ESAC Peer review of submission of external catch up validation study of *In vitro* Reconstructed Human Epidermis (RHE) assay for Skin Irritation Testing

The above submission was sent to ECVAM in the Test Submission Template for ECVAM/CORRELATE submissions recently developed by ECVAM and was reviewed by Peer Review Panel (PRP):

1. Independent scientific evaluation of the submitted Test Submission Template (TST) by three members of the ECVAM Scientific Advisory Committee (Dr M. Dambrova, Dr D. Jirova and Dr C. Westmoreland) was performed.

The study was evaluated according to the principles outlined in the ECVAM document 'Performance Standards for Applying Human Skin Models to in-vitro skin irritation testing' (1). These performance standards are used to evaluate the reliability and accuracy of the methods which are based on reconstructed human epidermis and which measure or predict the same biological or toxic effect as the fully validated and accepted reference method (see ESAC statement, 2007 and skin irritation validation study [SIVS] report) (2,4). In this context, there was a request from ECVAM/CORRELATE for additional data, i.e. retesting of diethyl phthalate at test laboratory 3 (Oroxcell Laboratory) because of one invalid run and for additional statistical evaluation. In addition, minor formal corrections/clarifications in the submision file were required.

1. Consensus views from scientific evaluation of submission

All procedures are clearly defined, explained and described in satisfactory details. The description of participating units and management structure of the study gives impression of a fully managed and coordinated study which afforded reliable data collection and statistical analysis. Following comments received from ECVAM in June 2008 regarding the original TST submission, a new version of the TST was made available to the peer review panel during the review process (TST V2.0) which integrated changes made to the original document. These changes included retesting of diethyl phthalate at test laboratory 3 (Oroxcell Laboratory) and formal corrections.

1.1 Data Collection

Are the data collection procedures and selection clearly defined ? YES

The presented data are obtained according to GLP compliant SOPs in test laboratories which either are GLP certified or fully trained for procedures and the documentation practice required for this study. Data of the submission are sufficient to assess the study goal. The study followed ECVAM SIVS Performance Standards and also reference data from ECVAM SIVS were employed.

The following comments were made regarding the conduct of the study. It was not felt by the peer review panel that these comments significantly altered the overall summary (Section 3) of the panel:

- Regarding the SOP for the IL 1α analysis, the Sections '5. Data Report and Calculations' including '5.1. Calculations' and '5.2. Assay acceptance criteria' of the IL 1α SOP were compiled as an Amendment only after the end of the study (26th March, 2008). They were missing in the original SOP for the IL 1α analysis.

- The real Project Plan, which should be elaborated before the start of the study, is missing. The document called 'Protocol Amendement No:1 – Multicenter Study Plan' (MSP) was submitted to ECVAM on March 26, 2008, which means it was compiled only in course of the study or even after the end of the study. It is written in a fast and non-systematic form, missing good structure, pages are not numbered. It is signed by the principal investigator, but not approved by study participants. The reason for the attachment of the sheet named 'Protocol Approval' (done by all laboratories) to the MSP is not clear. However, the 'Multicenter Study Plan' lists study participants, there is a reference to the company Vitroscreen responsible for coding/distribution of 20 reference substances and description of the way of distribution. Company Vitroscreen provided the decoding to the contract statistician company **before** the statistical analyses and sent also decoding to the participating labs on Feb. 21, 2008 (specified in the TST together with the type of statistical analyses applied). It means, that the statistical analysis was not performed blinded.

1.2 Goal of the study

1.2.1 Was the goal of the study clearly understandable? YES

The goal of the study is understandable and clearly defined as evaluation of the predictive capacity of the assay submitted to reliably discriminate skin irritants (I) from non-irritants (NI), as defined with EU risk phrases (R38; no label) according to the Dangerous Substances Directive, 67/548/EEC. The study is submitted as an "External" Catch up validation study and employs Reconstructed Human Epidermis (RHE) model provided by the biotech company SkinEthic^{TM} (Nice, France), which mimics the three-dimensional (3D) structure of the human skin. The method aims to demonstrate that RHE model as an *in vitro* skin irritation assay is similar/equivalent to the *in vivo* rabbit test for skin irritation.

