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 **21 April 2022**

Forty-second session
Geneva, 6-8 July 2022
Item 2 (d) of the provisional agenda
Classification of skin sensitizers using the results of local lymph node assays test methods in accordance with OECD Test Guideline 442B

Peer review panel report for GHS sub-categorisation based LLNA BrdU-ELISA

Transmitted by the expert from Japan

1. Japan wishes to thank the OECD secretariat and the experts for their extensive efforts for the Japanese proposal to clarify the criteria for classification for skin sensitization using animal studies.

2. This document contains the Peer Review Panel (PRP) report for GHS sub-categorisation based LLNA: BrdU-ELISA conducted as the OECD support work for this proposal.

3. The review work was conducted from June 2021 to October 2021 by the PRP, composed of five experts from the OECD Expert Group on Skin Sensitization. The results of PRP report were discussed at the Meeting of the Expert group on Alternative methods for skin sensitisation held on 28-29 October 2021.

4. Then the PRP concluded that the proposed criterion met the validation principles of OECD GD 34 based on the materials submitted, and the report was finalized at the meeting.

[OECD Test Guidelines Programme]

[LLNA:BrdU-ELISA criterion for UN GHS Skin Sensitisation Sub-Categorisation]

[Peer Review Report]

Table of Contents

1. Summary 3
2. Background 4
3. The peer review process 6
4. Evaluation Principle 1. The rational for the proposal should be available 8
5. Evaluation Principle 2. The relationship between the dataset/analysis and the (biological) phenomenon of interest should be described. 9
6. Evaluation Principle 3. A detailed protocol for the analysis should be available. 10
7. Evaluation Principle 4. Demonstration of the analysis should be based on the reference chemicals representative of the types of substances for which the sub-categorisation will be used. 12
8. Evaluation Principle 5. The performance of the analysis should have been evaluated in relation to the relevant information from the species of concern, and existing relevant toxicity testing data. 14
9. Evaluation Principle 6. Ideally, all data supporting the validity of the analysis should have been obtained in accordance with the principles of GLP. 15
10. Evaluation Principle 7. All data supporting the assessment of the validity of the analysis should be available for expert review. 16
11. Additional Remarks 17
12. Conclusions and recommendations 18
13. Acknowledgements 19
Annex 1. Peer Review Panel Composition 20
Annex 2. Charge Questions for the Peer Review of LLNA:BrdU-ELISA criterion for UN GHS Sub-categorisation 21
Annex 3. Response Compilation 22

1. Summary

1. This document contains the Peer Review Report (PRR) on the proposal from Japan to set criterion for UNGHS[1] subcategorisation for skin sensitisation (i.e., category 1A (Cat.1A) vs. category 1B (Cat.1B)) based on the LLNA:BrdU-ELISA data.

2. Currently, there are three skin sensitisation test methods to classify UNGHS Category 1 (Cat.1) chemicals: two guinea pig prediction tests[2] and the radioisotopic (RI) LLNA (LLNA-RI) test[3], which can be further applied for GHS sub-categorisation 1A/1B to provide information on the skin sensitisation potency of chemicals. The LLNA:BrdU-ELISA[4] is another reliable sensitisation test method using the same principle as the standard LLNA-RI, and can be used to classify GHS Cat.1 versus Not Classified (NC), but has not been used for further classification into Cat.1A and Cat.1B.

3. 32 skin sensitisers, that are classified as UNGHS Cat.1A or Cat.1B, had been used in the validation and peer review of LLNA:BrdU-ELISA and published in the ICCVAM Test Method Evaluation Report on the LLNA:BrdU-ELISA for skin sensitisers[5]. Japanese researchers conducted retrospective analysis on these chemicals, in an attempt to determine optimal criteria for GHS sub-categorisation for skin sensitisation. The analysis and its results were published in two peer-reviewed articles[6,7] and submitted for review, accompanying a proposal to apply the result to amend the texts of UN GHS Chapter 3.4[8], to an independent Peer Review Panel (PRP). The Panel was composed of members from OECD Expert Group in Skin Sensitisation, and the work of the Panel was coordinated by the OECD Secretariat.

4. The PRP was asked to evaluate how the proposed analysis and criterion for GHS sub-categorisation address the principle outlined in the OECD Guidance Document 34 (GD 34) on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment.

5. The PRP concluded that the proposed criterion, evaluated based on the materials submitted, meet the validation principles of OECD GD 34. This report elaborates the general and specific comments from PRP for each validation principle.

[1] United Nations (2019). Globally Harmonized System of Classification and Labelling of Chemicals (GHS) 8th revised edition. Chapter 3.4 Respiratory or skin sensitisation. United Nations Publications. New York.

[2] OECD (2021) Guideline for Testing of Chemicals No. 406: Skin Sensitisation Guinea Pig Maximisation Test and Buehler Test. Organization for Economic Co-operation and Development, Paris. Available at: [https://doi.org/10.1787/9789264070660-en].

[3] OECD (2010) Guideline for Testing of Chemicals No. 429: Skin Sensitisation: Local Lymph Node Assay. Organization for Economic for Co-operation and Development, Paris. Available at: [https://doi.org/10.1787/9789264071100-en]

[4] OECD (2018) Guideline for Testing of Chemicals No. 442B: Skin Sensitisation: Local Lymph Node Assay: BrdU-ELISA or –FCM. Organization for Economic for Co-operation and Development, Paris. Available at: [https://doi.org/10.1787/9789264090996-en]

[5] The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) (2010). ICCVAM Test Method Evaluation Report on the Murine Local Lymph Node Assay: BrdU-ELISA. A Nonradioactive Alternative Test Method to Assess the Allergic Contact Dermatitis Potential of Chemicals and Products. https://ntp.niehs.nih.gov/iccvam/docs/immunotox_docs/llna-elisa/tmer.pdf (accessed on January 12, 2021)

[6] Maeda Y, Takeyoshi M (2019) Proposal of GHS sub-categorisation criteria for LLNA:BrdU-ELISA (OECD TG442B). Regul.Toxicol.Pharmacol. 107:104409.

[7] Kobayashi T, Maeda Y, Kondo H, Takeyoshi M (2020). Applicability of the proposed GHS sub-categorisation criterion for LLNA:BrdU-ELISA (OECD TG442B) to the CBA/J strain mouse. J Appl Toxicol; 1-5. https://doi.org/10.002/jat.3996

[8] United Nations (2019). Globally Harmonised System of Classification and Labelling of Chemicals (GHS) 8th revised edition. Chapter 3.4 Respiratory or skin sensitisation. United Nations Publications. New York. pp. 161-170.

2. Background

6. Currently, there are three skin sensitisation test methods to classify UNGHS Category 1 (Cat.1) chemicals: two guinea pig prediction tests[9] and the radioisotopic (RI) LLNA (LLNA-RI) test[10], which can be further applied for GHS sub-categorisation 1A/1B to provide information on the skin sensitisation potency of chemicals. The LLNA:BrdU-ELISA[11] is another reliable sensitisation test method using the same principle as the standard LLNA-RI, and can be used to classify GHS Cat.1 versus Not Classified (NC), but has not been used for further classification into Cat.1A and Cat.1B.

7. 32 skin sensitizers, that are classified as UNGHS Category 1A (Cat.1A) or Category 1B (Cat.1B), had been used in the validation and peer review of LLNA:BrdU-ELISA and published in the ICCVAM Test Method Evaluation Report on the LLNA:BrdU-ELISA for skin sensitizers[12]. Japanese researchers conducted retrospective analysis on these chemicals, in an attempt to determine optimal criteria for GHS sub-categorisation for skin sensitisation. Consequently, the optimal criterion for the GHS sub-categorisation was determined to be a cut-off value of 6% when using EC1.6. Furthermore, this criterion represents similar GHS sub-categorisation performance to the existing criterion with the cut-off of 2% for EC3 values from LLNA-RI data[13]. Table 1 shows the proposed GHS sub-categorisation criterion for LLNA:BrdU-ELISA.

Table 1. Proposed GHS sub-categorisation for LLNA:BrdU-ELISA

Category	Criterion
Cat.1	SI ≥ 1.6
Cat.1A	EC1.6 value ≤ 6%
Cat.1B	EC1.6 value > 6%

8. Using this criterion, the correct outcomes for GHS Cat.1A and Cat.1B chemicals were 92.3% and 84.2%, respectively, for all 32 chemicals. Among the 4 mispredicted chemicals, 2-mercaptobenzothiazole was under-predicted, and 3-aminophenol, trimellitic anhydride and nickel sulfate were over-predicted. The under-prediction for 2-mercaptobenzothiazole was thought to be due to a

[9] OECD (2021) Guideline for Testing of Chemicals No. 406: Skin Sensitisation Guinea Pig Maximisation Test and Buehler Test. Organization for Economic Co-operation and Development, Paris. Available at: [https://doi.org/10.1787/9789264070660-en].

