

Unclassified

ENV/JM/MONO(2013)14

Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development

03-Jul-2013

English - Or. English

**ENVIRONMENT DIRECTORATE
JOINT MEETING OF THE CHEMICALS COMMITTEE AND
THE WORKING PARTY ON CHEMICALS, PESTICIDES AND BIOTECHNOLOGY**

**SUMMARY DOCUMENT ON THE STATISTICAL PERFORMANCE OF METHODS IN OECD TEST
GUIDELINE 431 FOR SUB-CATEGORISATION**

**Series on Testing and Assessment
No. 190**

JT03342677

Complete document available on OLIS in its original format

This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.



**ENV/JM/MONO(2013)14
Unclassified**

English - Or. English

OECD Environment, Health and Safety Publications

Series on Testing and Assessment

No. 190

**SUMMARY DOCUMENT ON THE STATISTICAL PERFORMANCE OF METHODS IN OECD
TEST GUIDELINE 431 FOR SUB-CATEGORISATION**

IOMC



INTER-ORGANIZATION PROGRAMME FOR THE SOUND MANAGEMENT OF CHEMICALS

A cooperative agreement among FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD

Environment Directorate

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

Paris 2013

ABOUT THE OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation in which representatives of 34 industrialised countries in North and South America, Europe and the Asia and Pacific region, as well as the European Commission, meet to co-ordinate and harmonise policies, discuss issues of mutual concern, and work together to respond to international problems. Most of the OECD's work is carried out by more than 200 specialised committees and working groups composed of member country delegates. Observers from several countries with special status at the OECD, and from interested international organisations, attend many of the OECD's workshops and other meetings. Committees and working groups are served by the OECD Secretariat, located in Paris, France, which is organised into directorates and divisions.

The Environment, Health and Safety Division publishes free-of-charge documents in 11 different series: **Testing and Assessment; Good Laboratory Practice and Compliance Monitoring; Pesticides; Biocides; Risk Management; Harmonisation of Regulatory Oversight in Biotechnology; Safety of Novel Foods and Feeds; Chemical Accidents; Pollutant Release and Transfer Registers; Emission Scenario Documents; and Safety of Manufactured Nanomaterials.** More information about the Environment, Health and Safety Programme and EHS publications is available on the OECD's World Wide Web site (www.oecd.org/chemicalsafety/).

This publication was developed in the IOMC context. The contents do not necessarily reflect the views or stated policies of individual IOMC Participating Organisations.

The Inter-Organisation Programme for the Sound Management of Chemicals (IOMC) was established in 1995 following recommendations made by the 1992 UN Conference on Environment and Development to strengthen co-operation and increase international co-ordination in the field of chemical safety. The Participating Organisations are FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD. The purpose of the IOMC is to promote co-ordination of the policies and activities pursued by the Participating Organisations, jointly or separately, to achieve the sound management of chemicals in relation to human health and the environment.

This publication is available electronically, at no charge.

Also published in the Series on Testing and Assessment [link](#)

**For this and many other Environment,
Health and Safety publications, consult the OECD's
World Wide Web site (www.oecd.org/chemicalsafety/)**

or contact:

**OECD Environment Directorate,
Environment, Health and Safety Division
2 rue André-Pascal
75775 Paris Cedex 16
France**

Fax: (33-1) 44 30 61 80

E-mail: ehscont@oecd.org

© OECD 2013

Applications for permission to reproduce or translate all or part of this material should be made to: Head of Publications Service, RIGHTS@oecd.org, OECD, 2 rue André-Pascal, 75775 Paris Cedex 16, France

FOREWORD

The original version of OECD Test Guideline 431, *In Vitro* Skin Corrosion: Human Skin Model Test, was adopted in 2004, including two test methods —EpiSkin[®] and EpiDerm[®]— able to categorise chemicals between corrosive and non-corrosive chemicals.

The Test Guideline was updated in 2013 to (i) include two additional test methods called SkinEthic[®], and EST-1000 renamed epiCS[®], and (ii) include sub-categorisation abilities of the test methods (i.e. sub-categorisation of the corrosive chemicals between 1A chemicals and 1BC chemicals, in line with the UN GHS for Classification, Labelling and Packaging). Candidates for categorisation and sub-categorisation are EpiSkin[®], EpiDerm[®] and SkinEthic[®], whereas epiCS[®] is a candidate only for categorisation. A consultant to the Secretariat performed a statistical analysis for an evaluation of the performance of the three methods for the sub-categorisation on a post-validation dataset of 80 chemicals. This analysis was reviewed and endorsed by the OECD expert group on skin irritation/corrosion in 2012. The statistical analysis was performed using the prediction models included in the Test Guideline, *i.e.*, not directly relying on results provided by test developers— and was therefore established by a strict application of the revised the Test Guideline.

The Working Group of the National Coordinators of the Test Guidelines Programme approved the summary report on the statistical analysis of the updated version of Test Guideline No. 431.

The Joint Meeting of the Chemicals Committee and the Working Party on Chemicals, Pesticides and Biotechnology agreed to its declassification on 14 June, 2013.

This document is published under the responsibility of the Joint Meeting of the Chemicals committee and the Working Party on Chemicals, Pesticides and Biotechnology.

TABLE OF CONTENTS

ABBREVIATIONS	9
INTRODUCTION	10
CONTINGENCY TABLES AND CALCULATIONS OF SENSITIVITY, SPECIFICITY AND ACCURACY VALUES PERFORMED ON A SET OF 80 CHEMICALS FOR SUB-CATEGORISATION 1A VERSUS 1BC VERSUS NON CORROSIVE CHEMICALS	11
Methodology:	11
Results:	12
Discussion	44
1- Discussion when the Entire Set of Chemicals is Addressed	44
2- Discussion on “MTT-reducers and Organic Bases Issues”	44
CONTINGENCY TABLES AND CALCULATIONS OF SENSITIVITY, SPECIFICITY AND ACCURACY VALUES PERFORMED ON THE SET OF 20 REFERENCE CHEMICALS FOR CATEGORISATION CORROSIVE VERSUS NON CORROSIVE CHEMICALS.....	47
CONTINGENCY TABLES AND CALCULATIONS OF SENSITIVITY, SPECIFICITY AND ACCURACY VALUES PERFORMED ON THE SET OF 30 REFERENCE CHEMICALS FOR SUB- CATEGORISATION 1A VERSUS 1BC VERSUS NON CORROSIVE CHEMICALS.....	49
CONCLUSION.....	52

LIST OF TABLES

Table 1.1: Results including accuracy values, correct classification, over-prediction and under-prediction rates for EpiDerm on entire set of chemicals.....	14
Table 1.2: Results including accuracy values, correct classification, over-prediction and under-prediction rates for EpiSkin on entire set of chemicals.....	16
Table 1.3: Results including accuracy values, correct classification, over-prediction and under-prediction rates for SkinEthic on entire set of chemicals.....	18
Table 2.1: Results including accuracy values, correct classification, over-prediction and under-prediction rates for EpiDerm after MTT-reducers removal	20
Table 2.2: Results including accuracy values, correct classification, over-prediction and under-prediction rates for EpiSkin after MTT-reducers removal	22
Table 2.3: Results including accuracy values, correct classification, over-prediction and under-prediction rates for SkinEthic after MTT-reducers removal	24
Table 3.1: Results including accuracy values, correct classification, over-prediction and under-prediction rates for EpiDerm on MTT-reducers only	26
Table 3.2: Results including accuracy values, correct classification, over-prediction and under-prediction rates for EpiSkin on MTT-reducers only	28
Table 3.3: Results including accuracy values, correct classification, over-prediction and under-prediction rates for SkinEthic on MTT-reducers only	30
Table 4.1: Results including accuracy values, correct classification, over-prediction and under-prediction rates for EpiDerm after organic bases removal.....	32
Table 4.2: Results including accuracy values, correct classification, over-prediction and under-prediction rates for EpiSkin after organic bases removal	34
Table 4.3: Results including accuracy values, correct classification, over-prediction and under-prediction rates for SkinEthic after organic bases removal.....	36
Table 5.1: Results including accuracy values, correct classification, over-prediction and under-prediction rates for EpiDerm on organic bases only	38
Table 5.2: Results including accuracy values, correct classification, over-prediction and under-prediction rates for EpiSkin on organic bases only	40
Table 5.3: Results including accuracy values, correct classification, over-prediction and under-prediction rates for SkinEthic on organic bases only	42
Table 6. Results including accuracy values, correct classification, over-prediction and under-prediction rates for EpiSkin [®] , EpiDerm [®] and SkinEthic [®] based on the set of 20 reference chemicals for categorisation	48
Table 7. Results including accuracy values, correct classification, over-prediction and under-prediction rates for EpiSkin [®] based on the set of 30 reference chemicals	51

