Comparative assessment of 24-hr primary skin irritation test and human patch test data with in vitro skin irritation tests according to OECD Test Guideline 439 (for quasi-drugs in⋯

Sugiyama, Mariko ; Akita, Masaharu ; Alepee, Nathalie ; Fujishiro, Miyuki ; Hagino, Shigenobu ; Handa, Yuki ; Ikeda, Hidefumi ; Imai,⋯

(URL)
https://hdl.handle.net/20.500.14094/90005643

Original Article

Comparative assessment of 24-hr primary skin irritation test and human patch test data with *in vitro* skin irritation tests according to OECD Test Guideline 439 (for quasi-drugs in Japan)

Mariko Sugiyama[1,2], Masaharu Akita[1,3], Nathalie Alépée[4], Miyuki Fujishiro[5],
Shigenobu Hagino[1,2], Yuki Handa[6], Hidefumi Ikeda[1,7], Noriyasu Imai[1,8], Setsuko Jitsukawa[9],
Masakazu Katoh[10], Koji Kurihara[11], Daiki Kyotani[11], Shigeyuki Nomura[8], Yuko Okamoto[1],
Hidenobu Okumura[1,12], Takashi Omori[1,13], Kenji Sugibayashi[1,14], Hiroaki Todo[1,14],
Akemi Toyoda[1,15] and Yasuo Ohno[1,16]

[1]Working Group on Alternatives to Primary Skin Irritation Test, Japanese Society of Alternatives to
Animal Experiments (JSAAE), 5-3-13 Otsuka, Bunkyo-Ku, Tokyo, 112-0012, Japan
[2]Shiseido Global Innovation Center, 2-2-1, Hayabuchi, Tsuzuki-ku, Yokohama, 224-8558, Japan
[3]Faculty of Family and Consumer Sciences, Kamakura Women's University, 6-1-3, Ofuna, Kamakura, Kanagawa, 247-8512, Japan
[4]L'Oréal Research & Innovation, 1, Avenue Eugène-Schueller, 93601, Aulnay-sous-bois, France
[5]Cosmos Technical Center Co., Ltd., 3-24-3, Hasune, Itabashi-ku, Tokyo, 174-0046, Japan
[6]Kurabo Industries Ltd., 14-5, Shimokida-cho, Neyagawa-shi, Osaka, 572-0823, Japan
[7]Mandom Corp., 5-12, Juniken-cho, Chuo-ku, Osaka, 540-8530, Japan
[8]KOSÉ Corp., Ltd., 1-18-4, Azusawa, Itabashi-ku, Tokyo, 174-0051, Japan
[9]L'Oréal Research & Innovation, 2-1, Sakado 3-chome, Takatsu-ku, Kawasaki-shi, Kanagawa, 213-0012, Japan
[10]Japan Tissue Engineering Co., Ltd., 6-209-1, Miyakitadori, Gamagori, Aichi, 443-0022, Japan
[11]Nikoderm Research Inc., 1-6-14, Azuchimachi, Chuo-ku, Osaka City, OSAKA, 541-0052, Japan
[12]Institute of Beauty and Health Science, 43-2, Imacho, Higashiomi, Shiga, 521-1211, Japan
[13]Division of Biostatistics, Department of Social/Comunity Medicine and Health Science, Kobe University School of Medicine,
7-5-2 Kusunoki-cho Chuo-ku, Kobe, 650-0017, Japan
[14]Faculty of Pharmacy and Pharmaceutical Sciences, Josai University, 1-1, Keyakidai, Sakado-shi, Saitama, 350-0295, Japan
[15]Pola Chemical Industries, Inc., 560, Kashio-cho, Totsuka-ku, Yokohama, 244-0812, Japan
[16]Kihara Memorial Yokohama Foundation, 1-6, Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan

ABSTRACT — The Organisation for Economic Co-operation and Development (OECD) Test Guideline (TG) 439 is an *in vitro* test method of reconstructed human *epidermis* (RhE), which was developed for hazard identification of irritating chemicals in accordance with a primary skin irritation test using rabbits with 4-hr exposure. A regulation for quasi-drugs in Japan requires data from primary skin irritation tests using rabbits to undergo 24-hr exposure, and this is used as an evidence for 24-hr closed patch tests in humans. In this study with the same chemicals, primary skin irritation test data using rabbits undergoing 24-hr exposure and a 24-hr occlusive human patch test data were analyzed by comparing the results obtained with four test methods adopted in OECD TG 439. The performances of *in vitro* test methods showed a positive predictive value of 72.7-85.7% to predict the results of 24-hr primary rabbit skin irritation test knowing that its positive predictive value was 57.1% against humans only. The prediction factors of *in vitro* test methods were higher for the human patch test data with a sensitivity reaching 60 to 80%. Three surfactants gave false negatives in some of the RhE methods evaluated with the human patch test, but in each case, they were correctly classified as positive when evaluated at double concentration. Therefore, the approach of setting the margin to 2 was effective in eliminating false negatives. This suggests that *in vitro* test methods are useful for assessing skin irritation potential without animal testing for the application of quasi-drugs in Japan.

Key words: Skin irritation, 24-hr Human patch test, Reconstructed human epidermis, OECD TG 439,
In vitro, Quasi-drugs

Correspondence: Mariko Sugiyama (E-mail: mariko.sugiyama@to.shiseido.co.jp)

M. Sugiyama *et al.*

INTRODUCTION

Skin irritation is an essential toxicological endpoint in assessing a cosmetic's safety. The rabbit-based Draize method (Draize *et al.*, 1944) has been developed and used for assessment of skin irritation, but several test conditions have been established for different purposes. Test methods for 4-hr exposure in rabbits, which are defined in OECD TG 404 (OECD, 2015a) are used when evaluating the hazards of chemicals and classifying the toxic effects according to United Nations (UN) Globally Harmonized System of Classification and Labelling of Chemicals (GHS) categories. When evaluating the safety of intentional applications to the skin on a daily basis, such as quasi-drugs, a 24-hr exposure study is required in Japan. In the safety data for quasi-drugs, primary skin irritation after 24-hr exposure in rabbits has been evaluated, and a 24-hr patch test with humans has been established after weak skin irritation was confirmed in animal studies (Japan Cosmetic Industry Association, 2015) (Guidebook for Quasi-drugs Application, 2017).

With regard to cosmetics, animal experiments were banned in the European Union in 2013 due to the growing movement for animal welfare, substituting by evaluations using alternative methods. As an alternative to the testing of skin irritation on animals, the OECD TG 439 using the reconstructed human epidermis (RhE) test methods was approved in 2010 and subsequently revised in 2013 and 2015 with the additional epidermis test methods (OECD, 2015b). This approach has been developed as a stand-alone alternative to OECD TG 404 for the assessment of chemical hazards and can be used to predict irritation above UN GHS category 2 in a 4-hr rabbit test.

Since it has been reported that skin reactions increase with prolonged exposure (Cruzan *et al.*, 1986; Gilman *et al.*, 1978), it was speculated that the test methods adopted in OECD TG 439 might not sufficiently sensitive to predict the results of a primary skin irritation study with 24-hr exposure using rabbits as required in the application for quasi-drugs.

