Section 4
Health effects

Guideline No. 497

Guideline on Defined Approaches for Skin Sensitisation

14 June 2021

OECD Guidelines for the Testing of Chemicals

OECD
BETTER POLICIES FOR BETTER LIVES

OCDE/OECD 497

Adopted
14 June 2021

OECD GUIDELINE FOR TESTING OF CHEMICALS

Defined Approaches for Skin Sensitisation

Table of Contents

Tables

Figures

1. Introduction

1.1. General Introduction

1. A skin sensitiser refers to a substance that will lead to an allergic response following repeated skin contact as defined by the United Nations Globally Harmonized System of Classification and Labelling of Chemicals (UN GHS) (1). There is general agreement on the key biological events underlying skin sensitisation. The current knowledge of the chemical and biological mechanisms associated with skin sensitisation initiated by covalent binding to proteins has been summarised as an Adverse Outcome Pathway (AOP) (2) that begins with a molecular initiating event, leading to intermediate key events, and terminating with the adverse effect, allergic contact dermatitis.

2. The skin sensitisation AOP focuses on chemicals that react with amino acid residues (*i.e.* cysteine or lysine) such as organic chemicals. In this instance, the molecular initiating event (*i.e.* the first key event), is the covalent binding of electrophilic substances to nucleophilic centres in skin proteins. The second key event in this AOP takes place in the keratinocytes and includes inflammatory responses as well as changes in gene expression associated with specific cell signaling pathways such as the antioxidant/electrophile response element (ARE)-dependent pathways. The third key event is the activation of dendritic cells, typically assessed by expression of specific cell surface markers, chemokines and cytokines. The fourth key event is T-cell proliferation, and the adverse outcome is presentation of allergic contact dermatitis.

3. The assessment of skin sensitisation has typically involved the use of laboratory animals. The classical methods that use guinea-pigs, the Guinea Pig Maximisation Test (GPMT) of Magnusson and Kligman and the Buehler Test (OECD TG 406) (3) assess both the induction and elicitation phases of skin sensitisation. The murine tests, such as the LLNA (OECD TG 429) (4) and its three non-radioactive modifications — LLNA: DA (OECD TG 442A) (5), LLNA: BrdU-ELISA, and BrdU-FCM (OECD TG 442B) (6) — all assess the induction response exclusively and have gained acceptance, since they provide an advantage over the guinea pig tests in terms of animal welfare together with an objective measurement of the induction phase of skin sensitisation.

4. Mechanistically-based *in chemico* and *in vitro* test methods (OECD TG 442C, 442D, 442E) (7, 8, 9) addressing the first three key events (KE) of the skin sensitisation AOP can be used to evaluate the skin sensitisation hazard potential of chemicals. None of these test methods are considered sufficient stand-alone replacements of animal data to conclude on skin sensitisation potential of chemicals or to provide information for potency sub-categorisation according to the UN GHS (sub-categories 1A and 1B). However, data generated with these *in chemico* and *in vitro* methods addressing multiple KEs of the skin sensitisation AOP are proposed to be used together, as well as with information sources such as *in silico* and read-across predictions from chemical analogues, within integrated approaches to testing and assessment (IATA) or defined approaches (DAs). Results from the individual information sources can only be used in DAs if the substances fall within the applicability domains of the methods (see "Initial Considerations, Applicability and Limitations" sections of respective methods (TG 442C, Appendix 1; TG 442D, Appendix 1A; TG 442E Annex 1) (7, 8, 9).

5. Results from multiple information sources can be used together in DAs to achieve an equivalent or better predictive capacity than that of the animal tests to predict responses in humans. A DA consists of a fixed data interpretation procedure (DIP) (*e.g.* a mathematical model, a rule-based approach) applied to data (*e.g in silico* predictions, *in chemico*, *in vitro* data) generated with a defined set of information sources to derive a prediction without the need for expert judgment. Individual DAs for skin sensitisation and their respective information sources were originally described in Guidance Document 256, Annex I/II (10) and a preliminary assessment was published in Kleinstreuer et al (11). The DAs use method combinations intended to overcome some of the limitations of the individual, stand-alone methods in order to provide increased confidence in the overall result obtained. The ultimate goal of DAs is to provide information that is equivalent to that provided by animal studies, *i.e.* information that can be used for hazard identification and/or potency categorisation.

6. Testing laboratories should consider all relevant available information on the test chemical prior to conducting the studies as directed by a DA. Such information could include, for example, the identity and chemical structure of the test chemical and its physico-chemical properties. Such information should be considered in order to determine whether the individual OECD test guideline methods under a specific DA are applicable for the test chemical.

7. When performing a hazard evaluation and/or potency sub-categorisation based on the output from an *in vivo* (LLNA or any other) test, from an *in chemico* test, from an *in vitro* test, from an *in silico* approach, from a DA, and any combination thereof, the same principles always apply, *i.e.* all available information relevant to the chemical in question should be taken into consideration as well as toxicological data on structurally related test chemicals if available.

8. This Guideline was developed with the input of an OECD Expert Group on Defined Approaches for Skin Sensitisation (EG DASS) comprised of scientific experts from regulatory agencies, validation bodies, non-governmental organisations, and industry.

9. Three rule-based DAs are included in this Guideline, and are described with respect to their intended regulatory purpose: hazard identification, *i.e.* discrimination between skin sensitisers and non-sensitisers (1.4.Part I), or potency sub-categorisation (1.6.Part IIPart II). The DAs included in Part II are also suitable for hazard identification. The evaluation and review of the DAs are described in detail in the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation.*

10. A comprehensive dataset of 196 chemicals with DA predictions, data on individual information sources, highly curated LLNA and Human Patch Predictive Test (HPPT) data, and physicochemical properties, was compiled and is attached as **Annex 2** to the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation*. Out of the 196 chemicals, 168 chemicals have LLNA classifications and 66 chemicals have HPPT classifications, which were all agreed upon by the EG DASS and used to evaluate the performance of the DAs. Due to the availability of data, this dataset contains mainly cosmetic ingredients but also other types of chemicals that are used across sectors such as preservatives, dyes, or food ingredients. The dataset is chemically diverse as shown by the physicochemical properties covered by these chemicals: it contains small and large molecules (molecular weight ranges from 30 to 512 g/mol), hydrophobic and hydrophilic substances (Log P ranges from -3.9 to 9.4), solids and liquids (melting point ranges from -122 to 253 ºC), volatile and non-volatile substances (boiling point ranges from -19 to 445 ºC). Further details on the chemical space characterization of the reference

database are available in **Section 4** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation.*

11. Other DAs may be included in this Guideline following future review and approval. DAs able to provide a quantitative measure of sensitisation potency, such as a point of departure which can be used for risk assessment, may be included in a new Part II to this Guideline in the future.

1.2. DAs and Use Scenarios included in the Guideline

12. The DAs currently described in this guideline are:

- The "2 out of 3" (2o3) defined approach to skin sensitisation hazard identification based on *in chemico* (KE1) and *in vitro* (KE2/KE3) data (12, 13). See Part I
- The integrated testing strategy (ITSv1) for UN GHS potency categorisation based on *in chemico* (KE1) and *in vitro* (KE3) data, and *in silico* (Derek Nexus) predictions (14, 15), with a DIP developed with expert group (EG DASS) input. See Part II Potency Categorisation.
- A modification of the integrated testing strategy (ITSv2) for UN GHS potency categorisation based on *in chemico* (KE1) and *in vitro* (KE3) data, and *in silico* (OECD QSAR Toolbox) predictions, with a DIP developed with expert group (EG DASS) input. See Part IIPotency Categorisation.

13. The DAs described in this guideline are based on the use of validated OECD test methods (DPRA, KeratinoSens™, h-CLAT), for which transferability, within- and between-laboratory reproducibility have been characterised in the validation phase (7, 8, 9).

14. The ITS DAs (ITSv1 and ITS v2) also make use of an *in silico* information source; Derek Nexus v6.1.0 (ITSv1), or OECD QSAR Toolbox v4.5 (ITSv2). Derek Nexus (referred to as Derek hereafter) is an expert knowledge-based tool which provides predictions of skin sensitisation potential using structural alerts, and OECD QSAR Toolbox (referred to as OECD QSAR TB hereafter) is a computational tool which uses an analogue-based read-across approach or structural alerts for protein binding identified by profilers to predict whether a chemical will be a sensitiser.

15. All DAs described in this guideline can each be used to address countries' requirements for discriminating between sensitisers (*i.e.* UN GHS Category 1) from non-sensitisers, though they do so with different sensitivities and specificities (detailed in the respective descriptions of each DA).

16. The ITS DAs (ITSv1 and ITS v2) can also be used to discriminate chemicals into three UN GHS potency categories (Category 1A = strong sensitisers; Category 1B = other sensitisers, and No Categorization (NC = not classified).

17. The known limitations and applicability domains of the individual information sources were used to design workflows for assigning confidence to each of the predictions produced by the DAs described in this guideline. In order to have a high confidence prediction, the underlying data must meet criteria in the respective test guidelines (see TG 442C, Appendix 1; TG 442D, Appendix 1A; TG 442E Annex 1 (7, 8, 9)), DA predictions with high confidence for hazard identification and/or potency are considered conclusive. DA predictions with low confidence are considered inconclusive for hazard identification and/or potency (see **Sections 2.1.4** and **3.1.4** for further information). These 'inconclusive'

predictions may nevertheless be considered in a weight-of-evidence approach and/or within the context of an IATA together with other information sources (*e.g.* demonstration of exposure to the test system, existing *in vivo* data, clinical data, read-across, other *in vitro* / *in chemico* / *in silico* data, etc.).

18. The performance of the DAs described in this guideline for discriminating between sensitisers and non-sensitisers was evaluated using 168 (135 GHS Skin Sens. Category 1, and 33 no classification) test chemicals for which DPRA, KeratinoSens™, h-CLAT, Derek, OECD QSAR TB predictions and classifications based on LLNA reference data agreed upon by the EG DASS are available (for additional details see **Section 2.1** and **Annex 3** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation*). For the purpose of evaluating the performance of the ITS DAs for predicting UN GHS classifications based on potency categorization (sub-category 1A, 1B, or "not classified" (NC)), 156 test chemicals (38 1A, 85 1B, and 33 NC) were used because for 12 test chemicals it was not possible to assign with sufficient confidence the potency sub-category 1A or 1B on the basis of LLNA data. Mixtures and botanicals with undefined structural composition were excluded from the curated LLNA reference data.

19. The performance of the three DAs (high confidence predictions only) against the LLNA reference data for predicting skin sensitisation hazard showed balanced accuracies (average of sensitivity and specificity; BA) in the range of 80-84%, with sensitivities of 82-93% and specificities of 67-85% (see **Table 1.1**). Note that specificity measures are more uncertain than sensitivities due to lower number of negative reference chemicals. Detailed performance statistics are reported in Part I (2o3 DA) and Part II (ITS DA). The performance of the ITSv1 and ITSv2 DAs for UN GHS classifications based on potency categorization (high confidence predictions only, sub-category 1A, 1B, or NC) when compared to the LLNA reference data yielded overall accuracies of 71%, overall balanced accuracies of 78% (ITSv1) or 77% (ITSv2), and balanced accuracies within a predicted sub-category or NC ranging from 72-81% (ITSv1) or 71-80% (ITSv2). There were no strong sensitisers (1A) that were incorrectly predicted as being a non-sensitiser (NC) or vice versa. Detailed performance statistics are reported in Part II and in **Section 5** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation.*

20. The performance of the DAs described in this guideline for discriminating between sensitisers and non-sensitisers was also evaluated using a set of 66, or 65 for 2o3, due to lack of assay data for one chemical, test chemicals (55 sensitisers and 11 non-sensitisers) for which classifications based on Human Predictive Patch Test (HPPT) data have been agreed upon by the EG DASS (for additional details see **Section 2.2** and **Annex 4** *of the Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation*). For the purpose of evaluating the performance of the ITS DAs for predicting UN GHS classifications based on potency categorization (sub-category 1A, 1B, or NC), 63 test chemicals were used (21 1A, 31 1B, and 11 NC) because for 3 test chemicals it was not possible to assign with sufficient confidence the potency sub-category 1A or 1B on the basis of human reference data. Mixtures and botanicals with undefined structural composition were excluded from the curated human reference data.

21. The performance of the DAs (high confidence predictions only) against the human reference data for predicting skin sensitisation hazard showed balanced accuracies in the range of 69-88%, with sensitivities of 89-94% and specificities of 44-88% (see **Table 1.1**). Note that specificity measures are more uncertain than sensitivities due to lower number of negative reference chemicals. Detailed performance statistics are reported in Part I(2o3

DA) and Part II (ITS DA). The performance of the ITSv1 and ITSv2 DAs for UN GHS skin sensitisation potency classification (high confidence predictions only, sub-category 1A, 1B and NC) when compared to the human reference data yielded overall balanced accuracies of 72% (ITSv1) or 73% (ITSv2), and balanced accuracies within a predicted sub-category or NC in the range of 68-79% (ITSv1) or 69-79% (ITSv2). Detailed performance statistics are reported in Part II and in **Section 5** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation.*

22. The overlap between the LLNA and human reference datasets was 56 chemicals for hazard and 47 chemicals for skin sensitisation potency categorisation, respectively, and the performance of the LLNA against the human reference data was evaluated using these chemicals as a basis for comparison. The performance of the LLNA against the human reference for predicting skin sensitisation hazard showed a balanced accuracy of 58%, with sensitivity of 94% and specificity of 22%. Note that the specificity measure is more uncertain than the sensitivity due to a lower number of negative reference chemicals. The performance of the LLNA for UN GHS potency classification when compared to the human reference data yielded an overall balanced accuracy of 64%, and balanced accuracies within a predicted sub-category or NC in the range of 59-73% There were no strong skin sensitisers (1A) in the human reference data that were incorrectly predicted by the DAs, or by the LLNA as not being a sensitiser (no classification) or vice versa. Detailed performance statistics are reported Part Iand Part II

Table 1.1. Summary of the DAs Included in this Guideline

DA/Method	Information Sources	Capability (Hazard and/or Potency)	Hazard Performance vs. LLNA	Hazard Performance vs. Human	Potency Performance vs. LLNA (Accuracy)	Potency Performance vs. Human (Accuracy)
2o3 DA	DPRA, KeratinoSens™, h-CLAT	Hazard	84% BA, 82% Sens, 85% Spec	88% BA, 89% Sens, 88% Spec	-	-
ITSv1 DA	DPRA, h-CLAT, DEREK Nexus v6.1.0	Hazard, Potency	81% BA, 92% Sens, 70% Spec	69% BA, 93% Sens, 44% Spec	70% NC, 71% 1B, 74% 1A	44% NC, 77% 1B, 65% 1A
ITSv2 DA	DPRA, h-CLAT, OECD QSAR Toolbox v4.5	Hazard, Potency	80% BA, 93% Sens, 67% Spec	69% BA, 94% Sens, 44% Spec	67% NC, 72% 1B, 72% 1A	44% NC, 80% 1B, 67% 1A
LLNA (provided for comparison)	*in vivo*	Hazard, Potency	-	58% BA, 94% Sens, 22% Spec	-	25% NC, 74% 1B, 56% 1A

Note: For hazard performance, sensitivity (Sens) is the true positive rate, specificity (Spec) is the true negative rate, and balanced accuracy (BA) is the average of sensitivity and specificity. Due to the imbalanced nature of the reference data, the measures of specificity are more uncertain than the measures of sensitivity. For potency performance, accuracy reflects correct classification rate within each UN GHS sub-category. Due to the imbalanced nature of the reference data, the measures of accuracy are more uncertain for smaller classes, *e.g.* for NC chemicals. Statistics reflect conclusive DA predictions only. This represents the data available at the time of initial guideline adoption.

