EUROPEAN COMMISSION
JOINT RESEARCH CENTRE

Institute for Health and Consumer Protection
European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM)

ECVAM
SCIENTIFIC
ADVISORY
COMMITTEE
(**ESAC**)

ESAC OPINION

on the validation study of the epiCS® test method based on the EURL ECVAM/OECD Performance Standards for in vitro skin irritation testing using Reconstructed human Epidermis (RhE)

ESAC Opinion Nr.	2014-01
Relevant ESAC request Nr.	2014-01
Date of opinion	17.11.2014

TABLE OF CONTENTS

EUROPEAN COMMISSION
JOINT RESEARCH CENTRE

Institute for Health and Consumer Protection
European Union Reference Laboratory for Alternatives to Animal Testing (EURL ECVAM)

Ispra, 17.11.2014

1. Summary of the ESAC Opinion

ESAC was requested to evaluate the quality of a Performance Standards-based validation study on the epiCS® test method for skin irritation testing. The epiCS® test method is based on Reconstructed human Epidermis (RhE) and sufficiently similar with regard to the essential test method components to validated reference methods (e.g. EpiDerm, OECD TG439). The study was conducted by the test method supplier (CellSystems, Germany) and submitted to EURL ECVAM for evaluation and ESAC peer review. In agreement with the Performance Standards, the study addressed within and between laboratory reproducibility and predictive capacity on the basis of the 20 Reference Chemicals. Like most methods covered by TG439, the epiCS® test method (initially trademarked and evaluated as EST1000) was developed to discriminate between irritant (Category 2) and non-classified chemicals in order to allow classification of chemicals according to the United Nations Globally Harmonized System (GHS) either as "Category 2" or "no Category" respectively.

The study was conducted in two phases: In 2011 data from a completed ring trial had been submitted to EURL ECVAM. While the study data met the predictive capacity target values (sensitivity = 93%, specificity = 73%), within laboratory reproducibility was not acceptable in the three laboratories involved (in particular in the naïve laboratory) and between laboratory reproducibility was slightly below the target. The submitter identified poor proficiency (no training phase had been conducted) and lack of stringency of SOP execution (in particular washing/rinsing) as a likely main reason for poor reproducibility; additionally, issues with shipment were identified in case of the overseas laboratory. After improved training of the participating laboratories, all 20 chemicals were tested again in the naïve laboratory to assess whether proficiency indeed had caused within laboratory reproducibility issues. Moreover, the other two laboratories also re-tested those chemicals that had shown non-concordant predictions in view of generating supplementary information on the plausibility of the problems identified and effectiveness of the measures taken to address the problems. This data set showed that the measures significantly improved WLR in all three laboratories. In case of the naïve laboratory where a complete matrix had been re-tested, WLR now met the acceptance value. The study was submitted in 2013 to EURL ECVAM and forwarded by EURL ECVAM to ESAC for scientific peer review in 2014.

At its 40th plenary meeting in October 2014, ESAC discussed the study design and data sets submitted in 2011 and in 2013 and concluded that Performance Standard acceptable values for BLR, Specificity, Sensitivity and Overall Accuracy were met in the validation study. However, only the naïve laboratory satisfied the performance values relating WLR when re-testing all 20 reference chemicals in 2013. The two other laboratories only met performance values when grouping data from the 2011 testing with data from the re-testing in 2013. Although these were generated under slightly different conditions with regard to exactitude of SOP execution, this is not *a priori* a reason to reject such grouping which can indeed be a valid approach in agreement with the modular approach, in particular with respect to use of retrospective data. However, in this particular case, ESAC is of the opinion that grouping the data from 2011 and 2013 may introduce bias into the data, since only non-concordant data points have been re-tested with the possibility of generating better (=concordant) or worse (=invalid) data, but that the other data points relating to the remaining 14 and 13 chemicals have not been re-tested and hence have not be subjected to the possibility of generating results of lesser quality (e.g. more non-concordant or invalid runs / run sequences). Therefore values for WLR

for these two laboratories can only be derived from the 2011 data set which leaves them below the acceptance threshold, although evaluation of all data together tends to support that performance values for WLR can be attained.

For the final evaluation of the WLR and in view of deriving final robust figures for WLR that will support regulatory use of the test method, ESAC recommends to re-test the 13 and 14 reference chemicals, not re-tested in 2013 in the two other participating laboratories (apart from the naïve one) in order to confirm WLR to meet also the Performance Standard acceptable values.