1.2.2 Is the scientific rationale given? YES

The scientific rationale of the assay is based on previously published data, which indicate that SkinEthic^{TM} RHE model is promising tool for the in vitro skin irritation and skin corrosion testing (publications by Kandárová et al., 2006). The SkinEthic^{TM} RHE model also presents a differentiation pattern comparable to EpiSkin^{TM} and EpiDerm^{TM} models and to normal human epidermis. Even though the submitted assay protocol has been developed recently, the RHE tissues are used routinely by several contract laboratories and companies, and the resulting data have been already used for assessing the skin irritancy potential of the test substances. By using SkinEthic^{TM} RHE model, the irritant substances are identified by their ability to decrease cell viability below the threshold of 50% level using the MTT assay. In addition, the inflammatory responses are evaluated by the quantification of the IL-1• assay.

The rationale is given and documented. The principle of the assay is based on the ability of irritant chemicals to penetrate the stratum corneum by diffusion and their subsequent cytotoxicity to living epidermal cells. Moreover, if cytotoxic effect is absent or weak, inflammatory mediators released might be considered in a tiered approach to increase the sensitivity of the test (similarly to the tiered approach scientifically validated for the EpiSkin^{TM} test assay, ESAC, May 2007). Irritant substances were identified by their ability to

decrease cell viability using the MTT below the defined 50% threshold level. The inflammatory responses were evaluated by the quantification of interleukin 1α using ELISA.

1.2.3 Is the regulatory rationale given? YES

The regulatory rationale is clearly defined, based on an urgent need for a full replacement for in vivo rabbit skin test according to Method B.4 of Annex V to Directive 67/548/ECC or OECD TG 404. The test is developed as a stand alone method for the assessment of the skin irritancy potential of the test substances and could be used for hazard identification and labelling of new and existing chemicals (including cosmetic raw materials and pharmaceuticals) according to the EU classification system (R38 or no label).

1.3. Test definition (Module 1)

1.3.1 Are the test and its purpose well defined? YES

The test and its purpose are clearly given . SkinEthic™ RHE assay is being proposed as a full replacement method of the in vivo rabbit skin irritation test for the assessment of dermal irritation according to the performance standards for applying human skin models to in vitro skin irritation testing, using the RHE model which is manufactured according to defined quality assurance procedures and commercialized by SkinEthic™ (Nice, France). SkinEthic™ RHE is a highly differentiated and stratified epidermis model comprising the main basal, supra basal, spinous and granular layers and a functional stratum corneum. The SkinEthic™ RHE model presents a histological morphology comparable to the *in vivo* human tissue and is used for skin irritation testing that involves topical application of test materials to the surface of the epidermis, and the subsequent assessment of their effects on cell viability.

The submitted SkinEthic™ RHE protocol employs a single exposure time of 42 minutes for both liquids and solids (powders) and a 42-hour post-exposure incubation before assessment of tissue viability. The twenty reference test substances were selected for test to give an adequate distribution of irritancy scores derived from *in vivo* rabbit skin irritation tests.The main endpoint used in the study is cell viability (measured as MTT reduction). Even though IL-1• was also measured, it did not turn out to improve the predictive ability of the assay. The provided SOP of the test method fully describes all procedures concerning tissue characteristics and handling, as well as application of test compounds and data processing. The mean relative tissue cell viability above 50 % was taken as a predictive measure for a non irritancy potential of the test substance.

The prediction model for the SkinEthic™ RHE assay is overtaken from the proposed Performance Standards for applying human skin models to *in vitro* skin irritation testing, in the ECVAM Skin Irritation Validation Study (4) and adequately serves the proposed purpose.