1.3. Limitations

23. **Table 1.1** provides an overview of the DAs included in this Guideline, their information sources used, whether they provide hazard and/or potency prediction, and

summarises their performance against the LLNA and human reference data. The LLNA (OECD TG 429) is included in **Table 1.1** as a basis for comparison. More details are provided in Part I and Part II of this Guideline, as well as in the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation*.

24. The identified limitations of the DAs and their individual components are summarised below.

1.3.1. Limitations of individual in chemico/in vitro information sources

25. Users should refer to the limitations of the individual in chemico/*in vitro* test methods as specified in their respective Test Guidelines, which are revised as new data become available and should be consulted regularly. The most up-to-date published version of the respective TGs should always be used. For example, some types of chemicals such as metals, inorganic compounds, UVCBs and mixtures, may not be within the applicability domain for certain test methods. Individual assay results within borderline ranges (**Annex 1**) may yield inconclusive DA predictions. The consideration of limitations of individual *in chemico*/*in vitro* test methods in each DA is detailed in **Section 2.1.4** (**Figure 2.1**) and **Section 3.1.4** (**Figure 3.1**).

1.3.2. Limitations of in silico information sources

26. Some DAs include *in silico* tools as an information source. These tools can either perform automated read-across or (Q)SAR predictions. (Q)SARs include both structure-activity relationship (SAR) models (*i.e.* structural alerts, expert systems) and quantitative structure-activity relationship (QSAR) models (*i.e.* statistical tools). (Q)SAR models should fulfil the OECD Principles for the Validation, for Regulatory Purposes, of (Q)SAR Models and be described in a QSAR Model Reporting Format (QMRF) document (21)[1]. One of the OECD QSAR validation principles refers to a defined domain of applicability. The defined domain of applicability reflects limitations beyond which less reliable predictions may be obtained (*e.g.* training set ranges of descriptors included in the model and types of chemical structures included in the training set). A given *in silico* model may be associated with more than one defined applicability domain, each of which is associated with its own reliability measures as established in the validation. Depending on the DIP, chemicals outside the applicability domain may result in DA predictions of low confidence that are considered inconclusive. Where a DA for skin sensitisation includes an *in silico* tool, users should refer to the limitations and applicability domain of the individual *in silico* tool. Two of the DAs covered in this Guideline, the ITSv1 and the ITSv2, rely upon the *in silico* tools Derek and OECD QSAR TB, respectively, and their specified limitations and applicability domains are detailed in **Annex 2** of this Guideline.

1.3.3. Limitations of DAs

27. The limitations of the DAs are based on the limitations of the individual *in chemico/in vitro/in silico* information sources. Details on using the limitations of individual information sources to determine confidence in DA predictions are provided in **Sections**

[1] The QMRF has been slightly adapted for reporting other *in silico* model predictions in the context of DASS. The adapted QPRF can be found on the OECD site for spreadsheets and software associated with OECD Test Guidelines on Health Effects: https://www.oecd.org/env/ehs/testing/section4software.htm.

2.1.4 and **3.1.4** and in the respective test guidelines (TG 442C, Appendix 1; TG 442D, Appendix 1A; TG 442E, Annex 1) (7, 8, 9).

28. During the evaluation of the DAs covered in this Guideline it was observed that, with respect to LLNA data, the DPRA (TG 442C), KeratinoSens™ (TG 442D), h-CLAT (TG 422E), as well as the proposed DAs, have lower sensitivity for test chemicals with Log P > 3.5 **(**for details see **Section 3.1.4** and **Annex 5** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation*). It was also noted that the LLNA test may produce a higher number of false positive results for these test chemicals when compared with human reference data, and supporting mechanistic information was provided **(**for details see **Section 3.2** and **Annex 6** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation*). Overall, the analyses and the number of reference chemicals with Log P > 3.5 are insufficient to draw firm conclusions. However, according to TG 442E, negative h-CLAT results for substances with Log P > 3.5 should not be considered, and this limitation is applied to the DAs as described in **Sections 2.1.4** and **3.1.4.**

29. For the 2o3 DA, borderline ranges (BRs) have been defined for the individual assays addressing the three KE of the DA, in order to define areas where lower confidence may exist (for details see **Section 2.1.4** and **Annex 1** of this Guideline, and **Section 3.3** and **Annex 7** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation*). Positive and/or negative test results falling within these BRs as well as individual assay limitations, *e.g.* negative h-CLAT results obtained for a chemical with Log P > 3.5 (according to TG 442E), have lower confidence and may result in inconclusive 2o3 DA predictions.

30. Inconclusive DA predictions may nevertheless be considered in a weight-of-evidence approach and/or within the context of an IATA together with other information sources (*e.g.* demonstration of exposure to the test system, existing *in vivo* data, clinical data, read-across, other *in vitro* / *in chemico* / *in silico* data, etc.).

1.4. References

1. United Nations (UN) (2019). Globally Harmonized System of Classification and Labelling of Chemicals (GHS). Eighth revised edition, New York and Geneva, United Nations Publications. Available at: [https://unece.org/ghs-rev8-2019]
2. OECD (2012). Series on Testing and Assessment No. 168. The Adverse Outcome Pathway for Skin Sensitisation Initiated by Covalent Binding to Proteins. Part 1: Scientific Evidence. Organisation for Economic Cooperation and Development, Paris. Available at: [https://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm]
3. OECD (1992). OECD Guidelines for the Testing of Chemicals No. 406. Skin Sensitisation. Organisation for Economic Cooperation and Development, Paris. Available at: [https://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm].
4. OECD (2010). OECD Guidelines for Chemical Testing No. 429. Skin sensitisation: Local Lymph Node assay. Organisation for Economic Cooperation and Development, Paris. Available at :(oecd-ilibrary.org).
5. OECD (2010). OECD Guidelines for Chemical Testing No. 442A.Skin sensitisation: Local Lymph Node assay: DA. Organisation for Economic Cooperation and Development, Paris. Available at:(oecd-ilibrary.org)..
6. OECD (2018). OECD Guidelines for Chemical Testing No. 442B. Skin sensitisation: Local Lymph Node assay: BrdU-ELISA or –FCM. Organisation for Economic Cooperation and Development, Paris. Available at: (oecd-ilibrary.org).
7. OECD (2020). OECD Guideline for the Testing of Chemicals No. 442C: *In chemico* Skin Sensitisation: Assays addressing the Adverse Outcome Pathway key even on covalent binding to proteins).*In chemico*. Paris, France: Organisation for Economic Cooperation and Development. Available at: (oecd-ilibrary.org).
8. OECD (2018), OECD Key Event based test Guideline 442D: *In vitro* Skin Sensitisation Assays Addressing AOP Key Event on Keratinocyte Activation. Organisation for Economic Cooperation and Development, Paris. Available at: (oecd-ilibrary.org).
9. OECD (2018). OECD Key event-based test Guideline 442E: *In vitro* Skin Sensitisation Assays Addressing the Key Event on Activation of Dendritic Cells on the Adverse Outcome Pathway for Skin Sensitisation. Organisation for Economic Cooperation and Development, Paris. Available at: (oecd-ilibrary.org).
10. OECD (2016). Series on Testing & Assessment No. 256: 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 Sensitisation, Annex 1 and Annex 2. Organisation for Economic Cooperation and Development, Paris. Available at: [https://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm].
11. Kleinstreuer N, Hoffmann S, Alepee N, et al. (2018). Non-Animal Methods to Predict Skin Sensitization (II): an assessment of defined approaches. Crit Rev Toxicol Feb 23:1-16. doi: 10.1080/10408444.2018.1429386

12. Bauch C, Kolle SN, Ramirez T, Eltze T, Fabian E, Mehling A, Teubner W, van Ravenzwaay B, Landsiedel R. (2012). Putting the parts together: combining *in vitro* methods to test for skin sensitizing potentials. Regul Toxicol Pharmacol, 63:489-504.

13. Urbisch D, Mehling A, Guth K, Ramirez T, Honarvar N, Kolle S, Landsiedel R, Jaworska J, Kern PS, Gerberick F, Natsch A, Emter R, Ashikaga T, Miyazawa M, Sakaguchi H. (2015). Assessing skin sensitization hazard in mice and men using non-animal test methods, Regul Toxicol Pharmacol, 71:337-51.

14. Nukada Y, Miyazawa M, Kazutoshi S, Sakaguchi H, Nishiyama N. (2013). Data integration of non-animal tests for the development of a test battery to predict the skinsensitizing potential and potency of chemicals. Toxicol *In vitro*, 27:609-18.

15. 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, 35:1318-32.

16. ICCVAM (1999). Report on The Murine Local Lymph Node Assay: A Test Method for Assessing the Allergic Contact Dermatitis Potential of Chemicals/Compounds. https://ntp.niehs.nih.gov/iccvam/docs/immunotox_docs/llna/llnarep.pdf

17. Dumont C, Barroso J, Matys I, Worth A, Casati S. (2016). Analysis of the Local Lymph Node Assay (LLNA) variability for assessing the prediction of skin sensitisation potential and potency of chemicals with non-animal approaches. Toxicol *In vitro*,34:220-228.

18. Hoffmann S. (2015). LLNA variability: An essential ingredient for a comprehensive assessment of non-animal skin sensitization test methods and strategies. ALTEX, 32:379-83.

19. Roberts DW, Api AM, Aptula AO. (2016). Chemical applicability domain of the LocalvLymph Node Assay (LLNA) for skin sensitisation potency. Part 2. The biological variability of the murine Local Lymph Node Assay (LLNA) for skin sensitisation. Regul Toxicol Pharmacol, 80:255-9.

20. Hoffmann, S., Kleinstreuer, N., Alépée, N., et al. (2018). Non-Animal Methods to Predict Skin Sensitization (I): the Cosmetics Europe database. Crit Rev Toxicol, 48(5):344-358.

21. ECHA (2008). see "CHAPTER R.6 – QSARS AND GROUPING OF CHEMICALS" *in* Guidance on Information Requirements and Chemical Safety Assessment. European Chemicals Agency [Guidance on Information Requirements and Chemical Safety Assessment - ECHA (europa.eu)].

31. Part I of this guideline applies to DAs that are intended solely for hazard identification, *i.e.* distinguishing between sensitisers and non-sensitisers. A summary of the DAs for hazard identification is provided below; additional detailed information can be found in the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation.*

1.5. "2 out of 3" Defined Approach

1.5.1. Summary

32. The 2 out of 3 (2o3) DA is intended for the identification of the skin sensitisation hazard of a chemical without the use of animal testing, *i.e.* UN GHS Cat. 1 vs. UN GHS NC. The data interpretation procedure (DIP) is currently not designed to provide information on the potency of a sensitiser.

33. The combination of test methods included in the 2o3 DA covers at least two of the first three KEs of the AOP leading to skin sensitisation as formally described by the OECD: KE1: protein binding (*i.e.* via the direct peptide reactivity assay (DPRA; OECD TG 442C)) (1); KE2: keratinocyte activation (*i.e.* KeratinoSens™; OECD TG 442D) (2); and KE3: dendritic cell activation (*i.e.* via the human cell line activation test (h-CLAT; OECD TG 442E)) (3).

34. The DIP entails that two concordant results obtained from methods addressing at least two of the first three KEs of the AOP determine the final classification. The 2o3 DA was compared to 168 chemicals with curated LLNA reference data agreed upon by the EG DASS and demonstrated an accuracy of 83% and a balanced accuracy of 84% (see **Table 2.1**). The 2o3 DA was also compared to 65 chemicals with curated human reference data agreed upon by the EG DASS and exceeded the accuracy, and balanced accuracy, of the LLNA for hazard identification. (see **Tables 2.1-2.2**). It should be noted that due to the imbalanced nature of the reference data (higher numbers of positives than negatives), the measures of balanced accuracy are more uncertain, particularly in the case of the human data comparison.

1.5.2. Data interpretation procedure

35. The data interpretation procedure (DIP) in the 2o3 DA is a transparent, rule-based approach requiring no expert judgment (4, 5, 6). The approach predicts skin sensitisation hazard by sequential testing, in an undefined order, in up to three of the following internationally accepted non-animal assays mapping to KE1-3 (*i.e.* DPRA, KeratinoSens™, h-CLAT). Assays are run for two KEs, and if these assays provide consistent results, then the chemical is predicted accordingly as sensitiser or non-sensitiser. If the first two assays provide discordant results, the assay for the remaining KE is run. The overall result is based on the two concordant findings taking into account the confidence on the obtained predictions as described in **Section 2.1.4**.

36. The performance of the 2o3 DA was found to be impacted by the consideration of borderline ranges for each of the methods, as described below in **Section 2.1.4**, and further detailed in **Section 3.3** and **Annex 7** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation.* A decision tree is provided in **Figure 2.1**

of **Section 2.1.4** to derive predictions for the 2o3 DA, with no modification of the 2o3 DA Data Interpretation Procedure.

1.5.3. Description and limitations of the individual information sources

37. The individual information sources in the DA are assays included in OECD KE-based test guidelines for skin sensitisation (OECD TG 442C, 442D, 442E) (1, 2, 3), and the protocols are detailed therein.