1. Mandate of the ESAC

During its 38th meeting in June 2013 the ESAC plenary unanimously decided to charge two ESAC members to act as rapporteurs to prepare a scientific review of a Performance Standards-based study on the epiCS® RhE skin irritation test method.

The EURL ECVAM Request for ESAC advice mandating ESAC to evaluate the quality of the submitted validation study on the epiCS® addressing, in agreement with the Performance Standards, the within- and between-laboratory reproducibility as well as the predictive capacity of the epiCS®, was discussed during the 39st ESAC plenary (11.-12.03.2014) and endorsed by ESAC.

In particular, ESAC was requested to evaluate the complete data set (as submitted to EVCAM in 2013) taking account of an earlier study submission (2011 – TM2009-09) and advise whether the additional training measures that were introduced before generating the second data set resolved problems with respect to (mainly) within-laboratory reproducibility reported in 2011.

Finally, the ESAC was requested to review whether the overall performance of the method appears to be sufficient in view of its potential use for routine dermal irritation testing and hazard classification for regulatory purposes.

2. Detailed opinion of the ESAC

2.1 Background, regulatory and scientific rationale

The study objective was to show that the epiCS® test method (initially trademarked and evaluated as EST1000) satisfactorily and reproducibly discriminates between irritant and non-classified chemicals in order to allow classification of chemicals according to the United Nations Globally Harmonized System (GHS) either as "Category 2" or "no category" respectively.

The test submitter had conducted a first Performance Standards-based validation study of the epiCS® test method (then known as EST1000) in 2011 and submitted the findings to EURL ECVAM. While the study data met the target values relating to predictive capacity as outlined in the Performance Standards, it did not satisfy the criteria established for reproducibility: Within-Laboratory Reproducibility (WLR) was 16% below the target value in the naïve laboratory, and 6% below the target value in the other two laboratories; Between Laboratory Reproducibility (BLR) was 2% below the target value (see Table 1).

Following remedial action by the test method developer, to improve the transport of the test kits and to improve the technical proficiency of the laboratories, supplementary blinded testing and independent statistical evaluation was undertaken intended to assess whether the potential factors responsible for the shortcomings of the original ring-trial had been addressed. These data were submitted to EURL ECVAM in 2013.

2.2   Design and conduct of the study

A Performance Standards-based ("catch-up") validation study based on the relevant Performance Standards (ECVAM 2009 Performance Standards for in vitro skin irritation test methods based on Reconstructed human Epidermis (RhE) and the Performance Standards in Annex 2 of OECD TG 439 2013) was planned and conducted involving three laboratories that received coded chemicals for blind testing followed by independent statistical analysis.

Using the Performance Standards as the benchmark, the number of participating laboratories, the number of reference chemicals (20 reference chemicals suitably representative and balanced with respect to different chemical classes and physical states), the use of positive and negative controls, and the criteria test run acceptance and repeating test runs meet the requirements for a Performance Standard-based "catch-up" validation study of this nature.

However, a major flaw of the study design appears that no training/transfer and proficiency phase had been included in the 2011 study - with poor proficiency then seeming to be a plausible explanation for the variability seen in the 2011 data set.

For supplementary testing the naïve laboratory re-tested the original 20 reference chemicals, the other two laboratories re-tested only the six chemicals responsible for most of the reproducibility errors observed, and a seventh chemical which generated invalid run sequences, during the 2011 ring-trial.

In the 2013 submission revised WLR and BLR were calculated by substituting the new data for all three laboratories for the results obtained for the same chemicals in the same laboratories as reported in the 2011 submission to EURL ECVAM. It is on that basis revised reproducibility and accuracy figures were calculated and supplied to EURL ECVAM by the test developer in 2013.

2.3   Study results and conclusion

Whilst the initial Performance Standard-based study findings (2011) satisfied the performance values relating to Predictive Capacity (specificity, sensitivity, and accuracy), the WLR acceptance criteria were not met by any of the participating laboratories (deviations were 16% for the naïve laboratory and 6% for the other two laboratories), and the BLR performance (deviating by 2%) fell short of the minimum acceptable value (see Table 1).

Table 1: Reliability and predictive capacity of the epiCS® based on 2011 data set submission

Data set 2011	Sensitivity [%]	Specifity [%]	Accuracy [%]	WLR [%]	BLR [%]
Laboratory 1 (naïve) Laboratory 2 Laboratory 3				*74* *84* *84*	
All Laboratories	93	78	85		78
Performance Standard Minimum Acceptable Value	90	70	75	90	80

Analysis of the 2011 results identified the most plausible reasons for the poor reproducibility to be (1) poor proficiency training in all laboratories (2) a lack of attention to detail when crucial elements of the SOP were performed (in particular washing/rinsing procedure after exposure of the tissue equivalents to test chemicals) and (3) shipment-related problems with the test kits concerning one laboratory located overseas.