ABBREVIATIONS

Cat.: Category

GHS: Global Harmonised System of Classification and Labelling of Chemicals

MTT: 3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide, Thiazolyl blue tetrazolium bromide (CAS number 298-93-1)

OECD: Organisation for Economic Co-operation and Development

RhE: Reconstructed human Epidermis

TG: Test Guideline

UN: United Nations

1A or 1A chemicals: 1A category of chemicals based on the UN GHS

1BC or 1BC chemicals: merged categories 1B and 1C of chemicals based on the UN GHS [stated in the guide as the “1B-and-1C category”]

NC or NC chemicals: Non Corrosive category of chemicals based on the UN GHS

INTRODUCTION

1. This document was prepared by Bertrand Desprez consultant for the OECD Secretariat. It aims to present the work performed—including statistical analysis—after September 2012 meeting on Skin Irritation/Corrosion. The expert group requested to:

- Perform an independent statistical analysis of post-validation data provided by test developers,
- Assess the ability of each test method to categorise test substances into corrosive and non-corrosive chemicals, as well as their ability to sub-categorise the identified corrosive chemicals in sub-categories, referring to the UN GHS,
- Develop a list of proficiency chemicals,
- Develop a list of reference chemicals,
- Develop and revise the performance standards,
- Revise TG 431 accordingly.

2. Furthermore, this document presents the main outcomes of the statistics based on data concerning EpiDerm[®], EpiSkin[®] and SkinEthic[®] test methods. Data were provided to the OECD Secretariat by test method developers. Those three test methods are *in vitro* alternative methods used in TG 431 to identify Corrosive Versus Non Corrosive chemicals, and to identify within the Corrosive chemicals category 1A chemicals versus category 1BC chemicals (see below).

3. It was decided at September meeting to establish a sub-group to perform this post-meeting work and to reach the above objectives. This sub-group included: Joao Barroso (SECAM), Chantra Eskes (SECAM), Claudius Griesinger (ECVAM), Jon Hamm (NIH/NIEHS), Emiel Rorije (RIVM), Hans Raabe (IIVS), and Bertrand Desprez (OECD Secretariat).

4. Three categories of chemicals can be identified by each test method. Those categories are based on those from the UN GHS classification:

- Corrosive chemicals:
 - **Category 1A chemicals,**
 - **Category 1BC chemicals,** corresponding to merged categories 1B and 1C (stated in the guideline as “1B-and-1C category”)
- **Category Non Corrosive chemicals.**

5. The statistical analysis was performed independently by the OECD Secretariat and reviewed at first by the sub-group, and then by the entire expert group on skin corrosion (test developers are part of the entire group).

6. Statistics consist in assessing the concordance between the *in vitro* test method results and the *in vivo* reference method (TG 404).

7. More over ‘**categorisation**’ term refers to the classification of **Corrosive chemicals Versus Non Corrosive chemicals**. Corrosive chemicals are Category 1A and Category 1BC. The term ‘**sub-**

categorisation' refers to the classification within the Corrosive chemicals and thus distinguishes **Cat. 1A chemicals Versus Cat. 1BC chemicals**.

8. The purpose of revision of TG 431 was mainly to include in the guideline test methods capabilities for sub-categorisation.

CONTINGENCY TABLES AND CALCULATIONS OF SENSITIVITY, SPECIFICITY AND ACCURACY VALUES PERFORMED ON A SET OF 80 CHEMICALS FOR SUB-CATEGORISATION 1A VERSUS 1BC VERSUS NON CORROSIVE CHEMICALS

9. This section displays several contingency tables and calculations performed on the set of 80 chemicals, tested over 3 runs, for each test method. Therefore, each method brought 240 classifications when using the entire set of chemicals.

Methodology:

10. Those data include a set of more than 80 chemicals tested 3 times (i.e. over 3 runs) for each method. Only chemicals tested in common by the three test methods, i.e. 80 chemicals, were retained so as to develop reliable comparisons of the three test methods.

11. Data provided by the test developers were Analysed using Stata® software and Excel® software too (to facilitate exchanges).

12. The sub-group hypothesised that MTT-reducers chemicals and organic base chemicals would have the potential to impair results provided by the test methods, and therefore those MTT-reducers and organic base issues were treated as well.

13. The statistics performed for each test method were thus:

- On the entire set of 80 chemicals
(n=80 chemicals tested over 3 runs, i.e. 240 classifications for the three test methods),
- Without MTT-reducers
(n=59/64/63 chemicals over 3 runs, i.e. 177/192/189 classifications respectively for EpiDerm/EpiSkin/SkinEthic),
- On MTT-reducers only, and thus sample size is small
(n=21/16/17 chemicals tested over 3 runs, i.e. 63/48/51 classifications respectively for EpiDerm/EpiSkin/SkinEthic),
- Without organic bases: at first 8 chemicals were identified as organic bases, but in a late stage 11 chemicals were identified as organic bases. Tables presented below consider 8 chemicals as organic bases.
(n=72 chemicals tested over 3 runs, i.e. 216 classifications for the three test methods),

- On organic bases only, and thus sample size is very small. At first 8 chemicals were identified as organic bases, but in a late stage 11 chemicals were identified as organic bases. Tables presented below consider 8 chemicals as organic bases.
(n=8 chemicals tested over 3 runs, i.e. 24 classifications for the three test methods),

14. In a first step, Stata[®] software was used to classify chemicals into Corrosive and Non Corrosive chemicals, as well as to split the Corrosive chemicals into 1A and 1BC chemicals (referring to the UN GHS classification). Stata[®] was able to provide contingency tables displaying the concordance between each *in vitro* test method and the *in vivo* reference method. Based on these tables, calculation of sensitivity, specificity and accuracy values was performed as well.

15. It is important to note that concordance of the *in vitro* methods with the *in vivo* method was assessed using the prediction models stated in TG 431—based on cell viabilities—and thus not directly relying on classification results provided by the test developers. Therefore, consistence of those results was independently assessed and based on strict application of the prediction models included in the test guideline. Stata[®] also helped in identifying and assessing misclassifications, and calculated their respective rates.

16. In a second step, contingency tables previously obtained were entered in Excel[®] software in order to get sheets to be exchanged within the sub-group and then shared with the entire expert group on skin corrosion. Excel[®] sheets developed were designed to enable automatic calculations of sensitivity, specificity and accuracy values. Under this step, misclassifications were assessed as well, i.e. under-classifications (e.g. 1A chemical under-classified as 1BC) and over-classifications (e.g. 1BC chemical over-classified 1A) rates.

17. All results are calculated in percentages and presented in summarizing tables. All numbers in tables cells are calculated based on formulas, and those are displayed when examining the Excel[®] sheet so as to be transparent for the sub-group and entire group reviewers.

18. Several commenting rounds were held within the sub-group including conference calls and numerous emails exchanges in order to review the results. Those results were finally proposed to the entire expert group on skin corrosion for a final commenting round which included a conference call and emails exchanges.

19. The sub-group used Fentem et al. publication *in Toxicology in vitro*, 12 (1998) 483-524 as a mark. This paper states on acceptable and unacceptable criteria for the validation of several *in vitro* test methods on skin corrosion, including the RhE test methods. The sub-group—and the entire expert group—took into account this study for its unacceptability criteria. However it was decided to mainly use this paper as a mark and that it should not bring any kind of constraint for the current work as data from 1998 are quite different from the current ones. Fentem's unacceptability criteria are included in all tables displayed (see Tables 1.1 to 5.3).