38. The following assays from those TGs have been characterised and included in the 2o3 DA.

- Direct Peptide Reactivity Assay (DPRA; OECD TG 442C; KE1) (1): Skin sensitisers are generally electrophilic and react with the nucleophilic moieties of proteins. The DPRA measures depletion of two peptides containing either cysteine or lysine residues due to covalent binding. A test chemical that induces 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%) is considered to be positive. In case borderline results are obtained for peptide depletion, additional testing should be conducted, as specified in OECD TG 442C and in **Annex 1**.
- KeratinoSens™ assay (*In vitro* Skin Sensitisation: ARE-Nrf2 Luciferase Test Method; OECD TG 442D; KE2) (2); Keratinocytes harbouring a reporter gene construct react to possible sensitisers via the Nrf2-Keap1 pathway. A test chemical that causes >1.5 fold luciferase induction, at viabilities > 70% when compared to the vehicle control, is considered to be positive. In case borderline results are obtained for luciferase induction, additional testing should be conducted, as specified in **Annex** 1.
- Human cell-line activation test (h-CLAT; OECD TG 442E; KE3) (3): Activation of antigen presenting cells is characterised by the up-regulation of CD86 and/or CD54. The h-CLAT is considered to be positive if CD86 induction exceeds 1.5-fold and/or CD54 exceeds 2-fold at viabilities > 50% when compared to the vehicle control. In case borderline results are obtained for CD54 and/or CD86 induction, additional testing should be conducted, as specified in **Annex 1**.

39. The current limitations of individual *in chemico* and *in vitro* test methods, such as limitations with respect to solubility, are described in the respective test guidelines (TG 442C, Appendix 1; TG 442D, Appendix 1A; TG 442E, Annex 1) and the validation studies cited therein (1, 2, 3).

1.5.4. Confidence in the 2o3 DA predictions

40. The first decision on whether each information element can be used is dictated by the limitations of the *in chemico* and *in vitro* methods (*e.g*. for substances that do not provide conclusive results in the individual methods due to solubility reasons) as found in in the respective test guidelines (TG 442C, Appendix 1; TG 442D, Appendix 1A; TG 442E, Annex 1) (1, 2, 3). Additionally, test results are subject to variation and these variations increase the uncertainty of a test result especially when close to a (classification) cut-off, *i.e.* in the borderline range. In order to define areas where lower confidence in the DA results may exist, borderline ranges (BRs) have been defined for output from the individual assays addressing the three KE of the 2o3 DA, (see **Annex 1** of this document, and **Section 3.3** and **Annex 7** of the *Supporting document to the Guideline (GL) on Defined Approaches*

(DAs) for Skin Sensitisation). The specific borderline ranges for each assay, as derived from their respective validation study data, are:

- DPRA BR: mean peptide depletion: 4.95% – 8.32%, Cys-only depletion (in the case of co-elution with lysine peptide): 10.56% – 18.47%;
- KeratinoSens™ BR: Imax: 1.35-fold – 1.67-fold;
- h-CLAT BR: RFI CD54: 157% – 255%; RFI CD86: 122% – 184%.

41. The incorporation of borderline ranges (BRs) into the prediction models (PM) for each of the individual information sources is are described in **Annex 1** of this guideline.

42. For the data with a single run as reported in the reference database, borderline cases in the DPRA are identified based on the borderline range for the mean peptide depletion or Cys-only depletion as described above. In case repeated runs are conducted, the PM in **Annex 1, Figure 1.1** shall be applied.

43. The prediction model of the KeratinoSens™ assay requires multiple runs. For the assessment of whether the outcome of repeated runs yields a positive, negative or borderline final outcome in KeratinoSens™, the PM in **Annex 1, Figure 1.2** shall be applied (adapted from the PM described in TG 442D to be used within the 2o3 DA to conclude on borderline cases). This prediction model introduces a third outcome (borderline) to be used within the 2o3 DA, based on the same decision cut-offs of the prediction model described in TG 442D. Thus, a negative in the original prediction model can only become negative or borderline, while a positive from the original prediction model can only become positive or borderline.

44. The prediction model of h-CLAT requires multiple runs. For the assessment of whether the outcome of repeated runs yields a positive, negative or borderline final outcome in the h-CLAT, the PM in **Annex 1**, **Figure 1.3** shall be applied (adapted from the PM described in TG 442E to be used within the 2o3 DA to conclude on borderline cases). This prediction model introduces a third outcome (borderline) to be used within the 2o3 DA, based on the same decision cut-offs of the prediction model described in TG 442E. Thus, a negative in the original prediction model can only become negative or borderline, while a positive from the original prediction model can only become positive or borderline.

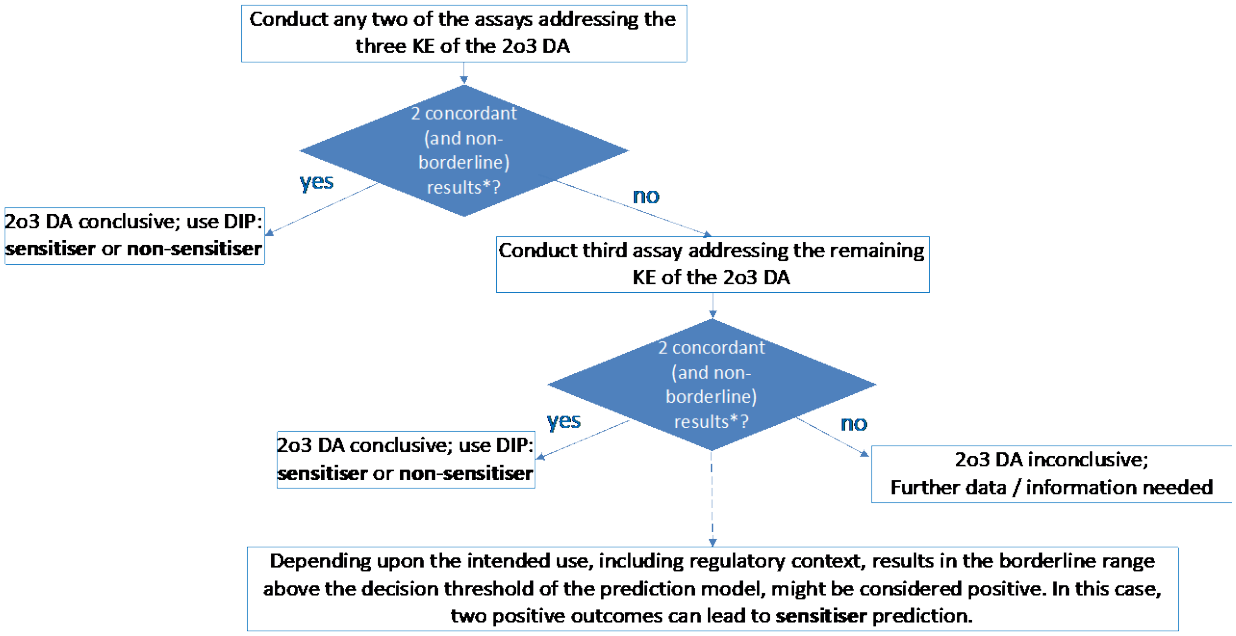
45. Positive and negative test results falling within these BRs as well as inconclusive results due to limitations in the *in chemico/in vitro* test guidelines are of lower confidence. For example, negative h-CLAT results obtained for a chemical with Log P > 3.5 (according to TG 442E (3)) are of lower confidence, and affect the outcome of the 2o3 DA as described below:

- In case the result of one of the 2o3 DA test methods falls into the respective test method's BR, a 2o3 DA prediction can still be made if the outcomes of the other two test methods composing the 2o3 DA are concordant and have high confidence (*i.e.*, results falling outside of the respective BRs).
- Similarly, in case a negative h-CLAT result is obtained for a chemical with Log P > 3.5, a 2o3 DA prediction can still be made if the outcomes of the other two test methods composing the 2o3 DA are concordant and have high confidence (*i.e.*, results falling outside of the respective BRs).
- However, if the result of one of the 2o3 DA test methods falls into the respective test method's BR or a negative h-CLAT result is obtained for a chemical with Log

P > 3.5, and the other two methods composing the 2o3 do not provide concordant and high confidence results, the 2o3 DA prediction is considered 'inconclusive'. These inconclusive predictions may nevertheless be considered in a weight-of-evidence approach and/or within the context of an IATA together with other information sources. Depending on the intended use, including regulatory context, results in the borderline range above the decision threshold of the prediction model might still be considered positive; in this case, two positive outcomes can lead to an overall positive (sensitiser) prediction.

46. These borderline considerations and their impact on the confidence of the 2o3 DA predictions are visualized in **Figure 2.1.** DA predictions with high confidence for hazard identification are considered conclusive. DA predictions with low confidence are considered inconclusive for hazard identification. These 'inconclusive' predictions may nevertheless be considered in a weight-of-evidence approach and/or within the context of an IATA together with other information sources.

Figure 0.1. Decision tree to be used for the 2o3 DA, taking into account borderline results



Note: Borderline results are determined based on workflows given in **Annex 1**.
* The use of information elements is dictated by the limitations as found in in the respective test guidelines (TG 442C, Appendix 1; TG 442D, Appendix 1A; TG 442E, Annex 1). For example, in case a negative h-CLAT result is obtained for a chemical with Log P > 3.5 (according to the limitation described in TG 442E (3)), a 2o3 DA prediction can only be made if the outcomes of the other two test methods composing the 2o3 DA are concordant and are non-borderline.

1.5.5. Predictive capacity of the 2o3 DA vs. the LLNA

47. The predictive capacity of the "2o3" DA is reported based on data generated by the LLNA (see **Table 2.1**), curated as agreed upon by the EG DASS (see **Section 2.1** and **Annex 3** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation)*. The borderline range analyses were applied as described above to assign confidence to the 2o3 DA predictions. Performance statistics are reported for conclusive (high confidence) predictions as compared to LLNA reference data, and

inconclusive (low confidence) results are indicated. DA predictions for specific chemicals and further details are available in **Section 5** and **Annex 2** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation.*

Table 0.2. Hazard identification performance of the "2o3" DA in comparison to LLNA reference data

	LLNA	
2o3 DA	Non	Sens
Non	22	19
Sens	4	89
Inconclusive	7	27

DA Performance vs. LLNA Data (N=134)	**2o3**
Accuracy (%)	83%
Sensitivity (%)	82%
Specificity (%)	85%
Balanced Accuracy (%)	84%

Note: Accuracy is the correct classification rate, sensitivity is the true positive rate, specificity is the true negative rate, and balanced accuracy is the average of sensitivity and specificity. Performance is reported based on DPRA, KeratinoSens™, and h-CLAT. Statistics reflect conclusive predictions only; inconclusive predictions are shown in grey. Additional performance characterisation is available in the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation.*

48. The application of the BR analyses and the designation of high/low confidence for the 2o3 DA predictions is applied as described above in **Section 2.1.4** and **Annex 1,** and further detailed in **Section 3.3** and **Annex 7** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation.*

49. Due to the imbalanced nature of the reference data, the measure of specificity (based on 26 LLNA negative chemicals) is more uncertain than the measure of sensitivity (based on 108 LLNA positive chemicals).

1.5.6. Predictive capacity of the 2o3 DA vs. Human Data

50. The predictive capacity of the "2o3" DA is also reported based on Human Predictive Patch Test (HPPT) data (see **Table 2.2**), curated as agreed upon by the EG DASS (see **Section 2.2** and **Annex 4** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation)*. The borderline range analyses were applied as described above to assign confidence to the 2o3 DA predictions. Performance statistics are reported for conclusive (high confidence) predictions as compared to human reference data, and inconclusive (low confidence) results are indicated. DA predictions for specific chemicals and further details are available in **Section 5 and Annex 2** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation.*

Table 0.3. Hazard identification performance of the "2o3" DA in comparison to human reference data

	Human	
2 of 3 DA	Non	Sens

Non	7	5
Sens	1	42
Inconclusive	3	7

DA Performance vs. Human Data (N=55)	2o3
Accuracy (%)	89%
Sensitivity (%)	89%
Specificity (%)	88%
Balanced Accuracy (%)	88%

Note: Accuracy is the correct classification rate, sensitivity is the true positive rate, specificity is the true negative rate, and balanced accuracy is the average of sensitivity and specificity with respect to HPPT data. Performance is reported based on DPRA, KeratinoSens™, and h-CLAT. Statistics reflect conclusive predictions only; inconclusive predictions are shown in grey. Additional performance characterisation is available in the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation*.

51. The application of the BR analyses and the designation of high/low confidence for the 2o3 DA predictions is applied as described above in **Section 2.1.4** and **Annex 1**, and further detailed in **Section 3.3 and Annex 7** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation.*

52. Due to the imbalanced nature of the reference data, the measure of specificity (based on 8 human negative chemicals) is more uncertain than the measure of sensitivity (based on 47 human positive chemicals).

1.5.7. Predictive capacity of the LLNA vs. Human Data

53. To provide a basis for comparison for the DA performance statistics given above, the predictive capacity of the LLNA is reported based on data from the Human Predictive Patch Test (see **Table 2.3**) curated as agreed upon by the EG DASS. Data for specific chemicals and further details are available in **Section 5** and **Annex 2** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation*.

Table 2.3. Hazard identification performance of the LLNA in comparison to Human reference data

	Human	
LLNA	Non	Sens
Non	2	3
Sens	7	44

LLNA Performance vs. Human Data (N=56)	LLNA
Accuracy (%)	82%
Sensitivity (%)	94%
Specificity (%)	22%
Balanced Accuracy (%)	58%

Note: Accuracy is the correct classification rate, sensitivity is the true positive rate, specificity is the true negative rate, and balanced accuracy is the average of sensitivity and specificity with respect to Human HPPT-based data. Additional performance characterisation is available in the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation.*

54. The hazard identification performance of the conclusive 2o3 DA predictions vs. human HPPT data was 89% accuracy, 89% sensitivity, 88% specificity, and 88% balanced accuracy, comparable to and/or exceeding the performance of the LLNA vs human HPPT data in every measure.

55. As previously noted, due to the imbalanced nature of the reference data, the measures of specificity are more uncertain than the measures of sensitivity.

1.5.8. Proficiency chemicals

56. The 2o3 DA relies on a simple, rule-based data interpretation procedure and requires no expert judgment. Proficiency chemicals for the individual information sources (KE1-3) are defined in the respective guidelines (1, 2, 3). Proficiency for the individual information sources demonstrates proficiency for the DA.

