**Committee of Experts on the Transport of Dangerous Goods and on the Globally Harmonized System of Classification and Labelling of Chemicals**

**Sub-Committee of Experts on the Globally Harmonized System of Classification and Labelling of Chemicals** **9 November 2023**

**Forty-fifth session**
Geneva, 6-8 December 2023
Item 2 (c) of the provisional agenda
**Work on the Globally Harmonized System of Classification and Labelling of Chemicals: Use of non-animal testing methods for classification of health and environmental hazards**

# Use of non-animal testing methods for classification of health and environmental hazards: progress report

**Transmitted by the experts from the United Kingdom and the Netherlands on behalf of the informal working group on the use of non-animal testing methods for classification of health and environmental hazards**

## Introduction

1. This informal document provides an update on the work performed by the Informal Working Group on "Use of non-animal testing methods for classification of health and environmental hazards" since the last update provided to the Sub-Committee in July 2023 (see informal document INF.13, 44th session)

## Background

2. At the forty-third session, the Sub-Committee agreed to keep the work on the use of non-animal testing methods for classification of health and environmental hazard classes on its programme of work for the 2023-2024 biennium.[1] Updates on the progress of the group's work are provided to the Sub-Committee at each session of the biennium.

3. The informal working group presently has approximately 60 members, reflecting the importance of, and interest in, this work. Its membership includes experts with specialised knowledge of test methods and their application to classification, and experts on national legislation that implements GHS. Discussions are often lively and detailed, but overall are propelled by a strong desire to make progress on the informal working group's mandate and ensure that non-animal testing methods are consistently incorporated in the GHS in a way that reflects their growing importance and scientific relevance, whilst recognising their limitations.

## Status report

4. Since the last update to the GHS Sub-Committee in July 2023, the informal working group has continued to work hard on the further revision of chapter 3.4 for skin sensitization in relation to mixtures for the inclusion of non-animal testing

[1] See informal document UN-SCEGHS-43-INF.16 and report SG/ST/AC.10/C.4/2022/86, both from the forty-third session

methods. This has been achieved via correspondence and virtual meetings (31 August 2023; and 17 October 2023), with a further webinar planned for 12 December 2023. After each meeting the Netherlands and the United Kingdom, the joint leads, with the assistance of the Joint Research Centre (JRC) have revised the draft text of chapter 3.4, drafted meeting notes and prepared papers, presentations and surveys on specific topics to take forward the discussions, taking into account written comments and information on specific topics provided by the members of the group.

5. The working draft (version 6.1) is provided in the annex to this document. It incorporates changes agreed by members during the October group meeting and amendments made by the project leads after the meeting, for instance, to address actions placed on the leads or make minor editorial changes. Version 6 of the working draft was considered by the working group via correspondence in October and November 2023 in preparation for further discussion on the document during the December working group meeting.

6. The group's key proposed changes (still under discussion) are:

(a) Replace the current text of 3.4.3.1 "Classification of mixtures when data are available for the complete mixture" to include non-animal test methods.

(b) Insert a new heading in 3.4.5.3 ("3.4.5.3.1 *Guidance on substances – skin sensitization"*) to clarify that the existing guidance pertains only to skin sensitization of substances, which also requires renumbering of the existing paragraphs. This was considered necessary given the group's current work on developing guidance of skin sensitization on mixtures and also that, in time, similar guidance sections could be added for the respiratory sensitization hazard class within chapter 3.4.

(c) Insert a new section "3.4.5.3.2 *Guidance on mixtures – skin sensitization".*

(d) Renumbering of the guidance section on weight of evidence (currently 3.4.5.3.7) to 3.4.5.3.3 to follow after the guidance on mixtures as it relates to both substances and mixtures for skin sensitization.

7. This working draft is presented so the Sub-Committee can see what has been achieved so far, and steer the working group as it considers appropriate.

# On-going work

8. The informal working group will continue its work on the revision of chapter 3.4 for skin sensitization in relation to mixtures to include non-animal testing methods during its next webinar meeting in early December 2023, followed by further webinar meetings in early 2024. There is tentative hope that it will be possible to finalise the revised chapter 3.4 in time for adoption by the Sub-Committee during the July 2024 session.

9. The Sub-Committee is invited to note the progress of the revision of chapter 3.4 (a snapshot is provided in the annex to this document) and the issues outlined in this informal document.

# Annex

# Working draft of chapter 3.4 (Version 6.1; 6 November 2023)

Black text (including deleted text which is shown in strikethrough) is from GHS chapter 3.4. (Rev.10).

Blue text is new in this chapter.

Red text requires further discussion.

**Note**: there are a number of sections within the chapter that have not been amended. To focus the readers review on what has been amended in the 'snapshot' below of the chapter, these sections are indicated by their section number and headings, together with the text *"- remains unchanged"*.