1.3.2 Are the proposed standardised protocol and prediction model adequate? YES

The detailed, GLP compliant SOP for MTT and IL 1α are provided. Training with 20 substances before the study is reported in the TST, but only partly documented in the Training Report of Coty and Oroxcell, no raw data are provided. The training report of Oroxcell is not adequately filled in. Training of L´Oréal laboratory is not documented at all.

The prediction model is fully adequate. The reconstructed human epidermal model (RHE) exhibits the structure of human epidermis. Human-derived epidermal keratinocytes have been cultured to form a differentiated model of human epidermis including basal, spinous and granular layers, and a multi-layered stratum corneum. RHE tissues contain 4 to 7 viable layers (approximate thickness of viable epidermis: 40-100 µm) which is typical for skin models and the thickness of the stratum corneum is approximately 20-60 µm. The SkinEthic RHE model contain 13.2 +/- 2.1% lipids. The stratum corneum (SC) and its lipid composition is sufficient to resist the rapid penetration of cytotoxic marker chemicals, e.g. SDS or Triton X-100. This property is estimated by the exposure time required to reduce cell viability by 50% (ET-50) upon application of 1% Triton X-100.

1.4 Data quality

The presented data are obtained according to GLP compliant SOPs in test laboratories which either are GLP certified or fully trained for procedures and the documentation practice required for this study. Data of the submission are sufficient to assess the study goal. The study followed ECVAM SIVS Performance Standards and also reference data from ECVAM SIVS were employed.

Evaluated data are of required quality, well described and discussed. Raw data from training experiments are not provided. Raw MTT and IL 1α data for 20 coded experimental substances are provided for all experimental runs of all laboratories, as well as for the transferability experiments (MTT endpoint) performed in week 48 with uncoded 20 substances.

1.4.1 Are they sufficient to assess the study goal? YES

The data presented in the TST for the SkinEthic RHE skin irritation test appear to be of very high quality. They are sufficient to assess the study goal.
The work was conducted to 'GLP-like' standards. The three of the participating institutions (L'Oreal, Coty Laboratories, Oroxcell) have either official GLP compliance as determined by AFSSAPS or have defined in-house quality systems. **Q.** It is unclear why the GLP compliance of the work at Oroxcell did not include evaluation of test substances and the report.

1.4.2 Quality of the reference data

Reference data from ECVAM SIVS were employed. Reference data, i.e. the classification data using the skin irritation effects in the rabbit, were used, as the rabbit represents the regulatory accepted species. Although the intention is to predict the human health hazard, neither in vivo nor in vitro skin irritation tests are currently calibrated to match human response.

1.5 Test materials

1.5.1 Is the number of evaluated substances sufficient ?

The 20 reference chemicals (10 skin irritants [R38 label], 10 non irritants [no label]) recommended by ECVAM as providing a representative distribution of the 58 chemicals used in the ECVAM international skin irritation validation study (Performance standards document: Skin irritation validation study, 2007) were used. This list allows comparison of results with those obtained originally with EpiSkin^{TM} in the Skin Irritation Validation study

(SIVS). The list includes two chemicals which were 'false positives' in the EpiSkin™ model (1-bromo-4-chlorobutane and 4-methyl-thio-benzaldehyde) and three 'false negatives' (hexyl salicylate, terpinyl acetate and dipropyl disulphide).

The independency in coding and distribution of substances were ensured in line with the study plan.

1.5.2 Are they representative of proposed applicability domain? YES

The selection and number of evaluated substances comply with the requirements of the ECVAM SIVS Performance Standards.

Applicability domain of the SkinEthic™ RHE assay is the same as for both models (EpiSkin™, EpiDerm™) that were the subject of validation under the ECVAM SIVS. Consequently, the information regarding the method applicability on volatiles, emulsions, mixtures, hydrolyzing and polymerizing chemicals, acids and bases is considered as insufficient. SkinEthic should not be considered validated for such classes of chemicals. For chemicals that directly reduce MTT, correction techniques were developed.