1.5.9. Reporting of the DA

57. The reporting of the DA application should follow the template described in OECD GD 255 (7), and should include at a minimum the following elements:

- Test chemical identification (*e.g.* chemical name, structural formula, composition, isomers, impurities including their quantities as available, CAS number, batch and lot number, and other relevant identifiers)
- Individual test reports performed per corresponding guideline (OECD TG 442C, 442D, 442E). Note that the chemical identity for each test report should match that above.
- Application of the individual prediction models adapted to be used within the 2o3 DA to determine borderline outcomes, as described in **Annex 1**
- Outcome of the DA application (hazard identification, *i.e.* skin sensitiser or not skin sensitiser or inconclusive result)
- Any deviation from or adaptation of the 2o3 DA
- Conclusion

1.6. References

1. OECD (2020). OECD Guideline for the Testing of Chemicals No. 442C: *In chemico* Skin Sensitisation: Assays addressing the Adverse Outcome Pathway key even on covalent binding to proteins).*In chemico*. Paris, France: Organisation for Economic Cooperation and Development. Available at: (oecd-ilibrary.org).
2. OECD (2018). OECD Key Event based test Guideline 442D: *In vitro* Skin Sensitisation Assays Addressing AOP Key Event on Keratinocyte Activation. Organisation for Economic Cooperation and Development, Paris. Available at: (oecd-ilibrary.org).
3. OECD (2018). OECD Key event based test Guideline 442E: *In vitro* Skin Sensitisation Assays Addressing the Key Event on Activation of Dendritic Cells on the Adverse Outcome Pathway for Skin Sensitisation. Organisation for Economic Cooperation and Development, Paris. Available at: (oecd-ilibrary.org).
4. OECD (2016). Series on Testing & Assessment No. 256: 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 Sensitisation, Annex 1 and Annex 2.. Organisation for Economic Cooperation and Development, Paris. Available at: [https://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm].
5. Bauch C, Kolle SN, Ramirez T, Eltze T, Fabian E, Mehling A, Teubner W, van Ravenzwaay B, Landsiedel R. (2012). Putting the parts together: combining *in vitro* methods to test for skin sensitizing potential. Regul Toxicol Pharmacol, 63:489-504.
6. Urbisch D, Mehling A, Guth K, Ramirez T, Honarvar N, Kolle S, Landsiedel R, Jaworska J, Kern PS, Gerberick F, Natsch A, Emter R, Ashikaga T, Miyazawa M, Sakaguchi H. (2015). Assessing skin sensitization hazard in mice and men using non-animal test methods, Regul Toxicol Pharmacol, 71:337-51.
7. OECD (2016). Series on Testing & Assessment No. 255: Guidance Document On The Reporting Of Defined Approaches To Be Used Within Integrated Approaches To Testing And Assessment. Organisation for Economic Cooperation and Development, Paris. Available at: [https://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm].

Part II. – Defined Approaches for Skin Sensitisation Potency Categorisation

58. Part II of the Guideline includes Defined Approaches that allow the allocation of skin sensitizers into UN GHS sub-category 1A, strong sensitizers, or sub-category 1B for other (moderate to weak) skin sensitizers, following the Globally Harmonised System for Classification and Labeling (GHS). These DAs may also be used for hazard identification, *i.e.* to distinguish between sensitisers (UN GHS Category 1) and non-sensitisers (no classification; NC). Currently the ITSv1 DA and ITSv2 DA are included in this section of the Guideline. Additional detailed information can be found in the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation.*

1.7. "Integrated Testing Strategy (ITS)" Defined Approach

1.7.1. Summary

59. This defined approach was constructed as an Integrated Testing Strategy (ITS) for prediction of the skin sensitisation hazard potential and potency sub-categorisation according to the UN GHS (sub-categories 1A and 1B) of a chemicals.

60. The ITS DA uses test methods that address key events (KEs) 1 and 3 in the Adverse Outcome Pathway (AOP) and includes an *in silico* prediction of skin sensitisation. Protein binding (KE1) is quantitatively evaluated using the Direct Peptide Reactivity Assay (DPRA; OECD TG 442C) (1). Dendritic cell activation (KE3) is quantitatively evaluated using the human cell line activation test (h-CLAT; OECD TG 442E) (2). The *in silico* prediction of skin sensitisation is provided by either Derek Nexus (ITSv1) or OECD QSAR Toolbox (ITSv2).

61. The ITSv1 DA was evaluated for hazard identification with 167 chemicals and for UN GHS sub-categorisation with 155 chemicals based on LLNA reference data curated as agreed upon by the EG DASS, and achieved accuracies equivalent to the LLNA (see **Tables 3.2-3.3**). The performance of the ITSv1 DA was compared to 64 chemicals with human reference data curated as agreed upon by the EG DASS (see **Tables 3.4-3.5**), and exceeded the accuracy of the LLNA in predicting the same human data for both hazard and potency categorisation.

62. The ITSv2 DA was evaluated for hazard identification for 167 chemicals and for UN GHS sub-categorisation for 153 chemicals based on LLNA reference data curated as agreed upon by the EG DASS, and achieved accuracies equivalent to the LLNA (see **Tables 3.6-3.7**). The performance of the ITSv2 DA was compared to 64 chemicals with human reference data curated as agreed upon by the EG DASS (see **Tables 3.8-3.9**), and exceeded the accuracy of the LLNA in predicting the same human data for both hazard and potency categorisation.

1.7.2. Data interpretation procedure

63. The ITS DIP uses scores assigned to the quantitative results from the h-CLAT (2) and the DPRA (1), and from either Derek Nexus v6.1.0 (2020, Lhasa Limited, https://www.lhasalimited.org/products/derek-nexus.htm) or OECD QSAR TB v4.5 (https://www.oecd.org/chemicalsafety/oecd-qsar-toolbox.htm) to discriminate chemicals into UN GHS category 1A (strong sensitiser); category 1B (other sensitiser), or Not Classified (non-sensitiser) (**Table 3.1**).

64. The DIP was amended from the original published version of the ITS (3) to change the cut-off for 1A sensitisers from a score of 7 to a score of 6 to optimize the ability of the DA to detect strong sensitisers and to extend the applicability of the ITS to chemicals for which *in silico* predictions cannot be generated. The DIP was also altered from the published version in that it was originally applied to ECETOC categories[2], and is here applied to the UN GHS subcategories.

65. The quantitative results of h-CLAT and DPRA are converted into a score from 0 to 3, as shown in **Table 3.1**. For h-CLAT, the minimum induction threshold (MIT) is converted to a score from 0 to 3 based on the cutoffs of 10 and 150 μg/ml. For DPRA, the mean percent depletion for the cysteine and lysine peptides is converted to a score from 0 to 3, based on the threshold values associated with reactivity classes described in OECD TG 442C (1). In cases where co-elution occurs only with the lysine peptide, the depletion for only cysteine peptides is converted to a score from 0 to 3. For the *in silico* prediction (Derek or OECD QSAR TB), a positive outcome is assigned a score of 1; a negative outcome is assigned a score of 0 (further details on the respective protocols are available in **Annex 2**). When these scores have been assessed, a total battery score ranging from 0 to 7, calculated by summing the individual scores, is used to predict the sensitising potential (hazard identification; UN GHS Cat. 1 vs. UN GHS NC) and potency (UN GHS Cat. 1A, Cat. 1B and NC). The positive criteria for identifying skin sensitisers (UN GHS Cat. 1) are set as a total battery score of 2 or greater. Based on the updated DIP, a total battery score is assigned into three ranks: score of 6-7 is defined as a strong (UN GHS Cat. 1A) sensitiser; score of 2-5 as moderate/weak (UN GHS Cat. 1B) sensitiser; score of 1 or 0, as not classified (*i.e.* a non-sensitiser).

[2] ECETOC Technical Report 087 (2003), Contact Sensitisation: Classification According to Potency. Available at: [https://www.ecetoc.org/publication/tr-087-contact-sensitisation-classification-according-to-potency/]

Table 0.4. Schematic of the ITS defined approach. The DA is a simple score-based system depending on assays from OECD TG 442E and 442C, and an *in silico* structure-based prediction, as shown.

Score	h-CLAT MIT µg/mL	DPRA mean Cysteine and Lysine% depletion	DPRA Cysteine % depletion*	*In silico* (ITSv1: DEREK; ITSv2: OECD TB)
3	≤10	≥42.47	≥98.24	
2	>10, ≤150	≥22.62, <42.47	≥23.09, <98.24	
1	>150, ≤5000	≥6.38, <22.62	≥13.89, <23.09	Positive
0	not calculated	<6.38	<13.89	Negative
	Potency	Total Battery Score		
	UN GHS 1A	6-7		
	UN GHS 1B	2-5		
	Not classified	0-1		

Source: Adapted from Takenouchi (6)

Note: UN GHS 1A correspond to strong sensitisers and UN GHS 1B correspond to other (moderate to weak) sensitisers. Not classified are considered non-sensitisers. *Cysteine-only depletion thresholds are used in the case of co-elution with the lysine peptide.

1.7.3. Description and limitations of the individual information sources

66. The individual *in chemico* and *in vitro* information sources are existing KE-based OECD test guidelines (OECD TG 442C, 442E) (1, 2), and the protocols are detailed therein.

67. The following assays from those TGs have been characterised and included in the ITS DA:

- Human cell-line activation test (h-CLAT; OECD TG 442E; KE3) (2): Activation of antigen presenting cells is characterised by the up-regulation of CD86 and/or CD54. The h-CLAT is considered to be positive if CD86 induction exceeds 1.5-fold and/or CD54 exceeds 2-fold at viabilities > 50% when compared to the vehicle control. From the experimental concentration-response curves, the median concentration(s) inducing 1.5- and/or 2-fold induction of CD86 and/or CD54 are calculated and the lowest of the two values is defined as the minimal induction threshold, MIT:

  $$MIT = \min(EC150\ CD86, EC200\ CD54)$$

  Test chemicals are assigned potency scores based on the MIT thresholds shown in **Table 3.1**.

- Direct Peptide Reactivity Assay (DPRA; OECD TG 442C; KE1) (1): Skin sensitisers are generally electrophilic and react with the nucleophilic moieties of proteins. The DPRA measures depletion of two peptides containing either cysteine or lysine residues due to covalent binding. A test chemical that induces 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%) is considered to be positive. In case borderline results are obtained for peptide depletion, additional testing should be conducted, as specified in OECD TG 442C. Test chemicals are assigned potency scores based on the mean peptide depletion thresholds shown in **Table 3.1**.

68. The limitations of the individual *in chemico* and *in vitro* test methods are described in the respective test guidelines and in the respective test guidelines (TG 442C, Appendix 1; TG 442E, Annex 1) (1, 2).

69. The *in silico* information source predictions for ITSv1 are derived from Derek, an expert, knowledge-based software tool comprising alerts on several toxicity endpoints, including skin sensitisation. Derek (Derek Nexus v.6.1.0, 2020, Lhasa Limited) fires alerts based on structural features *i.e.* whether a hapten has potential for electrophilic binding to skin proteins either directly or following metabolism/auto-oxidation. To each alert, a likelihood level is associated. Chemicals firing an alert with a likelihood of certain, probable, plausible, or equivocal are considered to be positive. Chemicals with a negative prediction of 'non-sensitiser with no misclassified or unclassified features' are considered to be negative (https://www.lhasalimited.org/products/skin-sensitisation-assessment-using-derek-nexus.htm#Negative%20Predictions). The approach for characterising the *in silico* applicability domain used in the ITSv1 and the protocol for generating Derek predictions are provided in **Annex 2** of this guideline.

70. The *in silico* information source predictions for ITSv2 are derived from the OECD QSAR TB automated workflow providing skin sensitiser hazard predictions (OECD QSAR TB v4.5). The target compound is profiled for protein binding alerts; auto-oxidation products and skin metabolites are generated and then profiled for protein binding alerts. In case a protein binding alert is identified in the parent or in its (a)biotic metabolites, the same alert is used to identify analogues with experimental skin sensitisation data. If no protein binding alert is identified, then structural profilers are used to identify analogue chemicals and the data gap is filled using read across or directly via profiler outcomes in case no suitable analogues are automatically identified. The approach for characterising the *in silico* applicability domain used in the ITSv2 and the protocol for generating OECD QSAR TB predictions are provided in **Annex 2** of this guideline.

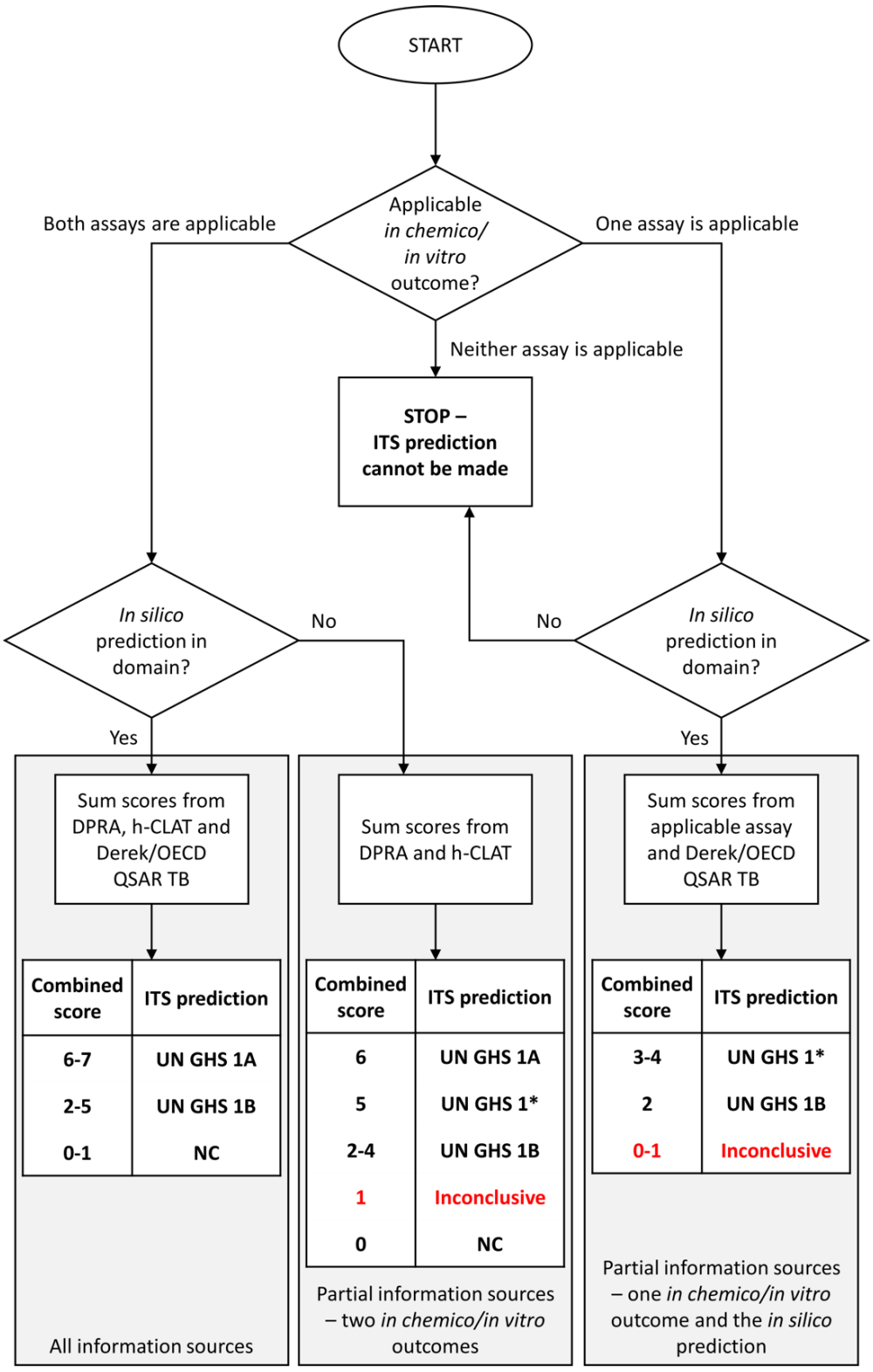
1.7.4. Confidence in the ITS DA predictions

71. The level of confidence of the ITS DA prediction is assigned based on the total DA score and applicability domain of the individual information sources, as shown via the flow chart in **Figure 3.1**. The first decision on whether all information elements can be used is dictated by the limitations of the *in chemico* and *in vitro* methods as found in TG 442C Appendix 1 and TG 442E Annex 1 (2) (*e.g.* for substances that do not provide conclusive results in the individual methods due to limited solubility or negative h-CLAT results for chemicals with Log P > 3.5 which are currently considered unrealiable), and by the applicability domain of the *in silico* prediction (**Annex 2**). Partial information sources (*i.e.* two *in chemico/in vitro* outcomes only, or one *in chemico/in vitro* outcome and an *in silico*

prediction) may be used to obtain a DA prediction as shown via the flow chart in **Figure 3.1**.