Results:

20. Results are summarised in the following tables and are included in Annex IV of the revised TG 431. The guideline states that all three methods are able to sub-categorise but highlights significant accuracy differences between EpiSkin[®] and the two test methods, EpiDerm[®] and SkinEthic[®].

21. Indeed, EpiSkin[®], EpiDerm[®] and SkinEthic[®] test methods are able to subcategorise chemicals (i.e. 1A Versus 1B-and-1C Versus NC) but significant differences are observed between EpiSkin[®] and the two other test methods, EpiDerm[®] & SkinEthic[®], for sub-categorisation:

- Results from EpiSkin[®] can be directly used as they are and allow direct sub-categorisation (1A Versus 1B-and-C Versus Non Corrosive) of the test chemicals,
- Results from EpiDerm[®] & SkinEthic[®] should take into account high over-classification rates¹ occurring when using those two test methods for 1B-and-1C category (see tables). Therefore, for EpiDerm[®] & SkinEthic[®], chemicals that are classified as 1B-and-1C can be considered as 1B-and-1C, but chemicals for which cell viability at 3 minutes is below 50% (and labeled as 1A) should be considered as 1, that is to say that either they could be claimed as 1A to as protective as possible, or they should undergo further testing to be possibly confirmed as 1B-and-1C.¹

¹ Over-prediction rates for 1B-and-C chemicals:
EpiSkin[®]: 21.51%; EpiDerm[®]:41.94%; SkinEthic[®]: 46.24%

Table 1.1: Results including accuracy values, correct classification, over-prediction and under-prediction rates for EpiDerm on entire set of chemicals

EpiDerm with entire set of chemicals (80 chemicals tested over 3 runs, i.e. 240 classifications)				
In vivo categories	Test method: EpiDerm			Sum
	Classified as Cat. 1A	Classified as Cat. 1BC	Classified as Cat. NC	
In vivo Cat. 1A	33	3	0	36
In vivo Cat. 1BC	39	54	0	93
In vivo Cat. NC	3	26	82	111
Sum	75	83	82	240

Calculations over the 3 chemicals categories:				
Within Corrosive: Cat. 1A Versus Cat. 1BC		Checking the misclassifications over the 3 categories:		
Sensitivity for 1A	91.67 %	Accuracy (Pred. C)	% OverClass 1BC as 1A	% OverClass NC as 1BC
Sensitivity for 1BC	58.06 %	70.42%	41.94%	23.42%
Accuracy	67.44 %		% UnderClass 1A as 1BC	% UnderClass 1A as NC
Cat. Corr. Versus Cat. Non Corr		Correctly Classified	8.33%	0.00%
		Cat.1A 91.67%	% of OverClass	% of UnderClass
		Cat.1BC 58.06%	28.33%	1.25%
		Cat. NC 73.87%		

Sensitivity for Corr.	100.00 %	Fentem's criteria		According to Fentem, test are UNACCEPTABLE if...	
Sensitivity for Non-Corr.	73.87 %	%OverClass NC as Corr. 26.13%	%OverClass 1BC as 1A 41.94%	If % OverClass NC as Corr $\geq 50\%$	If OverClass 1BC as 1A $\geq 50\%$
<i>Accuracy</i>	87.92 %	%UnderClass 1BC as NC 0.00%	%UnderClass 1A as NC 0.00%	If % UnderClass. 1BC as NC $\geq 30\%$	If UnderClass. 1A as NC $\geq 30\%$

Table 1.2: Results including accuracy values, correct classification, over-prediction and under-prediction rates for EpiSkin on entire set of chemicals

EpiSkin with entire set of chemicals (80 chemicals tested over 3 runs, i.e. 240 classifications)				
In vivo categories	Test method: EpiSkin			Sum
	Classified as Cat. 1A	Classified as Cat. 1BC	Classified as Cat. NC	
In vivo Cat. 1A	30	6	0	36
In vivo Cat. 1BC	20	71	2	93
In vivo Cat. NC	0	23	88	111
Sum	50	100	90	240

Within Corrosive: Cat. 1A Versus 1BC		Calculations over the 3 chemicals categories:			
Sensitivity for 1A	83.33 %	Accuracy (Pred. C) 78.75%	Checking the misclassifications over the 3 categories:		
Sensitivity for 1BC	78.02 %		% OverClass 1BC as 1A 21.51%	%OverClassNC as 1BC 20.72%	% OverClass NC as 1A 0.00%
Accuracy	79.53 %		% UnderClass 1A as 1BC 16.67%	% underClass 1A as NC 0.00%	% UnderClass 1BC as NC 2.15%
Cat. Corr. Versus Cat. Non Corr		Cat.1A 83.33% Cat.1BC 76.34% Cat. NC 79.28%	% of OverClass 17.92% % of UnderClass 3.33%		

Sensitivity for Corr.	98.45 %	Fentem's criteria		According to Fentem, tests are UNACCEPTABLE if...	
Sensitivity for Non Corr.	79.28 %	%OverClass NC as Corr. 20.72%	%OverClass 1BC as 1A 21.51%	If % OverClass NC as Corr $\geq 50\%$	If OverClass 1BC as 1A $\geq 50\%$
<i>Accuracy</i>	89.58 %	%UnderClass 1BC as NC 2.15%	%UnderClass 1A as NC 0.00%	If % UnderClass. 1BC as NC $\geq 30\%$	If UnderClass. 1A as NC $\geq 30\%$

Table 1.3: Results including accuracy values, correct classification, over-prediction and under-prediction rates for SkinEthic on entire set of chemicals

SkinEthic with entire set of chemicals (80 chemicals tested over 3 runs, i.e. 240 classifications)				
In vivo categories	Test method: SkinEthic			Sum
	Classified as Cat. 1A	Classified as Cat. 1BC	Classified as Cat. NC	
In vivo Cat. 1A	31	5	0	36
In vivo Cat. 1BC	43	43	7	93
In vivo Cat. NC	3	27	81	111
Sum	77	75	88	240

Calculations over the 3 chemicals categories:				
Within Corrosive: Cat. 1A Versus Cat. 1BC		Accuracy (Pred. C)		
Sensitivity for 1A	86.11 %	64.58%		
Sensitivity for 1BC	50.00 %			
Accuracy	60.66 %			
Corr. Versus Non Corr		Checking the misclassifications over the 3 categories:		
Sensitivity for Corr	94.57 %	% OverClass 1BC as 1A %OverClass NC as 1BC % OverClass NC as 1A 46.24% 24.32% 2.70%		
		% UnderClass 1A as 1BC % underClass 1A as NC % UnderClass 1BC as NC 13.89% 0.00% 7.53%		
		% of OverClass % of UnderClass 30.42% 5.00%		
		Cat.1A 86.11% Cat.1BC 46.24% Cat. NC 72.97%		
		Fentem's criteria According to Fentem, tests are UNACCEPTABLE if...		

Sensitivity for Non Corr	72.97 %	%OverClass NC as Corr. 27.03%	%OverClass 1BC as 1A 46.24%	If % OverClass NC as Corr $\geq$ 50%	If OverClass 1BC as 1A $\geq$ 50%
<i>Accuracy</i>	84.58 %	%UnderClass 1BC as NC 7.53%	%UnderClass 1A as NC 0.00%	If % UnderClass. 1BC as NC $\geq$ 30%	If UnderClass. 1A as NC $\geq$ 30%

Table 2.1: Results including accuracy values, correct classification, over-prediction and under-prediction rates for EpiDerm after MTT-reducers removal

EpiDerm Without its MTT Reducers (59 chemicals tested over 3 runs, i.e. 177 classifications)				
<i>In vivo</i> categories	Test method: EpiDerm			Sum
	Classified as Cat. 1A	Classified as Cat. 1BC	Classified as Cat. NC	
<i>In vivo</i> Cat. 1A	30	0	0	30
<i>In vivo</i> Cat. 1BC	27	36	0	63
<i>In vivo</i> Cat. NC	0	17	67	84
Sum	57	53	67	177