1.6 Within-laboratory variability (Module 2) – assessment of reproducibility of the data in the same laboratory

The within-laboratory variability was properly assessed. The within-laboratory variability was assessed for each laboratory by means of the assessment of the frequency of non qualified experimental runs, by one-way ANOVA statistics, by analysis of the within-laboratory standard deviation, by calculation of Bravais-Pearson correlation of the mean cell viability for the three pairs of runs and by evaluation of the proportion of identically classified test substances.

Regarding the MTT endpoint, only one laboratory (Oroxcell) for one test substance showed between-runs SD > 18 as unacceptable. Only in this laboratory was 1 out of 20 tested substances not consistently classified. For L´Oréal and COTY all 20 substances were consistently classified. Comparison of variability via the cumulative distribution of the SD of all runs in the three laboratories revealed the highest between-runs variability at Oroxcell and the lowest at COTY. However, the overall results for MTT endpoint for twenty reference test substances demonstrated a low variability for all laboratories. Additionally submitted data, that were generated in July 2008 on the ECVAM request for substance No.2 (diethyl phthalate), proved compliance with results from previous 2 valid runs. Data for 3 valid runs were recalculated. In this context, the PRP expressed the opinion, that for future validation studies a clear guidance should be given from ECVAM/CORRELATE on how many invalid runs can be repeated and not included in calculations.

Regarding the IL-1α, a substantial within-laboratory variability was observed within assay and between runs for experimental chemicals and for PC and NC. Between six to nine out of 20 test substances had a p-value below the level of 1% indicating significant differences between the independent runs. Overall, the results on the within-runs variability using the twenty reference test substances seemed to demonstrate variability for all laboratories using the RHE assay.

1.7 Transferability (Module 3) - how easy is it to transfer the tests to a second laboratory?

Transferability of the SkinEthic™ RHE assay was assessed during the training that included 3 test substances (2 liquids and 1 solid) according to template of method transfer (1 day or 3 days training) that was established within SkinEthic™. Both Oroxcell and L'Oréal laboratories were naïve, having never used the SkinEthic's protocol. Furthermore, Coty Laboratory routinely used the RHE model to assess skin irritancy potential, but was not familiar with SOP for the protocol of 42 min exposure time and 42-hour post-exposure incubation. The training exercise gave evidence that SkinEthic™ RHE assay can be efficiently transferred. Thus, the within-laboratory variability was acceptable (SD smaller than 18 for all tested substances, and positive and negative controls met the acceptance criteria in MTT assay). However, the IL-1• assay showed a highest variability. This allowed identifying MTT test cell viability parameter as a relevant endpoint, whereas the IL-1• might be a questionable endpoint for classification of substances.

1.8 Between-laboratory variability (Module 4) - assessment of reproducibility of the data in different laboratories

Using the proportion of identically classified tests substances as a measure of between laboratory reproducibility, the results showed that all of the 20 test substances were identified identically (R38 or no-label) in the three laboratories.

Between laboratory variability was also assessed using the ANOVA/t-test p values and the ANOVA sum of squares of the MTT results. 19 out of the 20 tested sustances produced results that were reproducible between laboratories. Between laboratory standard deviations for the MTT results for all 20 test chemicals were <18%. The data reproducibility was properly assessed.

Concerning the MTT endpoint, and using the 1-way ANOVA evaluation, 19 out of 20 tested chemicals were reproducible between the three laboratories. Analysis of SD betwen runs showed that none of the 20 substances exhibited a SD larger than 18. Results for all substances were highly reproducible between the three labs. All test substances were identically identified in the three labs in three valid runs.

Regarding the values obtained for IL-1• a lack of standardisation of this secondary endpoint was seen, irregular low values for irritants and high values for non-irritants can be seen in Table 43 (pg.78 of TST).

.