After remedial action by the manufacturer to improve the transport of the test kits to the test laboratories, and to improve the technical proficiency and compliance of the laboratories in particular with respect to how to execute crucial SOP steps (relating to chemical storage, test material application and, in particular, removal, i.e. rinsing procedure as well as appropriate separation of tissue kits exposed to different chemicals during the post-incubation period), supplementary testing of the full set of reference chemicals at the naïve laboratory, and re-testing of a subset of chemicals at the other laboratories, improved the reported reproducibility (see Table 2).

Table 2: Reliability and predictive capacity of the epiCS® based on 2013 data set submission

Data set 2013	Sensitivity [%]	Specifity [%]	Accuracy [%]	WLR [%]	BLR [%]
Laboratory 1 (naïve) Laboratory 2 Laboratory 3				95 100 95	
All Laboratories	97	80	88		95
Performance Standard Minimum Acceptable Value	90	70	75	90	80

For laboratory 1 re-testing appears to confirm that poor proficiency contributed to the poor WLR values in the 2011 report. For laboratory 2 and laboratory 3 re-testing appears to confirm that even experienced laboratories may require training and proficiency testing to ensure successful transfer of the protocol. Another confounding aspect in all laboratories may have been the improved post-treatment incubation step. During re-testing samples were separated for every test substance during

post incubation treatment in order to avoid contamination. Moreover, it is plausible that transport issues may have negatively impacted on the tissue quality of the 2011 study with regard to one laboratory located in the US, i.e. to which tissues needed to be shipped and imported.

However, in evaluating the 2013 submission, ESAC is concerned that the reported reproducibility and accuracy of the epiCS® test method set out in the 2013 Test Submission may not be the most reliable data set used to interpret the findings of this study with respect to reproducibility or reliability and that it is impossible to provide definitive figures for WLR and BLR that would appropriately characterise the performance of the epiCS® test method. When considering the available data generated under validation conditions, WLR of epiCS® may range from 74 to 100% and some uncertainty remains despite the re-testing due to possible bias introduced, i.e. the supplementary testing and data analysis conducted by the two "non- naïve" laboratories, may have introduced sources of bias into the data sets submitted to EURL ECVAM in 2013 as, effectively, good data were selected and only undesirable data subjected to additional testing. Interpreting the 2011 and 2013 data sets would have been considerably easier had the full set of chemicals been re-tested by all laboratories and not only by the naïve laboratory and would have allowed to determine a final finite figure to describe WLR.

The ESAC considers that a more appropriate data set on which to present and judge to the likely performance of the epiCS® method using the data available from this validation study is to combine and analyse the 2013 data from only the naïve laboratory (this element of the supplementary testing did address the most plausible avoidable errors introduced by the initial failure to incorporate formal training and proficiency testing and was justified) with the original 2011 data sets from the other two laboratories. On the basis of this re-calculation the only values which fall short of the minimum acceptable value (90%) set out in the Performance Standards are the WLR values produced for two of the three laboratories (84% and 84%) (see Fig. 3). However, supportive data is available also from these laboratories (i.e. from re-testing the six and seven chemicals) supporting the notion that, with appropriate training, appropriate levels of concordant predictions (WLR) can be attained by user laboratories.

Fig. 3: ESAC re- calculation based on complete test runs (data set from 2013 for laboratory 1, data set from 2011 for laboratory 2 and laboratory 3)

re-calculation 2011 / 2013	Sensitivity [%]	Specifity [%]	Accuracy [%]	WLR [%]	BLR [%]
Laboratory 1 (naïve) Laboratory 2 Laboratory 3				95 *84* *84*	
All Laboratories	95	79	87		88
Performance Standard Minimum Acceptable Value	90	70	75	90	80

Based on these considerations, the ESAC is of the opinion that WLR criteria may have been achieved had the original study design not been compromised by a failure to implement a transfer/training/proficiency phase, and had the full set of test chemicals been re-tested by all three laboratories after the potential proficiency issues been resolved.

On that basis ESAC accepts that the minimum Performance Standard acceptable values for WLR, BLR, Specificity, Sensitivity and Overall Accuracy can be met by trained and competent laboratories.