## "CHAPTER 3.4

## RESPIRATORY OR SKIN SENSITIZATION

**3.4.1 Definitions and general considerations** - *remains unchanged*

**3.4.2 Classification criteria for substances** - *remains unchanged*

**3.4.3 Classification criteria for mixtures**

**3.4.3.1 *Classification of mixtures when data are available for the complete mixture***

~~When reliable and good quality evidence from human experience or appropriate studies in experimental animals, as described in the criteria for substances, is available for the mixture, then the mixture can be classified by weight of evidence assessment of these data. Care should be exercised in evaluating data on mixtures that the dose used does not render the results inconclusive. (For special labelling required by some competent authorities, see the note to table 3.4.5 and 3.4.4.2).~~

3.4.3.1.1 In general, the mixture should be classified using the criteria for substances taking into account the tiered approach to evaluate data for this hazard class (as illustrated in Figure 3.4.1) and 3.4.3.1.2 below. If classification is not possible using the tiered approach, then the approach described in 3.4.3.2 (bridging principles), or, if that is not applicable, 3.4.3.3 (classification based on ingredients) should be followed. (For supplemental labelling required by some competent authorities, see the note to Table 3.4.5 and 3.4.4.2.)

3.4.3.1.2 Care should be exercised in evaluating data on mixtures that the dose used does not render the results inconclusive and that the test methods used to generate such results are appropriate for predicting the skin sensitizing properties of the mixture (see 3.4.5.3.2). Further, for both standard test methods (in vivo, in chemico, in vitro) and defined approaches, data can only be used for classification when all ingredients fall within their applicability domain. Specific limitations regarding applicability domains are described in the respective test methods and defined approaches and should be taken into consideration as well as any further information on such limitations from the published literature. A competent authority may decide which in chemico/in vitro test method or defined approach may be accepted for mixtures, (see 3.4.5.3.2.4 and 3.4.5.3.2.5). A more detailed overview of factors to consider in the classification of mixtures can be found in guidance section 3.4.5.3.2 and the test methods.

**3.4.3.2 *Classification of mixtures when data are not available for the complete mixture: bridging principles*** – *remains unchanged*

**3.4.3.3 *Classification of mixtures when data are available for all ingredients or only for some ingredients of the mixture*** – *remains unchanged*

**3.4.4 Hazard communication**- *remains unchanged*

**3.4.5 Decision logic and guidance**

**3.4.5.1 Decision logic 3.4.1 for respiratory sensitization** - *remains unchanged*

**3.4.5.2 Decision logic 3.4.2 for skin sensitization** - *remains unchanged*

**3.4.5.3 *Background guidance***

3.4.5.3.1 *Guidance on substances – skin sensitization*

3.4.5.3.1.1 *Relevant guidance documents*

Mechanistic information on the process of skin sensitization is available in the OECD document on the Adverse Outcome Pathway for skin sensitization (see OECD (2014)). This information can be helpful in understanding the value of the individual *in chemico* and in vitro methods compared to the in vivo methods.

3.4.5.3.1.2 *Guidance on the use of human data*

3.4.5.3.1.2.1 The classification of a substance can be based on human evidence generated from a variety of sources. These sources include human predictive patch testing, epidemiological studies, case studies, case reports or histories, diagnostic patch testing and medical surveillance reports, and poison control centre information. This data may have been generated for consumers, workers, or the general population. When considering human evidence, consideration should be given to the size, exposure level, and exposure frequency of the exposed population. Guidance for evaluating human evidence and the criteria in 3.4.2.2.2 is provided by some competent authorities (e.g. ECHA Guidance on the Application of the CLP Criteria, 2017).

3.4.5.3.1.2.2 Positive data from predictive patch testing (HRIPT or HMT) conducted through human experimental and clinical studies, showing allergic contact dermatitis caused by the test substance can be used to classify substances for skin sensitization These studies are generally conducted in controlled clinical settings and in general the larger the population size, the more reliable the study outcome is. Criteria for evaluating this data are provided in paragraphs 3.4.2.2.2.1 and 3.4.2.2.2.

3.4.5.3.1.2.3 Positive data from well-run epidemiological studies (in accordance with WHO CIOMS guidelines, 2009) can be used for classifying substances for skin sensitization. Some examples of epidemiological studies may include case control studies, cohort studies, cross-sectional studies, or longitudinal studies. These studies should have large sample sizes with well-documented exposures to a substance.