Calculations over the 3 chemicals categories:				
Accuracy (Pred. C)		Checking the misclassifications over the 3 categories:		
75.14%		% OverClass 1BC as 1A	%OverClassNC as 1BC	% OverClass NC as 1A
		42.86%	20.24%	0.00%
		% UnderClass 1A as 1BC	% underClass 1A as NC	% UnderClass 1BC as NC
		0.00%	0.00%	0.00%
Correctly Class				
Cat. 1A	100.00%			
Cat. 1BC	57.14%			
Cat. NC	79.76%			
		% of OverClass	% of UnderClass	
		24.86%	0.00%	

Within Corrosive: Cat. 1A Versus 1BC	
Sensitivity for 1A	100.00 %
Sensitivity for 1BC	57.14 %
<i>Accuracy</i>	70.97 %
Corr. Versus Non Corr	

Sensitivity for Corr	100.00 %	Fentem's criteria		According to Fentem, tests are UNACCEPTABLE if...	
Sensitivity for Non Corr	79.76 %	%OverClass NC as Corr. 20.24%	%OverClass 1BC as 1A 42.86%	If % OverClass NC as Corr $\geq$ 50%	If OverClass 1BC as 1A $\geq$ 50%
<i>Accuracy</i>	90.40 %	%UnderClass 1BC as NC 0.00%	%UnderClass 1A as NC 0.00%	If % UnderClass. 1BC as NC $\geq$ 30%	If UnderClass. 1A as NC $\geq$ 30%

Table 2.2: Results including accuracy values, correct classification, over-prediction and under-prediction rates for EpiSkin after MTT-reducers removal

EpiSkin Without its MTT reducers (64 chemicals tested over 3 runs, i.e. 192 classifications)				
<i>In vivo</i> categories	Test method: EpiSkin			Sum
	Classified as Cat. 1A	Classified as Cat. 1BC	Classified as Cat. NC	
<i>In vivo</i> Cat. 1A	28	2	0	30
<i>In vivo</i> Cat. 1BC	17	44	2	63
<i>In vivo</i> Cat. NC	0	20	79	99
Sum	45	66	81	192

Calculations over the 3 chemicals categories:				
Accuracy (Pred. C)		Checking the misclassifications over the 3 categories:		
78.65%		% OverClass 1BC as 1A	% OverClass NC as 1BC	% OverClass NC as 1A
		26.98%	20.20%	0.00%
		% UnderClass 1A as 1BC	% UnderClass 1A as NC	% UnderClass 1BC as NC
		6.67%	0.00%	3.17%
Correctly Class		% of OverClass		% of UnderClass
Cat. 1A	93.33%	19.27%		2.08%
Cat. 1BC	69.84%			
Cat. NC	79.80%			

Within Corrosive: Cat. 1A Versus 1BC	
Sensitivity for 1A	93.33 %
Sensitivity for 1BC	72.13 %
Accuracy	79.12 %
Corr. Versus Non Corr	

Sensitivity for Corr	97.85 %	Fentem's criteria		According to Fentem, test are UNACCEPTABLE if...	
Sensitivity for Non Corr	79.80 %	%OverClass NC as Corr. 20.20%	%OverClass 1BC as 1A 26.98%	If % OverClass NC as Corr $\geq$ 50%	If OverClass 1BC as 1A $\geq$ 50%
<i>Accuracy</i>	88.54 %	%UnderClass 1BC as NC 3.17%	%UnderClass 1A as NC 0.00%	If % UnderClass. 1BC as NC $\geq$ 30%	If UnderClass. 1A as NC $\geq$ 30%

Table 2.3: Results including accuracy values, correct classification, over-prediction and under-prediction rates for SkinEthic after MTT-reducers removal

SkinEthic Without its MTT Reducers(63 chemicals tested over 3 runs, i.e. 189 classifications)				
In vivo categories	Test method: SkinEthic			Sum
	Classified as Cat. 1A	Classified as Cat. 1BC	Classified as Cat. NC	
In vivo Cat. 1A	27	0	0	27
In vivo Cat. 1BC	27	35	4	66
In vivo Cat. NC	1	17	78	96
Sum	55	52	82	189

Within Corrosive: Cat. 1A Versus 1BC	
Sensitivity for 1A	100.00 %
Sensitivity for 1BC	56.45 %
Accuracy	69.66 %
Corr. Versus Non Corr	

Calculations over the 3 chemicals categories:				
Accuracy (Pred. C)		Checking the misclassifications over the 3 categories:		
74.07%		% OverClass 1BC as 1A	%OverClassNC as 1BC	% OverClass NC as 1A
		40.91%	17.71%	1.04%
		% UnderClass 1A as 1BC	% underClass 1A as NC	% UnderClass 1BC as NC
		0.00%	0.00%	6.06%
Correctly Class				
Cat.1A	100.00%			
Cat.1BC	53.03%	% of OverClass		
Cat. NC	81.25%	23.81%	% of UnderClass	
			2.12%	

Sensitivity for Corr	95.70 %	Fentem's criteria		According to Fentem, test are UNACCEPTABLE if...	
Sensitivity for Non Corr	81.25 %	%OverClass NC as Corr. 18.75%	%OverClass 1BC as 1A 40.91%	If % OverClass NC as Corr $\geq$ 50%	If OverClass 1BC as 1A $\geq$ 50%
<i>Accuracy</i>	88.36 %	%UnderClass 1BC as NC 6.06%	%UnderClass 1A as NC 0.00%	If % UnderClass. 1BC as NC $\geq$ 30%	If UnderClass. 1A as NC $\geq$ 30%

Table 3.1: Results including accuracy values, correct classification, over-prediction and under-prediction rates for EpiDerm on MTT-reducers only

EpiDerm Within MTT Reducers specific to EpiDerm only (21 chemicals tested over 3 runs, i.e. 63 classifications)				
<i>In vivo</i> categories	Test method: EpiDerm			Sum
	Classified as Cat. 1A	Classified as Cat. 1BC	Classified as Cat. NC	
<i>In vivo</i> Cat. 1A	3	3	0	6
<i>In vivo</i> Cat. 1BC	12	18	0	30
<i>In vivo</i> Cat. NC	3	9	15	27
Sum	18	30	15	63

Calculations over the 3 chemicals categories:				
Accuracy (Pred. C)		Checking the misclassifications over the 3 categories:		
57.14%		% OverClass 1BC as 1A	%OverClassNC as 1BC	% OverClass NC as 1A
		40.00%	33.33%	11.11%
		% UnderClass 1A as 1BC	% underClass 1A as NC	% UnderClass 1BC as NC
		50.00%	0.00%	0.00%
Correctly Class				
Cat.1A	50.00%			
Cat.1BC	60.00%			
Cat. NC	55.56%			
		% of OverClass	% of UnderClass	
		38.10%	4.76%	

Within Corrosive: Cat. 1A Versus 1BC	
Sensitivity for 1A	50.00 %
Sensitivity for 1BC	60.00 %
<i>Accuracy</i>	58.33 %
Corr. Versus Non Corr	

Sensitivity for Corr	100.00 %	Fentem's criteria		According to Fentem, test are UNACCEPTABLE if...	
Sensitivity for Non Corr <i>Accuracy</i>	55.56 % 80.95 %	%OverClass NC as Corr. 44.44%	%OverClass 1BC as 1A 40.00%	If % OverClass NC as Corr $\geq$ 50%	If OverClass 1BC as 1A $\geq$ 50%
		%UnderClass 1BC as NC 0.00%	%UnderClass 1A as NC 0.00%	If % UnderClass. 1BC as NC $\geq$ 30%	If UnderClass. 1A as NC $\geq$ 30%

Table 3.2: Results including accuracy values, correct classification, over-prediction and under-prediction rates for EpiSkin on MTT-reducers only

EpiSkin Within MTT Reducers specific to EpiSkin only (16 chemicals tested over 3 runs, i.e. 48 classifications)				
In vivo categories	Test method: EpiSkin			Sum
	Classified as Cat. 1A	Classified as Cat. 1BC	Classified as Cat. NC	
	2	4	0	
	3	27	0	
	0	3	9	
	5	34	9	
In vivo Cat. 1A				6
In vivo Cat. 1BC				30
In vivo Cat. NC				12
Sum				48

Calculations over the 3 chemicals categories:				
Accuracy (Pred. C)		Checking the misclassifications over the 3 categories:		
79.17%		% OverClass 1BC as 1A	%OverClassNC as 1BC	% OverClass NC as 1A
		10.00%	25.00%	0.00%
		% UnderClass 1A as 1BC	% underClass 1A as NC	% UnderClass 1BC as NC
		66.67%	0.00%	0.00%
Correctly Class				
Cat.1A	33.33%			
Cat.1BC	90.00%			
Cat. NC	75.00%			
		% of OverClass	% of UnderClass	
		12.50%	8.33%	
Fentem’s criteria		According to Fentem, test are UNACCEPTABLE if...		