1.9 Predictive capacity (Module 5)

1.9.1 Has the predictive capacity of the methods been properly assessed ? YES – but further statistical anaylsis is required

Regarding the MTT endpoint, three out of twenty test substances were consistantly misclassified in all laboratories in all runs. Two non-irritants (1-bromo-4-chlorobutane and 4-methyl-thio-benzaldehyde) were classified as irritants and one irritant was classified as non-irritant (hexyl salicylate). In addition, the Oroxcell laboratory misclassified one more

substance (allyl phenoxy-acetate) in one run with no impact on the median approach classification. Taking all individual classifications (n=93) into account or considering only three valid runs (n=90) a similar specificity of 80% and sensitivity of 90 % was achieved using the MTT endpoint.

Regarding the IL-1• endpoint, the predictive capacity was evaluated analysing the real IL-1• values, the IL-1• values corrected by NC and fold increase of the IL-1• values above 5. In general, none of the MTT non-irritant test substances exhibited a release of IL-1• above 60 pg/mL (resp. 50 pg/ml) and none of the MTT non-irritant test substances showed fold increase above 5 of released IL-1•. However, high between-run variability in IL-1• values and SD for non-irritants can be seen in Tabs.68-72 for the Oroxcell laboratory. In addition, one of the MTT classified non-irritant substances exhibited a fold increase above 5 in all 3 runs (Tab.74 of TST) in the Oroxcell laboratory. The irritant classified substances tended irregularly to induce higher IL-1• amount. No additional correctly classified irritants as positives were identified using the IL-1• endpoint. The specificity of 80% and sensitivity of 90% was recognized independently of the added IL-1• parameter compared to those obtained using the MTT parameter only. When considering the mean IL-1• fold increase release, two of the nine MTT identified non-irritants were missclassified as irritants in one laboratory in one run. For the combined prediction based on MTT and the IL-1• fold increase, the classification reached a sensitivity of 80% and specificity of 76.7%. On contrary to performance of the EpiSkin^{TM} test in the SIVS (3), the predictive ability of the SkinEthic^{TM} RHE assay was not improved by introduction of the IL-1 *a* endpoint.

1.9.2 Is the assay relevant for its stated purpose? Yes.

No problem was recognized using the two SOP protocols (MTT and IL-1• detailed procedures). Overall results of the within-laboratory variability show SD smaller than 18 for all tested substances using the MTT parameter whereas it was obvious that the IL-1• showed the highest variability. Concordance of results for positive and negative controls in all laboratories was shown, which met the acceptance criteria. The concordance in the reliability and predictivity using the MTT parameter was proved as well. The MTT cell viability parameter was identified as a relevant endpoint, whereas the IL-1• might be a questionable endpoint for classification of a given test substance according to the EU classification (R38, no label). The MTT results of the SkinEthic^{TM} assay are recognized as equivalent to the EpiSkin^{TM} test. The sensitivity of the SkinEthic^{TM} assay is equal or higher than 70%, and the specificity is equal or higher than 80%, shown for the EpiSkin^{TM} test.

10. Applicability domain (Module 6) - Is the proposed applicability domain well defined ? YES

The applicability domain comprises chemicals as liquids and solids, not gases, vapours or aerosols. Quickly hydrolysing or polymerising chemicals or chemicals with other type of degradation were not included in the ECVAM SIVS. Consequently, for this type of chemicals also the validity of the SkinEthic^{TM} method cannot be considered as assessed. Highly volatile substances and chemicals that react with the plastic material of the cell culture inserts may provide higher levels of variability and may give inconsistent results.

1.11 Performance standards (Module 7) - Have appropriate performance standards been defined for the test ? YES

The ECVAM SIVS Performance Standards Document was exactly followed in this study. However, the appearance of heptanal on the list of standards is surprising due to the fact that during the ECVAM SIVS it was tested in one laboratory only and only during the test optimization study on EpiSkin™. In addition, a possibly wrong calculation of the dominating rabbit skin irritation median value for heptanal was identified (3.35 instead of 4).