3.4.5.3.1.2.4 A specific type of epidemiological study (such as randomized control studies or trials) may include information from diagnostic patch testing. Diagnostic patch testing is considered by some competent authorities to be the gold standard in diagnosing contact allergy in dermatitis patients (Johansen et al, 2015). Importantly, due consideration needs to be given to the appropriate selection of vehicle, substance and patch test concentrations for the purpose of not causing false negatives, false positives, irritant reactions or inducing contact allergy (skin sensitization). Positive data from experimental/clinical/diagnostic studies in humans and/or well-documented episodes of allergic contact dermatitis may be used to classify substances for skin sensitization, when it can be assumed with sufficient likelihood that the tested substance was indeed the most likely cause for induction of sensitization. Therefore, it should be established that there is at least a general likelihood that the respective patient(s) had been previously exposed to the substance. On the other hand, negative results from such tests are not sufficient to prove that the test substance should not be classified as a skin sensitizer.

3.4.5.3.1.2.5 Human data not generated in controlled experiments with volunteers for the purpose of hazard classification (e.g. case studies, case reports and case histories, and poison control centre information) can be used with caution. Consideration should be given to the frequency of cases, the inherent properties of the substances, as well as factors such as the exposure situation, bioavailability, individual predisposition, cross-reactivity and preventive measures taken.

3.4.5.3.1.2.6 Special consideration should be given to negative human data as full dose-response information is generally not available. For example, a negative result in an HRIPT or HMT at a low concentration may not allow for the conclusion that the substance does not have skin sensitizing properties as such effect at a higher concentration may not be excluded. In addition, negative human data should not

necessarily be used to negate positive results from animal studies and/or defined approaches, but can be used as part of a weight of evidence assessment. For both animal and human data, consideration should be given to the impact of the vehicle (e.g. Wright et al, 2001 and Kligman, 1966).

3.4.5.3.1.2.7 For example, negative results from substances tested in a predictive patch test at a DSA (dose per skin area) < 500 µg/cm$^2$ imply that a classification for skin sensitization might not be needed at all, however, classification as sub-category 1A or 1B cannot be ruled out, because the concentration tested was not high enough to exclude these possibilities. The same holds for test results for which it is unknown whether the test concentration corresponded to a DSA < 500 µg/cm$^2$. Negative results from substances tested at a DSA ≥ 500 µg/cm$^2$ suggest that classification might not be needed. However, while classification as sub-category 1A can be ruled out, classification as sub-category 1B cannot, because a higher test concentration might have resulted in a positive test result. However, a negative test result at a concentration of 100% can justify no classification (based on this test). Nevertheless, negative results at low concentrations may be informative for mixtures containing the substance at similar or lower concentrations.

3.4.5.3.1.3 *Guidance on the use of standard animal data*

3.4.5.3.1.3.1 The most common assays used for dermal sensitization testing in animals are the Local Lymph Node Assay (LLNA, OECD test guidelines 429 and 442A and 442B), the Guinea Pig Maximization Test (GPMT, OECD Test Guideline 406) and the Buehler test (OECD Test Guideline 406). When evaluating the quality of the study, consideration should be given, as relevant, to the strain of the mouse and guinea pig used, the number, age, and sex of the animals, and the test conditions used (e.g. preparation of patch test site, dose level selection, chemical preparation, positive and negative test controls).

3.4.5.3.1.3.2 OECD test guidelines for the LLNA include the radioactive assay (OECD Test Guideline 429) and non-radioactive assays (OECD test guidelines 442A and 442B; LLNA:DA, LLNA:BrdU-ELISA, and LLNA:BrdU-FCM). In these tests, sensitizers are characterized by increasing the group mean stimulation index ("SI", a measure of lymph node proliferation) in treated groups versus concurrent vehicle controls by more than a predefined critical value which is different for each form of the LLNA (e.g. SI ≥ 3 for the radioactive LLNA, SI ≥ 1.6 for the LLNA:BrdU-ELISA). For sensitizers, subcategorization is performed based on the effective concentration (EC) causing an increase in SI of exactly the critical magnitude (e.g. the EC3 under OECD Test Guideline 429 is the concentration leading to an exactly threefold increase in group mean SI versus control).

3.4.5.3.1.3.3 The respective OECD test guidelines for the different LLNA variants specify that a pre-screen test should be undertaken to determine the highest concentration to be tested. If such a test has not been performed and the LLNA was carried out with a test concentration < 100%, a rationale (e.g. based on solubility, local or systemic toxicity, see OECD test guidelines 429, and 442A and 442B) needs to be provided that the highest test concentration represents the maximum testable concentration. Otherwise, the reliability of a negative test result has to be considered compromised.