In Japan, the Review Committee on the Proposed Application of Data on the Safety of Quasi-drugs for Manufacturing Approval Applications was organized between 2007 and 2009, and the use of alternatives in the application for marketing approval for quasi-drugs was considered. However, it was concluded that as a primary skin irritation assessment for quasi-drugs application dossiers, it is not sufficiently predictable to use the alternatives recommended at the moment (MHLW Grants System, 2011), but it is preferable to use the alternatives from the standpoint of animal welfare. Judgment on whether to use animal experiments or alternatives should be left to the applicant. Even if a test chemical has been identified as not classified using alternative methods, human patch testing is required for confirmation. Thus, some ethical consideration should be addressed.

As a result, OECD TG 439 has not been applied as a standalone for the assessment of skin irritation in the application for attaining marketing approval for quasi-drugs in Japan.

On the other hand, there are few data that sufficiently examine the predictability of the test methods adopted in OECD TG 439 for the results of primary skin irritation test and human patch test with 24-hr exposure as test samples. Imai *et al.* has been evaluated the correlation between the *in vitro* OECD TG 439 test method, Epi-Derm™ SIT (EPI-200) and primary skin irritation tests of 24-hr exposure in rabbits using the same test chemicals as in this report (Imai *et al.*, 2018). It is important to note that only one adopted RhE test method was evaluated and no correlations were investigated with human patch tests. Therefore, the irritation potential of some ingredients was evaluated according to the four OECD TG 439 test methods and compared with the results of a primary skin irritation test and a 24-hr human occlusive patch test. In this study, using *in vitro* methods listed in OECD TG 439, the correlation of *in vivo* (24-hr exposure, rabbits, humans) and *in vitro* were evaluated and the usefulness of the *in vitro* test as a prediction method for the 24-hr human patch test was examined.

MATERIALS AND METHODS

Chemicals

The test chemicals used for assessment are ingredients used in cosmetics. Table 1 shows the International Nomenclature of Cosmetic Ingredients (INCI) names of the test chemicals and their characterization, i.e., the degree of lipophilicity categories of the chemicals, CAS Registration Number, Log Kow values, obtained as predicted values using EPI Suite™ (ver.4.1) as well as the supplier's details of each chemicals being tested using the four RhE test methods. Test chemicals with Log Kow > 3.5 are judged as oil soluble chemicals. Characteristics of the 40 test chemicals are shown in Table 2, summarizing their physical categories and their corresponding *in vivo* irritancy potential. The test chemicals included 12 surfactants and 23 oils, representing respectively 30% and 58% of the test set. The remaining 5 test chemicals were acids, polymers, and powders. Irritation categories were also widely distributed from none to severe scores.

Table 1. Test Chemicals.

No.	INCI	CAS Registry Number™	Logkow*	Solubility**	Major classification	Minor classification	Supplier for *in vitro* test			
							EpiSkin™ SIT	SkinEthic™RHE SIT	EpiDerm™SIT (EPI-200)	LabCyte EPI-MODEL24 SIT
1	PEG-400	25322-68-3	-2.3	-	Polymer	-	Sigma	Sigma	Sigma-Aldrich	Wako
2	Benzalkonium Chloride	8001-54-5	2.93	-	Surfactant	Cation	Acros Organic	Acros Organic	Sigma-Aldrich	Wako
3	Xanthan Gum	11138-66-2	unpredictable	-	Polymer	-	Internal Batch	Internal Batch	DANISCO	Tokyo Chemical Industry Co., Ltd.
4	Squalane	111-01-3	14.63	oil soluble	Oil	Hydrocarbon	Aldrich	Aldrich	Sigma-Aldrich	Wako
5	Talc	14807-96-6	-1.5	-	Powder		Daito Kasei	Daito Kasei	Matsumura Sangyo Co., Ltd.	Wako
6	Barium Sulfate	7727-43-7	0.63	-	Powder	-	Sigma	Sigma	Sigma-Aldrich	Kanto Chemical
7	Lactic Acid	50-21-5	-0.65	-	Acid	-	Sigma	Sigma	Nacalai tesque	Wako
8	Diisopropyl Sebacate	7491-02-3	5.17	oil soluble	Oil	Ester oil	Interchim Wako	Interchim Wako	Wako	Wako
9	Cyclotetrasiloxane	556-67-2	6.79	oil soluble	Oil	Silicone	Acros Organic	Acros Organic	Tokyo Chemical Industry Co., Ltd.	Wako
10	Acetyl Tributyl Citrate	77-90-7	4.29	oil soluble	Oil	Ester oil	Aldrich	Aldrich	Morimura Bros., Inc.	Wako
11	Mineral Oil	8042-47-5	unpredictable	oil soluble	Oil	Hydrocarbon	Sigma-Aldrich	Sigma-Aldrich	Sigma-Aldrich	Wako
12	Myristic Acid	544-63-8	5.98	oil soluble	Oil	Higher fatty acid	Sigma	Sigma	Emery Oleochemicals	Wako
13	Diethyl Sebacate	110-40-7	4.33	oil soluble	Oil	Ester oil	Aldrich	Aldrich	Kanto Chemical Co., Inc.	Wako
14	Isopropyl Myristate	110-27-0	7.17	oil soluble	Oil	Ester oil	Aldrich	Aldrich	Sigma-Aldrich	Wako
15	Octyldodecyl Myristate	22766-83-2	15.52	oil soluble	Oil	Ester oil	COSMOS	COSMOS	Wako	Not applicable
16	Lecithin	8002-43-5	13.41	oil soluble	Surfactant	Ampholytic	Not applicable	Not applicable	J-OIL MILLS. Inc.	Wako
17	Ethyl Oleate	111-62-6	8.51	oil soluble	Oil	Ester oil	Aldrich	Aldrich	NOF Co.	Wako
18	Cetyl Ethylhexanoate	59130-69-7	10.61	oil soluble	Oil	Ester oil	COSMOS	COSMOS	Kao	Wako
19	Lauric Acid	143-07-7	5.00	oil soluble	Oil	Higher fatty acid	Sigma-Aldrich	Sigma-Aldrich	SAFC	Wako
20	Olea Europaea (Olive) Fruit Oil	8001-25-0	unpredictable	oil soluble	Oil	Ester oil	Sigma	Sigma	Wako	Wako
21	Ethyhexyl Stearate	22047-49-0	11.59	oil soluble	Oil	Ester oil	CRODA	CRODA	Kao	Wako
22	Ethylhexyl Palmitate	29806-73-3	10.61	oil soluble	Oil	Ester oil	BASF	BASF	Wako	Wako
23	Laureth-3	3055-94-5	3.95	oil soluble	Surfactant	Anion	Tokyo Chemical Industry Co., Ltd.	Tokyo Chemical Industry Co., Ltd.	Tokyo Chemical Industry Co., Ltd.	Not applicable
24	Octyldodecanol	5333-42-6	8.63	oil soluble	Oil	Higher alcohol	COGNIS	COGNIS	BASF	Tokyo Chemical Industry Co., Ltd.
25	TEA-Laureth Sulfate	27028-82-6	1.83	-	Surfactant	Anion	COSMOS	COSMOS	Kao	Wako
26	Rozmariuns Officinalis (Rosemary) Leaf Oil	84604-14-8/8000-25-7	undetectable	oil soluble	Oil	Essential oil	Aldrich	Aldrich	ChromaDex	Sigma Aldrich

M. Sugiyama *et al.*

Table 1. (Continued).