72. DA predictions with high confidence for hazard identification and potency are considered conclusive. DA predictions with low confidence are considered inconclusive for hazard identification and/or potency. These 'inconclusive' predictions may nevertheless be considered in a weight-of-evidence approach and/or within the context of an IATA together with other information sources. Details including applicability domain and confidence considerations are provided in **Annex 2**.

Figure 0.2. Decision tree for assigning confidence to the ITS DA predictions



***Conclusive for hazard, inconclusive for potency**

1.7.5. Predictive capacity of the ITSv1 DA vs the LLNA

73. The predictive capacity of ITSv1 using Derek is reported based on data from the LLNA (see **Tables 3.2-3.3**), curated as agreed upon by the EG DASS (see **Section 1.1** and **Annex 3** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation)*. The workflow shown in **Figure 3.1** was applied to assign confidence to the ITSv1 DA predictions. The designation of conclusive/inconclusive for the ITSv1 DA predictions is further detailed in **Annex 2**. Performance statistics are reported for conclusive predictions as compared to LLNA reference data, and inconclusive results are indicated. DA predictions for specific chemicals and further details are available in **Section 5** and **Annex 2** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation.*

Table 0.5. Hazard identification performance of the ITSv1 DA in comparison to LLNA reference data

	LLNA	
ITSv1 DA	Non	Sens
Non	21	11
Sens	9	118
Inconclusive	3	6

DA Performance vs. LLNA Data (N=159)	**ITSv1**
Accuracy (%)	87%
Sensitivity (%)	92%
Specificity (%)	70%
Balanced Accuracy (%)	81%

Note: Accuracy is the correct classification rate, sensitivity is the true positive rate, specificity is the true negative rate, and balanced accuracy is the average of sensitivity and specificity with respect to LLNA data. Statistics reflect high confidence predictions only; inconclusive predictions are shown in grey. Additional performance characterisation is available in the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation.*

74. Due to the imbalanced nature of the reference data, the measure of specificity (based on 30 LLNA negative chemicals) is more uncertain than the measure of sensitivity (based on 129 LLNA positive chemicals).

Table 0.6. Potency categorisation performance of the ITSv1 DA in comparison to LLNA reference data, based on the UN GHS 1A/1B sub-categorisation

	LLNA		
ITSv1 DA	NC	1B	1A
NC	21	11	0
1B	9	55	10
1A	0	12	28
Inconclusive	3	7	0

71% correct classification overall

ITSv1 vs. LLNA reference data: Statistics based on the UN GHS 1A/1B sub-categorisation

Performance (N=146)	NC (N=30)	1B (N=78)	1A (N=38)
Correct classification (%)	70%	71%	74%
Underpredicted (%)	NA	14% (NC)	0% (NC); 26% (1B)
Overpredicted (%)	30% (1B); 0% (1A)	15% (1A)	NA

Note: Statistics reflect high confidence predictions only; inconclusive predictions are shown in grey. For more details on within-class performance (sensitivity, specificity, and balanced accuracy), please see **Section 5** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation.*

75. The designation of high/low confidence for the ITSv1 DA predictions is applied as described above in **Figure 3.1** and further detailed in **Annex 2**.

1.7.6. Predictive capacity of the ITSv2 DA vs the LLNA

76. The predictive capacity of ITSv2 using OECD QSAR TB is reported based on data from the LLNA (see **Tables 3.4-3.5**), curated as agreed upon by the EG DASS (see **Section 2.1** and **Annex 3** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation)*. The workflow shown in **Figure 3.1** was applied to assign confidence to the ITSv2 DA predictions. The designation of high/low confidence for the ITSv2 DA predictions is further detailed in **Annex 2**. Performance statistics are reported for high confidence predictions as compared to LLNA reference data, and inconclusive results are indicated. DA predictions for specific chemicals and further details are available in **Section 5** and **Annex 2** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation.* **Table 0.4**. Hazard identification performance of the ITSv2 DA in comparison to LLNA reference data.

Table 0.7. Hazard identification performance of the ITSv2 DA in comparison to LLNA reference data.

	LLNA	
ITSv2 DA	Non	Sens
Non	20	9
Sens	10	117
Inconclusive	3	9

DA Performance vs. LLNA Data (N=156)	ITSv2
Accuracy (%)	88%
Sensitivity (%)	93%
Specificity (%)	67%
Balanced Accuracy (%)	80%

Note: Accuracy is the correct classification rate, sensitivity is the true positive rate, specificity is the true negative rate, and balanced accuracy is the average of sensitivity and specificity with respect to LLNA data. Statistics reflect conclusive predictions only; inconclusive predictions are shown in grey. Additional performance characterisation is available in the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation.*

77. Due to the imbalanced nature of the reference data, the measure of specificity (based on 30 LLNA negative chemicals) is more uncertain than the measure of sensitivity (based on 126 LLNA positive chemicals).

Table 0.8. Potency categorisation performance of the ITSv2 DA in comparison to LLNA reference data, based on the UN GHS 1A/1B sub-categorisation

	LLNA		
ITSv2 DA	NC	1B	1A
NC	20	9	0
1B	10	54	10
1A	0	12	26
Inconclusive	3	10	2

71% correct classification overall

ITSv2 vs. LLNA reference data: Statistics based on the UN GHS 1A/1B sub-categorisation

Performance (N=141)	**NC (N=30)**	**1B (N=75)**	**1A (N=36)**
Correct classification (%)	67%	72%	72%
Underpredicted (%)	NA	12% (NC)	0% (NC); 28% (1B)
Overpredicted (%)	33% (1B); 0% (1A)	16% (1A)	NA

Note: Statistics reflect conclusive predictions only; inconclusive predictions are shown in grey. For more details on within-class performance (sensitivity, specificity, and balanced accuracy), please see **Section 5** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation.*

78. The designation of conclusive/inconclusive for the ITSv2 DA predictions is applied as described above in **Figure 3.1** and further detailed in **Annex 2**.

1.7.7. Predictive capacity of the ITSv1 DA vs Human Data

79. The predictive capacity of ITSv1 using Derek is reported based on data from the Human Predictive Patch Test (see **Tables 3.6-3.7**), curated as agreed upon by the EG DASS. The designation of high/low confidence for the ITSv1 DA predictions is further detailed in **Annex 2**. Performance statistics are reported for high confidence predictions as compared to human reference data, and inconclusive results are indicated. DA predictions for specific chemicals and further details are available in **Section 5** and **Annex 2** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation.*

Table 0.9. Hazard identification performance of the ITSv1 DA in comparison to Human reference data

	Human	
ITSv1 DA	Non	Sens
Non	4	4
Sens	5	51
Inconclusive	2	0

DA Performance vs. Human Data (N=64)	**ITSv1**
Accuracy (%)	86%
Sensitivity (%)	93%
Specificity (%)	44%
Balanced Accuracy (%)	69%

Note: Accuracy is the correct classification rate, sensitivity is the true positive rate, specificity is the true negative rate, and balanced accuracy is the average of sensitivity and specificity with respect to Human HPPT-based data. Statistics reflect conclusive predictions only; inconclusive predictions are shown in grey. Additional performance characterisation is available in the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation.*

80. Due to the imbalanced nature of the reference data, the measure of specificity (based on 9 Human negative chemicals) is more uncertain than the measure of sensitivity (based on 55 Human positive chemicals).

Table 0.10. Potency categorisation performance of the ITSv1 DA in comparison to Human reference data, based on the UN GHS 1A/1B sub-categorisation

	Human		
ITSv1 DA	NC	1B	1A
NC	4	4	0
1B	5	24	7
1A	0	3	13
Inconclusive	2	0	1

68% correct classification overall

ITSv1 vs. Human reference data: Statistics based on the UN GHS 1A/1B sub-categorisation

Performance (N=60)	**NC (N=9)**	**1B (N=31)**	**1A (N=20)**
Correct classification (%)	44%	77%	65%
Underpredicted (%)	NA	13% (NC)	0% (NC); 35% (1B)
Overpredicted (%)	56% (1B); 0% (1A)	10% (1A)	NA

Note: Statistics reflect conclusive predictions only; inconclusive predictions are shown in grey. For more details on within-class performance (sensitivity, specificity, and balanced accuracy), please see **Section 5** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation.*

81. The designation of conclusive/inconclusive for the ITSv1 DA predictions is applied as described above in **Figure 3.1** and further detailed in **Annex 2**.

82. Due to the imbalanced nature of the reference data and the small numbers of chemicals, the measures of accuracy are more uncertain for smaller classes, *e.g.* for NC chemicals.

1.7.8. Predictive capacity of the ITSv2 DA vs Human Data

83. The predictive capacity of ITSv2 using OECD QSAR Toolbox is reported based on data from the Human Predictive Patch Test (see **Tables 3.8-3.9**), curated as agreed upon by the EG DASS. The designation of high/low confidence for the ITSv2 DA predictions is further detailed in **Annex 2**. Performance statistics are reported for conclusive predictions as compared to human reference data, and inconclusive results are indicated. DA predictions for specific chemicals and further details are available in **Section 5** and **Annex 2** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation.*

Table 0.11. Hazard identification performance of the ITSv2 DA in comparison to Human reference data

	Human	
ITSv2 DA	Non	Sens
Non	4	3
Sens	5	50
Inconclusive	2	2

DA Performance vs. Human Data (N=62)	**ITSv2**
Accuracy (%)	87%
Sensitivity (%)	94%
Specificity (%)	44%
Balanced Accuracy (%)	69%

Note: Accuracy is the correct classification rate, sensitivity is the true positive rate, specificity is the true negative rate, and balanced accuracy is the average of sensitivity and specificity with respect to Human HPPT-based data. Statistics reflect conclusive predictions only; inconclusive predictions are shown in grey. Additional performance characterisation is available in the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation.*

84. Due to the imbalanced nature of the reference data, the measure of specificity (based on 9 Human negative chemicals) is more uncertain than the measure of sensitivity (based on 53 Human positive chemicals).

Table 0.12. Potency categorisation performance of the ITSv2 DA in comparison to Human reference data, based on the UN GHS 1A/1B sub-categorisation

	Human		
ITSv2 DA	NC	1B	1A
NC	4	3	0
1B	5	24	6
1A	0	3	12
Inconclusive	2	1	3

70% correct classification overall

ITSv2 vs. Human reference data: Statistics based on the UN GHS 1A/1B sub-categorisation

Performance (N=57)	NC (N=9)	1B (N=30)	1A (N=18)
Correct classification (%)	44%	80%	67%
Underpredicted (%)	NA	10% (NC)	0% (NC); 33% (1B)
Overpredicted (%)	56% (1B); 0% (1A)	10% (1A)	NA

Note: Statistics reflect conclusive predictions only; inconclusive predictions are shown in grey. For more details on within-class performance (sensitivity, specificity, and balanced accuracy), please see **Section 5** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation.*

85. The designation of conclusive/inconclusive for the ITSv2 DA predictions is applied as described above in **Figure 3.1** and further detailed in **Annex 2**.

86. Due to the imbalanced nature of the reference data and the small numbers of chemicals, the measures of accuracy are more uncertain for smaller classes, *e.g.* for NC chemicals.

1.7.9. Predictive capacity of the LLNA vs. Human Data

87. To provide a basis for comparison for the DA performance, the predictive capacity of the LLNA is reported based on data from the Human Predictive Patch Test (see **Tables 3.10-3.11**) curated as agreed upon by the EG DASS. Data for specific chemicals and further details are available in **Section 5** and **Annex 2** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation*.

Table 0.13. Hazard identification performance of the LLNA in comparison to Human reference data

	Human	
LLNA	Non	Sens
Non	2	3
Sens	7	44

LLNA Performance vs. Human Data (N=56)	LLNA
Accuracy (%)	82%
Sensitivity (%)	94%
Specificity (%)	22%
Balanced Accuracy (%)	58%

Note: Accuracy is the correct classification rate, sensitivity is the true positive rate, specificity is the true negative rate, and balanced accuracy is the average of sensitivity and specificity with respect to Human HPPT-based data. Additional performance characterisation is available in the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation.*

88. The hazard identification performance of the conclusive ITSv1 DA predictions vs. human data was 86% accuracy, 93% sensitivity, 44% specificity, and 69% balanced accuracy, comparable to and/or exceeding the performance of the LLNA in every measure.

89. The hazard identification performance of the conclusive ITSv2 DA predictions vs. human data was 87% accuracy, 94% sensitivity, 44% specificity, and 69% balanced accuracy, comparable to and/or exceeding the performance of the LLNA in every measure.

90. As previously noted, due to the imbalanced nature of the reference data, the measures of specificity are more uncertain than the measures of sensitivity.