Regarding the number of acceptable invalid runs and the possibility to repeat the invalid run, the PRP expressed the opinion, that for future validation studies a clear guidance should be given from ECVAM/CORRELATE on how many and how many times the invalid runs can be repeated and not included in calculations. Such specification should be included in "modified" ECVAM SIVS Performance Standards Document.

It is suggested, that existing human data should not be ignored, as they provide valuable toxicological information. A considerable number of chemicals classified R38 by the rabbit test do not cause irritation in human skin in vivo. Reconstructed human skin models consist of cells of human origin. They were developed to predict human skin hazard. That is why they classify some of rabbit irritants as human non-irritants (e.g. Hexyl salicylate is one of the R38 reference chemicals of the PS document which was repeatedly shown to be non-irritating to humans).

Although volatile and rapidly changing/decomposing chemicals should be excluded from the list of standards, some of the standard substances exhibit these characteristics (e.g. 1-bromohexane, 4-methyl-thio-benzaldehyde, a-terpineol).

1.12 Readiness for regulatory purposes

The method is ready for regulatory purposes using as the main endpoint the cell viability (MTT reduction), with a threshold of 50% viability. Interleukin 1 alpha (IL-1 *a*) was also measured to determine if this additional endpoint would improve the predictive ability of the assay. For SkinEthic™ RHE model, a sensitivity of 90% and specificity of 80% was identified for the MTT only. It was not improved by IL-1 *a* endpoint. The overall accuracy was 85%. It can be suggested as a stand alone in vitro toxicity test to replace the Draize rabbit skin irritation test for classification of skin irritants and non-irritants.

2. Further statistical analysis requested from ECVAM

Within ECVAM document 'Performance Standards for Applying Human Skin Models to in-vitro skin irritation testing' there is a requirement that 'the sensitivity of a 'me too' test must be higher than 70%, and the specificity must be equal to or higher than 80%'. These figures are based on the analysis of the 20 reference chemicals in the SIVS with EpiSkin™ using median test results from the dataset.

It is therefore important when quoting sensitivity and specificity results for a new test that the same statistical methodology is used as was used when defining these requirements within the Performance Standards for Applying Human Skin Models to in-vitro skin irritation testing.

The peer review panel therefore asked ECVAM to recalculate sensitivity, specificity and accuracy figures for the SkinEthic™ RHE test using exactly the same methodology (i.e. using median test results) to that used for the same 20 chemicals in the SIVS using the EpiSkin™

test. These results are presented in Appendix 1 (EpiSkin™) and 2 (SkinEthic™ RHE test) and summarised below

	EpiSkin™ (SIVS)	SkinEthic™ RHE test
Sensitivity	70%	90%
Specificity	80%	80%
Accuracy	75%	85%

3. Overall Summary

Given the consistently positive comments within the initial scientific peer review of the submission of the follow-up validation study of the SkinEthic™ RHE assay t together with the additional statistical analysis of sensitivity, specificity and accuracy, the combined veiws of the peer review panel can be summarised as:

The performance of the SkinEthic™ RHE assay met the criteria outlined to be considered to have sufficient accuracy and reliability for prediction of R38 skin irritating and no-label (non-skin irritating) test substances compared to the validated and accepted method. Limitations associated with the previously validated and accepted in-vitro reference method 1 (EpiSkin™) for skin irritation e.g. applicability domain (1, 4) also apply to the SkinEthic™ RHE assay.