No.	INCI	CAS Registry Number™	Logkow*	Solubility**	Major classification	Minor classification	Supplier for *in vitro* test			
							EpiSkin™ SIT	SkinEthic™RHE SIT	EpiDerm™SIT (EPI-200)	LabCyte EPI-MODEL24 SIT
27	Oleth-2	9004-98-2	6.13	oil soluble	Surfactant	Nonion	Aldrich	Aldrich	Tokyo Chemical Industry Co., Ltd.	Wako
28	Isopropyl Palmitate	142-91-6	8.16	oil soluble	Oil	Ester oil	Alfa Aesar	Alfa Aesar	Internal Batch	Wako
29	Lauryl Alcohol	112-53-8	4.77	oil soluble	Oil	Higher alcohol	Alfa Aesar	Alfa Aesar	Sigma-Aldrich	Tokyo Chemical Industry Co., Ltd.
30	Steareth-4	59970-10-4	6.62	oil soluble	Surfactant	Nonion	Sigma	Sigma	Wako	Wako
31	Cetylpyridinium Chloride	123-03-5	2.80	-	Surfactant	Cation	Aldrich	Aldrich	Sigma-Aldrich	Wako
32	Dicaprylyl Ether	629-82-3	6.94	oil soluble	Oil	Ether	Aldrich	Aldrich	Tokyo Chemical Industry Co., Ltd.	Tokyo Chemical Industry Co., Ltd.
33	Oleic Acid	112-80-1	7.73	oil soluble	Oil	Higher fatty acid	Sigma	Sigma	Wako	Wako
34	Oleyl Alcohol	143-28-2	7.5	oil soluble	Oil	Higher alcohol	Sigma	Sigma	CRODA	Wako
35	Myristyl Betaine	2601-33-4	3.46	-	Surfactant	Ampholytic	Aldrich	Aldrich	Affymetrix	Not applicable
36	Steartrimonium Chloride	112-03-8	4.17	oil soluble	Surfactant	Cation	Sigma	Sigma	Tokyo Chemical Industry Co., Ltd.	Wako
37	Citrus Aurantium Dulcis (Orange) Oil	8008-57-9	4.83	oil soluble	Oil	Essential oil	Internal Batch	Internal Batch	Wako	Wako
38	Sodium Lauryl Sulfate	151-21-3	1.69	-	Surfactant	Anion	Aldrich	Aldrich	Sigma-Aldrich	Wako
39	Lauryl Betaine	683-10-3	2.47	-	Surfactant	Ampholytic	CHEMOS	CHEMOS	Wako	Wako
40	Laurtrimonium Chloride	112-00-5	1.22	-	Surfactant	Cation	Acros Organic	Acros Organic	Tokyo Chemical Industry Co., Ltd.	Wako

*Log Kow was calculated with EPI Suite™ (ver.4.1) .
**Chemicals which are LogKow > 3.5 and oils were classified as oil soluble chemicals.
INCI, International Nomenclature of Cosmetic Ingredients; CAS Registry Number™, Chemical Abstructs Service Registry Number.

Table 2. Categorization by physical properties and irritancy of test chemicals

Major classification	Irritation Category				No. of chemicals	ratio
	severe	moderate	mildly	none		
Surfactant	5	4	2	1	12	30%
Acid	0	0	1	0	1	3%
Oil	1	6	15	1	23	58%
Polymer	0	0	0	2	2	5%
Powder	0	0	0	2	2	5%
	6	10	18	6	40	100%
	Irritant:16		Non-Irritant:24		40	100%

Irritation was classified as follows based on the Draize score (Rabbits, 24 hr).
none (0), mild irritant (> 0 but ≤ 2), moderate irritant (> 2 but ≤ 5), severe irritant (>5 but ≤ 8)

The existing *in vivo* dataset

Preparation of dataset for primary skin irritation test

The primary skin irritation test dataset is shown in Table 3. In this study, the results of previously existing primary skin irritation tests with 24-hr rabbit exposure were used in the interest of animal welfare. Data that had been produced and stored in companies were collected upon request of the Japan Cosmetic Industry Association (JCIA) and used as the *in vivo* dataset. No *in vivo* data were generated for this review, as historical data were considered as such. Therefore, information such as manufacturer names and trade names was treated as confidential and the reports remained unavailable upon request. However, the method is briefly described below (Draize, 1959).

In the primary skin irritation test, the test chemical was applied to rabbits (n = 3-6) for 24 hr (0.5 mL or 0.5 g / 6 cm^2), it was removed, and erythema (score 0-4) and edema (score 0-4) were macroscopically determined at 24 and 72 hr after application. All data were obtained by applying test samples to intact skin, summing up the score for each animal, and dividing the score by the number of animals and the number of reading to obtain a mean score (Primary Irritation Indices: P.I.I.). The results were classified according to the P.I.I. values as an irritant (I) if they exceeded 2, a non-irritant (NI) if they are equal or less (≤) than 2. The P.I.I. and irritation potential (class) for each test chemical are presented in Table 3. In the primary skin irritation test, 16 irritant chemicals and 24 non-irritant chemicals were used.

The test results were also classified according to the category of irritant values (none (0), mild (0 < P.I.I. ≤ 2), moderate (2 < P.I.I. ≤ 5), and severe (5 < P.I.I. ≤ 8). Datasets were characterized by substance category (surfactant, acid, polymer, oil) and stimulus intensity (Table 2).

The existing dataset for human 24-hr closed patch testing

In this study, the results from human patch tests were obtained and compiled by the JCIA for the same INCI chemicals as those from the primary skin irritation test. No *in vivo* data were generated for this review, as historical data were considered as such. For 31 test chemicals, 24-hr human occlusive patch test data were collected from companies affiliated with the JCIA. Human studies were conducted in accordance with the Declaration of Helsinki (World Medical Association Declaration of Helsinki-Ethical Principles for Medical Research Involving Human Subjects) after review by each company's ethics council. The number of datasets was 31 instead of 40 data for the primary skin irritation test, as human studies were not conducted when primary skin irritation tests revealed more than a moderate irritation effect. In addition, for similar reasons, test chemicals were occasionally found to be tested at lower concentrations than in the primary skin irritation test with rabbits. Similarly, to the *in vivo* primary skin irritation test dataset, manufacturer names and trade names were treated as confidential and were not disclosed.

In human patch tests, 31-55 healthy adult subjects were used to apply the test sample under occlusion for 24 hr, and test samples were judged by Japanese criteria (6-point evaluation; -, ±, +, ++, +++, ++++) 1-3 hr after adhesive removal (Kawamura *et al.*, 1970). Adhesive plasters used were either Finn Chambers (φ = 8 mm) or Torii-Patch test adhesive plasters (small). Irritation indices were determined by dividing the number of positive subjects (+ or higher score) by the number of test subjects and by the irritation index based on test subjects' readings in accordance with the method of Sugai (Sugai, 1995) (Table 3). Sugai's Irritation Indices (S.I.I.) were calculated by giving each of the judgments a rating score (-: 0, ±: 0.5, +: 1, +++: 2, +++: 3, ++++: 4), summing the scores of the subjects, dividing the score by the number of subjects, and multiplying by 100. Sugai classified irritant indi-

Table 3. *In vitro* and *in vivo* results.