[10] OECD (2010) Guideline for Testing of Chemicals No. 429: Skin Sensitisation: Local Lymph Node Assay. Organization for Economic for Co-operation and Development, Paris. Available at: [https://doi.org/10.1787/9789264071100-en]

[11] OECD (2018) Guideline for Testing of Chemicals No. 442B: Skin Sensitisation: Local Lymph Node Assay: BrdU-ELISA or –FCM. Organization for Economic for Co-operation and Development, Paris. Available at: [https://doi.org/10.1787/9789264090996-en]

[12] The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) (2010). ICCVAM Test Method Evaluation Report on the Murine Local Lymph Node Assay: BrdU-ELISA. A Nonradioactive Alternative Test Method to Assess the Allergic Contact Dermatitis Potential of Chemicals and Products. https://ntp.niehs.nih.gov/iccvam/docs/immunotox_docs/llna-elisa/tmer.pdf (accessed on January 12, 2021)

[13] Maeda Y, Takeyoshi M (2019) Proposal of GHS sub-categorisation criteria for LLNA:BrdU-ELISA (OECD TG442B). Regul.Toxicol.Pharmacol. 107:104409.

CBA/JN mouse strain-specific low-response in the assay used for the analysis; in further examination[14] using CBA/J strain mouse to confirm the applicability of the proposed criterion, 15 chemicals were tested for GHS sub-categorisation, and all chemicals were correctly predicted, including 2-mercaptobenzothiazole; for the analysis, a different round-off of the values was introduced compared to the original one (see paragraph 13).

9. According to these scientific grounds, Japan proposed to add a new skin sensitisation sub-categorisation criterion to the UN-GHS document at the 39th meeting of GHS sub-committee in December 2020. The proposal was included in the two-year programme of work of the GHS sub-committee for 2021-2022, and the sub-committee invited the experts from Japan to report to the 40th meeting of GHS sub-committee on the review and discussion of their proposal and outcome at OECD level.

[14] Kobayashi T, Maeda Y, Kondo H, Takeyoshi M. (2020) Applicability of the proposed GHS sub-categorisation criterion for LLNA:BrdU-ELISA (OECD TG442B) to the CBA/J strain mouse. J Appl Toxicol; 1-5. https://doi.org/10.1002/jat.3996

3. The peer review process

10. The proposal from Japan was introduced at the 33rd meeting of the Working Group of the National Coordinators for the Test Guidelines Programme (WNT) in April 2021 to involve OECD Expert Group on skin sensitisation in the review of LLNA BrdU-ELISA data in support of a criterion for the sub-categorisation of skin sensitisers. The WNT approved the request to involve the OECD Expert Group on skin sensitisation, and welcomed the collaboration with the UN GHS Sub-Committee of Experts to further develop criteria to make best use of data generated using enhanced methods.

11. Following the outcome of the 33rd WNT meeting, the secretariat contacted the members of the OECD Expert Group to establish the Peer Review Panel (PRP) that can provide an independent review of the proposal to use LLNA: BrdU-ELISA data for UN GHS sub-categorisation. The panel was established in June 2021. The selected members of the Panel are listed in Annex 1.

12. The first PRP meeting took place virtually on 12 July 2021. Japan was invited to present an overview of the proposal in the beginning of the meeting. After the presentation, a closed PRP discussion was held on the presentation received and on the charge questions to evaluate the proposal based on validation principles outlined in OECD GD 34.

13. Panel members were asked to base their review on following materials:

- A document from Japan describing the enclosed dataset.
- Proposal of GHS sub-categorisation criteria for LLNA: BrdU-ELISA (OECD TG442B). Regulatory Toxicology and Pharmacology. 107: 104409 (2019)
- Applicability of the proposed GHS sub-categorisation for LLNA: BrdU-ELISA (OECD TG442B) to the CBA/J strain mouse. Journal of Applied Toxicology; 1-5. (2020)
- Table 3-1 from ICCVAM Test Method Evaluation Report[15] that lists the 32 chemicals used for LLNA:BrdU-ELISA sub-categorisation analysis
- Two excel files: one containing the original EC1.6 values and one containing data corrected by excluding a round-off function

14. The charge to the Panel was to assess to what extent the OECD validation principle set out in the OECD GD 34 had been met. The validation principle to assess the intra-, and inter-laboratory reproducibility was excluded from the list of charge questions, as it was deemed not applicable given the proposal is based on retrospective analysis of existing data from another validation study[15]. The charge questions are listed in Annex 2.

15. Each Panel member provided written responses to the charge questions to the Secretariat by 7 September 2021. For transparency, the individual comments from the Panel members are provided anonymously in Annex 3.

16. Based on these response, the Secretariat prepared a draft summary of the PRP's responses to the individual questions and the initial report was circulated to the Panel on 14 September 2021. The issues identified were discussed at a teleconference on 24 September 2021.

17. The initial Peer Review Report was updated based on the discussions at the second teleconference and circulated to the Panel for review and comments on 29 September 2021. The draft

[15] The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) (2010). ICCVAM Test Method Evaluation Report on the Murine Local Lymph Node Assay: BrdU-ELISA. A Nonradioactive Alternative Test Method to Assess the Allergic Contact Dermatitis Potential of Chemicals and Products.

revised based on the feedback from the Panel was submitted to the Expert Group in Skin Sensitisation on 14 October 2021 for a discussion at the virtual meeting taking place in 28 – 29 October 2021.

18. This report presents the summary of the assessment of the proposal to use LLNA:BrdU-ELISA for UN-GHS skin sensitisation sub-categorisation and the resulting agreed responses of the PRP to each of the charge questions.

4. Evaluation Principle 1. The rational for the proposal should be available

Charge Question 1: *Do you consider that the rationale for the proposal is clearly elaborated in the request for review from Japan together with the supporting publications?*

19. Overall, the PRP considers that the rational for using LLNA: BrdU-ELISA dataset for UNGHS skin sensitisation sub-categorisation (i.e., Category 1A (Cat.1A) vs. Category 1B (Cat.1B)) is clearly elaborated in the proposal for review requested from Japan, supporting publications as well as in the presentation to the PRP on 12 July 2021 by the lead.

20. The PRP considers that the scientific validation for which the proposal is established on is well understood in the following contexts:

- LLNA:BrdU-ELISA was endorsed as OECD TG 442B in 2010, but the test method has not been yet applied for UN GHS skin sensitisation sub-categorisation.
- For the standard radioisotopic (RI) LLNA (OECD TG 429), EC3 with cut-off value of 2% was used for Cat.1A and Cat.1B classification. The proposal is based on the retrospective re-analysis of existing data of 32 sensitisers classified as Cat.1A and Cat.1B (ICCVAM, 2010), identifying the optimal criteria for GHS sub-categorisation using LLNA:BrdU-ELISA (OECD TG 442B) with EC1.6 cut-off value of 6% (Cat.1A: EC1.6≤ 6%, Cat.1B: EC1.6≥ 6%)
- Given that RI-LLNA and LLNA:BrdU-ELISA are mechanistically based upon the same biological process and differ solely on the endpoint detection methodology (incorporation of BrdU as a surrogate to ^{3}H-thymidine incorporation), which was demonstrated originally in the prevalidation and validation of LLNA:BrdU-ELISA (ICCVAM, 2010), it is reasonable hypothesis that the LLNA:BrdU-ELISA could further be used to discriminate between GHS Cat.1A and Cat.1B substances in a similar manner as for the RI-LLNA.
- In order to identify the optimal criterion for GHS sub-categorisation, a list of 32 chemicals that were previously identified as Cat.1A or Cat.1B chemicals based on RI-LLNA was re-analysed using LLNA:BrdU-ELISA data with EC1.6 cut-off value of 6%. Although the number of chemicals was somewhat limited, the outcomes from the re-analysis for Cat.1A and Cat.1B chemicals were 92.3% and 84.2% for all 32 chemicals, respectively (Maeda and Takeyoshi, 2019)[16]

21. Based upon the findings originally presented by Maeda and Takeyoshi (2019)[16] and followed upon by Kobayashi et al. (2020)[17], the rationale for initiating the analyses, as well as the follow up study to evaluate the subcategorization using the standard strain of CBA/J mice was clearly presented.

Overall, the PRP agreed that the Evaluation Principle 1 has been met.

[16] Maeda Y, Takeyoshi M (2019) Proposal of GHS sub-categorisation criteria for LLNA:BrdU-ELISA (OECD TG442B). Regul.Toxicol.Pharmacol. 107:104409.