Dr M. Dambrova,
Dr D. Jirova
Dr C. Westmoreland

Date : November 20, 2008

References

1. ECVAM (2007) *Performance Standards for Applying Human Skin Models to In Vitro Skin Irritation Testing.* Online: http://ecvam.jrc.ec.europa.eu
2. ECVAM (2007) *Statement of the ECVAM Scientific Advisory Committee (ESAC) on the Validity of In Vitro Tests for Skin Irritation.* Online: http://ecvam.jrc.ec.europa.eu
3. ECVAM (2007) *Skin Irritation Validation Study Phase II: Analysis of the primary endpoint MTT and the secondary endpoint IL-1•* . Online: http://ecvam.jrc.ec.europa.eu.
4. Spielmann, H., Hoffmann, S., Liebsch, M., Botham, P., Fentem, J., Eskes, C., Roguet, R., Cotović, J., Cole, T., Worth, A., Heylings, J., Jones, P., Robles, C., Kandárová, H., Gamer, A., Remmele, M., Curren, R., Raabe, H., Cockshott, A., Gerner, I. and Zuang, V. (2007) *The ECVAM International Validation Study on In Vitro Tests for Acute Skin Irritation: Report on the Validity of the EPISKIN and EpiDerm Assays and on the Skin Integrity Function Test.* ATLA **35,** 559-601.

Appendix 1: Statistical analysis of sensitivity, specificity and accuracy of EpiSkin™ test using 20 reference chemicals tested in the full prospective exposure study

Individual Laboratory Predictions for the 20 Reference Chemicals ONLY. Used for calculating the predictive values (sensitivity, specificity, accuracy)

Nr.	Chemical	EU Class	L'Oréal	Unilever	Sanofi	Median	In Vitro Prediction	
2	**1-bromo-4-chlorobutane**	no label	1	1	1	1	I	FP
22	Diethyl phthalate	no label	0	0	0	0	NI	
24	di-propylene glycol	no label	0	0	0	0	NI	
41	Naphthalene acetic acid	no label	0	0	0	0	NI	
11	Allyl phenoxy-acetate	no label	0	0	0	0	NI	
36	Isopropanol	no label	0	0	0	0	NI	
8	**4-methyl-thio-benzaldehyde**	no label	0	1	1	1	I	FP
39	Methyl stearate	no label	0	0	0	0	NI	
10	Allyl heptanoate	no label	0	0	0	0	NI	
33	Heptyl butyrate	no label	0	0	0	0	NI	
34	**Hexyl salicylate**	R38	0	0	0	0	NI	FN
55	Terpinyl acetate	R38	0	1	0	0	NI	FN
58	Tri-isobutyl phosphate	R38	1	1	1	1	I	
4	1-decanol	R38	1	1	1	1	I	
20	Cyclamen aldehyde	R38	1	1	1	1	I	
3	1-bromohexane	R38	1	1	1	1	I	
15	a-terpineol	R38	1	1	1	1	I	
23	**di-n-propyl disulphide**	R38	0	1	0	0	NI	FN
18	Butyl methacrylate	R38	1	1	1	1	I	
XXX	Heptanal	R38	1	.	.	1	I	

1) CALCULATION ON THE BASIS OF ALL LABORATORY PREDICTIONS

30 individual predictions for NEGATIVES
but only 28 for positives = total of 58 predictions

			absolut	percent
SENSITIVITY				
Sensitivity = TP / (TP+FN)		TP	21	75
Sensitivity	**75**	FN	7	**25**
		Sum	**28**	
SPECIFICITY				
Specificity = TN / (TN + FP)		TN	25	83.3
Specificity	**83.3**	FP	5	**16.7**
		Sum	**30**	
ACCURACY				
Sum (TP + TN) / (TP+TN+FN+FP)				
Accuracy	**79.3**			

1a) CALCULATION ON THE BASIS OF ALL LABORATORY PREDICTIONS

Including 3 concordant laboratory predictions for Heptanal (same weighting as for the other chemicals)

30 individual predictions per class = total of 60 predictions

			absolut	percent
SENSITIVITY				
Sensitivity = TP / (TP+FN)		TP	23	76.7
Sensitivity	**76.7**	FN	7	**23.3**
		Sum	**30**	
SPECIFICITY				
Specificity = TN / (TN + FP)		TN	25	83.3
Specificity	**83.3**	FP	5	**16.7**
		Sum	**30**	
ACCURACY				
Sum (TP + TN) / (TP+TN+FN+FP)				
Accuracy	**80.0**			