Within Corrosive: Cat. 1A Versus 1BC	
Sensitivity for 1A	33.33%
Sensitivity for 1BC	90.00%
Accuracy	80.56%
Corr. Versus Non Corr	
Sensitivity for Corr	100.00%

Sensitivity for Non Corr	75.00%	%OverClass NC as Corr. 25.00%	%OverClass 1BC as 1A 10.00%	If % OverClass NC as Corr $\geq$ 50%	If OverClass 1BC as 1A $\geq$ 50%
<i>Accuracy</i>	93.75%	%UnderClass 1BC as NC 0.00%	%UnderClass 1A as NC 0.00%	If % UnderClass. 1BC as NC $\geq$ 30%	If UnderClass. 1A as NC $\geq$ 30%

Table 3.3: Results including accuracy values, correct classification, over-prediction and under-prediction rates for SkinEthic on MTT-reducers only

SkinEthic Within MTT Reducers specific to SkinEthic only (17 chemicals tested over 3 runs, i.e. 51 classifications)				
<i>In vivo</i> categories	Test method: SkinEthic			Sum
	Classified as Cat. 1A	Classified as Cat. 1BC	Classified as Cat. NC	
<i>In vivo</i> Cat. 1A	4	5	0	9
<i>In vivo</i> Cat. 1BC	16	8	3	27
<i>In vivo</i> Cat. NC	2	10	3	15
Sum	22	23	6	51

Calculations over the 3 chemicals categories:				
Accuracy (Pred. C)		Checking the misclassifications over the 3 categories:		
29.41%		% OverClass 1BC as 1A	% OverClass NC as 1BC	% OverClass NC as 1A
		59.26%	66.67%	13.33%
Correctly Class		% UnderClass 1A as 1BC	% underClass 1A as NC	% UnderClass 1BC as NC
Cat.1A 44.44%		55.56%	0.00%	11.11%
Cat.1BC 29.63%				
Cat. NC 20.00%		% of OverClass	% of UnderClass	
		54.90%	15.69%	

Within Corrosive: Cat. 1A Versus 1BC	
Sensitivity for 1A	44.44%
Sensitivity for 1BC	33.33%
Accuracy	36.36%
Corr. Versus Non Corr	

Sensitivity for Corr	91.67 %	Fentem's criteria		According to Fentem, test are UNACCEPTABLE if...	
Sensitivity for Non Corr	20.00 %	%OverClass NC as Corr.	%OverClass 1BC as 1A	If % OverClass NC as Corr $\geq$ 50%	If OverClass 1BC as 1A $\geq$ 50%
<i>Accuracy</i>	70.59 %	80.00%	59.26%		
		%UnderClass 1BC as NC	%UnderClass 1A as NC	If % UnderClass. 1BC as NC $\geq$ 30%	If UnderClass. 1A as NC $\geq$ 30%
		11.11%	0.00%		

Table 4.1: Results including accuracy values, correct classification, over-prediction and under-prediction rates for EpiDerm after organic bases removal

EpiDerm Without Organic Bases (72 chemicals tested over 3 runs, i.e. 216 classifications)				
<i>In vivo</i> categories	Test method: EpiDerm			Sum
	Classified as Cat. 1A	Classified as Cat. 1BC	Classified as Cat. NC	
<i>In vivo</i> Cat. 1A	30	0	0	30
<i>In vivo</i> Cat. 1BC	33	48	0	81
<i>In vivo</i> Cat. NC	3	24	78	105
Sum	66	72	78	216

Calculations over the 3 chemicals categories:				
Accuracy (Pred. C)		Checking the misclassifications over the 3 categories:		
72.22%		% OverClass 1BC as 1A	% OverClass NC as 1BC	% OverClass NC as 1A
		40.74%	22.86%	2.86%
		% UnderClass 1A as 1BC	% underClass 1A as NC	% UnderClass 1BC as NC
		0.00%	0.00%	0.00%
Correctly Class		% of OverClass		% of UnderClass
Cat.1A	100.00%	27.78%		0.00%
Cat.1BC	59.26%			
Cat. NC	74.29%			

Within Corrosive: Cat. 1A Versus 1BC	
Sensitivity for 1A	100.00 %
Sensitivity for 1BC	59.26 %
Accuracy	70.27 %
Corr. Versus Non Corr	

Sensitivity for Corr	100.00 %	Fentem's criteria		According to Fentem, test are UNACCEPTABLE if...	
Sensitivity for Non Corr	74.29 %	%OverClass NC as Corr. 25.71%	%OverClass 1BC as 1A 40.74%	If % OverClass NC as Corr $\geq$ 50%	If OverClass 1BC as 1A $\geq$ 50%
<i>Accuracy</i>	87.50 %	%UnderClass 1BC as NC 0.00%	%UnderClass 1A as NC 0.00%	If % UnderClass. 1BC as NC $\geq$ 30%	If UnderClass. 1A as NC $\geq$ 30%

Table 4.2: Results including accuracy values, correct classification, over-prediction and under-prediction rates for EpiSkin after organic bases removal

EpiSkin Without Organic Bases (72 chemicals tested over 3 runs, i.e. 216 classifications)				
<i>In vivo</i> categories	Test method: EpiSkin			Sum
	Classified as Cat. 1A	Classified as Cat. 1BC	Classified as Cat. NC	
<i>In vivo</i> Cat. 1A	28	2	0	30
<i>In vivo</i> Cat. 1BC	17	62	2	81
<i>In vivo</i> Cat. NC	0	22	83	105
Sum	45	86	85	216

Calculations over the 3 chemicals categories:				
Accuracy (Pred. C)		Checking the misclassifications over the 3 categories:		
80.09%		% OverClass 1BC as 1A	%OverClassNC as 1BC	% OverClass NC as 1A
		20.99%	20.95%	0.00%
		% UnderClass 1A as 1BC	% underClass 1A as NC	% UnderClass 1BC as NC
		6.67%	0.00%	2.47%
Correctly Class		% of OverClass		% of UnderClass
Cat. 1A	93.33%	18.06%		1.85%
Cat. 1BC	76.54%			
Cat. NC	79.05%			

Within Corrosive: Cat. 1A Versus 1BC	
Sensitivity for 1A	93.33 %
Sensitivity for 1BC	78.48 %
Accuracy	82.57 %
Corr. Versus Non Corr	

Sensitivity for Corr	98.20 %	Fentem's criteria		According to Fentem, test are UNACCEPTABLE if...	
Sensitivity for Non Corr	79.05 %	%OverClass NC as Corr. 20.95%	%OverClass 1BC as 1A 20.99%	If % OverClass NC as Corr $\geq$ 50%	If OverClass 1BC as 1A $\geq$ 50%
<i>Accuracy</i>	88.89 %	%UnderClass 1BC as NC 2.47%	%UnderClass 1A as NC 0.00%	If % UnderClass. 1BC as NC $\geq$ 30%	If UnderClass. 1A as NC $\geq$ 30%

Table 4.3: Results including accuracy values, correct classification, over-prediction and under-prediction rates for SkinEthic after organic bases removal

SkinEthic Without Organic Bases (72 chemicals tested over 3 runs, i.e. 216 classifications)				
<i>In vivo</i> categories	Test method: SkinEthic			Sum
	Classified as Cat. 1A	Classified as Cat. 1BC	Classified as Cat. NC	
<i>In vivo</i> Cat. 1A	30	0	0	30
<i>In vivo</i> Cat. 1BC	34	43	4	81
<i>In vivo</i> Cat. NC	2	25	78	105
<i>Sum</i>	66	68	82	216