[17] Kobayashi T, Maeda Y, Kondo H, Takeyoshi M. (2020) Applicability of the proposed GHS sub-categorisation criterion for LLNA:BrdU-ELISA (OECD TG442B) to the CBA/J strain mouse. J Appl Toxicol; 1-5. https://doi.org/10.1002/jat.3996

5. Evaluation Principle 2. The relationship between the dataset/analysis and the (biological) phenomenon of interest should be described.

Charge Question 2: *Is the relationship between the dataset/analysis and the (biological) phenomenon of interest (skin sensitisation sub-categorisation) described adequately, using the relevant scientific references?*

22. Overall, the PRP considers that the relationship between the outcome of interest, skin sensitisation classification and the test and analytical methods were clearly described.

23. The ICCVAM Test Method Evaluation Report (2010) on the LLNA:BrdU-ELISA[18] was referenced to provide background information on the LLNA:BrdU-ELISA and EC3 data for the test chemical dataset used for the analysis.

24. In the current analyses for sub-categorisation, two publications by Maeda and Takeyoshi (2019) and Kobayashi et al. (2020) are referenced to present on the rationale for conducting the analyses.

25. The publication by Kobayashi et al. (2020) provided information on 15 chemicals tested as part of the development of the LLNA performance standard. This data was generated in the CBA/J mouse strain as a comparison to data in the Maeda and Takeyoshi (2019) studies, which evaluated 32 chemicals that has been tested using the CBA/JN strain that is not as common in hypersensitivity testing *in vivo*.

26. It was noted, however, that whilst the biological relationship between the RI-LLNA and LLNA:BrdU-ELISA for skin sensitisaton hazard assessment was originally presented in the ICCVAM test method report (2010), the references by Maeda and Takeyoshi (2019) and Kobayashi et al. (2020) do not present initially the biological relationship between the RI-LLNA and LLNA:BrdU-ELISA for skin sensitisation sub-categorisation, and whether the two test methods should provide the same discrimination. It would be useful to introduce and justify the biological relationship not only for identification of skin sensitisation hazard, but also to provide a hypothesis for the intended purpose of sub-categorisation.

27. More importantly, some discussion of the potential for differences in the two nucleotide incorporation endpoints should be presented with respect to potential impact upon discriminating among skin sensitisers to justify the analyses. It would have been useful to present the peak threshold for the fold change in the LLNA: BrdU-ELISA, assuming that it was defined in the initial hazard assessment. Fold change could play a role in terms of what EC concentration is resulting and it might be found that the ^{3}H-thymidine incorporation and BrdU incorporation are at a comparable concentration if the fold increase was evaluated, going from RI-LLNA fold increase to which that is measured by LLNA:BrdU-ELISA. Although this is not critical for the scope of this review, it would be of a concern if the fold changes are to be used for quantitative risk assessment. For example, whilst the SI values are similar for concentrations required for folding changes for some chemicals, they are not for others.

Overall, the PRP agreed that the Evaluation Principle 2 has been met.

[18] The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) (2010). ICCVAM Test Method Evaluation Report on the Murine Local Lymph Node Assay: BrdU-ELISA. A Nonradioactive Alternative Test Method to Assess the Allergic Contact Dermatitis Potential of Chemicals and Products.

6. Evaluation Principle 3. A detailed protocol for the analysis should be available.

Charge Question 3: *Do you consider that the process of setting the sub-categorisation criterion described in the supporting publication (Maeda and Takeyoshi, 2019) is sufficiently detailed (for example, description of what is measured, how it is measured, appropriate data analysis, decision criteria for evaluation of data and what are the criteria for acceptable test performance)? Does the proposed criterion provide a well-defined applicability domain? Is exclusion of data from the evaluation, if any, well justified?*

28. Overall, the PRP considers the process of setting the sub-categorisation criterion was well described. However, a number of shortcomings were identified by some of the reviewers.

29. The publication of Maeda and Takeyoshi (2019) did not present on any prospective procedures, or criteria for establishing the optimal cut-off, nor established criteria for optimal accuracy, sensitivity, specificity, false negative or false positive outcomes.

30. The publication did provide details on which cut-off values would be evaluated (from an empirical review of the existing dataset) and that the 10 cut-off values selected would be individually evaluated in 1% intervals to determine correct and incorrect outcomes, and from those analysis a determination was made which cut-off value was most ideal. No further algorithms were applied to correlate the RI-LLNA data with the LLNA:BrdU-ELISA data.

31. Regarding the mis-predicted chemicals (i.e., 2-mercaptobenzothiazole which was under-predicted, and 3-aminophenol, trimellitic anhydride and nickel sulfate, which were over-predicted), while it was considered appropriate to evaluate the cause for mis-predictions, and in certain predefined circumstances to exclude data, no similar analyses on the correctly predicted substances near the cut-off was conducted, which should have been done. An analysis to determine the range of EC3 values for all of the substances near the putative cut-off should be conducted to best fit the available LLNA:BrdU-ELISA EC1.6 values with the ranges of RI-LLNA EC3 values.

32. Furthermore, a more sophisticated analyses should be conducted by evaluating not only the mean data driving the single EC3 and EC1.6 values presented for each substance in the tables, but rather an analysis of the impact of variability of the individual animals.

33. Regarding the borderline results, while the possibility of obtaining borderline positive values (i.e., SI 1.6-1.9) with the LLNA:BrdU-ELISA is recognized by the proposal, the potential impact of borderline positive values on the potency sub-categorisation is not fully described.

34. The proposed criterion expands the applicability domain to include potency sub-categorisation. However, there is no guidance on the types of chemicals that the protocol is suitable for. ICCVAM considers the applicability domain on the LLNA: BrdU-ELISA to be the same as the traditional RI-LLNA. Importantly, though, there are known limitations for the RI-LLNA (i.e., metals, mixtures, aqueous solutions) and results from the RI-LLNA and LLNA:BrdU-ELISA do not always align. For example, unlike the RI-LLNA, the LLNA:BrdU-ELISA can be used for testing nickel compounds. The proposal should include more information on the types of chemicals that can be evaluated using this specific assay.

35. Regarding the applicability domain, it was concerned that not all chemical categories are represented, which brings questions to the possible limitations of using the proposed criterion. It was recommended to include more than one chemical in each chemical class to ensure the complete applicability domain.

Overall, the PRP agreed that the Evaluation Principle 3 has been met. However, the PRP made a number of recommendations for further clarifications on the borderline results, applicability domain and on the impact of variability of individual animals used for the study.

7. Evaluation Principle 4. Demonstration of the analysis should be based on the reference chemicals representative of the types of substances for which the sub-categorisation will be used.

Charge Question 4: *Are the reference chemicals used to demonstrate the performance of this analysis representative of the types of substance for which the sub-categorisation criterion will be used? Do you consider that the number of chemicals used for setting the sub-categorisation criterion is sufficient?*

36. The majority of PRP considers that the overall number of chemicals used for the analysis to be limited and that further assessment with a statistical tool (i.e., power analysis) to determine whether the number of chemicals was sufficient may be desirable. However, the PRP acknowledges that the proposed criteria is based on the most complete and well-documented dataset available (ICCVAM 2010).

37. The expressed concerns on the limited number of chemicals used for the analysis were:

- Compared to the number of chemicals used for setting the sub-categorisation criterion for the RI-LLNA (Cat.1A= 21; Cat.1B =49), which is not endorsed for sub-categorisation by ICCVAM, the number of chemicals used for the analysis in the proposal is rather small (i.e., Cat.1A = 13; Cat.1B = 19). Recognizing that the RI-LLNA and the LLNA:BrdU-ELISA are not always concordant (e.g., nickel compounds) and the number of chemicals used for the analysis is small, providing data for more chemicals is preferred.
- The overall number of chemical tested using different mouse strains is limited: 15 chemicals for CBA/J strain and no data for CBA/Ca strain. In Maeda and Takeyoshi (2019), CBA/JN mice were used, which is a strain unavailable from the majority of providers of experimental animals at present.
- It would be helpful to have power calculations or comparison with other OECD analytical strategies to determine if the current number of chemicals is sufficient.