No.	INCI name	Conc.	solvent	*in vivo* (rabbit 24 hr)		*in vivo* (human 24 hr Patch Test)			*in vitro* (OECD TG439) [3]			
				P.I.I.	class [1]	Positive rate	S.I.I.	Class [2]	EpiSkin™ SIT	SkinEthic™RHE SIT	EpiDerm™SIT (EPI-200)	LabCyte EPI-MODEL24 SIT
1	PEG-400	100.0%		0.0	NI	0.0%	0.0	NI	NC 101.4 ± 1.9%	NC 93.4 ± 2.5 %	NC 99.8 ± 2.6 %	NC 98.2 ± 5.4 %
2	Benzalkonium Chloride	0.01%	D.W.	0.0	NI				NC 105.8 ± 3.5%	NC 93.2 ± 4.3 %	NC	NC 96.6 ± 8.2 %
		0.1%	D.W.			28.9%	43.4	I	C 102.8 ± 3.2%	C 77.4 ± 5.5 %	C 36.6 ± 7.6 %	C 43.4 ± 11.9 %
3	Xanthan Gum	10.0%	D.W.	0.0	NI	0.0%	4.3	NI	NC 101.3 ± 3.3%	NC 98.4 ± 7.6 %	NC	NC 94.0 ± 2.6 %
4	Squalane	100.0%		0.0	NI	0.0%	0.0	NI	NC 95.8 ± 7.3%	NC 105.3 ± 2.5 %	NC 96.1 ± 3.0 %	NC 92.5 ± 4.1 %
5	Talc	100.0%		0.0	NI	0.0%	0.0	NI	NC 99.1 ± 12.3%	NC 103.8 ± 4.7 %	NC 97.4 ± 1.3 %	NC 107.6 ± 8.8 %
6	Barium Sulfate	65.0%	D.W.	0.0	NI	0.0%		NI	NC 102.1 ± 7.1%	NC 97.7 ± 2.6 %	NC	NC 92.9 ± 2.5 %
		100.0%		0.0	NI	0.0%	3.1	NI	NC 95.3 ± 9.1%	NC 99.4 ± 2.5 %	NC 99.0 ± 3.6 %	
7	Lactic Acid	10.0%	D.W.	0.1	NI	N.D.: pH < 2			C 34.2 ± 4.5%	C 39.6 ± 0.2 %	C	C 30.1 ± 21.8%
8	Diisopropyl Sebacate	100.0%		0.2	NI	0.0%	0.0	NI	NC 91.1 ± 3.0%	NC 91.0 ± 5.7%	NC	NC 104.6 ± 9.4%
9	Cyclotetrasiloxane	100.0%		0.3	NI	0.0%	0.9	NI	NC 101.5 ± 12.8%	NC 97.5 ± 1.3%	NC 96.3 ± 1.0%	NC 96.6 ± 3.2%
10	Acetyl Tributyl Citrate	2.0%	D.W.	0.3	NI	0.0%	0.9	NI	NC 101.3 ± 2.1%	NC 98.7 ± 13.6%	NC	NC
		100.0%		0.3	NI				NC 110.0 ± 7.8%	NC 92.8 ± 3.6%	NC	NC 116.4 ± 5.1%
11	Mineral Oil	100.0%		0.5	NI	0.0%	0.0	NI	NC 101.1 ± 7.6%	NC 95.8 ± 3.4%	NC 99.5 ± 2.6%	NC 84.5 ± 3.0%
12	Myristic Acid	20.0%	Squalane	0.6	NI	0.0%	0.0	NI	NC 105.8 ± 9.8%	NC 91.0 ± 6.0%	NC	NC 72.9 ± 7.6%
13	Diethyl Sebacate	100.0%		0.7	NI	0.0%	0.0	NI	NC 104.2 ± 2.0%	NC 99.6 ± 5.2%	NC 103.6 ± 4.8%	NC 96.5 ± 3.9%
14	Isopropyl Myristate	100.0%		0.8	NI	0.0%	0.0	NI	NC 102.8 ± 3.2%	NC 113.3 ± 1.0%	NC 100.5 ± 3.5%	NC 99.3 ± 5.8%
15	Octyldodecyl Myristate	100.0%		0.8	NI	0.0%	0.0	NI	NC 102.5 ± 2.6%	NC 96.9 ± 0.6%	NC	NC
16	Lecithin	10.0%	D.W.	1.3	NI	0.0%	0.0	NI	NC	NC	NC	NC 96.2 ± 9.0%
17	Ethyl Oleate	100.0%		1.3	NI	0.0%	0.0	NI	NC 104.3 ± 1.3%	NC 96.5 ± 5.3%	NC	NC 97.0 ± 5.4%

Table 3. (Continued).

No.	INCI name	Conc.	solvent	*in vivo* (rabbit 24 hr) P.I.I.	class [1]	*in vivo* (human 24 hr Patch Test) Positive rate	S.I.I.	Class [2]	*in vitro* (OECD TG439) [3] EpiSkin™ SIT	SkinEthic™RHE SIT	EpiDerm™SIT (EPI-200)	LabCyte EPI-MODEL24 SIT
18	Cetyl Ethylhexanoate	100.0%		1.3	NI	0.0%	1.3	NI	NC 111.6 ± 2.6%	NC 87.0 ± 4.3%	NC 103.4 ± 2.1%	NC 103.1 ± 2.7%
19	Lauric Acid	20.0%	Squalane	1.4	NI				NC 93.5 ± 9.0%	C 33.3 ± 2.6%	C	NC 103.6 ± 2.2
		100.0%				0.0%	0.0	NI	NC 103.8 ± 2.3%	NC 56.6 ± 39.4%	C 19.6 ± 3.6%	NC 88.8 ± 14.0
20	Olea Europaea (Olive) Fruit Oil	100.0%		1.4	NI	2.0%	4.0	NI	NC 94.9 ± 1.0%	NC 104.1 ± 6.9%	NC 107.9 ± 4.1%	NC 101.2 ± 9.4%
21	Ethylhexyl Stearate	100.0%		1.7	NI	0.0%	0.0	NI	NC 95.0 ± 11.6%	NC 91.2 ± 7.1%	NC 81.3 ± 15.1%	NC 99.5 ± 4.6%
22	Ethylhexyl Palmitate	100.0%		2.0	NI	0.0%	0.0	NI	NC 95.7 ± 6.6%	NC 91.0 ± 7.7%	NC 111.9 ± 5.1%	NC 97.4 ± 2.5%
23	Laureth-3	10.0%	D.W.	2.0	NI				C 16.9 ± 12.7%	C 44.9 ± 1.1%	C	
24	Octyldodecanol	50.0%	Petrolatum	2.0	NI	0.0%	0.0	NI	NC 93.5 ± 0.0%	NC 90.5 ± 2.8%	NC	NC 94.9 ± 3.6%
25	TEA-Laureth Sulfate	10.0%	D.W.	2.1	I				C 28.9 ± 7.8%	NC 76.2 ± 2.0%	C 3.3 ± 0.40%	NC 82.4 ± 6.0%
26	Rozmarinum Officinalis (Rosemary) Leaf Oil	10.0%	Squalane		I	0.0%	0.0	NI	C[6] 20.1 ± 5.1%	C[6] 18.5 ± 9.5%	C	
		100.0%		2.2	I				C 35.2 ± 14.9%	C 2.1 ± 1.8%	NC	NC 57.4 ± 5.3%
27	Oleth-2	5.0%	Petrolatum	2.3	I				NC 93.2 ± 2.1%	NC 94.4 ± 13.0%	NC 95.2 ± 4.2%	NC 101.7 ± 5.6%
28	Isopropyl Palmitate	100.0%		2.3	I				C 97.1 ± 0.8%	NC 102.1 ± 2.6%	NC	NC 101.8 ± 9.9%
29	Lauryl Alcohol	100.0%		2.8	I				C 7.1 ± 1.3%	C 2.4 ± 0.1%	NC[6] 93.9% ± 5.7% (C 25.0 ± 9.6%)	NC 93.1 ± 4.8%
30	Steareth-4	5.0%	Mineral Oil	2.9	I	0.0%	2.8	NI	NC 101.2 ± 11.1%	NC 95.3 ± 4.2%	NC 100.7 ± 3.1%	NC 77.5 ± 6.1%
31	Cetylpyridinium Chloride	1.0%	D.W.		I	65.8%	105.3	I	C 14.7 ± 10.7%	C 1.7 ± 0.4%	C 3.7 ± 0.3%	C 42.3 ± 11.5%
		10.0%	D.W.	3.3	I			(I)[4]	C 4.3 ± 0.8%	C 1.1 ± 0.3%	C 4.6 ± 0.9%	C 28.1 ± 2.3%
32	Dicaprylyl Ether	100.0%		3.5	I				NC 97.2 ± 10.3%	C 36.1 ± 11.6%	NC 100.8 ± 2.5%	NC 81.7 ± 12.5%