Table 0.14. Potency categorisation performance of the LLNA in comparison to Human reference data, based on the UN GHS 1A/1B sub-categorisation

Additional performance characterisation is available in the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation.*

	Human		
LLNA	NC	1B	1A
NC	2	3	0
1B	6	17	7
1A	0	3	9

60% correct classification overall

LLNA vs. Human reference data: Statistics based on the UN GHS 1A/1B sub-categorisation

Performance (N=47)	**NC (N=8)**	**1B (N=23)**	**1A (N=16)**
Correct classification (%)	25%	74%	56%
Underpredicted (%)	NA	13% (NC)	0% (NC); 44% (1B)
Overpredicted (%)	75% (1B); 0% (1A)	13% (1A)	NA

Note: For more details on within-class performance (sensitivity, specificity, and balanced accuracy), please see **Section 5** of the *Supporting document to the Guideline (GL) on Defined Approaches (DAs) for Skin Sensitisation.*

91. The performance of the conclusive ITSv1 DA predictions vs. human data for potency sub-categorisation showed 68% correct classification overall, with accuracies of 44% for NC, 77% for 1B, and 65% for 1A, comparable to and/or exceeding the performance of the LLNA in every measure.

92. The performance of the conclusive ITSv2 DA predictions vs. human data for potency sub-categorisation showed 70% correct classification overall, with accuracies of 44% for NC, 80% for 1B, and 67% for 1A, comparable to and/or exceeding the performance of the LLNA in every measure.

93. As previously noted, due to the imbalanced nature of the reference data and the small numbers of chemicals, the measures of accuracy are more uncertain for smaller classes, *e.g.* for NC chemicals.

1.7.10. Proficiency chemicals

94. The ITS DA relies on a simple, rule-based data interpretation procedure and no expert judgment is required. Proficiency chemicals for the individual *in chemico* and *in vitro* information sources (KE1 and KE3) are defined in the respective guidelines (OECD TG 442C, 442E) (1, 2). The protocol details for the *in silico* information source options, Derek and OECD QSAR Toolbox, are included in **Annex 2** of this guideline. Proficiency has been demonstrated for Derek Nexus v6.1.0 and OECD QSAR Toolbox v4.5, and these are the software versions that are intended for use in the ITSv1 and ITSv2 DAs, respectively. Proficiency for the individual information sources demonstrates proficiency for the DA.

1.7.11. Reporting of the DA

95. The reporting of the ITS DA should follow the template described in OECD GD 255 (4), and should include at a minimum the following elements:

- Test chemical identification (*e.g.* chemical name, structural formula, composition, isomers, impurities including their quantities as available, CAS number, batch and lot number, and other relevant identifiers)
- Individual test reports for the individual tests performed per corresponding guideline (OECD TG 442C, 442E). Note that the chemical identity for each test report should match that above.
- Description of protocol used for *in silico* prediction **(Annex 2)** and outcome, *e.g.* reported via a QPRF (5).
- Outcome of the DA application (hazard identification and potency categorisation according to UN GHS categories, or inconclusive result)
- Any deviation from the ITS DA
- Conclusion

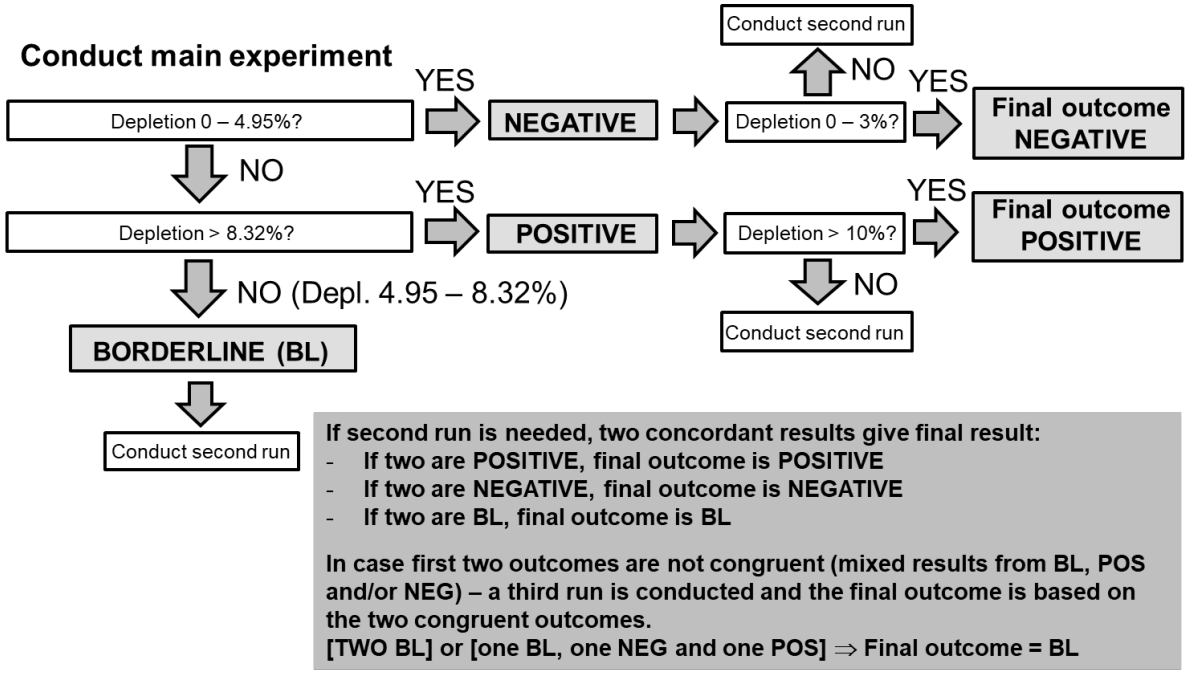
1.8. References

1. OECD (2020). OECD Guideline for the Testing of Chemicals No. 442C: *In chemico* Skin Sensitisation: Assays addressing the Adverse Outcome Pathway key even on covalent binding to proteins).*In chemico*. Paris, France: Organisation for Economic Cooperation and Development. Available at:(oecd-ilibrary.org).
2. OECD (2018). OECD Key event based test Guideline 442E: *In vitro* Skin Sensitisation Assays Addressing the Key Event on Activation of Dendritic Cells on the Adverse Outcome Pathway for Skin Sensitisation. Organisation for Economic Cooperation and Development, Paris. Available at: (oecd-ilibrary.org).
3. OECD (2016). Series on Testing & Assessment No. 256: 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 Sensitisation, Annex 1 and Annex 2. ENV/JM/HA(2016)29. Organisation for Economic Cooperation and Development, Paris. Available at: [https://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm].
4. OECD (2016). Series on Testing & Assessment No. 255: Guidance Document On The Reporting Of Defined Approaches To Be Used Within Integrated Approaches To Testing And Assessment. ENV/JM/HA(2016)28. Organisation for Economic Cooperation and Development, Paris. Available at: [https://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm].
5. ECHA (2008). see "CHAPTER R.6 – QSARS AND GROUPING OF CHEMICALS" *in* Guidance on Information Requirements and Chemical Safety Assessment. European Chemicals Agency *[*Guidance on Information Requirements and Chemical Safety Assessment - ECHA (europa.eu)]

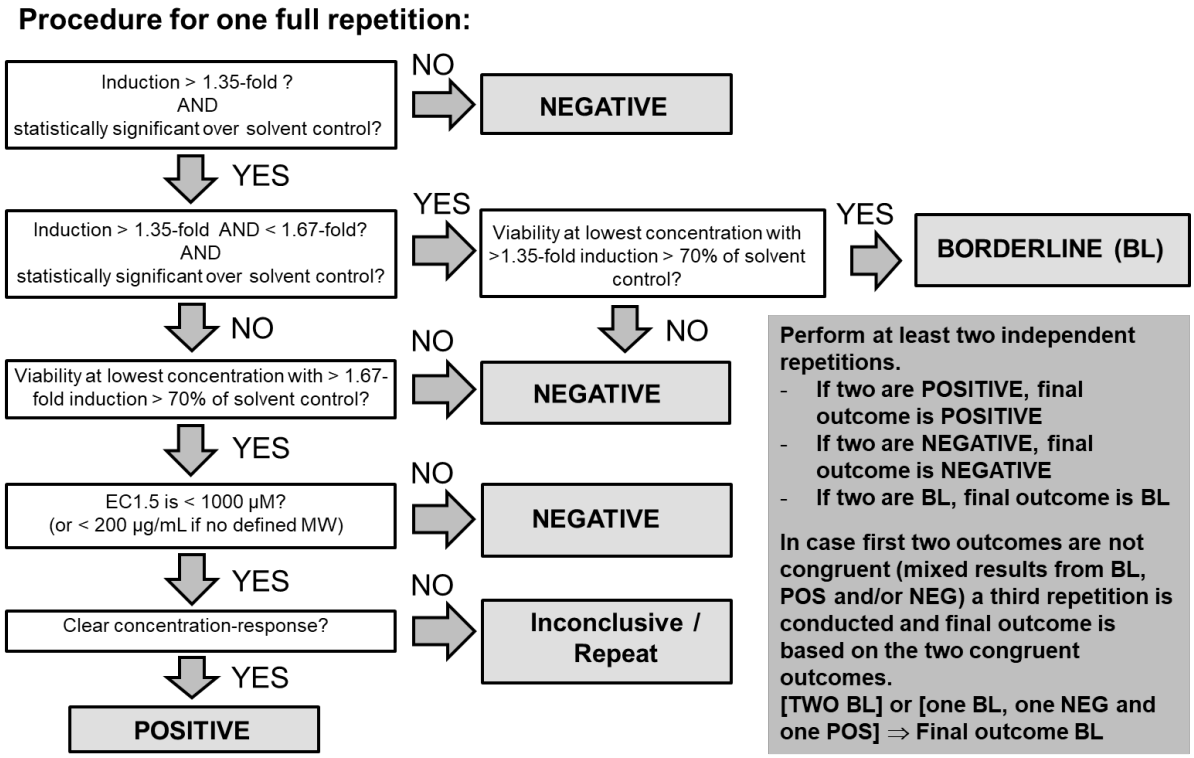
Annex 1: Prediction model for the individual *in chemico*/*in vitro* tests with multiple runs for use in 2o3 DA

96. The individual prediction models of h-CLAT and KeratinoSens™ require multiple runs (independent repetitions). An adaptation of the prediction model was used to determine borderline cases in the individual runs for the purpose of making predictions within the 2o3 DA. These adaptations (Figures 1.2. and 1.3) below should be used in these methods to come to the final conclusion of the individual tests.

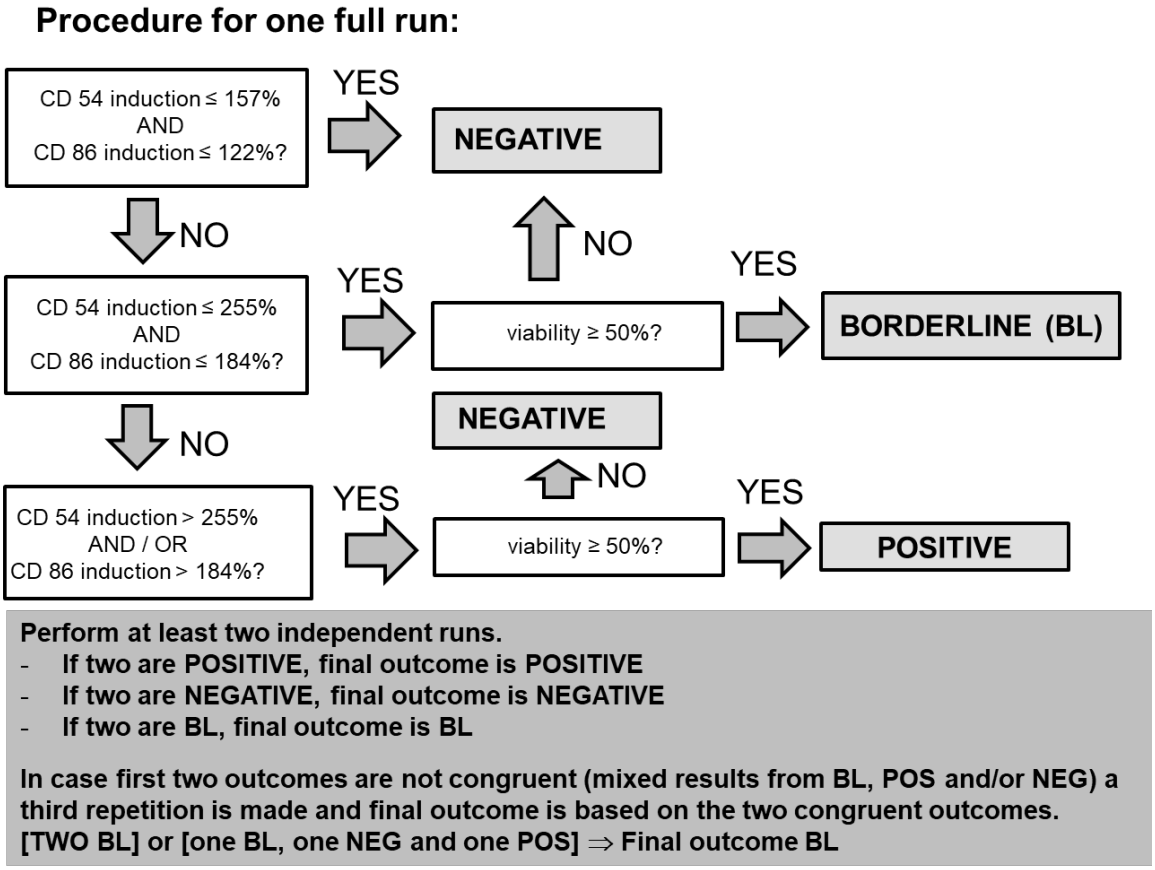
97. For the DPRA, repeated runs are required to be conducted if average depletion is within the range 3 - 10% (9 – 17% in case of Cysteine only depletion model is used). For this adaptation, the flowchart in Figure 1.1 is used to decide on run repetition and borderline assessment within the 2o3 DA.



Annex 1, Figure 1.1. Flow-chart of the DPRA prediction model (mean depletion) taking into borderline ranges and multiple runs conclude on borderline results within the 2o3 DA. The original threshold for a positive classification is 6.38%, and the statistically derived borderline range around this threshold is 4.95% - 8.32%. The same flowchart applies to the cysteine-only prediction model, whereby the following thresholds apply: 9% instead of 3%, >17 % instead of >10%, 10.56 % instead of 4.95% and > 18.47 % instead of >8.32%.



Annex 1, Figure 1.2. Flow-chart of the KeratinoSens™ prediction model taking into account borderline ranges and multiple runs to conclude on borderline results within the 2o3 DA. The original threshold for a positive classification is 1.5-fold induction, and the statistically derived borderline range around this threshold is 1.35 – 1.67-fold. Note: An independent run is referred to as 'repetition' in 442D, while it is called a 'run' in 442C and 442E; these nomenclatures do mean the same thing.