38. On the other hand, one reviewer considers that a sufficient number of chemicals were selected for the analysis, and that chemicals selected for the analysis adequately reflected the chemical class domain represented by the LLNA test method performance standard reference chemicals based on the following rationale:

- The analysis to define the threshold for discrimination between GHS Cat.1A and 1B (per Maeda and Takeyoshi, 2019) utilised all of the existing data presented in Table 3-1 of the 2010 ICCVAM Test Method Evaluation Report on the LLNA:BrdU-ELISA for skin sensitisers, and thus all of the same chemistries used originally to define the applicability domain were included in the current analysis. The 2010 ICCVAM Test Method Evaluation report "considered the database of substances tested in the LLNA:BrdU-ELISA to be representative of a sufficient range of chemicals typically tested for skin sensitisation potential." Among this dataset are 14 reference substances from 22 ICCVAM-recommended LLNA performing standards reference substances (ICCVAM 2009)[19], of which 6 are identified as GHS Cat.1A skin sensitisers and 8 as Cat.1B skin sensitisers. It should be stated for the record that certain chemistries are typically associated with reactive events resulting in skin sensitisation. Whereas the original analysis included a wide range of chemistries to cover the spectrum of skin sensitisers and non-

[19] The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) (2009). ICCVAM Test Method Evaluation Report. The Reduced Murine Local Lymph Node Assay: An Alternative Test Method Using Fewer Animals to Assess the Allergic Contact Dermatitis Potential of Chemicals and Products

sensitisers, the current analysis is focused solely upon discriminating amongst skin sensitisers and thus the chemical representation is decidedly limited to this domain.

- A total of 32 sensitisers comprised of 13 Cat.1A and 19 Cat.1B chemicals were included in the analyses (per Maeda and Takeyoshi, 2019). The original analyses reported in the 2010 ICCVAM report included a total of 43 chemicals, comprised of 32 GHS Cat.1 sensitisers and 11 non-categorised for skin sensitisation, which were used to evaluate the ability of the test method to discriminate between sensitisers and non-sensitisers. The number of chemicals used in the current analyses to discriminate between GHS Cat.1A (n=13) and Cat.1B (n=19) were both greater than the number of non-sensitisers chemicals used in the original hazard analysis (n=11), and thus the current analysis has a somewhat greater number of chemicals used in establishing the discrimination threshold.
- In the follow-up study to evaluate the performance of the LLNA:BrdU-ELISA in the frequently-used CBA/J strain of mice (Kobayashi, et al., 2020), a total of 15 skin sensitisers, comprised of 6 GHS Cat.1A and 9 GHS Cat.1B, from the ICCVAM-recommended LLNA performance standards reference substances (ICCVAM 2009) were subsequently tested in CBA/J mice.
- When comparing the data of the LLNA:BrdU-ELISA relative to the reference test method RI-LLNA, 3 Cat.1B chemicals were over-predicted and 1 Cat.1B chemical was under-predicted. These four chemicals were clustered quite close to the reference method EC3 cut-off of 2%, and their EC3 values ranged from 1.7% to 4.8%, suggesting a limit to the ability to discriminate between the two hazard classes. This range represents <1% of the total dynamic range of RI-LLNA 2010 dataset (i.e., 0.009% - 47.5%). Two other correctly predicted chemicals were also within this narrow range, thus presenting that a total of 6 of the 32 chemicals (18.8%) were within the RI-LLNA EC3 ranges of 1.7% to 4.8%

39. It was commented that adding more number of chemicals at the targeted threshold could provide a better confidence on the threshold prediction. In addition, instead of using mean values of the study, understanding variability of animals to compare individual outcomes for the prediction would be useful to evaluate whether the threshold was established properly or not.

Overall, the PRP agrees that review for the Evaluation Principle 4 for the assessment on whether sufficient number of chemicals was used for the analysis cannot be performed decisively based on the information provided. Whilst the majority of the panel considers that the number of chemicals used was insufficient, the PRP also acknowledges that the analysis is based on the most complete and well-documented dataset available.

8. Evaluation Principle 5. The performance of the analysis should have been evaluated in relation to the relevant information from the species of concern, and existing relevant toxicity testing data.

Charge Question 5: *In the context of the retrospective assessment, do you consider the approach employed to assess the performance of the analysis is relevant? Do you consider that the performance of the sub-categorisation criterion has been evaluated in relation to relevant information from the species of concern, and existing relevant toxicity testing data? Do you consider the experimental data obtained for assessment of the sub-categorisation criterion in the supporting publication (Kobayashi et al., 2020) are valid?*

40. Overall, the PRP considers the approach employed to assess the performance of the analysis is relevant.

41. The performance of sub-categorisation criteria was considered to have been evaluated in relation to relevant information from the species of concern (mice), and existing relevant toxicity testing data. The proposal was supported based on its focus on chemicals for which there are data from both the RI-LLNA and LLNA:BrdU-ELISA. Recognizing a deficiency in the data/approach (i.e., mouse strain used for method validation is no longer available), the lead conducted additional laboratory studies using a widely accessible mouse strain. While this action strengthened the proposal, it was noted that the number of chemicals evaluated with the newly recommended mouse strain is highly limited (i.e., 15 chemicals). The experimental data obtained for assessment of the sub-categorisation criterion in the supporting publication (Kobayashi et al., 2020) are valid. However, testing more chemicals would bolster the proposal.

42. It is worth noting that ideally unrounded numbers would have been used in the analyses.

Overall, the PRP agreed that the Evaluation Principle 5 has been met.

9. Evaluation Principle 6. Ideally, all data supporting the validity of the analysis should have been obtained in accordance with the principles of GLP.

Charge Question 6: *Have all the data supporting the validity of the anlaysis obtained in accordance with the principles of GLP? If not, has an adequate consideration been given to the potential impact on the evaluation status of the proposed sub-categorisation criterion?*

43. The PRP considers that it is difficult to know to which extent the data supporting the validity of the analysis have been obtained in accordance with the principles of Good Laboratory Practices (GLP). However, the PRP agrees that it is reasonable to assume that data used and generated were following good scientific practices and that they do not suggest any bias in the execution of the analysis and experiment.

44. The data used to support the analysis in the proposal were taken from the 2010 ICCVAM LLNA:BrdU-ELISA test method validation report, which states that its studies were not fully GLP compliant. However, the quality of the data used to support this analysis were carefully reviewed prior to inclusion in the ICCVAM report, which was carefully constructed and extensively reviewed. For this reason, there are no major concerns regarding the selection of data used to support this analysis.

45. The supporting publications by Maeda and Takeyoshi (2019) and Kobayashi et al. (2020) do not make any claims of GLP compliance in the analyses, nor do they present any essential components of GLP principles which would include a detailed protocol describing the goals and procedures of the analyses, what specific data analyses and algorithms would be utilized, what criteria for a validation evaluation would be employed, and what analyses review and data integrity procedures would be employed.

46. It was noted that two transcription errors were found in Table 1 of the supporting publication of Maeda and Takeyoshi (2019); namely, the EC1.6 (%) values presented for 2-mercaptobenzothiazole and isopropyl myristate were incorrectly presented relative to the data originally presented in Table 3-1 of the Test Method Evaluation report on the LLNA:BrdU-ELISA (ICCVAM, 2010). For 2-mercaptobenzothiazole, an EC1.6 value of 12.097 was presented in Table 3-1 of the ICCVAM report, while a value of 12.907 was presented in Table 1 of the supporting publication of Maeda and Takeyoshi (2019). For isopropyl myristate, an EC1.6 value of 9.404 was presented in Table 3-1 of the ICCVAM report, while a value of 9.440 was presented in Table 1 of the supporting publication of Maeda and Takeyoshi (2019).

47. Whereas the error for the isopropyl myristate is unlikely to have affected the selection of the LLNA:BrdU-ELISA cut-off value, the value presented for 2-mercaptobenzothiazole was indeed near the EC1.6 threshold for Cat.1A and 1B discrimination. It is unclear whether the transcription error was an error in presentation in Table 1, or if the error proceeded initiation of the overall data analyses. The transcription errors call into question whether independent audits of the data analyses were conducted in accordance of the principles of the GLPs.

Overall, the PRP agreed that the Evaluation Principle 6 cannot be properly evaluated based on the limited information provided. However, the PRP acknowledges the credibility of the work presented given that it is published in two peer-reviewed articles.

10. Evaluation Principle 7. All data supporting the assessment of the validity of the analysis should be available for expert review.

Charge Question 7: *Do you consider that all the data supporting the assessment of the validity of analysis are easily available for expert review?*

48. The PRP agrees that all of the data used for the analyses are publically available and presented in Test Method Evaluation report on the LLNA:BrdU-ELISA (ICCVAM, 2010) and in the supporting publications of Maeda and Takeyoshi (2019) and Kobayashi et al (2020)

Overall, the PRP agreed that the Evaluation Principle 7 has been met.

11. Additional Remarks

Charge Question 8: *Do you have any suggestions or remarks to share?*

49. It was commented that conducting a formal evaluation based on only peer-reviewed publications was unusual. It would have been useful for the Panel to have the proposal in a format that addresses the evaluation principles and provides the lead's self-assessment on how those principles were met. Some evaluation principles (such as #6 on GLP) cannot be easily evaluated from publications.