Calculations over the 3 chemicals categories:				
Accuracy (Pred. C)		Checking the misclassifications over the 3 categories:		
69.91%		% OverClass 1BC as 1A	% OverClass NC as 1BC	% OverClass NC as 1A
		41.98%	23.81%	1.90%
		% UnderClass 1A as 1BC	% UnderClass 1A as NC	% UnderClass 1BC as NC
		0.00%	0.00%	4.94%
Correctly Class		% of OverClass		% of UnderClass
Cat. 1A	100.00%	28.24%		1.85%
Cat. 1BC	53.09%			
Cat. NC	74.29%			

Within Corrosive: Cat. 1A Versus 1BC	
Sensitivity for 1A	100.00 %
Sensitivity for 1BC	55.84 %
Accuracy	68.22 %
Corr. Versus Non Corr	

Sensitivity for Corr	96.40 %	Fentem's criteria		According to Fentem, test are UNACCEPTABLE if...	
Sensitivity for Non Corr	74.29 %	%OverClass NC as Corr. 25.71%	%OverClass 1BC as 1A 41.98%	If % OverClass NC as Corr $\geq$ 50%	If OverClass 1BC as 1A $\geq$ 50%
<i>Accuracy</i>	85.65 %	%UnderClass 1BC as NC 4.94%	%UnderClass 1A as NC 0.00%	If % UnderClass. 1BC as NC $\geq$ 30%	If UnderClass. 1A as NC $\geq$ 30%

Table 5.1: Results including accuracy values, correct classification, over-prediction and under-prediction rates for EpiDerm on organic bases only

EpiDerm Within Organic Bases (8 chemicals tested over 3 runs, i.e. 24 classifications)				
<i>In vivo</i> categories	Test method: EpiDerm			Sum
	Classified as Cat. 1A	Classified as Cat. 1BC	Classified as Cat. NC	
<i>In vivo</i> Cat. 1A	3	3	0	6
<i>In vivo</i> Cat. 1BC	6	6	0	12
<i>In vivo</i> Cat. NC	0	2	4	6
Sum	9	11	4	24

Calculations over the 3 chemicals categories:				
Accuracy (Pred. C)		Checking the misclassifications over the 3 categories:		
54.17%		% OverClass 1BC as 1A	%OverClassNC as 1BC	% OverClass NC as 1A
		50.00%	33.33%	0.00%
Correctly Class		% UnderClass 1A as 1BC	% underClass 1A as NC	% UnderClass 1BC as NC
Cat. 1A 50.00%		50.00%	0.00%	0.00%
Cat. 1BC 50.00%		% of OverClass		% of UnderClass
Cat. NC 66.67%		33.33%		12.50%
Fentem's criteria		According to Fentem, test are UNACCEPTABLE if...		

Within Corrosive: Cat. 1A Versus 1BC	
Sensitivity for 1A	50.00 %
Sensitivity for 1BC	50.00 %
Accuracy	50.00 %
Corr. Versus Non Corr	
Sensitivity for Corr	100.00 %

Sensitivity for Non Corr	66.67 %	%OverClass NC as Corr. 33.33%	%OverClass 1BC as 1A 50.00%	If % OverClass NC as Corr $\geq$ 50%	If OverClass 1BC as 1A $\geq$ 50%
<i>Accuracy</i>	91.67 %	%UnderClass 1BC as NC 0.00%	%UnderClass 1A as NC 0.00%	If % UnderClass. 1BC as NC $\geq$ 30%	If UnderClass. 1A as NC $\geq$ 30%

Table 5.2: Results including accuracy values, correct classification, over-prediction and under-prediction rates for EpiSkin on organic bases only

EpiSkin Within Organic Bases (8 chemicals tested over 3 runs, i.e. 24 classifications)				
In vivo categories	Test method: EpiSkin			Sum
	Classified as Cat. 1A	Classified as Cat. 1BC	Classified as Cat. NC	
In vivo Cat. 1A	2	4	0	6
In vivo Cat. 1BC	3	9	0	12
In vivo Cat. NC	0	1	5	6
Sum	5	14	5	24

Calculations over the 3 chemicals categories:				
Accuracy (Pred. C) 66.67%		Checking the misclassifications over the 3 categories:		
		% OverClass 1BC as 1A	%OverClassNC as 1BC	% OverClass NC as 1A
		25.00%	16.67%	0.00%
Correctly Class Cat.1A 33.33% Cat.1BC 75.00% Cat. NC 83.33%		% UnderClass 1A as 1BC	% underClass 1A as NC	% UnderClass 1BC as NC
		66.67%	0.00%	0.00%
		% of OverClass	% of UnderClass	
		16.67%	16.67%	

Within Corrosive: Cat. 1A Versus 1BC	
Sensitivity for 1A	33.33 %
Sensitivity for 1BC	75.00 %
Accuracy	61.11 %
Corr. Versus Non Corr	
Sensitivity for Corr	100.00 %

Fentem's criteria		According to Fentem, test are UNACCEPTABLE if...	
-------------------	--	--	--

Sensitivity for Non Corr	83.33 %	%OverClass NC as Corr. 16.67%	%OverClass 1BC as 1A 25.00%	If % OverClass NC as Corr $\geq$ 50%	If OverClass 1BC as 1A $\geq$ 50%
<i>Accuracy</i>	95.83 %	%UnderClass 1BC as NC 0.00%	%UnderClass 1A as NC 0.00%	If % UnderClass. 1BC as NC $\geq$ 30%	If UnderClass. 1A as NC $\geq$ 30%

Table 5.3: Results including accuracy values, correct classification, over-prediction and under-prediction rates for SkinEthic on organic bases only

SkinEthic Within Organic Bases (8 chemicals tested over 3 runs, i.e. 24 classifications)				
In vivo categories	Test method: SkinEthic			Sum
	Classified as Cat. 1A	Classified as Cat. 1BC	Classified as Cat. NC	
In vivo Cat. 1A	1	5	0	6
In vivo Cat. 1BC	9	0	3	12
In vivo Cat. NC	1	2	3	6
Sum	11	7	6	24

Calculations over the 3 chemicals categories:				
Accuracy (Pred. C)		Checking the misclassifications over the 3 categories:		
16.67%		% OverClass 1BC as 1A	% OverClass NC as 1BC	% OverClass NC as 1A
		75.00%	33.33%	16.67%
Correctly Class		% UnderClass 1A as 1BC	% underClass 1A as NC	% UnderClass 1BC as NC
Cat.1A 16.67%		83.33%	0.00%	25.00%
Cat.1BC 0.00%		% of OverClass		% of UnderClass
Cat. NC 50.00%		50.00%		33.33%
Fentem's criteria		According to Fentem, test are UNACCEPTABLE if...		

Within Corrosive: Cat. 1A Versus 1BC	
Sensitivity for 1A	16.67 %
Sensitivity for 1BC	0.00%
Accuracy	6.67%
Corr. Versus Non Corr	
Sensitivity for Corr	83.33 %

Sensitivity for Non Corr	50.00 %	%OverClass NC as Corr. 50.00%	%OverClass 1BC as 1A 75.00%	If % OverClass NC as Corr $\geq$ 50%	If OverClass 1BC as 1A $\geq$ 50%
<i>Accuracy</i>	75.00 %	%UnderClass 1BC as NC 25.00%	%UnderClass 1A as NC 0.00%	If % UnderClass. 1BC as NC $\geq$ 30%	If UnderClass. 1A as NC $\geq$ 30%

Discussion

1- Discussion when the Entire Set of Chemicals is Addressed

22. Correct classifications were examined (e.g. an *in vivo* 1A chemical correctly classified as 1A *in vitro* by a given test method) and correct classifications rates were calculated as well, for each test method. Over- and Under- classifications were examined (e.g. respectively an *in vivo* 1BC chemical classified as 1A; an *in vivo* 1BC chemical classified as NC chemical) and over-, under- prediction rates were calculated as well, for each test method.