3.4.5.3.1.3.4 EC values are normally obtained by interpolation between adjacent test concentrations, i.e. between the highest test concentration causing an SI below, and the lowest test concentration causing an SI above the critical value. However, care must be taken when the EC value falls below the lowest concentration tested and can therefore only be estimated by extrapolation, which is associated with additional uncertainty. In some cases, the SI at the highest concentration tested falls only slightly below the critical SI value, which raises the question of upward extrapolation (unless the maximum testable concentration has been applied). These and other issues regarding the reliability of LLNA results are further discussed in Ryan et al. (2007) and Annex 3 of OECD Series on Testing and Assessment No. 336 (Supporting Document to OECD Guideline Document 497), which also provides a highly curated database of test guidelines 429 LLNA EC3 values.

3.4.5.3.1.3.5 Further limitations have been identified for the radioactive and non-radioactive LLNAs. For example, substances containing certain functional groups may interfere with the accuracy of the assay. These limitations as well as the possibility of borderline positive results are described in OECD test guidelines 429, and 442A and 442B. Variability in EC values for the same substance may also be the result of the vehicle used. For example, analysis has shown an underestimation of potency (i.e. higher EC3 values) with predominantly aqueous vehicles or propylene glycol (see Jowsey, 2008).

3.4.5.3.1.3.6 For OECD Test Guideline 406, the concentration of test chemical used for each induction exposure should be systemically well-tolerated using the highest dose to cause mild-to-moderate skin irritation. The concentration used for the challenge exposure should be the highest non-irritant dose. A positive result in a

guinea pig test is defined as a grade above zero according to the applicable grading scale such as the Magnusson and Kligman grading scale for OECD Test Guideline 406 at one or more of the two observation time-points. A grade of 0.5, which is sometimes reported, is therefore also considered a positive result.

3.4.5.3.1.4 *Guidance on the use of defined approaches*

Defined approaches validated according to international procedures and described in OECD Guideline 497 have been characterized for the level of confidence that can be assigned to the predictions based on the applicability domain of the individual information sources used and the data interpretation procedure applied (see table 3.4.7). Other defined approaches under consideration but not yet validated according to international procedures and described in OECD Guidance Document 256 according to internationally agreed criteria for their reporting (OECD Guidance Document 255) may be accepted by some competent authorities.

3.4.5.3.1.5 *Guidance on the use of non-stand-alone in chemico/in vitro methods*

Individual in chemico/in vitro methods such as those reported in OECD test guidelines 442C, 442D and 442E, due to their limited mechanistic coverage, cannot be used on their own to conclude on Category 1 or no classification according to the criteria defined in table 3.4.8 and further data are necessary for classification in tier 2. In addition, although some of these methods provide quantitative information, these cannot be used for the purposes of subcategorization into sub-categories 1A and 1B since the criteria have not been validated according to international procedures. Nevertheless, such quantitative information may be accepted by a competent authority when used in a weight of evidence assessment under tier 2 for the purpose of subcategorization. This is also in line with the statement in these test guidelines that "*Depending on the regulatory framework, positive results generated with these methods may be used on their own to classify a chemical into UN GHS Category 1*." Therefore, the GHS also allows a competent authority to decide that a positive result with one of these non stand-alone *in chemico*/in vitro methods, may be used on its own to classify in Category 1 and whether test guideline 442C (Appendix III) kinetic Direct Peptide Reactivity Assay (kDPRA) can be used to differentiate between sub-category 1A and no sub-category 1A.

3.4.5.3.1.6 *Guidance on the use of non-standard data*

3.4.5.3.1.6.1 Validated but not yet adopted *in chemico*/in vitro methods such as those reported under 3.4.5.3.6.2 as well as in vivo test methods which do not comply with internationally agreed guidelines for the identification of skin sensitizers or the assessment of skin sensitizing potency may provide supportive evidence when used in an overall weight of evidence assessment (i.e. tier 3).

3.4.5.3.1.6.2 A non-exhaustive list of other validated *in chemico*/in vitro test methods accepted by some competent authorities but not adopted as OECD test guidelines is provided below. A competent authority may decide which classification criteria, if any, should be applied for these test methods:

(a) The Genomic Allergen Rapid Detection (GARD) potency is a transcriptomics-based in vitro assay addressing the third key event of the skin sensitization Adverse Outcome Pathway (activation of dendritic cells) similar to the GARDskin but uses a different gene signature that provides subcategorization of skin sensitizers (Gradin et al., 2020; Zeller et al., 2017; Corsini et al. 2021);

(b) The SENS-IS assay is a genomic approach applied to a Reconstructed Human Epidermis (RhE) (Cottrez et al., 2015; Cottrez et al., 2016);

(c) The Epidermal Sensitization Assay (EpisensA) is based on the measurement of the upregulation of four genes in a reconstructed human epidermis (RhE) to discriminate between sensitizers and non-sensitizers (Saito et al., 2017).