50. It was questioned whether the acceptance of the proposal could lead to confusion for risk assessment. For example, if there are data from both the RI-LLNA and the LLNA:BrdU-ELISA, it is not clear which threshold value should be used to support risk assessment (i.e., EC3≤ 2% vs. EC1.6≤ 6% for Cat.1A sensitisers).

51. It would be interesting to compare the data with the recently revised allergen classification done by the OECD and see if the correct Cat.1A and 1B classification is supported.

12. Conclusions and recommendations

52. Overall, the PRP concluded that the proposal for new LLNA:BrdU-ELISA criterion for UN GHS skin sensitisation sub-categorisation is based on work that has been performed according to the validation principles described in OECD GD 34.

53. The PRP considers that the rational for using LLNA:BrdU-ELISA dataset for UN GHS skin sensitisation sub-categorisation is clearly elaborated in the proposal and supporting publications. The publications by Maeda and Takeyoshi (2019) and Kobayashi et al (2020) clearly presented the rationale for initiating the analyses as well as the follow-up study to evaluate the subcategorisation using the standard strain of CBA/J mice.

54. The PRP considers that the relationship between the outcome of interest, skin sensitisation classification and the test and the analytical methods were clearly described. However, given that the validation of using LLNA:BrdU-ELISA for sub-categorisation relies on its biological relationship with RI-LLNA, it was recommended to include some discussion of the potential for differences in the two nucleotide incorporation endpoints with respect to potential impact upon discriminating among skin sensitisers to justify the analyses.

55. The PRP considers that the process of setting the sub-categorisation criterion was well described. However, the PRP made recommendations for further clarifications on the borderline results, applicability domain and on the impact of variability of individual animals used for the study.

56. The PRP considers that the overall number of chemicals used for the analysis to be limited and that further assessment with a statistical tool (i.e., power analysis) to determine whether the number of chemicals was sufficient may be desirable. It was also recommended to employ individual data, instead of using mean values of the study, to understand variability of animals, which would help to further evaluate whether the proposed cut-off for sub-categorisation criterion was established properly or not. Although the PRP found the number of chemicals used for the analysis to be limited, it does acknowledge that the proposed criterion is based on the most complete and well-documented dataset available.

57. The PRP considers the approach employed to assess the performance of the analysis is relevant.

58. The PRP considers that it is difficult to know to which extent the data supporting the validity of the analysis have been obtained in accordance with the principles of Good Laboratory Practices (GLP). However, the PRP agrees that it is reasonable to assume that data used and generated were following good scientific practices and that they do not suggest any bias in the execution of the analysis and experiment.

59. The PRP agrees that all the data used for the analyses are publically available and presented in Test Method Evaluation report on the LLNA:BrdU-ELISA (ICCVAM, 2010) and in the supporting publications of Maeda and Takesyohi (2019) and Kobayashi et al. (2020).

13. Acknowledgements

The OECD Secretariat thanks the Peer Review Panel for their review and valuable discussions and comments.

Annex 1. Peer Review Panel Composition

Emanuela Corsini	Laboratory of Toxicology Department of Pharmacological and Biomolecular Sciences University of Milan Italy
Dori Germolec	Systems Toxicology Branch National Toxicology Program National Institute of Environmental Health Science U.S. National Institutes of Health
David Lehmann	Office of Research and Development U.S. Environmental Protection Agency
Hans Raabe	Institute for In Vitro Sciences, Inc. U.S.
M. Pilar Vinardell	Department of Biochemistry and Physiology Faculty of Pharmacy and Food Sciences University of Barcelona, Spain

Peer Review Manager	Eugene Choi (OECD Secretariat)

Annex 2. Charge Questions for the Peer Review of LLNA:BrdU-ELISA criterion for UN GHS Sub-categorisation

Charge Question 1: *Do you consider that the rationale for the proposal is clearly elaborated in the request for review from Japan together with the supporting publications?*
Charge Question 2: *Is the relationship between the dataset/analysis and the (biological) phenomenon of interest (skin sensitisation sub-categorisation) described adequately, using the relevant scientific references?*
Charge Question 3: *Do you consider that the process of setting the sub-categorisation criterion described in the supporting publication (Maeda and Takeyoshi, 2019) is sufficiently detailed (for example, description of what is measured, how it is measured, appropriate data analysis, decision criteria for evaluation of data and what are the criteria for acceptable test performance)? Does the proposed criterion provide a well-defined applicability domain? Is exclusion of data from the evaluation, if any, well justified?*
Charge Question 4: *Are the reference chemicals used to demonstrate the performance of this analysis representative of the types of substance for which the sub-categorisation criterion will be used? Do you consider that the number of chemicals used for setting the sub-categorisation criterion is sufficient?*
Charge Question 5: *In the context of the retrospective assessment, do you consider the approach employed to assess the performance of the analysis is relevant? Do you consider that the performance of the sub-categorisation criterion has been evaluated in relation to relevant information from the species of concern, and existing relevant toxicity testing data? Do you consider the experimental data obtained for assessment of the sub-categorisation criterion in the supporting publication (Kobayashi et al., 2020) are valid?*
Charge Question 6: *Have all the data supporting the validity of the anlaysis obtained in accordance with the principles of GLP? If not, has an adequate consideration been given to the potential impact on the evaluation status of the proposed sub-categorisation criterion?*
Charge Question 7: *Do you consider that all the data supporting the assessment of the validity of analysis are easily available for expert review?*
Charge Question 8: *Do you have any suggestions or remarks to share?*

Annex 3. Response Compilation

Peer Review of LLNA:BrdU-ELISA Criterion for UN GHS Skin Sensitisation Sub-Categorisation

Responses to the charge questions were collected on July – September 2021.

Charge Question 1: *Do you consider that the rationale for the proposal is clearly elaborated in the request for review from Japan together with the supporting publications?*
PR1: Yes, the rationale for the proposal is clearly elaborated in the request for review from Japan and supporting publications.
PR2: The rational of the proposal is clearly described in the two papers published in 2019 and 2020 provided at support. Although the LLNA: BrdU-ELISA was endorsed as OECD TG442B in 2010, this test method is not yet applied for GHS sub-categorization because the criteria for sub-categorization have not been proposed so far. Similarly, to what has been done for the standard radioisotopic LLNA (OECD TG429), where EC3 of 2% is the cut off value for 1A and 1B classification, based on re-analyse of the existing data of 32 sensitizers classified in the 1A or 1B categories of the GHS, the optimal criteria for GHS sub-categorization using LLNA: BrdU-ELISA (OECD TG442B) of skin sensitizers were identified (Maeda and Takeyoshi, 2019). In detail, criteria are: Category 1A: EC1.6 ≤6%, and Category 1B: EC1.6 >6%, showing the correct outcomes (%) for GHS 1A and GHS 1B category chemicals were 92.3 and 84.2 for all 32 chemicals, respectively (Maeda and Takeyoshi, 2019).
PR3: The rationale was clearly elaborated in the presentation made by Japan to the review group and in the supporting publications. The proposal describes the use of standardized criteria to determine GHS sub-categorization using the LLNA: BrdU-ELISA (OECD TG442B). Dr. Takeyoshi and colleagues have analyzed data for 32 sensitizers and using specific criteria (Category 1A: EC1.6 ≤6%, and Category 1B: EC1.6 >6%) have classified them into GHS category 1A or 1B. The results of these analyses were compared with previously identified GHS outcomes to obtain information on the accuracy of prediction and correlation with EC3 values. Although the data set was somewhat limited, correct outcomes (%) for GHS 1A and GHS 1B category chemicals were 92.3 and 84.2 for all 32 chemicals, respectively (Maeda and Takeyoshi, 2019).
PR4: The rationale for the proposal is clearly elaborated. The use of EC1.6 is well justified and allow the subcategorization of different sensitizers.
PR5: The rationale for conducting the retrospective evaluation of existing data to determine an appropriate cutoff for subcategorization was reasonably presented in the supporting publication Maeda and Takeyoshi (2019), as well as in the presentation to the Peer Review Panel on 12 July 2021 by Dr. Takeyoshi. Given that the RI-LLNA and LLNA:BrdU-ELISA are mechanistically based upon the same biological process and differ solely on the endpoint detection methodology (incorporation of BrdU as a surrogate to ^{3}H-thymidine incorporation), which was demonstrated originally in the prevalidation and validation of the LLNA:BrdU-

ELISA (ICCVAM, 2010), it is a reasonable hypothesis that the LLNA:BrdU-ELISA could further be used to discriminate between GHS Cat 1A and 1B substances in a similar manner as for the RI:LLNA.
Based upon the findings originally presented by Maeda and Takeyoshi (2019) and followed upon by Kobayashi et al. (2020), the rationale for initiating the analyses, as well as the follow up study to evaluate the subcategorization using the standard strain of CBA/J mice was clearly presented.