Table 3. (Continued).

No.	INCI name	Conc.	solvent	in vivo (rabbit 24 hr) P.I.I.	class [1]	in vivo (human 24 hr Patch Test) Positive rate	S.I.I.	Class [2]	in vitro (OECD TG439) [3] EpiSkin™ SIT	SkinEthic™RHE SIT	EpiDerm™SIT (EPI-200)	LabCyte EPI-MODEL24 SIT
33	Oleic Acid	100.0%		3.6	I	2.2%	6.5	NI	NC 101.7 ± 9.1%	NC 93.3 ± 2.5%	NC 92.0 ± 13.2%	NC 88.6 ± 7.4%
34	Oleyl Alcohol	100.0%		4.7	I	0.0%	0.0	NI	NC 93.7 ± 7.5%	NC 106.4 ± 12.3%	NC	NC 107.1 ± 1.4%
35	Myristyl Betaine	10.0%	D.W.	5.3	I				C 6.7 ± 0.3%	C 7.3 ± 6.2%	C	
36	Steartrimonium Chloride	100.0%		5.8	I				C 19.7 ± 10.0%	C 12.1 ± 5.9%	C 7.4 ± 0.3%	C 24.9 ± 4.9%
37	Citrus Aurantium Dulcis (Orange) Oil	1.0%	Squalane			0.0%	1.0	NI	NC 103.7 ± 8.7%	NC 101.1 ± 15.9%	NC 91.9 ± 3.4%	
		100.0%		6.0	I				C 18.7 ± 15.6%	C 1.6 ± 0.2%	NC	C 22.3 ± 2.7%
38	Sodium Lauryl Sulfate	1.0%	D.W.			44.7%	85.5	I	C 25.8 ± 14.1%	C 1.1 ± 0.2%	C 2.5 ± 0.07%	NC 54.7 ± 11.2%
		5.0%	D.W.	6.3	I			(I) [4]	C 5.8 ± 1.3%	C 0.9 ± 0.1%	C 2.1 ± 0.04%	C 7.8 ± 1.4%
39	Lauryl Betaine	1.0%	D.W.			39.5%	50.0	I	C 39.9 ± 0.6%	C 37.3 ± 3.9%	NC 72.2 ± 20.7%	NC [6] 67.2 ± 13.7% (C 47.7 ± 1.1%)
		100.0%		6.8	I			(I) [4]	C 9.1 ± 2.8%	C 1.7 ± 0.7%	C 2.6 ± 0.3%	C 25.2 ± 1.3%
40	Lautrimonium Chloride	1.0%	D.W.			15.8%	34.2	I	C 5.1 ± 0.9%	C 1.0 ± 0.3%	C 1.3 ± 1.4%	C 20.6 ± 0.8%
		10.0%	D.W.	6.9	I			(I) [4]	C 5.4 ± 0.1%	C 0.9 ± 0.3%	C 3.8 ± 0.2%	C 8.2 ± 1.6%

1) I (Irritant: moderate or severe irritant): P.I.I. (Primary Irritation Index) > 2.0
   NI (Non-Irritant: none or mild irritant): P.I.I. ≤ 2.0
2) I (Irritant): S.I.I. (Sugai's Irritation Index) ≥30
   NI (Non-Irritant): S.I.I. < 30
3) C (Classified): Viability ≤ 50%
   NC (Not Classified): Viability > 50 %
4) Prediction from low concentration test results
5) When evaluation result is C and NC, it is treated as NC
6) Evaluation was performed more than once in several RhE.
INCI, International Nomenclature of Cosmetic Ingredients; D.W., Distilled Water; OECD TG, Organisation for Economic Co-operation and Development Test Guideline

ces of 5 or less as "safe products", 5-15 as "acceptable", 15-30 as "requiring improvement", and +30 as "risk products". In the human patch test, there is no general criterion that indicates the presence or absence of irritation, so in this study, an S.I.I. of 30 or more is regarded as an irritant (I), and an S.I.I. of less than 30 is regarded as a non-irritant (NI). These criteria were decided by referring to the opinion of the dermatologist on the S.I.I., in order not to apply any irritant to subjects, from an ethical view point.

In vitro evaluation by OECD TG 439 test methods

In vitro skin irritation test methods assess skin irritation potential by measuring the viability of the epidermal cells of an RhE model upon application of the test chemical, which diffuses through the stratum corneum and penetrates into the epidermal layer. The tested chemicals can be distinguished as an irritant (classified as UN GHS category 2) if the viability of epidermal cells measured by 3-(4,5-Dimethylthiazol-2-yl)- 2,5-diphenyltetrazolium bromide (MTT) is 50% or less, or as a non-irritant if it is greater than 50% (UN GHS No category). The current study was carried out on the commercially available RhE models (EpiSkin™ SM, SkinEthic™ RHE, EpiDerm EPI-200, and LabCyte EPI-MODEL24) with a protocol respectively defined that differ for example in the volume and the duration of the application of test chemicals (EpiSkin™, 2018; EpiDerm™, 2009; SkinEthic™, 2018; LabCyte, 2011).

Regarding *in vitro* data, the providers/developers of the test methods performed the testing according to OECD TG 439 protocols. Reagents were purchased and tests were conducted by L'Oréal Research & Innovation for EpiSkin™ SIT and SkinEthic™ RHE SIT (hereinafter abbreviated as SkinEthic™SIT), Kurabo Industries Ltd.

for EpiDerm™ SIT (EPI-200) (hereinafter abbreviated as EpiDerm™SIT), and Japan Tissue Engineering Co., Ltd. for LabCyte EPI-MODEL24 SIT (hereinafter abbreviated as LabCyte SIT). The applied concentration and vehicle for test samples were based on the available data (primary skin irritation test and human patch test).