23. The sub-group assessed that accuracy value is the greatest in EpiSkin® (78.75%), then EpiDerm® (70.42%) and then SkinEthic® (64.58%). However accuracy values for EpiDerm® and SkinEthic® are comparable, while accuracy value for EpiSkin® was significantly higher than in the two others.

24. It was assessed that over-prediction rate of 1BC chemicals—i.e. over-classified as 1A— is the lowest in EpiSkin® (21.51%), then EpiDerm® (41.94%) and then SkinEthic® (46.24%). However over-prediction rates in EpiDerm® and SkinEthic® are comparable, while over-prediction rate in EpiSkin® is significantly lower than in the two others.

25. The sub-group was also concerned about under-prediction rates of the chemicals in a public health scope. None of Category 1A chemicals were classified as Non Corrosive in any of the methods. Concerning Category 1BC chemicals that were under-predicted as NC there are 0/240 (0.00%), 2/240 (0.83%), 7/240 (2.92%) under-predictions respectively for EpiDerm®, EpiSkin®, SkinEthic®. Therefore, regarding those very low rates the sub-group felt that the three test methods were sufficiently protective for human health.

2- Discussion on “MTT-reducers and Organic Bases Issues”

26. During the sub-group work it was hypothesised that MTT-reducers and organic bases would both have potential to impair test results, i.e. would be more likely to provide high over-prediction rates. Therefore, statistics were performed for each test method on non-MTT-reducers chemicals and MTT-reducers only, as well as non-organic bases and organic bases only.

27. In general, it was shown that accuracy values are greater when MTT-reducers were removed, compared to the accuracy values obtained when the entire set of chemicals is kept. Similarly, it was shown also that accuracy values are greater when organic bases were removed, compared to the accuracy values when the entire set of chemicals is kept. Over-predictions rates decrease when either MTT-reducers or organic bases are removed.

28. In parallel, accuracy values decrease when focusing only either on MTT-reducers or organic bases, and over-predictions rates are increased. However accuracy values and over-prediction rates calculated either on MTT-reducers only or on organic-bases only are related both to very small samples. The sub-group believed therefore that information of accuracy values and over-predictions rates, based on

those small samples, could not be considered at this stage very informative. It was thus felt that focusing on MTT-reducers only, or organic bases only, could not lead to any strong and definitive conclusion.

29. Noteworthy while EpiDerm[®] and SkinEthic[®] were comparable for the MTT-reducers and organic bases issues, EpiSkin[®] resisted better than the two other test methods in misclassifications due to MTT-reducers and organic bases.

30. To determine whether or not MTT-reducers and organic bases were able to impair test results, and how they could do so, the Secretariat proposed to focus on the endpoint which is cell viability and to compare cell viabilities in MTT-reducers and non-MTT-reducers, as well as cell viability between organic bases and non-organic bases.

31. For that purpose, several statistical assessments were performed to compare cell viabilities values between those groups for each test method. Cell viabilities values were examined and if cell viability is significantly lower in a group (e.g. organic bases group) compared to another group (e.g. non-organic bases) it would be likely to observe higher over-prediction rates in the group where cell viability is the lowest. However the final prediction does not depend only on cell viabilities values, but also on prediction models (i.e. the way to interpret the endpoints) and on intrinsic properties of the test system (i.e., physiological cell properties).

32. Despite the impossibility to derive conclusions at this stage—because of small sample size of MTT-reducers and organic bases sets—on the non-applicability of the test methods for MTT-reducers and organic bases, the sub-group supported by the entire expert group recommended to include all the tables, including those on MTT-reducers and organic bases.

33. Therefore the test guideline was revised and states, following the previous outcomes, that:

- Results from EpiSkin[®] can be directly used as they are and allow direct sub-categorisation (1A Versus 1B-and-C Versus Non Corrosive) of the test chemicals,
- Results from EpiDerm[®] & SkinEthic[®] should take into account high over-classification rates* occurring when using those two test methods for 1B-and-1C category (see tables). Therefore, for EpiDerm[®] & SkinEthic[®], chemicals that are classified as 1B-and-1C can be considered as 1B-and-1C, but chemicals for which cell viability at 3 minutes is below 50% (labeled as 1A) should be considered as 1, that is to say that either under the prediction principle they could be claimed as 1A or they should undergo further testing to be possibly confirmed as 1B-and-1C.²

² Over-prediction rates for 1B-and-C chemicals:
EpiSkin[®]: 21.51%; EpiDerm[®]: 41.94%; SkinEthic[®]: 46.24%

CONTINGENCY TABLES AND CALCULATIONS OF SENSITIVITY, SPECIFICITY AND ACCURACY VALUES PERFORMED ON THE SET OF 20 REFERENCE CHEMICALS FOR CATEGORISATION CORROSIVE VERSUS NON CORROSIVE CHEMICALS

34. This section presents the contingency tables and calculations of sensitivity, specificity and accuracy values based on the set of 20 reference chemicals developed for categorisation Corrosive versus Non Corrosive chemicals.

35. EpiSkin, EpiDerm and SkinEthic gave same results for categorisation of the 20 reference chemicals between Corrosive and Non Corrosive chemicals. For each test method those 20 reference chemicals were tested over 3 runs, and thus provided 60 classifications. Table 6 of this document displays those results. The results are:

- All Corrosive chemicals were correctly classified, that is to say that none of the Corrosive chemicals were under-predicted as Non Corrosive in any of the test methods. Therefore, the three methods are protective in terms of human health; sensibility reaches thus 100%—which corresponds to 30 correct predictions of Corrosive chemicals out of 30.
- 80% of the Non Corrosive chemicals were correctly classified, and 20% were over-predicted as Corrosive. That is to say that specificity reaches 80%—which corresponds to 24 correct predictions of Non Corrosive chemicals out of 30.
- When using this set of 20 reference chemicals, as sensitivity—i.e. ability to correctly classify Corrosive chemicals— is 100% and as specificity —i.e. ability to correctly classify Non Corrosive chemicals, is 80% — therefore all test methods are protective for human health.

Table 6. Results including accuracy values, correct classification, over-prediction and under-prediction rates for EpiSkin[®], EpiDerm[®] and SkinEthic[®] based on the set of 20 reference chemicals for categorization

EpiSkin with reference chemicals with the set of 20 reference chemicals for categorisation			
<i>In vivo</i> categories	Test methods: EpiSkin EpiDerm SkinEthic		Sum
	Classified as Corrosive	Classified as Cat. NC	
<i>In vivo</i> Cat. Corrosive	30	0	30
<i>In vivo</i> Cat. NC	6	24	30
Sum	36	24	60

Calculations over the 2 Corrosive and Non-Corrosive categories:			
Accuracy (Pred. C) 90.00%		Checking the misclassifications:	Predictive Values:
Correctly Class		% OverClass Non-Corr. as Corr. 20.00%	Probability that an <i>in vitro</i> Corr. Is really Corr. i.e. Positive Predictive Value 83.33%
Correctly Classified as Cat. Corr, i.e. Sensitivity	100.00%	% UnderClass Corr. As Non-Corr. 0.00%	Probability that an <i>in vitro</i> Non-Corr. Is really Non-Corr. i.e. Negative Predictive Value 100.00%
Correctly classified as Cat. NC, i.e. Specificity	80.00%		

CONTINGENCY TABLES AND CALCULATIONS OF SENSITIVITY, SPECIFICITY AND ACCURACY VALUES PERFORMED ON THE SET OF 30 REFERENCE CHEMICALS FOR SUB-CATEGORISATION 1A VERSUS 1BC VERSUS NON CORROSIVE CHEMICALS

36. This section presents the contingency tables and calculations of sensitivity, specificity and accuracy values based on the set of 30 reference chemicals developed for sub-categorisation 1A versus 1BC versus Non Corrosive chemicals. This set includes 10 Cat. 1A chemicals, 10 Cat. 1BC chemicals, and 10 Cat. Non Corrosive chemicals. Those chemicals were tested over 3 runs and provided therefore 90 classifications. All results and contingency are provided in Table 7 of this document.