3.4.5.3.2 *Guidance on mixtures – skin sensitization*

3.4.5.3.2.1 Most of the standard animal test methods, defined approaches, in vitro and in chemico methods were developed and formally validated for identifying sensitizing substances and not mixtures. Nevertheless they are technically applicable to mixtures. However, there is limited data indicating whether there is a difference in the predictive capacity between standard animal test methods and defined approaches for the classification of mixtures. Sometimes, standard animal tests (see 3.4.2.2.3) on mixtures are required by

competent authorities or applied voluntarily and the results are internationally accepted for classification. Therefore, the results of standard animal test methods can be used for the classification of mixtures. The defined approaches were developed recently and no international consensus has yet been reached whether these approaches can be accepted for classification of mixtures. Human data can also be used for the classification of mixtures (see 3.4.5.3.2.2).

3.4.5.3.2.2 Guidance on the use of human data

3.4.5.3.2.2.1 The classification of substances and mixtures can be based on human evidence generated from a variety of sources. These sources include human predictive patch testing, epidemiological studies, case studies, case reports or histories, diagnostic patch testing and medical surveillance reports, and poison control centre information. This data may have been generated for consumers, workers, or the general population. Guidance for evaluating human evidence and the criteria in 3.4.2.2.2 is provided by some competent authorities (e.g., ECHA Guidance on the Application of the CLP Criteria, 2017). Further valuable information which should be considered for classification purposes (e.g., on use of appropriate concentrations and vehicles, as well as mixture evaluation) is also available (Consumer Product Safety Commission, 20xx; European Society of Contact Dermatitis guidance, 2015; Frosch et al., 201x).

3.4.5.3.2.2.2 When using human epidemiological data for classification, consideration should be given to the available data from a number of sources: (a) well-conducted clinical and diagnostic studies; (b) epidemiological studies, either general population studies or occupational studies; (c) cross-reactivity data; (d) case histories. Positive data from well-run epidemiological studies (which should also comply with WHO CIOMS guidelines, 2009) can be used for classifying substances and mixtures for skin sensitization.

3.4.5.3.2.2.3 General population epidemiological studies may be preferred over occupational epidemiological studies because the degree of sensitization in the workplace is likely greater than that of the general population due to greater exposure (both in time and concentration) to the sensitizing agent. Although providing helpful information regarding the potential sensitizing strength of a chemical, occupational data could exaggerate the estimation of the sensitizing strength of a chemical to the consumer scenario because of potentially higher exposure levels or frequency (Consumer Product Safety Commission, 20xx).

3.4.5.3.2.2.4 If population data are lacking, worker sensitization prevalence could be used to estimate prevalence in the exposed general population by taking exposure and dose-response relationships into account. “Case histories” are studies typically on a single or few individuals and are less helpful in providing information on sensitization rates in the general population (Consumer Product Safety Commission, 20xx).

3.4.5.3.2.2.5 When evaluating existing data, its quality should be taken into consideration. Criteria for a “well conducted” study would include validated outcomes, relevant dosing and route of administration and use of appropriate controls. Special attention should be applied to ascertain that exposure to the relevant substance or mixture is established with sufficient reliability. Studies should, where applicable, be carried out according to national and/or international test guidelines and according to good laboratory practice (GLP), compliance with good clinical practice (GCP), and good epidemiological practice (GEP) (Consumer Product Safety Commission, 20xx; Hoffman, 2019; Alba, 2020; WHO CIOMS, 2009).

3.4.5.3.2.2.6 A specific type of epidemiological study (such as randomized control studies or trials) may include information from diagnostic patch testing. Diagnostic patch testing is considered by some competent authorities to be the gold standard in diagnosing contact allergy in dermatitis patients (Johansen et al, 2015; Frosch et al., 201x). Importantly, due consideration needs to be given to the appropriate selection of vehicle, test material composition, and patch test concentrations for the purpose of not causing false negatives, false positives, irritant reactions or inducing contact allergy (skin sensitization). Positive data from experimental, clinical or diagnostic studies in humans and/or well-documented episodes of allergic contact dermatitis may be used to classify substances and mixtures for skin sensitization, when it can be assumed with sufficient confidence that the tested substance or mixture was indeed the most likely cause for induction of sensitisation. Therefore, it should be established that there is at least a general likelihood that the respective patient(s) had been previously exposed to the substance or mixture. On the other hand, negative results from such tests are not sufficient to prove that the test substance or mixture should not be classified as a skin sensitizer.

3.4.5.3.2.2.7 For some substances and mixtures, predictive patch test data in human volunteers are available (e.g. Strickland et al., 2023). Two test designs for predicting whether the substance or mixture will

induce sensitization are the Human Maximization Test (HMT) and the Human Repeated Insult Patch Tests (HRIPT).