RESULTS

Correlation between *in vivo* studies

Data from the primary skin irritation test and human patch test are shown in Table 3. Of the test chemicals of which primary skin irritation test data (40 chemicals) and human patch test data (31 chemicals) were available, 23 chemicals were found to have been tested at the same concentration both in a primary skin irritation test and in the patch test, and 9 were tested at different concentrations. For four of the surfactants [1% Cetylpyridinium Chloride (#31), 1% Sodium Lauryl Sulfate (#38), 1% Lauryl Betaine (#39), and 1% Lautrimonium Chloride (#40)], human skin reactions were extrapolated to be positive at the same concentration as in the primary skin irritation test because the results of the human patch tests were positive at lower concentrations than in the primary skin irritation test on the rabbits. Predicted results for these chemicals were added to Table 3 in brackets. This resulted in 26 test samples whose results could be compared to the results of the primary skin irritation test and human patch test. Tables 4-1 through 4-4 show the concordance of the results of the primary skin irritation test in rabbits, the human patch test, and the *in vitro* assessment using the RhE models. Assessments were organized in a 2 × 2-contingency table and then sensitivity, specificity and accuracy were calculated according to the ECVAM SIVS study (Balls *et al.,* 1990; Spielmann

Table 4-1. Contingency table for *in vivo* results.

2x2 contingency table and predictive capacity for *in vivo* (rabbit)

		in vivo (rabbit)						
		I	NI	Σ	sensitivity	100.0%	(4/4)
in vivo	I	(4)*	0	(4)*	specificity	86.4%	(19/22)
(human)	NI	3	19	22	accuracy	84.6%	(22/26)
	Σ	7	19	26	PPV	57.1%	(4/7)
					NPV	100%	(19/19)

*:prediction

sensitivity: number of irritants correctly classified / total number of irritants
specificity: number of non-irritants correctly classified / total number of non-irritants
accuracy: number of correctly classified irritants and non-irritants / total number of substances
PPV (positive predictive value): number of irritants correctly classified / total number of substances predicted as irritants
NPV (negative predictive value): number of non-irritants correctly classified / total number of substances predicted as non-irritants
in vivo (rabbit); I (Irritant): P.I.I. (Primary Irritation Index) > 2.0, NI (Non-Irritant): P.I.I. ≤ 2.0
in vivo (human); I (Irritant): S.I.I. (Sugai's Irritation Index) ≥ 30, NI (Non-Irritant): S.I.I. < 30

M. Sugiyama *et al.*

Table 4-2. Contingency tables for RhE models.

a) 2x2 contingency table and predictive capacity for EpiSkin

		in vitro EpiSkin™ SIT					
		C	NC	Σ	sensitivity	62.5%	(10/16)
in vivo	I	10	6	16	specificity	91.3%	(21/23)
(rabbit)	NI	2	21	23	accuracy	79.5%	(31/39)
	Σ	12	27	39	PPV	83.3%	(10/12)
					NPV	77.8%	(21/27)

b) 2x2 contingency table and predictive capacity for SkinEthic

		in vitro SkinEthic™ RHE SIT					
		C	NC	Σ	sensitivity	62.5%	(10/16)
in vivo	I	10	6	16	specificity	87.0%	(20/23)
(rabbit)	NI	3	20	23	accuracy	76.9%	(30/39)
	Σ	13	26	39	PPV	76.9%	(10/13)
					NPV	76.9%	(20/26)

c) 2x2 contingency table and predictive capacity for EpiDerm

		in vitro EpiDerm™ SIT (EPI-200)					
		C	NC	Σ	sensitivity	50.0%	(8/16)
in vivo	I	8	8	16	specificity	87.5%	(21/24)
(rabbit)	NI	3	21	24	accuracy	72.5%	(29/40)
	Σ	11	29	40	PPV	72.7%	(8/11)
					NPV	72.4%	(21/29)

d) 2x2 contingency table and predictive capacity for LabCyte

		in vitro LabCyte EPI-MODEL24 SIT					
		C	NC	Σ	sensitivity	40.0%	(6/15)
in vivo	I	6	9	15	specificity	95.5%	(21/22)
(rabbit)	NI	1	21	22	accuracy	73.0%	(27/37)
	Σ	7	30	37	PPV	85.7%	(6/7)
					NPV	70.0%	(21/30)

sensitivity: number of irritants correctly classified / total number of irritants
specificity: number of non-irritants correctly classified / total number of non-irritants
accuracy: number of correctly classified irritants and non-irritants / total number of substances
PPV (positive predictive value): number of irritants correctly classified / total number of substances predicted as irritants
NPV (negative predictive value): number of non-irritants correctly classified / total number of substances predicted as non-irritants
in vitro; C (Classified): Viability ≤ 50%, NC (Not Classified): Viability > 50 %
in vivo (rabbit); I (Irritant): P.I.I. (Primary Irritation Index) > 2.0, NI (Non-Irritant): P.I.I. ≤ 2.0

et al., 2007). Table 4-1 shows the correlation of the 26 test samples compared between rabbits and humans. The sensitivity was 100%, suggesting that all test samples that were positive in humans were predictable in rabbits. However, the positive predictive value (PPV) is 57.1%, indicating that the primary skin irritation test is significantly over-predicted against the human patch test.

The scatter plot using P.I.I in the primary skin irritation test and S.I.I. in human patch test shown in Fig. 1 demonstrates the correlation of skin response between rabbits and humans. The scatter plots were plotted separately for oil-soluble chemicals and other chemicals.

Fig. 1 shows that in oil-soluble chemicals there are no correlations between rabbits and humans, and these included chemicals such as Oleyl Alcohol (OA in Fig. 1) that do not react to human skin even when P.I.I. was 4.7. Of the test chemicals evaluated as irritant in rabbits, three [5% Steareth-4 (#30), 100% Oleic Acid (#33), and 100% Oleyl Alcohol (#34)] were evaluated as non-irritants in human patch tests, and were oil-soluble chemicals. For chemicals other than oil-soluble chemicals, four surfactants which were evaluated as irritant in rabbits were correctly predicted for human patch test (Fig. 1, True Positive).

Comparative study of *in vivo* and *in vitro* to predict skin irritation

Table 4-3. Contingency tables for RhE models.

a) 2x2 contingency table and predictive capacity for EpiSkin

		in vitro EpiSkin™ SIT						
		C	NC	Σ		sensitivity	80.0%	(4/5)
in vivo	I	4	1	5		specificity	96.0%	(24/25)
(human)	NI	1	24	25		accuracy	93.3%	(28/30)
	Σ	5	25	30		PPV	80.0%	(4/5)
						NPV	96.0%	(24/25)

b) 2x2 contingency table and predictive capacity for SkinEthic

		in vitro EpiDerm™ SIT (EPI-200)						
		C	NC	Σ		sensitivity	80.0%	(4/5)
in vivo	I	4	1	5		specificity	96.0%	(24/25)
(human)	NI	1	24	25		accuracy	93.3%	(28/30)
	Σ	5	25	30		PPV	80.0%	(4/5)
						NPV	96.0%	(24/25)

c) 2x2 contingency table and predictive capacity for EpiDerm

		in vitro EpiDerm™ SIT (EPI-200)						
		C	NC	Σ		sensitivity	80.0%	(4/5)
in vivo	I	4	1	5		specificity	92.3%	(24/26)
(human)	NI	2	24	26		accuracy	90.3%	(28/31)
	Σ	6	25	31		PPV	66.7%	(4/6)
						NPV	96.0%	(24/25)