37. It corresponds to the values provided in table 3 in the Annex 1—Performance Standards—of the revised test guideline. EpiSkin® is the method which gave the best results on the set of 30 reference chemicals. As performance standards are intended to be used by test developers and as those test developers should develop test methods with the highest accuracy values to comply with a validation process, it is the reason why EpiSkin® was chosen to provide those targets values.

38. The 30 reference chemicals were tested over 3 runs and thus provided 90 classifications.

- Correctly classified chemicals included 77 classifications out of 90 classifications, i.e. correct prediction rate is 85.56% (29 1A correctly classified + 24 1BC correctly classified + 24 Non Corrosive correctly classified),
- None of the Corrosive chemicals were under-predicted as Non Corrosive, i.e. under-prediction rate of Corrosive as Non Corrosive is 0.00%. Therefore, potential new test methods that would apply for validation in the future should comply with this 100% rate of protection for human health,
- Furthermore there's only 1 under-classification out of 90 which concerns a 1A chemical under-classified as 1BC, i.e. under-prediction rate of 1A as 1BC is 1.11%,
- Over-prediction rates concern either 1BC chemicals over-predicted as 1A, or Non Corrosive chemicals over-predicted as 1BC. For each of those over-predictions it corresponds to 6 over-predictions out of 90, i.e. 6.66%. The total amount of over-predictions is thus 12 out of 90 classifications, i.e. 13.33%.
- In the end:
 - correct classification rate for 1A chemicals is 96.67% (29 out of 30 classifications), under-prediction rate for 1BC chemicals is 3.33% (1 out of 30 classifications) and therefore under-prediction rate as Non Corrosive chemicals is 0.00% (none of the 1A chemicals were classified as Non-Corrosive),
 - correct classification rate for 1BC chemicals is 80.00% (24 out of 30 classifications), over-prediction rate as 1A chemicals is 20.00% (6 out of 30 classifications) and therefore under-prediction rate is 0% (none of the 1BC chemicals were classified as Non Corrosive),

- correct classification rates for Non Corrosive chemicals is 80% (24 out of 30 classifications), over-prediction rate as 1BC chemicals is 20.00% (6 out of 30 classifications).

Table 7. Results including accuracy values, correct classification, over-prediction and under-prediction rates for EpiSkin® based on the set of 30 reference chemicals

EpiSkin testing the set of 30 reference chemicals over 3 runs, i.e. providing 90 classifications)					
In vivo categories	Test method: EpiSkin			Sum	
	In vitro Classified as Cat. 1A	In vitro Classified as Cat. 1BC	In vitro Classified as Cat. NC		
In vivo Cat. 1A	29	1	0	30	
In vivo Cat. 1BC	6	24	0	30	
In vivo Cat. NC	0	6	24	30	
Sum	35	31	24	90	

Calculations over the 3 chemicals categories:					
Within Corrosive: Cat. 1A Versus 1BC		Checking the misclassifications over the 3 categories:			
		Accuracy (Pred. C)	% OverClass 1BC as 1A	%OverClassNC as 1BC	% OverClass NC as 1A
Sensitivity for 1A	96.67%	85.56%	20.00%	20.00%	0.00%
Sensitivity for 1BC	80.00%				
Accuracy	88.33%				
Cat. Corr. Versus Cat. Non Corr		Correctly Class	% UnderClass 1A as 1BC	% underClass 1A as NC	% UnderClass 1BC as NC
		Cat.1A 96.67%	3.33%	0.00%	0.00%
		Cat.1BC 80.00%			
		Cat. NC 80.00%	% of OverClass 13.33%	% of UnderClass 1.11%	
Sensitivity for Corr.		Fentem's criteria			
		According to Fentem, tests are UNACCEPTABLE if...			
Sensitivity for Non Corr.	93.33%	%OverClass NC as Corr. 20.00%	%OverClass 1BC as 1A 20.00%	If % OverClass NC as Corr≥50%	If OverClass 1BC as 1A≥50%
Accuracy	80.00%	%UnderClass 1BC as NC 0.00%	%UnderClass 1A as NC 0.00%	If % UnderClass. 1BC as NC≥30%	If UnderClass. 1A as NC≥30%

CONCLUSION

39. Test Guideline No. 431 “*In Vitro* Skin Corrosion: Reconstructed Human *Epidermis* (RhE) Test Method” was updated and revised following the results of the sub-group work and was agreed by the entire expert group on skin corrosion. Issues related to the impact of MTT-reducers and organic bases chemicals were assessed and discussed during the process through several commenting rounds. It was reached that it was not possible at this stage to state on the non-applicability of the guideline to those types of chemicals.

40. Currently, the revised guideline states that:

- Results from EpiSkin[®] can be directly used as they are and allow direct sub-categorisation (1A Versus 1B-and-C Versus Non Corrosive) of the test chemicals,
- Results from EpiDerm[®] & SkinEthic[®] should take into account high over-classification rates* occurring when using those two test methods for 1B-and-1C category (see tables). Therefore, for EpiDerm[®] & SkinEthic[®], chemicals that are classified as 1B-and-1C can be considered as 1B-and-1C, but chemicals for which cell viability at 3 minutes is below 50% (labeled as 1A) should be considered as 1, that is to say that either under the prediction principle they could be claimed as 1A or they should undergo further testing to be possibly confirmed as 1B-and-1C.³

41. A list of 20 reference chemicals was developed for categorisation between Corrosive versus Non Corrosive chemicals. It is included in the performance standards of the guideline (Annex 1 of TG 431). The three test methods gave the same results of categorisation when testing those 20 reference chemicals. All Corrosive chemicals were correctly classified as Corrosive. New test methods that would be proposed in the future for validation by test developers (so called me-too tests) should therefore comply with this feature of 100% protective for human health.

42. A list of 30 reference chemicals was developed for sub-categorisation between Cat. 1A chemicals Versus Cat. 1BC chemicals Versus Non Corrosive chemicals. It is included in the performance standards of the guideline (Annex 1 of TG 431). EpiSkin[®] test methods gave the best results in terms of accuracy values for sub-categorisation when testing those 30 reference chemicals. In terms of public health it's noteworthy to bear in mind that all Corrosive chemicals were correctly classified as Corrosive.

43. New test methods that would be proposed in the future for validation by test developers (so called me-too tests) should therefore comply with this feature of 100% protective for human health. Target values for the validation of a new test method applying for sub-categorisation should be:

- For Cat. 1A chemicals: 96.67% (i.e. 29 out of 30) and none of them are under-classified as Non Corrosive —see the row “*in vivo* Cat. 1A” in Table 7;

³ Over-prediction rates for 1B-and-C chemicals:

EpiSkin[®]: 21.51%; EpiDerm[®]: 41.94%; SkinEthic[®]: 46.24%

- For Cat. 1BC chemicals: 80.00% (i.e. 24 out of 30) and none of them are under-classified as Non-Corrosive —see the row “*in vivo* Cat. 1BC” in Table 7;
- For Non Corrosive chemicals: 80.00% (i.e. 24 out of 30) and the 20.00% of those Non Corrosive chemicals are classified as 1BC but never 1A —see the row “*in vivo* Cat. NC” in Table 7.

44. Therefore when using this list of 30 reference chemicals, a new test developer that would get at least the same results as EpiSkin[®] would therefore comply with the following minimum probabilities and requirements for sub-categorisation:

- When a chemical is *in vitro* predicted as 1A, the probability that it really corresponds to a 1A chemical is 82.86% (i.e. 29 out of 35) —and the probability that it corresponds to a real 1BC chemical is 17.14%; the probability that it corresponds to a real Non Corrosive chemical is 0.00%—see the column “*in vitro* Classified as Cat. 1A” in Table 7.
- When a chemical is *in vitro* predicted as 1BC, the probability that it really corresponds to a 1BC chemical is 77.42% (i.e. 24 out of 31) —and the probability that it corresponds to a real 1A chemical is 3.23%; the probability that it corresponds to a real Non Corrosive chemical is 19.35%— see the column “*in vitro* Classified as Cat.1BC” in Table 7.
- When a chemical is *in vitro* predicted as Non Corrosive, the probability that it really corresponds to a real Non Corrosive chemical is 100% (i.e. 24 out of 24) —see the column “*in vitro* Classified as Cat. Non Corrosive” in Table 7.