Charge Question 2: *Is the relationship between the dataset/analysis and the (biological) phenomenon of interest (skin sensitisation sub-categorisation) described adequately, using the relevant scientific references?*

PR1: Yes, the relationship between the dataset/analysis and skin sensitization sub-categorization is adequately described.

PR2: Both data set used and analysis are clearly described. As data set the ICCVAM Test Method Evaluation Report on the Murine Local Lymph Node Assay: BrdU-ELISA A Nonradioactive Alternative Test Method to Assess the Allergic Contact Dermatitis Potential of Chemicals and Products (ICCVAM, 2010) was used. In addition, 15 chemicals from the LLNA performance standard were tested (Kobayashi et al., 2020). To determine the optimal cutoff (%) for sub-categorization, the percentages of correct outcomes (%) and incorrect outcomes (%) obtained by following formulas for each candidate cutoff (%) value were examined from 1% to 10% by 1% interval.

PR3: The relationship between the outcome of interest, skin sensitization classification and the test method and analytical methods were clearly described. The ICCVAM Test Method Evaluation Report on the Murine Local Lymph Node Assay: BrdU-ELISA A Nonradioactive Alternative Test Method to Assess the Allergic Contact Dermatitis Potential of Chemicals and Products (ICCVAM, 2010) was cited to provide background information on the LLNA:BrdU assay and EC3 data for the test chemical data set. One of the publications submitted for the evaluation provided information on 15 chemicals tested as part of the development of the LLNA performance standard (Kobayashi et al., 2020). This data was generated in the CBA/J mouse strain as a comparison to data in the Maeda and Takeyoshi, (2019) studies, which evaluated 32 chemicals, but used the CBA/JN strain that is not as common in hypersensitivity testing in vivo.

PR4: It is adequately described

PR5: The biological relationship between the RI-LLNA and LLNA:BrdU-ELISA for skin sensitization hazard assessment was originally presented in the test method report on the prevalidation and validation of the LLNA:BrdU-ELISA (ICCVAM, 2010). The two methods are mechanistically based upon the same biological process and differ in the endpoint detection methodology (incorporation of BrdU as a surrogate to ^{3}H-thymidine incorporation). In the current analyses for subcategorization, Maeda and Takeyoshi (2019) and Kobayashi et al. (2020) present on the rationale for conducting the analyses, but they do not present initially the biological relationship between the RI-LLNA and LLNA:BrdU-ELISA for skin sensitization subcategorization, and whether the two test methods should provide the same discrimination. It would be useful to introduce and justify the biological relationship not only for identification of skin sensitization hazard, but also to provide a hypothesis for the intended purpose of subcategorization. Furthermore, some

discussion of the potential for differences in the two nucleotide incorporation endpoints should be presented with respect to potential impact upon discriminating among skin sensitizers to justify the analyses.

Charge Question 3: *Do you consider that the process of setting the sub-categorisation criterion described in the supporting publication (Maeda and Takeyoshi, 2019) is sufficiently detailed (for example, description of what is measured, how it is measured, appropriate data analysis, decision criteria for evaluation of data and what are the criteria for acceptable test performance)? Does the proposed criterion provide a well-defined applicability domain? Is exclusion of data from the evaluation, if any, well justified?*

PR1: The process of setting the sub-categorization criterion (description of what is measured, how it is measured, appropriate data analysis, decision criteria for evaluation of data, and the criteria for acceptable test performance) are provided. However, while the possibility of obtaining borderline positive values (i.e., SI 1.6 – 1.9) with the LLNA: BrdU ELISA is recognized by the requestors, the potential impact of borderline positive values on potency sub-categorization is not fully understood.

The proposed criterion expands the applicability domain to include potency sub-categorization. However, there is no guidance on the types of chemicals that the protocol is suitable for. ICCVAM considers the applicability domain of the LLNA: BrdU ELISA to be the same as the traditional radioisotope LLNA. Importantly, though, there are known limitations of the radioisotope LLNA (i.e., metals, mixtures, aqueous solutions) and results from the radioisotope LLNA and LLNA: BrdU ELISA do not always align. For example, unlike the traditional LLNA, the LLNA: BrdU-ELISA can be used for testing nickel compounds. The test guideline should include more information on the types of chemicals that can be evaluated using this specific assay.

The requestors were transparent about how they selected data for their evaluation (i.e., class 1A/1B sensitizers included in the ICCVAM 2010 report on the LLNA: BrdU ELISA).

PR2: The process of setting the sub-categorisation criterion is clearly described. In Maeda and Takeyoshi (2019), a total of 32 chemicals were included in the analysis: 13 1A, and 19 1B. Based on the analysis:

Determination of the optimal cutoff (%) for GHS sub-categorization

Proposed Cut-off (%)	GHS sub-category							
	1A				1B			
	CORRECT		INCORRECT		CORRECT		INCORRECT	
1	61.5%	(8/13)	38.5%	(5/13)	94.7%	(18/19)	5.3%	(1/19)
2	76.9%	(10/13)	23.1%	(3/13)	89.5%	(17/19)	10.5%	(2/19)
3	76.9%	(10/13)	23.1%	(3/13)	84.2%	(16/19)	15.8%	(3/19)
4	76.9%	(10/13)	23.1%	(3/13)	84.2%	(16/19)	15.8%	(3/19)
5	84.6%	(11/13)	15.4%	(2/13)	84.2%	(16/19)	15.8%	(3/19)
6	**92.3%**	**(12/13)**	**7.7%**	**(1/13)**	**84.2%**	**(16/19)**	**15.8%**	**(3/19)**
7	92.3%	(12/13)	7.7%	(1/13)	84.2%	(16/19)	15.8%	(3/19)
8	92.3%	(12/13)	7.7%	(1/13)	78.9%	(15/19)	21.1%	(4/19)
9	92.3%	(12/13)	7.7%	(1/13)	73.7%	(14/19)	26.3%	(5/19)
10	92.3%	(12/13)	7.7%	(1/13)	68.4%	(13/19)	31.6%	(6/19)

Ref.: Maeda Y, Takeyoshi M. (2019). Proposal of GHS sub-categorization criteria for LLNA: BrdU-ELISA (OECD TG442B). Regul Toxicol Pharmacol. 107:104409.

The cut off of 6% was chosen. The value of 2% resulted in a lower performance compared to the 6%.

In a subsequent study (Kobayashi et al., 2020), 15 GHS 1A/1B chemicals (6 1A and 9 1B) listed in the LLNA performance standard were tested in the accordance with OECD TG 442B using CBA/J mice, and the EC1.6 calculated and sub-categorized into GHS 1A or 1B using the 6% cut off. Using the round-off function of the EXCEL software for raw data, all 15 chemicals were correctly classified. Without round-off of raw data, 22% of 1B chemicals were misclassified as 1A (2 out of 9), while all 1A were correctly classified. Finally, the correlation between EC1.6 and EC3 was also calculated: strong correlation between the two ($r = 0.9076$, $P < .0001$) was confirmed. Among the chemicals with important difference diphenylcyclopropenone (EC3=0.05%, EC1.6=0.45%), 2-Mercaptobenzothiazole (EC3=1.7% and EC1.6=12.9%), cyclamen aldehyde (EC3=22.3%, EC1.6=41.5), and isopropyl myristate (EC3=44%, EC1.6=9.4%) can be mentioned. This difference maybe relevant when performing quantitative risk assessment for skin sensitization. Of course, this is not relevant for the current discussion on GHS classification. Among these chemicals, only one GHS 1A category chemical, 2-mercaptobenzothiazole, was assigned in GHS sub-category of 1B. The incorrect classification of this chemical may be due to the use of CBA/JN mice. Indeed, in the second paper (Kobayashi et al., 2020) where chemicals were tested in a different strain (CBA/J), 2-mercaptobenzothiazole was correctly classified as 1A (EC1.6=5.03%). When comparing the results of the 13 common chemicals tested with CBA/JN and CBA/J, subcategorization results using the proposed criterion were consistent, excluding 2-mercaptobenzothiazole, which may cause a strain-specific low response and was misclassified into GHS 1B in CBA/JN mouse (Maeda and Takeyoshi, 2019). The 2-mercaptobenzothiazole could be correctly subcategorized in GHS 1A when using the CBA/J mouse. Finally, three GHS 1B category chemicals, 3-aminophenol, trimellitic anhydride and nickel sulfate, were incorrectly assigned in GHS 1A category. This should not be considered relevant as in safety evaluation overprediction is better than underprediction. Overall, regarding the applicability of the criterion (EC1.6= 6%) seems to hold among different strains of mice.
PR3: The process of setting the sub-categorisation criterion is sufficiently detailed and well described. In Maeda and Takeyoshi (2019), the authors outline their methods for determining the optimal cutoff (%) for sub-categorization, the percentages of correct outcomes (%) and incorrect outcomes (%) for each candidate cutoff (%) value. They describe the systematic assessment of differing 1% intervals and how setting the criteria at specific values affected the performance. This data is very transparently summarized in a table in the paper. Using the suggested cut off value of 6%, the authors further examined the criteria by testing 15 GHS 1A/1B chemicals (6 1A and 9 1B) in accordance with OECD TG 442B using CBA/J mice. A high degree of correlation ($r = 0.9076$, $P < .0001$) was observed between EC3 values obtained in the LLNA-RI and the EC1.6 values obtained for the LLNA-BrdU ELISA and all 15 chemicals were correctly classified into GHS subcategory when the raw data were rounded off in EXCEL. In general, the results between the two different sets of analyses were consistent. The only discordant chemical was 2-mercaptobenzothiazole which was misclassified as GHS 1B in the 2019 paper. The authors suggest that this could be due to a strain-specific low response in the CBA/JN mice. The compound was correctly categorized as 1A in the second data set generated from CBA/J mice. Further discussion on the applicability

