs-notes-guidance-testing-cosmetic-ingredients-and-their-safety-evaluation-12th-revision_en.

Selvestrel G, Robino F, Russo MZ. 2022. In silico models for skin sensitization and irritation. Methods Mol Biol 2425:291-354. https://doi.org/10.1007/978-1-0716-1960-5_13.

Simon T, Zhu Y, Dourson ML, Beck NB 2016. Bayesian methods for uncertainty factor application for derivation of reference values. Regul Toxicol Pharmacol 80:9-24. https://doi.org/10.1016/j.yrtph.2016.05.018.

Strickland J, Daniel AB, Allen D, Aguila C, Ahir S, Bancos S, Craig E, Germolec D, et al. 2019. Skin sensitization testing needs and data uses by U.S. regulatory and research agencies. Arch Toxicol 93(2):273-291. https://doi.org/10.1007/s00204-018-2341-6.

Strickland J, Zang Q, Kleinstreuer N., Paris M, Lehmann DM, Choksi N, Matheson J, Jacobs A., et al. 2017. Integrated decision strategies for skin sensitization hazard. J Appl Toxicol 36(9):1150-1162. https://doi.org/10.1002/jat.3281.

Strickland J, Abedini J, Allen DG, Gordon J, Hull, V, Kleinstreuer NC, Ko HS, Matheson J, et al. 2023. A database of human predictive patch test data for skin sensitization. Archives Toxicol 97(11):2825-2837. https://doi.org/10.1007/s00204-023-03530-3.

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. 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. J Appl Toxicol 35(11):1318-32. doi: 10.1002/jat.3127. https://doi.org/10.1002/jat.3127.

Tourneix F, Alépée N, Detroyer A, Eilstein J, Ez-Zoubir M, Teissier SM, Noçairi H, Piroird C, Basketter D, Del Bufalo A. 2020. Skin sensitisation testing in practice: Applying a stacking meta model to cosmetic ingredients. Toxicol in Vitro 66:104831-104844. https://doi.org/10.1016/j.tiv.2020.104831.

Tourneix F, Carron L, Jouffe L, Hoffmann S, Alépée N. 2024. Deriving a continuous point of departure for skin sensitization risk assessment using a bayesian network model. Toxics 12:536-552. https://doi.org/10.3390/toxics12080536.

UN (United Nations). 2023. Globally Harmonized System of Classification and Labelling of Chemicals (GHS) (10th Rev. Ed.). Retrieved June 3, 2025 from: https://unece.org/transport/documents/2023/07/standards/ghs-rev10.

Wikoff D, Lewis RJ, Erraguntla N, Franzen, A, Foreman, J. 2020. Facilitation of risk assessment with evidence-based methods—A framework for use of systematic mapping and systematic reviews in determining hazard, developing toxicity values, and characterizing uncertainty. Regul Toxicol Pharmacol 118:104790. https://doi.org/10.1016/j.yrtph.2020.104790.

Wilm A, Kühnl J, Kirchmair J. 2018. Computational approaches for skin sensitization prediction. Crit Rev Toxicol 48(9):738-760. https://doi.org/10.1080/10408444.2018.1528207.

# Appendix A. *In chemico* and *in vitro* skin sensitization methods with applicability domain information

Table A1. *In Chemico* and *In Vitro* Skin Sensitization Methods with Applicability Domain Information

| Test Method | Description | Applicability Domain | Test Guideline |
|---|---|---|---|
| DPRA | Quantifies residual concentration of cysteine- or lysine-containing synthetic peptide after exposure to substance | - Generally applicable to a range of functional groups and physicochemical properties<br>- Not applicable for testing metals<br>- Substances that require metabolic activation may produce false negative results; substances that require abiotic transformation are correctly predicted in most cases<br>- Substances that promote oxidation could lead to false positive predictions | OECD TG 442C |
| ADRA | Quantifies residual concentration of a cysteine or lysine derivative after exposure to substance | - Generally applicable to a range of functional groups and physicochemical properties<br>- Not applicable for testing metals<br>- Substances that require metabolic activation may produce false negative results; substances that require abiotic transformation ware correctly predicted in most cases<br>- Substances that promote oxidation could lead to false positive predictions | OECD TG 442C |
| kDPRA | Quantifies residual concentration of a cysteine-containing synthetic peptide after exposure to substance in a time and concentration dependent manner | - Generally applicable to a range of functional groups and physicochemical properties<br>- Not applicable for testing metals or substances that exclusively have lysine activity<br>- Substances that require metabolic activation may produce false negative results<br>- Substances that promote oxidation could be underpredicted | OECD TG 442C |
| ARE-Nrf2 luciferase KeratinoSens™ test method (KeratinoSens$^{TM}$) | Quantifies changes in luciferase induction in the KeratinoSens$^{TM}$ transgenic cell line, which is under the transcriptional control of a promoter fused with the antioxidant response | - Applicable to test substances in solution or form stable dispersion<br>- Generally applicable to a range of functional groups and physicochemical properties<br>- False negative results may occur with substances that exclusively react with lysine-residues<br>- Substances that require metabolic activation may produce false negative results due to limited metabolic capacity of cells | OECD TG 442D |