2.4   Recommendations

At its 40th plenary meeting in October 2014, ESAC discussed the study design and data sets submitted in 2011 and in 2013 and concluded that Performance Standard acceptable values for BLR, Specificity, Sensitivity and Overall Accuracy were met in the validation study. However, only the naïve laboratory satisfied the performance values relating WLR when re-testing all 20 reference chemicals in 2013. The two other laboratories only met performance values when mixing data from the 2011 testing with data from the re-testing in 2013, generated under different testing conditions. ESAC is of the opinion that mixing of data from 2011 and 2013 may introduce sources of bias into the data set submitted to EURL ECVAM in 2013. Therefore values for WLR for these two laboratories can only be derived from the 2011 data set and are considered to be below the acceptance threshold, although evaluation of all data together tends to support that performance values for WLR can be attained. For the final evaluation of the WLR, ESAC recommends to re-test also the remaining 13 and 14 chemicals so that, ultimately, the full set of chemicals is available also from laboratory 2 and laboratory 3 in view of assessing the final WLR values and judging whether these meet the Performance Standard acceptable value of 90%.

ESAC also suggests to amend the SOP in view of stressing

     (1)   the importance of an accurate rinsing procedure;
     (2)   to improve the description of the washing procedure in the SOP;
     (3)   that tissues treated with different test chemicals should not be placed next to each other neither during exposure nor during post-incubation.

3.   References

1. OECD, 2002. OECD Guideline for Testing of Chemicals, No. 404: Acute Dermal Irritation/Corrosion. Organisation for Economic Cooperation and Development, Paris.

2. Draize J.H., Woodand G. & Calvery H.O. 1944. Methods for the study of irritation and toxicity of substances applied topically to the skin and mucous membranes.Journal of Pharmacology and Experimental Therapeutics. 82, pp 337-390.

3. Botham, P.A., Earl, L.K., Fentem, J.H., Roquet, R., Van DE Sandt, J.J.M., 1998. Alternative methods for skin irritation testing: the current status, ECVAM Skin Irritation Task Force Report 1. ATLA 26, 195-211.

4. Van de Sandt, J., Roguet, R., Cohen, C., Esdaile, D., Ponec, M., Corsini, E., Barker, C., Fusenig, N., Liebsch, M., Benford, D., de Brugerolle deFraissinette, A., Fartasch, M., 1999 The Use of Human Keratinocytes and Human Skin Models for Predicting Skin Irritation. The Report and Recommendations of ECVAM Workshop 38. ATLA 27, 723-743.

5. Fentem, J.H., Briggs, D., Chesné, C., Elliott, G.R., Harbell, J.W., Heylings, J.R., Portes, P., Roguet, R., Van De Sandt, J.J.M., Botham, P.A., 2001 A prevalidation study on in vitro tests for acute skin irritation: results and evaluation by the Management Team. Toxicoloy in Vitro 15, 57-93.

6. ECETOC, 1995. Skin Irritation and Corrosion: Reference Chemicals Data Bank ECETOC Technical Report No. 66. European Centre for Ecotoxicology and Toxocology of Chemicals, Brussels.

7. Robinson, M.K., Cohen, C., de Brugerolle de Fraissinette, A., Ponec, M., Whittle, E., Fentem, J.H., 2002 Non-animal testing strategies for assessment of the skin corrosion and skin irritation potential of ingredients and finished products. Food and Chemical Toxicology 40, 573-592.

8. Zuang, V., Balls, M., Botham, P.A., Coquette, A., Corsini, E., Curren, R.D., Elliott, G.R., Fentem, J.H., Heylings, J.R., Liebsch, M., Medina, J., Roguet, R., Van De Sandt, J.J.M., Wiemann, C., Worth, A.P., 2002 Follow up to the ECVAM prevalidation study on in vitro tests for acute skin irritation, ECVAM Skin Irritation Task Gorce Report 2. ATLA 30, 109-129.

9. Portes, P., Grandidier, M.-H., Cohen, C., Roguet, R., 2002 Refinement of the Episkin® protocol for the assessment of acute skin irritation of chemicals: follow-up to the ECVAM prevalidation study. Toxicology in Vitro 16, 765-770.

10. Griesinger, C., Barroso, J., Zuang, V., Cole, T., Genschow, E. and Liebsch M. 2009. Explanatory background document to the OECD draft test guideline on in vitro skin irritation.
Online:
http://ihcp.jrc.ec.europa.eu/our_labs/eurl-ecvam/validation-regulatory-acceptance/topical-toxicity/skin-irritation
http://ihcp.jrc.ec.europa.eu/our_labs/eurl-ecvam/validation-regulatory-acceptance/docs-skin-irritation-1/DOC10-background%20document.pdf

11. Eskes, C., Cole, T., Hoffmann, S., Worth, A., Cockshott, A., Gerner, I. and Zuang, V. 2007. The ECVAM international validation study on in vitro tests for acute skin irritation: selection of test chemicals. ATLA 35, 603-619

12. Cotovio, J., Grandidier, M-H., Portes, P., Roguet, R., Rubinstenn, G. 2005. The In Vitro Acute Skin Irritation of Chemicals:Optimisation of the EPISKIN Prediction Model within the Framework of the ECVAM Validation Process. ATLA 33, 329–349.