3.4.5.3.2.2.8 Positive data from predictive patch testing (HRIPT or HMT) conducted through human experimental and clinical studies, showing allergic contact dermatitis caused by the test substance or mixture can be used to classify for skin sensitization. These studies are generally conducted in controlled clinical settings and in general the study outcome is considered more reliable the larger the test panel size. Criteria for evaluating these data are provided in paragraph 3.4.2.2.2.2 and 3.4.2.2.2.3. When evaluating the data from HRIPT, consideration should be given to the appropriate use of vehicle as this can affect the outcome of testing (European Society of Contact Dermatitis guidance, 2015; Frosch et al., 201x).

3.4.5.3.2.2.9 The HMT is no longer in use, due to ethical concerns about its potential to create adverse health consequences for the person being tested. In cases where such data exist, nevertheless they can be used for classification.

3.4.5.3.2.2.10 Special consideration should be given to negative human data as full dose-response information is generally not available. For example, a negative result in an HRIPT or HMT at a low concentration may not allow for the conclusion that the substance or mixture does not have skin sensitizing properties, as such effect at a higher concentration may not be excluded. In addition, negative human data should not necessarily be used to negate positive results from animal studies and/or defined approaches but can be used as part of a weight of evidence assessment. For both animal and human data, consideration should be given to the impact of the vehicle (e.g. Wright et al, 2001 and Kligman, 1966).

3.4.5.3.2.2.11 For example, negative results from substances or mixtures tested in a predictive patch test at a DSA (dose per skin area) of $< 500$ µg/cm$^2$ imply that a classification for skin sensitization might not be needed at all, however, classification as category 1A or 1B cannot be ruled out, because the concentration tested was not high enough to exclude these possibilities. The same holds for test results for which it is unknown whether the test concentration corresponded to a DSA $< 500$ µg/cm$^2$. Negative results from substances or mixtures tested at a DSA $\geq 500$ µg/cm$^2$ suggest that classification might not be needed. However, while classification as category 1A can be ruled out, classification as category 1B cannot, because a higher test concentration might have resulted in a positive test result. However, a negative test result at a concentration of 100% (i.e. the mixture as it is) can justify no classification (based on this test). Nevertheless, negative results at low concentrations may be informative for the safety assessment of mixtures containing the substance or mixture at similar or lower concentrations.

3.4.5.3.2.3 Guidance on the use of standard animal data

3.4.5.3.2.3.1 Animal tests have been developed to identify sensitizing substances and not mixtures. Therefore, the results obtained on mixtures need to be evaluated with care. The following considerations can be relevant for mixtures because of dilution effects, in particular for borderline cases, but can also be applicable for substances.

3.4.5.3.2.3.2 For example, a stimulation index of 3 or more in the traditional radioactive local lymph node assay (LLNA) (OECD Test Guideline 429) should be seen as a regulatory threshold for identification of a sensitizing mixture rather than as a threshold for sensitization as such. If a sensitizing substance is present at a low concentration in a mixture, a stimulation index of 3 may not be reached in the LLNA, but the substance in that mixture may still act as a sensitizer at population level. For this reason, a conclusion on the absence of sensitizing potential of a mixture based on the negative outcome in a test must be taken with great caution.

3.4.5.3.2.3.3 Where the mixture is tested undiluted, contains sensitizing ingredients and there is an increase in positive animals (Buehler, guinea pig maximisation test (GPMT)) or in the response (LLNA) which does not fulfil the criteria for a positive result, an overall weight of evidence assessment is required including the indicators included in Tier 3. This should also include available data on the sensitizing ingredient(s) regarding their potency, bioavailability, accumulation in the skin and interaction with the other ingredients. When the result is inconclusive, where applicable the bridging principles should be applied and otherwise the ingredient-based approach should be followed.

3.4.5.3.2.3.4 Test data on a mixture takes into account effects of possible interactions of its components. For instance, it is known that the presence of a vehicle may significantly influence the skin sensitizing potency, by altering the penetration of the sensitizing component(s) through the skin, (Basketter et al. 2001, Dearman et al. 1996, Heylings et al. 1996) or through other mechanisms involved in the induction of sensitization (Cumberbatch et al. 1993; Dearman et al. 1996). These mechanisms may differ between animals and humans. Especially where differences are known or suspected that could lead to the underestimation of sensitization, negative outcomes may not be reliable.