2) CALCULATION ON THE BASIS OF THE FINAL DECISION MAKING (MEDIAN)

SENSITIVITY

Sensitivity = TP / (TP+FN)

Sensitivity **70** ***-> led to PS Value for me-too's***

SPECIFICITY

Specificity = TN / (TN + FP)

Specificity **80** ***-> led to PS Value for me-too's***

ACCURACY

Sum (TP + TN) / (TP+TN+FN+FP)

Accuracy **75**

10 individual predictions per class

	absolut	percent
TP	7	70
FN	3	**30**
Sum	**10**	

TN	8	80
FP	2	**20**
Sum	**10**	

Appendix 2: Statistical analysis of sensitivity, specificity and accuracy of the SkinEthic™ RHE Test

Individual Laboratory Predictions for the 20 Reference Chemicals used for calculating the predictive values (sensitivity, specificity, accuracy)

No	Chemical	*In vivo* score[§]	EU label	EU GHS	Lab1	Lab 2	Lab 3	Overall	
1	1-bromo-4-chlorobutane	0	0	0	1	1	1	1	FP
2	diethyl phthalate	0	0	0	0	0	0	0	
3	di-propylene glycol	0	0	0	0	0	0	0	
4	naphthalene acetic acid	0	0	0	0	0	0	0	
5	allyl phe0xy-acetate	0.3	0	0	0	0	0	0	
6	isopropa0l	0.3	0	0	0	0	0	0	
7	4-methyl-thio-benzaldehyde	1	0	0	1	1	1	1	FP
8	methyl stearate	1	0	0	0	0	0	0	
9	allyl hepta0ate	1.7	0	0	0	0	0	0	
10	heptyl butyrate	1.7	0	0	0	0	0	0	
11	hexyl salicylate	2	1	0	0	0	0	0	FN
12	terpinyl acetate	2	1	0	1	1	1	1	
13	tri-isobutyl phosphate	2	1	0	1	1	1	1	
14	1-deca0l	2.3	1	1	1	1	1	1	
15	cyclamen aldehyde	2.3	1	1	1	1	1	1	
16	1-bromohexane	2.7	1	1	1	1	1	1	
17	a-terpineol	2.7	1	1	1	1	1	1	
18	di-n-propyl disulphide	3	1	1	1	1	1	1	
19	butyl methacrylate	3	1	1	1	1	1	1	
20	heptanal*	4	1	1	1	1	1	1	

1) CALCULATION ON THE BASIS OF ALL LABORATORY PREDICTIONS

40 individual predictions per class

				absolut	percent	
SENSITIVITY						
Sensitivity = TP / (TP+FN)			TP	27	90	
Sensitivity	**90**	***ok (• 70%)***	FN	3	**10**	***ok (• 30%)***
			Sum	**30**		
SPECIFICITY						
Specificity = TN / (TN + FP)			TN	24	80	
Specificity	**80**	***ok (• 80%)***	FP	6	**20**	***ok (• 20%)***
			Sum	**30**		
ACCURACY						
Sum (TP + TN) / (TP+TN+FN+FP)						
Accuracy	**85**	***no values provided in the PS***				

2) CALCULATION ON THE BASIS OF THE FINAL DECISION MAKING (MEDIAN)

10 individual predictions per class

				absolut	percent	
SENSITIVITY						
Sensitivity = TP / (TP+FN)			TP	9	90	
Sensitivity	**90**	***ok (• 70%)***	FN	1	**10**	***ok (• 30%)***
			Sum	**10**		
SPECIFICITY						
Specificity = TN / (TN + FP)			TN	8	80	
Specificity	**80**	***ok (• 80%)***	FP	2	**20**	***ok (• 20%)***
			Sum	**10**		
ACCURACY						
Sum (TP + TN) / (TP+TN+FN+FP)						
Accuracy	**85**	***no values provided in the PS***				