d) 2x2 contingency table and predictive capacity for LabCyte

		in vitro LabCyte EPI-MODEL24 SIT						
		C	NC	Σ		sensitivity	60.0%	(3/5)
in vivo	I	3	2	5		specificity	100.0%	(20/20)
(human)	NI	0	20	20		accuracy	92.0%	(23/25)
	Σ	3	22	25		PPV	100%	(3/3)
						NPV	90.9%	(20/22)

sensitivity: number of irritants correctly classified / total number of irritants
specificity: number of non-irritants correctly classified / total number of non-irritants
accuracy: number of correctly classified irritants and non-irritants / total number of substances
PPV (positive predictive value): number of irritants correctly classified / total number of substances predicted as irritants
NPV (negative predictive value): number of non-irritants correctly classified / total number of substances predicted as non-irritants
in vitro; C (Classified): Viability ≤ 50%, NC (Not Classified): Viability > 50 %
in vivo (human); I (Irritant): S.I.I. (Sugai's Irritation Index) ≥ 30, NI (Non-Irritant): S.I.I. < 30

Correlation of OECD TG 439 adopted test methods with *in vivo* tests

The *in vitro* test results (cell viability and UN GHS categorization) using four RhE models (EpiSkin™ SM, SkinEthic™ RHE, EpiDerm EPI-200, and LabCyte EPI-MODEL24) are shown in the Table 3.

Table 4-2 summarizes the existing results of the *in vivo* (primary skin irritation in rabbits) and the obtained results *in vitro* as well as the sensitivity, specificity, accuracy, PPV and the negative predictive value (NPV). In the Epi-Derm™ SIT, 100% Lauryl Alcohol (#29) assessment was rated as "Classified" and "Not Classified" in two runs, so the result was considered as Not Classified in the Tables and Figures (most conservative approach).

The correlations between *in vivo* (rabbits) and *in vitro* results were weak using all RhE test methods, with a sensitivity ranging from 40.0 to 62.5% (EpiSkin™ SIT: 62. 5%, SkinEthic™ SIT 62. 5%, EpiDerm™ SIT: 50. 0% and LabCyte SIT: 40. 0%), a specificity ranging from 87.5 to 95.5% (EpiSkin™ SIT: 91. 3%, SkinEthic™ SIT: 87. 0%, EpiDerm™ SIT: 87. 5%, and LabCyte SIT: 95. 5%), an accuracy ranging from 72.5 to 79.5% (EpiSkin™ SIT:

M. Sugiyama *et al.*

Table 4-4. Contingency tables for RhE models (without oil soluble chemicals).

a) 2x2 contingency table and predictive capacity for EpiSkin

		in vitro EpiSkin™ SIT					
		C	NC	Σ	sensitivity	100.0%	(6/6)
in vivo	I	6	0	6	specificity	83.3%	(5/6)
(rabbit)	NI	1	5	6	accuracy	91.7%	(11/12)
	Σ	7	5	12	PPV	85.7%	(6/7)
					NPV	100.0%	(5/5)

b) 2x2 contingency table and predictive capacity for SkinEthic

		in vitro SkinEthic™ RHE SIT					
		C	NC	Σ	sensitivity	83.3%	(5/6)
in vivo	I	5	1	6	specificity	83.3%	(5/6)
(rabbit)	NI	1	5	6	accuracy	83.3%	(10/12)
	Σ	6	6	12	PPV	83.3%	(5/6)
					NPV	83.3%	(5/6)

c) 2x2 contingency table and predictive capacity for EpiDerm

		in vitro EpiDerm™ SIT (EPI-200)					
		C	NC	Σ	sensitivity	100.0%	(5/5)
in vivo	I	5	0	5	specificity	83.3%	(5/6)
(rabbit)	NI	1	5	6	accuracy	90.9%	(10/11)
	Σ	6	5	11	PPV	83.3%	(5/6)
					NPV	100.0%	(5/5)

d) 2x2 contingency table and predictive capacity for LabCyte

		in vitro LabCyte EPI-MODEL24 SIT					
		C	NC	Σ	sensitivity	80.0%	(4/5)
in vivo	I	4	1	5	specificity	83.3%	(5/6)
(rabbit)	NI	1	5	6	accuracy	81.8%	(9/11)
	Σ	5	6	11	PPV	80.0%	(4/5)
					NPV	83.3%	(5/6)

sensitivity: number of irritants correctly classified / total number of irritants
specificity: number of non-irritants correctly classified / total number of non-irritants
accuracy: number of correctly classified irritants and non-irritants / total number of substances
PPV (positive predictive value): number of irritants correctly classified / total number of substances predicted as irritants
NPV (negative predictive value): number of non-irritants correctly classified / total number of substances predicted as non-irritants
in vitro; C (Classified): Viability ≤ 50%, NC (Not Classified): Viability > 50 %
in vivo (rabbit); I (Irritant): P.I.I. (Primary Irritation Index) > 2.0, NI (Non-Irritant): P.I.I. ≤ 2.0

79.5%, SkinEthic™ SIT 76.9%, EpiDerm™ SIT: 72.5% and LabCyte SIT: 73.0%), and a PPV ranging from72.7 to 85.7% (EpiSkin™ SIT: 77.8%, SkinEthic™ SIT: 76.9%, EpiDerm™ SIT: 72.7%, and LabCyte SIT: 85.7%).

The scatter plot shown in Fig. 2 demonstrates the relationship between primary skin irritation indices (P.I.I.) and cellular viability. This Figure shows that when the P.I.I. *in vivo* test is greater than or equal to 5.3 (Myristyl Betaine: MB in Fig. 2), the viability *in vitro* of all chemicals is lower than 50%. On the other hand, it is difficult to adjust the cut-off value (cell viability %) of the *in vit-ro* test to identify a Draize score greater than or equal to a moderate irritant.

In at least one of the four RhE test methods, false negatives against *in vivo* rabbit results were observed with 10 oil-soluble chemicals: 7 oil formulations [100% Rozmarinum Officinalis Rosemary leaf oil (#26; Rosemary Oil), 100% Isopropyl Palmitate (#28), 100% Lauryl Alcohol (#29), 100% Dicaprylyl Ether (#32), 100% Oleic Acid (#33), 100% Oleyl Alcohol (#34), and 100% Citrus Aurantium Dulcis (Orange) oil (#37; Orange oil)], 1 anionic surfactant [10% TEA-Laureth Sulfate (#25)] and 2 nonionic
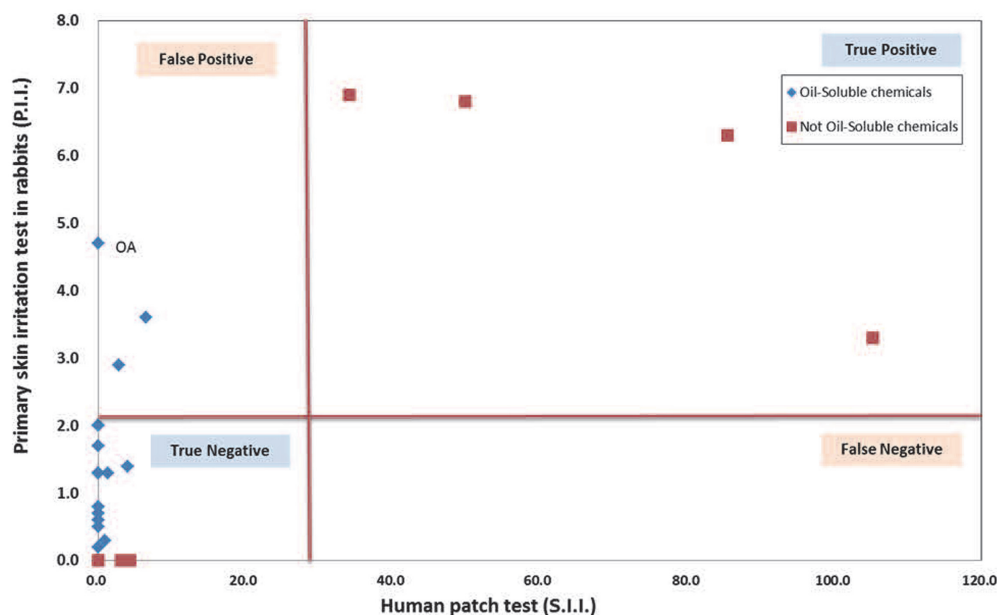
Comparative study of *in vivo* and *in vitro* to predict skin irritation