3.4.5.3.2.3.5 Repeated exposure to mixtures that are non-sensitizing in a standard LLNA might induce skin sensitization, if the sensitizing component in the mixture has sufficient accumulation potential in the skin to reach the minimum effect concentration (De Jong et al. 2007). Uncertainty also exists about the effect of such a mixture after exposure of a larger skin area. Therefore, additional information is important, if the outcome of sensitization tests on mixtures contrasts with the classification based on the content of a sensitizing ingredient. For example, the validity of a well-conducted LLNA on a mixture with a negative outcome can scientifically be confirmed by spiking the test mixture with another sensitizer (positive control) at different concentrations, or by showing a dose response relationship.

3.4.5.3.2.3.6 These considerations are not only relevant for mixtures but also for substances. However, they are more relevant to mixtures, as due to the dilution of the sensitizing ingredient(s), results are more likely to be close to the border between a positive and a negative result.

3.4.5.3.2.3.7 Where the mixture contains corrosives or potent irritants resulting in unacceptable irritation in the pilot study with the mixture, either a dilution has to be used or the results may be a false positive. If a dilution is tested, the lower tested dose of the potential sensitizer(s) in the mixture may lead to false negative results for classification. In such cases, where applicable the bridging principles and otherwise the ingredient-based approach should be applied unless evidence is provided that the negative result is not caused by the dilution. This could for example be shown by testing the mixture without the corrosive or irritant ingredients at the actual concentration. Also, the validity of a well conducted LLNA on a mixture with a negative outcome can scientifically be confirmed by spiking the test mixture with another sensitizer (positive control) at different concentrations, or by showing a dose-response relationship.

3.4.5.3.2.4 Guidance on the use of defined approaches

3.4.5.3.2.4.1 Defined approaches may not have been formally validated for mixtures according to international procedures. Several defined approaches require upfront consideration to whether such testing will yield results that are predictive of the skin sensitizing properties of the mixture (see 3.4.5.3.2.7). This upfront consideration could include a comparison of the classification based on the results of a defined approach with existing classifications of similar mixtures. Where the comparison shows that the defined approach is predictive of a certain type of mixture, the outcome of the defined approach can be used for other mixtures of the same type for classification.

3.4.5.3.2.4.2 In chemico and in vitro methods used in defined approaches do not account for dermal penetration. Therefore, results from defined approaches may lead to false positive predictions compared to the standard animal tests that account for dermal penetration.

3.4.5.3.2.4.3 Also, it is necessary to exercise care when evaluating whether the dose used will yield results that are predictive of the skin sensitizing properties of the mixture. For example, when testing mixtures in the DPRA (Direct Peptide Reactivity Assay, OECD Test Guideline 442C) a specific molar ratio between the tested chemical and the cysteine or the lysine peptides that react with it has to be used. Otherwise, no valid outcome can be obtained.

3.4.5.3.2.4.4 In some methods, e.g. in silico predictions in the defined approaches for skin sensitization integrated testing strategies (ITSv1 and ITSv2) listed in OECD Test Guideline 497, all ingredients have to be assessed individually and the outcome from the in silico component of the defined approach is considered positive, if one ingredient is positive. However, it is noted that this may provide overly conservative or false positive predictions, as the in silico methods currently do not take into account the concentration at which the ingredient is present in the mixture.

3.4.5.3.2.4.5 Where the mixture contains potent irritants/cytotoxic agents, the tested dose of the potential sensitizer(s) in the mixture may not lead to conclusive results for classification. In such cases, the ingredient-based approach should be applied (see section 3.4.5.3.2.4.7).

3.4.5.3.2.5 Guidance on the use of non-stand-alone in chemico/in vitro methods

3.4.5.3.2.5.1 In chemico/in vitro methods may not have been formally validated for mixtures according to international procedures. This requires upfront consideration of whether the results of such testing will yield results that are predictive of the skin sensitizing properties of the mixture (see 3.4.5.3.2.7). Such upfront consideration should be taken into account when assessing the outcome of in chemico/in vitro methods for mixtures and should justify that the results are scientifically meaningful for classification.

3.4.5.3.2.5.2 In general, in chemico/in vitro methods can only be used for classification when all ingredients fall within their applicability domain. Specific limitations regarding applicability domains are described in the respective methods and should be taken into consideration as well as any further information on such limitations from the published literature. Specifically, kinetic interactions should be taken into account. In contrast to in vivo testing, dermal absorption is not considered in chemico/in vitro, whereby substance interactions may be observed, which would not occur under physiological conditions. This may lead to an overestimation of the sensitizing potential in case of additive or synergistic effects or, more rarely, its underestimation in case of decreased bioavailability, e.g. by binding or degradation of the sensitizer. Where the potential sensitizer is known and data on substance bioavailability and interaction are available it should be taken into account to predict the uncertainty of a positive or negative test outcome.

3.4.5.3.2.5.3 Where the mixture contains potent irritants/cytotoxic agents, the tested dose of the potential sensitizer(s) in the mixture may not lead to conclusive results for classification. In such cases, the ingredient-based approach should be applied (see also section 3.4.5.3.2.4.7).