across strains is provided and suggests that EC values obtained in the CBA/Ca mouse strain are "almost equivalent" to those in the CBA/J strain when using the LLNA-RI, suggesting that subcategorization using the criteria for the LLNA-BrdU would be applicable to the CBA/Ca mouse strain.
PR4: The process of setting the sub-categorisation criterion is sufficiently detailed, except the exact criteria for the selection of the two or three doses used in the study of Kobayashi in the CBA/J strain. The only point is that not all the chemical categories are represented and I have concerns related to the complete applicability domain and possible limitations. To be sure about the applicability domain more than one chemical should be included in each chemical class.
PR5: The process of establishing the sub-categorisation cutoff was not adequately described as a prospectively designed protocol in the supporting publication Maeda and Takeyoshi (2019), as the cutoff was based upon empirical analyses of the existing data from the ICCVAM Test Method report on the LLNA:BrdU-ELISA (ICCVAM, 2010a). The publication of Maeda and Takeyoshi (2019) did not present on any prospective procedures or criteria for establishing the optimal cutoff, nor established criteria for optimal accuracy, sensitivity, specificity, false negative or false positive outcomes. The publication did provide details on which cutoff values would be evaluated (from an empirical review of the existing data set) and that the 10 cutoff values selected would be individually evaluated in 1% intervals to determine correct and incorrect outcomes, and from those analyses a determination was made which cutoff value was most ideal. No further algorithms were applied to correlate the RI LLNA data with the LLNA:BrdU-ELISA data. After initially identifying a cutoff of either 6% or 7% which presumably was selected to minimize under-predictions of the Cat 1A substances, Maeda and Takeyoshi (2019) presented on an evaluation of the impacts of four substances incorrectly predicted by the 6% cutoff value (i.e., 2-mercaptobenzothiazole which was under-predicted, and 3-aminophenol, trimellitic anhydride and nickel sulfate, which were over-predicted) and provided various suggested reasons for potential misclassification by the RI-LLNA. The publication referenced other data sets suggesting that the three over-predicted substances were either close to the EC3 cutoff of the RI-LLNA or incorrectly presented as a Cat 1B material in the current dataset. It was suggested that the under-predicted substance 2-mercaptobenzothiazole, was incorrectly predicted due to mouse strain differences impacting the LLNA:BrdU-ELISA. Whereas the reviewer considers it appropriate to evaluate the causes for mis-predictions, and in certain predefined circumstances exclude data, no similar analyses on the correctly predicted substances near the cutoff was conducted, which indeed should have been done. An analysis to determine the range of EC3 values for all of the substances near the putative cutoff should be conducted to best fit the available LLNA:BrdU-ELISA EC1.6 values with the ranges of RI-LLNA EC3 values. Furthermore, a more sophisticated analyses should be conducted by evaluating not only the mean data driving the single EC3 and EC1.6 values presented for each substance in the tables, but rather an analysis of the impact of variability of the individual animals should be performed.

Charge Question 4: *Are the reference chemicals used to demonstrate the performance of this analysis representative of the types of substance for which the sub-categorisation criterion will be used? Do you consider that the number of chemicals used for setting the sub-categorisation criterion is sufficient?*
PR1: Yes, the reference chemicals used to demonstrate the performance of this analysis representative of the types of substance for which the sub-categorisation criterion will be used. Per the 2010 ICCVAM validation report, "…considered the database of substances tested in the LLNA: BrdU-ELISA to be representative of a sufficient range of chemicals typically tested for skin sensitization potential.". The requestors used the most complete and well-documented dataset available (i.e., ICCVAM 2010). Still, compared to the number of chemicals used for setting the sub-categorization criterion for the radioisotope LLNA (category 1A = 21; category 1B = 49), which is not endorsed for sub-categorization by ICCVAM, the number of chemicals used for this analysis is rather small (i.e., category 1A = 13; category 1B = 19). Recognizing that the radioisotope LLNA and the LLNA: BrdU ELISA are not always concordant (e.g., nickel compounds) and the number of chemicals used for the analysis is small, providing data for more chemicals is preferred.
PR2: The chemicals used to demonstrate the perform of this analysis are representative of different chemistry and range of potency, thus, acceptable is the distribution among 1A and 1B sensitizers. At a glance, they include all sensitizers listed in the ICCVAM-recommended performance standards. However, the overall number of chemicals is limited as limited is the number of chemicals tested using CBA/J (n=15) or CBA/Ca (no data) mice. In Maeda and Takeyoshi (2019), CBA/JN mice were used, strain unavailable from the majority of providers of experimental animals at present. I am not a statistician, whom may provide an accurate estimation of the number of chemicals necessary (power analysis). However, even if further work may be desirable and more chemicals are always better, the criterion proposed is promising for GHS sub-categorization using TG442B.
PR3: The range of chemicals used to demonstrate the performance of these classification criteria are representative of different chemical classes and potencies. The overall number of chemicals represented in the analyses seems limited and it would be helpful to have power calculations or comparisons with other OECD analytical strategies to determine if the current number is sufficient.
PR4: The number of chemicals seems to be poor, specially for different chemical classes. The number of 1A and 1B should be more equilibrated. What happens with the borderline chemicals?

PR5: The analysis to define the threshold for discrimination between GHS Cat 1A and 1B (per Maeda and Takeyoshi, 2019) utilised all of the existing data presented in Table 3-1 of the 2010 ICCVAM Test Method Evaluation Report on the LLNA:BrdU-ELISA for skin sensitizers, and thus all of the same chemistries used originally to define the applicability domain were included in the current analysis. The 2010 ICCVAM Test Method Evaluation report "considered the database of substances tested in the LLNA: BrdU-ELISA to be representative of a sufficient range of chemicals typically tested for skin sensitization potential." Among this data set are 14 reference substances from the 22 ICCVAM-recommended LLNA performance standards reference substances (ICCVAM 2009), of which 6 are identified as GHS Cat 1A skin sensitizers and 8 as Cat 1B skin sensitizers. It should be stated for the record that certain chemistries are typically associated with reactive events resulting in skin sensitization. Whereas the original analysis included a wide range of chemistries to cover the spectrum of skin sensitizers and non-sensitizers, the current analysis is focused solely upon discriminating amongst skin sensitizers and thus the chemical representation is decidedly limited to this domain.

A total of 32 sensitizers comprised of 13 Cat 1A and 19 Cat 1B chemicals were included in the analyses (per Maeda and Takeyoshi, 2019). The original analyses reported in the 2010 ICCVAM report included a total of 43 chemicals, comprised of 32 GHS Cat 1 sensitizers and 11 non-categorized for skin sensitization, which were used to evaluate the ability of the test method to discriminate between sensitizers and non-sensitizers. The number of chemicals used in the current analyses to discriminate between GHS Cat 1A (13) and Cat 1B (19) were both greater than the number of non-sensitizer chemicals used in the original hazard analysis (11), and thus the current analysis has a somewhat greater number of chemicals used in establishing the discrimination threshold.

In the follow-on study to evaluate the performance of the LLNA: BrdU-ELISA in the frequently-used CBA/J strain of mice (Kobayashi, et al., 2020), a total of 15 skin sensitizers, comprised of 6 GHS Cat 1A and 9 GHS Cat 1B, from the ICCVAM-recommended LLNA performance standards reference substances (ICCVAM 2009) were subsequently tested in CBA/J mice.