Annex 1, Figure 1.3. Flow-chart of the h-CLAT prediction model taking into account borderline ranges and multiple runs to conclude on borderline results within the 2o3 DA. The original threshold for a positive classification is 150% induction of CD86 with a statistically derived borderline range around this threshold of 122 – 184% and 200% induction of CD54 with a statistically derived borderline range around this threshold of 157 – 255%.

Annex 2: Defining the applicability domain and assessing confidence in DASS ITS predictions and protocols for generating *in silico* predictions

Introduction

98. As described in **Section 3.1** of the ***Guideline for Defined Approaches for Skin Sensitisation*** the ITS defined approaches (DAs) are based on three information sources: two *in chemico/in vitro* assays (DPRA; OECD TG 442C (OECD, 2015) and h-CLAT; OECD TG 442E (OECD, 2018)) and one *in silico* tool (prediction from either Derek Nexus (ITSv1) or OECD QSAR Toolbox (ITSv2) (referred to hereafter as *in silico*)). For each information source a score is given depending on the outcome of the individual assay and/or prediction, that is then summed to obtain the DA prediction.

Applicability domain of the individual information sources

In chemico/in vitro information source (DPRA and h-CLAT)

99. A test chemical is considered to be within the *in chemico*/*in vitro* domain (i.e. applicable) of DPRA and/or h-CLAT if it can be tested according to the individual protocols, taking into account the technical and chemical type limitations of each assay (as defined in the respective test guidelines OECD TG 442C and OECD TG 442E (OECD, 2015, 2018)). The *in chemico*/*in vitro* results are considered applicable, in case there are no technical or chemical space specific limitations and no reason why the results obtained from the assay cannot be considered.

In silico information source

100. The ITS DAs use *in silico* information sources that are based on chemical structures. These *in silico* sources rely on molecular representation of the chemicals: input usually by drawing the chemical structure, or by entering the Simplified Molecular-Input Line-Entry System (SMILES) or the IUPAC International Chemical Identifier (InChi). As a single chemical can be represented by several CAS or EC numbers (due to differences in composition e.g. stereochemical differences, present as varied salt forms, present as the main componenet in a mixture), it is important to specify the exact structure if possible. Resources such as the US EPA CompTox Chemicals Dashboard (https://comptox.epa.gov/dashboard) or NIH PubChem (https://pubchem.ncbi.nlm.nih.gov/) may be useful in mapping chemical names or structures to SMILES or InChi format. Available guidance can be consulted regarding minimum purity level of substances used in *in silico* predictions based on molecular structure.[34]

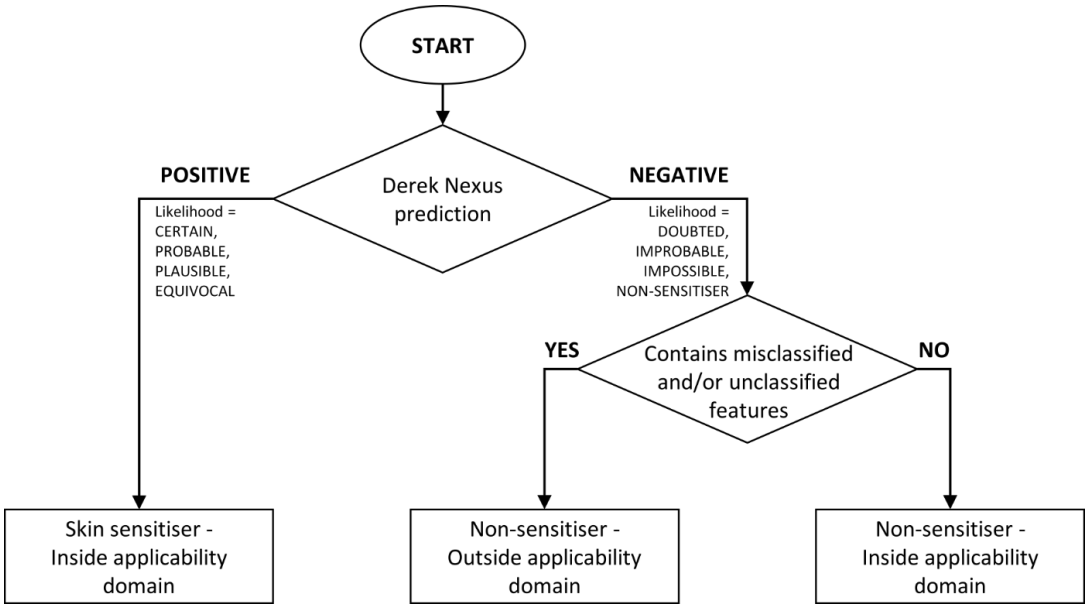
[3] OECD (2017), *Guidance on Grouping of Chemicals, Second Edition*, OECD Series on Testing and Assessment, No. 194, OECD Publishing, Paris, https://doi.org/10.1787/9789264274679-en.

[4] ECHA (2008) CHAPTER R.6 – QSARS AND GROUPING OF CHEMICALS *in* Guidance on Information Requirements and Chemical Safety Assessment. European Chemicals Agency [Guidance on Information Requirements and Chemical Safety Assessment - ECHA (europa.eu)

Derek Nexus (ITSv1)

101. Skin sensitisation predictions from Derek Nexus v6.1.0 are used in ITSv1. The protocol for running Derek Nexus (Derek) predictions is defined in **Appendix 1** of this document. All positive predictions (likelilhood = certain, probable, plausible or equivocal) are considered to be inside the applicability domain. Negative predictions (likelihood = doubted, improbable, impossible or non-sensitiser) are also considered to be in the applicability domain unless they contain misclassified and/or unclassified features. A prediction of non-sensitiser with misclassified features indicates the presence of a fragment that has been observed exclusively in known sensitisers which Derek fails to alert for. A prediction of non-sensitiser with unclassified features indicates the presence of a fragment that has not been observed in publicly available data (although Derek may have seen this in proprietary data) (Chilton et al., 2018). Usually expert review is recommended for predictions containing these features but as a fixed data interpretation procedure, required in a DA, does not permit expert review these are best considered as out of domain for use in ITSv1 (**Figure A2.1**).

Figure A2.1. Applicability domain for Derek Nexus skin sensitisation predictions used in ITSv1.



QSAR Toolbox (ITSv2)

102. Skin sensitisation predictions from the QSAR Toolbox automated workflow "Skin sensitisation for defined approaches" (Yordanova et al., 2019) are used in ITS v2. The protocol for running QSAR Toolbox predictions is defined in **Appendix 2** of this document.

103. The calculation of the applicability domain of the predictions is automatically provided by Toolbox when running DASS AW predictions and consists of three layers: structural, parametric and mechanistic. The applicability domain layers considered for each individual prediction depend on the type and outcome of the prediction, as summarised in Table A2.1. A detailed description of the three layers and the rationale for their selection is

explained in **Appendix 3** of this document. Toolbox results within applicability domain are considered as applicable in the DA.

Table A2.1. Applicability domain layers for the QSAR Toolbox automated workflow "Skin sensitisation for defined approaches" predictions.

Toolbox DASS AW outcome		Applicability domain layer		
		Structural	Parametric	Mechanistic
Positive	Read-across	Not considered	Not considered	**Considered**
	Profiling	Not considered	Not considered	**Met by definition**
Negative	Read-across	Not considered	Not considered	**Considered**
	Profiling	**Considered**	**Considered**	**Met by definition**

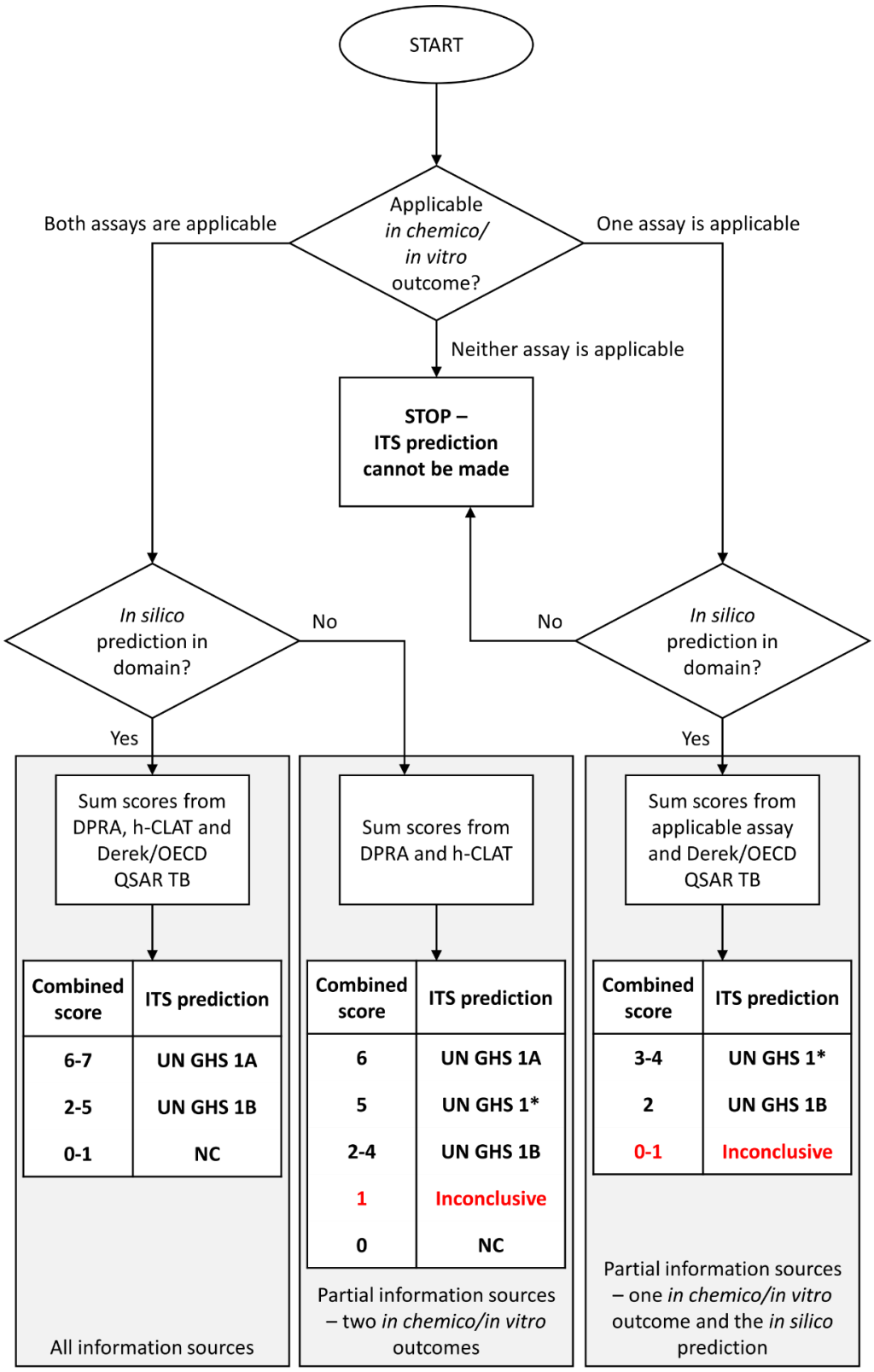
Confidence in ITS predictions

104. The applicability domain of the individual information sources used in the ITS DA are assessed and this determines whether the ITS predictions can be considered conclusive (i.e. high confidence) or inconclusive (i.e. low confidence) for hazard identification and/or potency.

How to apply the data interpretation procedure (DIP) for the ITS

105. The ITS was originally developed to use three information sources (DPRA, h-CLAT, and an *in silico* tool (Derek Nexus or OECD QSAR Toolbox)). Where all three information sources are applicable, a conclusive ITS prediction can be made. In some cases, a conclusive ITS prediction can be made, if there are two information sources with applicable results (**Figure A2.2**).

Figure A2.2. Workflow for data interpretation procedure for the ITS.



***Conclusive for hazard, inconclusive for potency**

106. Depending on the applicability of the individual information sources, three different scenarios for the ITS DA are possible (see Figure A2.2 and Table A2.2). In Scenario 1, all three information sources are applicable. In Scenarios 2 and 3, only two information sources are applicable. Details are provided below:

107. Scenario 1: all of the information sources i.e. *in chemico/in vitro* outcomes are applicable and can be considered (as prescribed in each individual assay) and the *in silico* prediction is in domain. The obtained ITS DA prediction is conclusive and of high confidence

108. Scenario 2: *in silico* prediction out of domain, however *in chemico*/*in vitro* methods are in domain and provide conclusive predictions (i.e. *in chemico*/*in vitro* methods are applicable).

- Combined DA score of 0, 2, 3, 4 or 6, *in silico* prediction out of *in silico* domain: DA conclusion is possible based on the two *in chemico/in vitro* outcomes. Conclusive prediction as the *in silico* prediction would not lead to a different DA prediction.
- Combined DA score of 5, *in silico* prediction out of *in silico* domain: DA conclusion possible for hazard identification (conclusive positive DA prediction for hazard identification). DA conclusion not possible for potency (inconclusive DA prediction for potency).
- Combined DA score of 1, *in silico* prediction out of *in silico* domain: DA conclusion not possible. Inconclusive DA prediction for hazard identification and potency.

109. Scenario 3: one *in chemico/in vitro* method out of domain or the result of that method cannot be considered (inapplicable):

- Combined DA score of 2 based on one *in chemico/in vitro* and *in silico* prediction: DA conclusion possible. Conclusive DA prediction as UN GHS 1B, as the outcome of the other *in chemico/in vitro* method would not to a different DA prediction.
- Combined DA score of 3 or 4, based on one *in chemico/in vitro* and *in silico* prediction: DA conclusion possible for hazard identification (conclusive positive DA prediction for hazard identification). DA conclusion not possible for potency (inconclusive DA prediction for potency).
- Combined DA score of 0 or 1, one *in chemico/in vitro* and *in silico* prediction: DA conclusion not possible. Inconclusive prediction for hazard identification and potency.

Table A2.2. Applicability domain and confidence of the ITS.

Scenario	Combined score[5]	ITS prediction	Confidence	DA prediction including confidence considerations
1	0-1	NC	High	Conclusive prediction Not Classified (NC).
	2-5	UN GHS 1B	High	Conclusive prediction UN GHS 1B.
	6-7	UN GHS 1A	High	Conclusive prediction UN GHS 1A.
2	0	NC	High	Conclusive prediction NC.
	1	Inconclusive	Low	Inconclusive prediction whether positive or negative.
	2-4	UN GHS 1B	High	Conclusive prediction UN GHS 1B.
	5	UN GHS 1	High	Conclusive positive prediction for hazard identification.
			Low	Inconclusive prediction for potency.
	6	UN GHS 1A	High	Conclusive prediction UN GHS 1A.
3	0-1	Inconclusive	Low	Inconclusive prediction whether positive or negative.
	2	UN GHS 1B	High	Conclusive prediction UN GHS 1B.
	3-4	UN GHS 1	High	Conclusive positive prediction for hazard identification.
			Low	Inconclusive prediction for potency.