3.4.5.3.~~7~~.3 *Guidance on the weight of evidence assessment for classifying substances and mixtures for skin sensitization*

3.4.5.3.~~7~~.3.1 There may be situations where results from tests and/or non-test methods are available but disagree with each other with respect to the classification. In these situations, the tiered approach to classification for skin sensitization requires a weight of evidence assessment consistent with the principles elaborated in sections 1.3.2.4.2 and 1.3.2.4.9 on test data quality and weight of evidence, respectively. In addition, some guidance on the weight of evidence assessment specific for skin sensitization is provided below which can be applied when the general principles do not result in a conclusion on the classification. It should be noted that human and animal results for a substance obtained at low concentrations may still be informative for classifying a mixture containing the substance at similar or lower concentrations.

3.4.5.3.~~7~~.3.2 Mutual compatibility of study results

3.4.5.3.~~7~~.3.2.1 In cases where results are in disagreement with each other (e.g. not classified versus Category 1, sub-category 1A or 1B; sub-category 1A versus 1B), a weight of evidence assessment becomes necessary. However, less obvious situations may also occur such as where certain studies may point to not classified or sub-category 1B, while it cannot be excluded that a stricter classification might have resulted under a different dosing regime. For example, a negative HMT result at a dose per skin area of 100 µg/cm$^2$ cannot exclude that a positive result might have been obtained at e.g. 300 µg/cm$^2$ (sub-category 1A) or 700 µg/cm$^2$ (sub-category 1B). The same holds for LLNA test results obtained from tests which have not been carried out using the highest possible test concentration (see OECD Test Guideline 429 for details).

3.4.5.3.~~7~~.3.2.2 In the following ambiguous cases, study results for substances and mixtures would not be in disagreement with another study result pointing at that stricter classification:

(a) A not classified result obtained at a lower test concentration does not exclude the possibility of a sub-category 1B outcome at a higher test concentration. Therefore, a not classified result obtained at a low concentration is compatible with other not classified outcomes, or with category 1 and sub-category 1B outcomes obtained at higher test concentrations.

(b) A not classified result at a very low-test concentration does not even exclude a possible outcome of sub-category 1A at a higher test concentration. Therefore, a not classified outcome obtained at a very low-test concentration is compatible with all possible classification outcomes (i.e. not classified, category 1, sub-category 1A or 1B) obtained at higher test concentrations.

(c) A sub-category 1B result at a higher test concentration does not exclude a sub-category 1A outcome at a lower test concentration. Therefore, a sub-category 1B classification tested at a high-test concentration is compatible with other outcomes of sub-category 1B, or even sub-category 1A, obtained at lower test concentrations.

3.4.5.3.~~7~~.3.2.3 If at least one unambiguous study result allows for subcategorization of a substance or mixture and all other study results are not in disagreement (see above), then it can be classified into a sub-category. For example, if all study results are in the same sub-category (i.e. sub-category 1A or 1B), or with at least one study permitting subcategorization (i.e. either sub-category 1A or 1B) and all other studies classified into Category 1 without subcategorization, then the substance or mixture can be subcategorized.

3.4.5.3.~~7~~.3.3 Weight of evidence considerations for giving one study result more weight than another

3.4.5.3.~~7~~.3.3.1 Some classifiers or competent authorities may take various approaches to evaluate study results given the required level of expert judgement (see 1.3.2.4.8) required to perform a weight of evidence assessment. Competent authorities may specify their preferred approach in their own guidance. For example, through:

(a) Applying a precautionary approach, giving more weight to studies resulting in the stricter classification outcome;

(b) Giving human data higher weight than animal or non-test data;

(c) Giving certain animal data (e.g. LLNA data) more weight than other animal data (e.g. Buehler test data).

3.4.5.3.~~7~~.3.3.2 Often, several results (of the same or different type) may have to be considered in the weight of evidence assessment. There are no generally recognized rules for this situation, however, possible solutions to integrating several results of the same type may include, for example:

(a) A precautionary approach where the strictest classification outcome from all studies of sufficient quality is assigned as the overall classification outcome.

(b) Averaging the obtained dose descriptors (e.g. LLNA EC3 values) or classification outcomes (no classification, Category, 1, 1A, 1B). A detailed discussion of such approaches can be found in Annex 3 (on LLNA data) and Annex 4 (on HMT/HRIPT data) of OECD Series on Testing and Assessment No. 336 (Supporting document to OECD Guideline Document 497).

**Table 3.4.7: Criteria for defined approaches** – *remains unchanged*

**Table 3.4.8: Criteria for individual *in chemico*/in vitro methods** – *remains unchanged*