When comparing the data of the LLNA: BrdU-ELISA relative to the reference test method Radioisotope LLNA, three Cat 1B chemicals were over-predicted and 1 Cat 1A chemical was under-predicted. These four chemicals were clustered quite close to the reference method EC3 threshold of 2%, and their EC3 values ranged from 1.7% to 4.8%, suggesting a limit to the ability to discriminate between the two hazard classes. This range represents <1% of the total dynamic range of the Radioisotope LLNA from the lowest value of 0.009% to the highest value of 47.5% represented in the ICCVAM 2010 data set. Two other correctly predicted chemicals were also within this narrow range, thus presenting that 6 of the 32 chemicals (18.8%)

Accordingly, it is the reviewer's opinion that a sufficient number of chemicals were selected for the analysis, and that chemicals selected for the analysis adequately reflected the chemical class domain represented by the LLNA test method performance standard reference chemicals.

Charge Question 5: *In the context of the retrospective assessment, do you consider the approach employed to assess the performance of the analysis is relevant? Do you consider that the performance of the sub-categorisation criterion has been evaluated in relation to relevant information from the species of concern, and existing relevant toxicity testing data? Do you consider the experimental data obtained for assessment of the sub-categorisation criterion in the supporting publication (Kobayashi et al., 2020) are valid?*
PR1: Yes, the approach employed to assess the performance of the analysis is relevant. Yes, the performance of the sub-categorisation criterion has been evaluated in relation to relevant information from the species of concern, and existing relevant toxicity testing data. To support their proposal, the requestors focused on chemicals for which there are data from both the radioisotope LLNA and LLNA: BrdU ELISA. Recognizing a deficiency in the data/approach (i.e., mouse strain used for method validation is no longer available), the requestors conducted additional laboratory studies using a widely accessible mouse strain. While this action strengthened their proposal, the number of chemicals evaluated with the newly recommended mouse strain is highly limited (i.e., 15 chemicals). Yes, the experimental data obtained for assessment of the sub-categorisation criterion in the supporting publication (Kobayashi et al., 2020) are valid. However, testing more chemicals would bolster their proposal.
PR2: In both studies, relevant reference sources were used (ICCVAM and the LLNA performance standard). It would be interesting to compare the data with the recently revised allergen classifications done by the OECD.
PR3: Yes, to all the above. My only concern is with the limited number of chemicals evaluated as compared to the total of existing data.
PR4: Number of 1A and 1B chemicals should be equilibrated.

Charge Question 6: *Have all the data supporting the validity of the analysis obtained in accordance with the principles of GLP? If not, has an adequate consideration been given to the potential impact on the evaluation status of the proposed sub-categorisation criterion?*
PR1: The data used to support this analysis were taken from the 2010 ICCVAM LLNA: BrdU ELISA test method validation report. As stated in the aforementioned report, the studies were not fully GLP compliant. While that is the case, the quality of the data used to support this analysis were carefully reviewed prior to

inclusion in the ICCVAM report, a report that was carefully constructed and extensively reviewed. For this reason, there are no major concerns regarding the selection of data used to support this analysis.
PR2: Even if GLP is not mentioned in Kobayashi et al. (2020), there is nothing that suggests bias in the execution of the experiments. Articles have been published in Peer reviewed journals, and they should considered acceptable.
PR3: There is no discussion of the principles of GLP in either supporting manuscript or consideration elsewhere. However, based on the available data, I do not think that it would impact the evaluation status of the proposed sub-categorization criterion.
PR4: There is no information about the following of the principles of GLP
PR5: The GLPs govern the both the generation of the data as well as the subsequent data analyses. It is not clear to the reviewer whether the original data presented in the Test Method Evaluation report on the LLNA: BrdU-ELISA (ICCVAM, 2010) were generated and reported in compliance with the GLPs, or at least in accordance with the principles of GLPs. The supporting publications by Maeda and Takeyoshi (2019) and Kobayashi, et al. (2020) do not make any claims of GLP compliance in the analyses, nor do they present any essential components of GLP principles which would include a detailed protocol describing the goals and procedures of the analyses, what specific data analyses and algorithms would be utilized, what criteria for a valid evaluation would be employed, and what analyses review and data integrity procedures would be employed. It is worth noting that two transcription errors were found in Table 1 of the supporting publication of Maeda and Takeyoshi (2019); namely, the EC1.6 (%) values presented for 2-mercaptobenzothiazole and isopropyl myristate were incorrectly presented relative to the data originally presented in Table 3-1 of the Test Method Evaluation report on the LLNA: BrdU-ELISA (ICCVAM, 2010). For the substance 2-mercaptobenzothiazole an EC1.6 value of 12.097 was presented in Table 3-1 of the ICCVAM report, while a value of 12.907 was presented in Table 1 of the supporting publication of Maeda and Takeyoshi (2019). For the substance isopropyl myristate an EC1.6 value of 9.404 was presented in Table 3-1 of the ICCVAM report, while a value of 9.440 was presented in Table 1 of the supporting publication of Maeda and Takeyoshi (2019). Whereas the error for the isopropyl myristate is unlikely to have affected the selection of the LLNA: BrdU-ELISA cutoff value, the value presented for 2-mercaptobenzothiazole was indeed near the EC1.6 threshold for Cat 1A and 1B discrimination. It is unclear whether the transcription error was an error in presentation in Table 1, or if the error preceded initiation of the overall data analyses. The transcription errors call into question whether the data analyses were conducted in accordance of the principles of the GLPs.

Charge Question 7: *Do you consider that all the data supporting the assessment of the validity of analysis are easily available for expert review?*

PR1: Yes, the data supporting the assessment of the validity of the analysis are easily available for expert review.
PR2: Yes, there is total transparency in the data presented and analysis conducted to me are valid.
PR3: Yes, all the data supporting the assessment were readily available to the panel in the publications and supplemental materials.
PR4: Yes
PR5: It appears that given the simplicity of the data analyses employed that all of the data used are publically available and presented in Test Method Evaluation report on the LLNA: BrdU-ELISA (ICCVAM, 2010) and in the supporting publications of Maeda and Takeyoshi (2019) and Kobayashi, et al. (2020).

Charge Question 8: *Do you have any suggestions or remarks to share?*
PR1: I am somewhat concerned that there will be some confusion if this proposal is accepted. For example, assuming there are data from both the radioisotope LLNA and the LLNA:BrdU ELISA, which threshold value should be used to support risk assessment (i.e., EC3 $\leq$ 2% vs EC1.6 $\leq$ 6% for 1A sensitizers)?
PR2: I have only two remarks: 1. Even if further work may be desirable as more chemicals are always better, the criterion proposed is promising for GHS sub-categorization using TG442B and it should be adopted proving that the number of tested compounds is sufficient (power analysis). 2. It would be interesting to compare the data with the recently revised allergen classifications done by the OECD and see if the correct 1A/1B classification is maintained.
PR3: While I have limited experience with these types of reviews, it seems unusual to base the evaluation only on peer reviewed publications. It would be useful for the panel to have the proposal in a format that somewhat addresses the evaluation principles and to have information from the submitting individual or group on how they feel they have addressed these principles. Some questions (such as #6) are not easily evaluated from publications.
PR4: As l mentioned before the number of chemicals and categories seems not enough. As is pointed in the ICCVAM Test Method Evaluation Report on the Murine Local Lymph Node Assay: BrdU-ELISA. A Nonradioactive Alternative Test Method to Assess the Allergic Contact

Dermatitis Potential of Chemicals and Products in 2010: Efforts should be made to identify additional human data
PR5: Overall, this reviewer agrees with the proposal to utilize the existing RI-LLNA and LLNA: BrdU-ELISA data to establish a relationship and determine empirically a cutoff for Cat 1A and 1B discrimination. The biological relationship between the RI-LLNA and LLNA:BrdU-ELISA for skin sensitization hazard assessment was originally presented in the test method report on the prevalidation and validation of the LLNA:BrdU-ELISA (ICCVAM, 2010), and thus provides a sound rationale for the subsequent analyses presented in the supporting publications. It is the reviewer's opinion that the use of the existing data provides a sound foundation to establish a putative cutoff value for the LLNA:BrdU-ELISA application. It is recommended that some discussion of the potential for differences in the two nucleotide incorporation endpoints be presented with respect to potential impact upon discriminating among skin sensitizers to justify the analyses. An analysis to determine the range of EC3 values for all of the substances near the putative cutoff should be conducted to best fit the available LLNA:BrdU-ELISA EC1.6 values with the ranges of RI-LLNA EC3 values. Furthermore, a more sophisticated analyses should be conducted by evaluating not only the mean data driving the single EC3 and EC1.6 values presented for each substance in the tables, but rather an analysis of the impact of variability of the individual animals should be performed.