13. Kanďárová H., Liebsch M., Genschow E., Ingrid Gerner, Traue D., Slawik B., Spielmann H. 2004: Optimisation of the EpiDerm Test Protocol for the upcoming ECVAM Validation Study on In Vitro Skin Irritation Tests. ALTEX 21, 3/04, 107-114.

14. Kanďárová, H., Liebsch, M., Gerner, I., Schmidt, E., Genschow, E., Traue, D., Spielmann, H., 2005. EpiDerm Skin Irritation Test Protocol – Assessment of the performance of the optimised test. ATLA 33 (4), p. 351-367.

15. Heylings, J.R., Diot,., Esdaile,D.J., Fasano,W.J., Manning, L.A., Owen H.M. 2003. A prevalidation study on the in vitro skin irritation function test (SIFT) for prediction of acute skin irritation in vivo: Results and evaluation of ECVAM Phase III. Toxicology in Vitro 17, 123–138.

16. Hoffmann J., Heisler E., Karpinski S., Losse J., Thomas D., Siefken W., Ahr HJ., Vohr HW., Fuchs HW 2005. Epidermal-skin-test 1000 (EST-1000)--a new reconstructed epidermis for in vitro skin corrosivity testing. Toxicol In Vitro. 2005 Oct;19(7):925-9. Epub 2005 Aug 2

17. Kandárová H., Hayden P., Spiller E., Klausner M., Kubilus J. and Sheasgreen J. 2007: In Vitro Skin Irritation Test: Increasing the Sensitivity of the EpiDerm Skin Irritation Protocol Evaluated in the ECVAM Skin Irritation Validation study. In: Toxicology Letters, Abstracts of the 44th Congress of the European Societies of Toxicology Volume 172, Supplement 1, 7 October 2007, Page 81.

18. Mossman, T., 1983. Rapid colorimetric assay for cell growth and survival application to proliferation and cytotoxicity assays. Journal of Immunological Methods 65, 55-63.

19. Faller, C., Bracher, N., Dami, N., Roguet 2002 Predictive ability of reconstructed human epidermis equivalents for assessment of skin irritation of cosmetics. Toxicology in Vitro 16 (5), 557-572.

20. Faller, C., Bracher, M., 2002. Reconstructed skin kits: Reproducibility of cutaneous irritating testing. Skin Pharmacol Appl Skin Physiol 15 (suppl 1), 74-91.

21. 22. EURL ECVAM, 2009. Performance Standards For In-Vitro Skin Irritation Test Methods Based On Reconstructed Human Epidermis. Online: http://ihcp.jrc.ec.europa.eu/our_labs/eurl-ecvam/validation-regulatory-acceptance/docs-skin-irritation-1/DOC8-updated_ECVAM_PS_2009.pdf

23. OECD, 2005 OECD Guidance Document 34 — Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment.

24. OECD, 2010. OECD Guideline for Testing of Chemicals, No. 439: In Vitro Skin Irritation: Reconstructed Human Epidermis Test Method

ESAC and ESAC Working Group charged with the scientific review

ECVAM Scientific Advisory Committee (ESAC)

- Prof. Jürgen BORLAK
- Dr. Neil CARMICHAEL
- Dr. Edward CARNEY
- Dr. Harvey CLEWELL
- Prof. Lucio G. COSTA
- Dr. Kristina KEJLOVÁ
- Prof. David John KIRKLAND
- Prof. Annette KOPP-SCHNEIDER
- Dr. Renate KRÄTKE
- Prof. Claus-Michael LEHR
- Dr. José Maria NAVAS
- Prof. Aldert PIERSMA
- Dr. Jonathan RICHMOND
- Dr. Erwin L ROGGEN
- Dr. Dorothea SESARDIC

ESAC Working Group:

- Dr. Jon Richmond (ESAC Member, WG Chair)
- Dr. Renate Krätke (ESAC Member)

EURL ECVAM:

- Dr. Claudius GRIESINGER (ESAC Coordinator)
- Dr. Michael SCHÄFFER