| Test Method | Description | Applicability Domain | Test Guideline |
|---|---|---|---|
| | element and indicates activation of the Nrf2 gene | | |
| ARE-Nrf2 luciferase LuSens test method (LuSens) | Quantifies changes in luciferase induction in the LuSens transgenic cell line, which is under the transcriptional control of a promoter fused with the antioxidant response element and indicates activation of the NQO1 gene | - Applicable to test substances in solution or form stable dispersion<br>- Generally applicable to a range of functional groups and physicochemical properties<br>- False negative results may occur with substances that exclusively react with lysine-residues<br>- Substances that require metabolic activation may produce false negative results due to limited metabolic capacity of cells | OECD TG 442D |
| Epidermal sensitization assay (EpiSensA) | Quantifies changes in expression of four genes associated with keratinocyte activation in the reconstructed human epidermis model | - Applicable to test substances in solution or those that form stable dispersion at 0.0122%<br>- Generally applicable to a range of functional groups and physicochemical properties<br>- Surfactants may produce a false positive result | OECD TG 442D |
| Human cell line activation test (h-CLAT) | Quantifies changes in the expression of cell surface markers CD86 and CD57, in the human monocytic leukemia cell line THP-1, which is associated with activation of monocytes and dendritic cells after exposure to test substance | - Applicable to test substances in solution or those that form stable dispersion<br>- Substances with log Kow >3.5 tend to produce false negative responses<br>- Substances that require metabolic activation may produce false negative results due to limited metabolic capacity of cells<br>- Generally applicable to range of functional groups and physicochemical properties | OECD TG 442E |
| U937 cell activation test (U-SENS™) | Quantifies changes in the expression of cell surface marker CD86, in the human histiocytic lymphoma cell | - Applicable to test substances in solution or those that form stable dispersion<br>- Substances with membrane disrupting capacity may produce false positive results | OECD TG 442E |

| Test Method | Description | Applicability Domain | Test Guideline |
|---|---|---|---|
| | line U937, after exposure to test substance | - Substances with surfactant properties may produce false responses<br>- Generally applicable to range of functional groups and physicochemical properties | |
| Interleukin-8 reporter gene assay (IL-8 Luc assay) | Quantifies changes in cytokine IL-8 expression in the THP-G8 cell line, which is associated with the activation of DC | - Applicable to test substances in solution or those that form stable dispersion<br>- False negatives may occur with anhydrides<br>- False negatives more likely to occur with substances showing low/moderate skin sensitization potency (e.g., UN GHS 1B)<br>- Generally applicable to range of functional groups and physicochemical properties | OECD TG 442E |
| GARD$^{TM}$skin | Evaluates transcriptional patterns of an endpoint-specific biomarker signature in SenzaCell cell line | - Autofluorescent test substances may interfere with assay<br>- Generally applicable to range of functional groups and physicochemical properties | OECD TG 442E |

# Appendix B. Historically applied Safety Assessment Factors (SAFs)

Table B1. Historically Applied Safety Assessment Factors (SAFs)

| **Product type** | **Overall SAF** |
|---|---|
| Products applied to lips | 100 |
| Products applied to underarm | 300 |
| Products applied to anogenital area | 300 |
| Products applied to face | 100 |
| Products applied to body (rinse off) | 300 |
| Products applied to body (leave on) | 100 |
| Hair products (no hand exposure) | 30 |

Values based upon those published in Api et al. (2020).