References

Chilton, M. L., Macmillan, D. S., Steger-Hartmann, T., Hillegass, J., Bellion, P., Vuorinen, A., Etter, S., Smith, B. P. C., White, A., Sterchele, P., De Smedt, A., Glogovac, M., Glowienke, S., O'Brien, D., & Parakhia, R. (2018). Making reliable negative predictions of human skin sensitisation using an in silico fragmentation approach. *Regulatory Toxicology and Pharmacology*, *95*, 227–235. https://doi.org/10.1016/j.yrtph.2018.03.015

OECD. (2015). *Test No. 442C: In Chemico Skin Sensitisation: Direct Peptide Reactivity Assay (DPRA).* OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264229709-en

[5]Total scores calculated only from information sources that are applicable/in domain.

OECD. (2018). *Key Event Based Test Guideline 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. OECD Publishing, Paris.

Yordanova, D., Schultz, T. W., Kuseva, C., Tankova, K., Ivanova, H., Dermen, I., Pavlov, T., Temelkov, S., Chapkanov, A., Georgiev, M., Gissi, A., Sobanski, T., & Mekenyan, O. G. (2019). Automated and standardized workflows in the OECD QSAR Toolbox. *Computational Toxicology*, *10*, 89–104. https://doi.org/10.1016/j.comtox.2019.01.006

Appendix 1: Protocol for Derek Nexus predictions

110. The following protocol may be used to generate predictions for skin sensitisation hazard using Derek Nexus v.6.1.0 with Derek Knowledge Base (KB) 2020 1.0 to be used as the *in silico* information source for the ITSv1 defined approach.

Protocol for generating predictions for skin sensitisation hazard using Derek Nexus v.6.1.0 with Derek KB 2020 1.0

Single chemical

1. Open Nexus
2. Input structure using one of the following options:
    a. Input structure manually by drawing on the canvas
    b. Go to File>Open Structure(s) to input a single structure from a file (.mol, .sdf, .smi, .csv, .cdx (file list not exhaustive))
    c. Go to File>Type Chemistry to enter or paste SMILES, InChi or MOL file
    d. Go to File>New Structure to input structure by drawing a structure
3. Set up prediction
    a. Go to Prediction>Derek Prediction>Derek Prediction Setup
4. Apply processing constraints
    a. Knowledge Bases
        i. For Nexus v6.1.0, ensure Derek KB 2020 1.0 is selected
        ii. For newer releases, use the default Derek KB supplied
    b. Perception
        i. Ensure ‘Perceive tautomers’ and Perceive mixtures’ are selected
        ii. Ensure ‘Match alerts without rules’ is unselected
    c. Species
        i. Select ‘mammal’
    d. Endpoints
        i. Click ‘Deselect all’ then expand ‘Skin sensitisation (ALL)’ to view ‘Photoallergenicity’ and ‘Skin sensitisation’. Select ‘Skin sensitisation’
    e. Structure properties
        i. Ensure the ‘Overwrite’ box(es) for logP, logKp, and average molecular mass are unselected to use the values calculated by Derek Nexus, otherwise, check the ‘Overwrite’ box(es) to input own values.
5. Generate prediction
    a. Click ‘Start Prediction’

b. If an alert is fired: Knowledge base, endpoint, species, reasoning level, alert fired, EC3 prediction (if applicable), and example matched (if applicable) are shown in the prediction navigator.
    i. Click the likelihood (certain, probable, plausible, equivocal) to view the reasoning rules leading to the likelihood level.
    ii. Click the Alert in the prediction navigator to view alert match(es), description image, comments, validation comments, endpoint, references, patterns, and examples associated with the alert.
c. If no alert is fired, a negative prediction is generated: Knowledge base, endpoint, species and negative prediction reasoning (non-sensitiser) and negative prediction overview (absence or presence of misclassified and/or unclassified features) are shown in the prediction navigator.
    i. Click the negative prediction overview ('No misclassified or unclassified features', 'Contains misclassified/unclassified features') to view information about the negative prediction. Similar nearest neighbours are available to view for misclassified features.
d. Use the Derek likelihood to classify each compound as positive or negative (alert fired with certain, probable, plausible, or equivocal is classified as positive, alert fired with doubted, improbable, impossible, or a negative prediction of non-sensitiser with no misclassified or unclassified features is classified as negative).
    i. Negative predictions of non-sensitiser with misclassified and/or unclassified features are of lower confidence and are not used in ITSv1.
    ii. In cases where more than one alert is fired or structures in a mixture generate different likelihoods, the most conservative classification is applied (positive > negative).
    iii. A positive outcome from Derek is scored as 1 in the ITSv1 and a negative outcome is scored as 0.

Multiple chemicals

1. Open Nexus
2. Input structures
    a. Go to File>Open Structure(s) to input a file containing multiple structures (.mol, .sdf, .smi, .csv, .cdx (file list not exhaustive))
    b. Select the fields from the file which will be mapped to structure properties used during the prediction (Name, Average Molecular Mass, LogP, LogKp). If left unchanged then the values set by Derek will be used.
3. Set up batch prediction
    a. Go to Prediction>Derek Prediction>Derek Batch Setup
4. Apply processing constraints
    a. Knowledge Bases
        i. For Nexus v6.1.0, ensure Derek KB 2020 1.0 is selected
        ii. For newer releases, use the default Derek KB supplied
    b. Perception
        i. Ensure 'Perceive tautomers' and Perceive mixtures' are selected

ii. Ensure ‘Match alerts without rules’ is unselected

c. Species
   i. Select ‘mammal’

d. Endpoints
   i. Click ‘Deselect all’ then expand ‘Skin sensitisation (ALL)’ to view ‘Photoallergenicity’ and ‘Skin sensitisation’. Select ‘Skin sensitisation’

e. Report configuration
   i. Directory - Leave as default directory or map to preferred location.
   ii. Pick type - Select report for batch (left side icon)
   iii. Pick format - Select desired file type (e.g. Excel)
   iv. Pick design - Select desired design (e.g. Tabular Report)
   v. Filename - input desired filename

f. Report display options
   i. Ensure ‘Show predictions of at least impossible’ is selected
   ii. Select ‘Show Negative Predictions’
   iii. Select ‘Filter All Nearest Neighbours by Misclassified Features’
   iv. Select ‘Show Open Likelihood’
   v. Select ‘Show Rapid Prototypes’

5. Generate batch prediction
   a. Click ‘Start Batch Prediction’
      i. Once the batch prediction is finished, select the ‘Open Report Directory’ when prompted
   b. Use the Derek likelihood to classify each compound as positive or negative (alert fired with certain, probable, plausible, or equivocal is classified as positive, alert fired with doubted, improbable, impossible, or a negative prediction of non-sensitiser with no misclassified or unclassified features is classified as negative).
      i. Negative predictions of non-sensitiser with misclassified and/or unclassified features are of lower confidence and are not used in ITSv1.
      ii. In cases where more than one alert is fired or structures in a mixture generate different likelihoods, the most conservative classification is applied (positive > negative).
   c. A positive outcome from Derek is scored as 1 in the ITSv1 and a negative outcome is scored as 0.

Appendix 2: Protocol for OECD QSAR Toolbox predictions

111. The following protocol may be used to generate predictions for skin sensitisation hazard using OECD QSAR Toolbox v.4.5 with the automated workflow for defined approaches for skin sensitisation (DASS AW) to be used as the in silico information source for the ITSv2 defined approach.

Protocol for generating predictions for skin sensitisation hazard using DASS AW in Toolbox 4.5.

Step 1: Input the chemical in the "Input module". SMILES is the preferred way to input the structure. (If other identifiers such as the CAS number are used as input, the Toolbox will assign the SMILES based on its internal database. In this case, the user needs to make sure that Toolbox identifies and consequently uses for the prediction the correct structure.)

Step 2: Go to the "Data gap filling module" and click on "Automated" button. Select "EC3 from LLNA or Skin sensitization from GPMT assays for defined approaches" and click OK. The scheme with the implemented logic will be shown.

Step 3: Click the Run button -  or press F5 key of the keyboard and confirm with "Yes". The workflow will run automatically.

Step 4: If a substance is predicted "positive" or "negative" as a result of read-across, the prediction will appear on the data matrix with "R" in front of the result (e.g. "R: Negative). If a substance is predicted "positive" or "negative" as a result of profiling, then the result will appear next to the name of the customized profiler "Skin sensitization for DASS".

Step 5: Affiliation of the substance to the domain of the automated workflow for DASS will be automatically determined and presented.

Appendix 3: Information on applicability domain for OECD QSAR Toolbox

Technical aspects

112. The Toolbox prediction used by DA ITS v.2 is calculated using the DASS automated workflow (DASS AW) included in OECD QSAR Toolbox v.4.5. The workflow also includes the automatic calculation of the applicability domain of Derek skindescribed below.

Calculation of the in silico domain of Toolbox

113. Applicability domain of the QSAR Toolbox Skin sensitisation predictions for use in the ITS defined approach approaches automated workflow (DASS AW) is defined by based on the training set substances of the same automated workflow. The training set (TS) consists of 2268 substances having LLNA and/or GPMT skin sensitisation experimental data[6](the full list of substances can be consulted in the QSAR Toolbox). The TS substances are part of the following OECD QSAR Toolbox databases:

- Skin sensitisation;
- REACH Skin sensitisation (normalized) databases.

114. Based on the correctly predicted training set substances, three layers of applicability domain are automatically calculated by the Toolbox: 1) parametric; 2) structural and 3) mechanistic layers. Depending on the Toolbox prediction approach (read-across or profiling predictionss) and prediction outcomes (positive or negative), one or more of these layers are taken into account to establish the overall Toolbox domain of the specific prediction.

115. The applicability domain layers considered for different types of Toolbox predictions are summarised in the table here:

Toolbox DASS AW outcome		Applicability domain layer		
		Structural	Parametric	Mechanistic
Positive	Read-across	Not considered	Not considered	**Considered**
	Profiling	Not considered	Not considered	**Met by definition**
Negative	Read-across	Not considered	Not considered	**Considered**
	Profiling	**Considered**	**Considered**	**Met by definition**

116. Explanation and rationale for the use of different domain layers:

1. Positive predictions (both by read-across and profiling): the presence of an alert (which is the requirement for positive Toolbox prediction to be considered within in the mechanistic domain) is sufficient to consider the prediction to be within the Toolbox domain. Substances triggering an alert are considered as in domain because they contain the toxicophore that has been observed experimentally in skin sensitisers. No further checks are needed in this context to consider the prediction within the Toolbox *in silico* domain.

[6] In case of multiple data points for one substance, the most conservative scenario is taken into account.

2. Negative predictions by read-across: the structural and parametric domains are not taken into account because the Toolbox has already ensured some level of similarity with other substances in its training set that met the requirements to be selected as suitable analogues for read-across (these requirements are explained in detail in the DASS AW description).
3. Negative prediction by profiling predictions: all domain layers are taken into account to ensure the highest possible reliability level for the Toolbox prediction. Stricter requirements are needed mainly for two reasons: 1. lack of alerts is not equal to proof of lack of sensitisation potential and 2. to apply a cautious approach since acceptance of negative predictions may lower the human health protection level risk in case of a false negative predictions.

Calculation of applicability domain layers

1. Parametric layer

Four physico-chemical parameters of the substances are taken into consideration: logKow, molecular weight, vapour pressure and water solubility[7]. The ranges of variation for the selected parameters are defined based on the training set substances that are correctly predicted by the DASS AW.

A substance is considered within the parametric domain of the DASS AW if its physico-chemical parameter values as calculated by the QSAR Toolbox fall into the ranges of variation given in the table below. It is noted that the ranges include parametric values calculated using EPISuite models implemented in Toolbox that in some cases are wider than that covered by existing test methods.

Physico-chemical parameter	Calculated Parameter range
logKow	-9.66 ÷ 18.6
Molecular weight	16 Da ÷ 2290 Da
Vapour pressure*	0 Pa ÷ 3.45 x 10^7 Pa
Water solubility	2.48 x 10^{-15} mg/L ÷ 1.00 x 10^6 mg/L

*EPIWIN Vapor Pressure (Antoine method) is used for calculation

2. Structural layer

The structural layer is defined based on the atom centred fragments (ACF) derived from the structural characteristics of the TS substances that are correctly predicted[8] by the DASS AW.

The ACF are defined according to the following Toolbox default values for ACF:

- Any atom distance = 1

[7] QSAR Toolbox is used for the calculation of the physico-chemical properties.

[8] All ACF that are extracted from the correctly predicted TS test chemicals "good space". The "bad space" is formed from the ACF present in the incorrectly predicted test chemicals. The default QSAR Toolbox settings for ACF are used. Supplementary file with the ACF forming the good and the bad space are available.

- Heteroatom distance = 1
- Extract C (sp^3) fragments = YES
- Include whole aromatic rings = NO

For each substance, the following values are calculated:

- % Correct fragments: percentage of ACF occurring in correctly predicted structures in the training set
- % incorrect fragments: percentage of ACF occurring in incorrectly predicted structures in the training set
- % unknown fragments: percentage of ACF not occurring in the training set.

A substance is considered within the structural domain of the DASS AW if 100% of its ACF belong to the correct fragments.

3. Mechanistic layer

The predicted capability of a substance to interact with the skin proteins without and after (a)biotic activation is taken into consideration. The Toolbox endpoint-specific profiler *Protein binding for skin sensitization by OASIS* and two metabolic simulators – *Autoxidation simulator* and *Skin metabolism simulator* are used to predict such interaction.

A positive prediction is considered within the mechanistic domain if the substance triggers "*Protein binding for skin sensitization by OASIS*" alerts without or after (a)biotic activation.

A negative prediction is considered within the mechanistic domain if the substance does not permit expert review these are best considered as out of domain for use in the ITS "*trigger Protein binding for skin sensitization by OASIS*" without or after (a)biotic activation.

117. Note that predictions obtained by profiling results will meet the mechanistic layer requirements by definition because positive Toolbox predictions by profiler are triggered exactly by the presence of alert. If the test chemical cannot be tested or the outcome/prediction cannot be considered in at least two of the information sources (*in chemico/in vitro and/or in silico)* then the DA cannot be applied.