Fig. 1. Correlation of primary skin irritation in rabbits (P.I.I.) with human patch test (S.I.I.). Primary skin irritation test in rabbits; Positive (Irritant): P.I.I. (Primary Irritation Index) > 2.0, Negative (Non-Irritant): P.I.I. ≤ 2.0. Human patch test; Positive (Irritant): S.I.I. (Sugai's Irritation Index) ≥ 30, Negative (Non-Irritant): S.I.I. < 30.
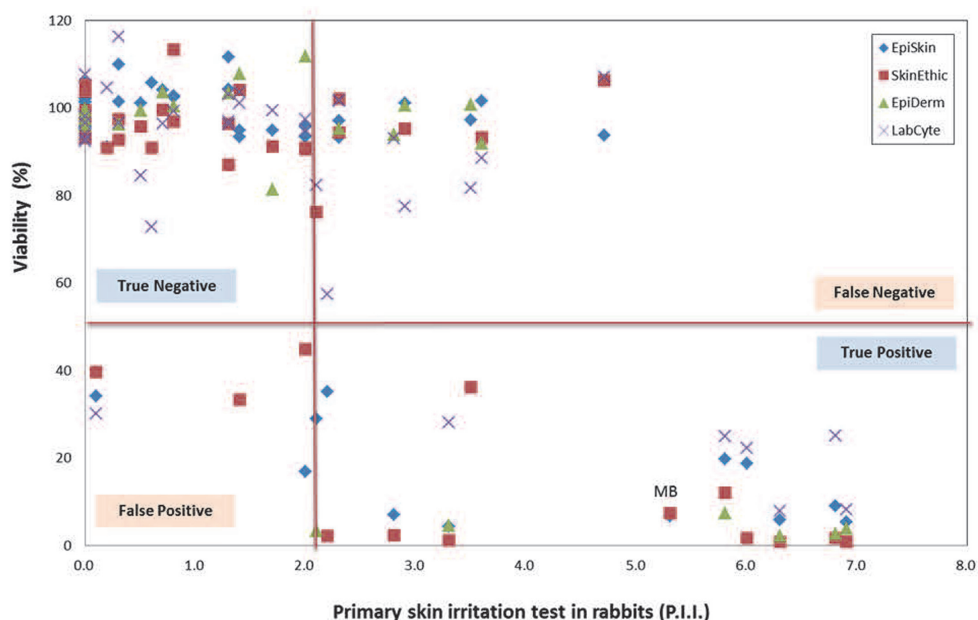

Fig. 2. Correlation of cell viability % data of the 4 adopted OECD TG 439 test methods with primary skin irritation test in rabbits (P.I.I.). EpiSkin: EpiSkin™ SIT, SkinEthic: SkinEthic™ RHE SIT, EpiDerm: EpiDerm™ SIT (EPI-200), LabCyte: LabCyte EPI-MODEL24 SIT. *In vitro* method using OECD TG439; Positive (Classified): Viability ≤ 50%, Negative (Not Classified): Viability > 50%. Primary skin irritation test in rabbits; Positive (Irritant): P.I.I. (Primary Irritation Index) > 2.0, Negative (Non-Irritant): P.I.I. ≤ 2.0.

surfactants [5% Oleth-2 (#27), and 5% Steareth-4 (#30)]. False-negative results of *in vitro data versus in vivo* primary test were presumed to be chemicals that caused mild irritation at 4 hr, and showed increased irritation when applied for 24-hr. Knowing that it is known that the 4-hr primary skin irritation test already over predicts, increasing the application up to 24-hr would also have an impact of the relevance of these positive results obtained in rabbits against the observation in humans.

The true positive chemicals predicted to be irritating in both *in vivo* and *in vitro* assessments are those expected to exceed 2.3 (UN GHS category 2) in the Draize score of rabbits at 4 hr of application. There were 11 chemicals which were determined as "Classified" using any of these RhE test methods: 7 surfactants [TEA-Laureth Sulfate (#25), 10% Cetylpyridinium Chloride (#31), 10% Myristyl Betaine (#35), 100% Steartrimonium Chloride (#36), 5% Sodium Lauryl Sulfate (#38), 100% Lauryl Betaine (#39), and 10% Lautrimonium Chloride (#40)], and 4 oils [100% Rosemary Oil (#26), 100% Lauryl Alcohol (#29), 100% Dicaprylyl Ether (#32), and 100% Orange Oil (#37)].

On the other hand, three false positive chemicals were detected in this study, which are 10% Lactic Acid (#7), 20% Lauric Acid (#19), and 10% Laureth-3 (#23). Of these, 10% Lactic Acid was considered to be influenced by low pH because the pH was less than 2. Chemicals with extreme pH are likely to damage the integrity of the cells upon contact with tissues, such as skin, and thus may be classified as skin corrosives (Cat. 1). For chemicals with pH $\leq$ 2.0 or pH $\geq$ 11.5, skin corrosion could be expected. However, it may also be important to take into consideration the acid/alkaline reserve (a measure of buffering capacity of a chemical) on skin (Young *et al*., 1988). The observation of only mild irritancy in rabbits may be the result of reduction by buffering ability on skin.

To assess the relevance of the *in vitro* test methods adopted in OECD TG 439, the concordance of the results with the results of 24-hr human patch test were examined. Table 4-3 summarizes the concordance between *in vivo* (human patch test) and *in vitro* (OECD TG 439 test methods) results using sensitivity, specificity, accuracy, PPV and NPV. The numbers of test chemicals for which correlations between the results of the *in vitro* assessment and *in vivo* (human) could be compared were 30, 30, 31, and 25 for the EpiSkin™ SIT, the SkinEthic™ SIT, the EpiDerm™ SIT, and the LabCyte SIT, respectively. In the LabCyte SIT, the results of 1% Lauryl Betain (#39) were rated as classified and not classified in two runs; thus the result was evaluated as not classified in the tables and graphs.

The correlation between the results of *in vitro* and *in vivo* (human) was a sensitivity of 60.0-80.0% (EpiSkin™ SIT: 80. 0%, SkinEthic™ SIT: 80. 0%, EpiDerm™ SIT: 80. 0%, LabCyte SIT: 60. 0%), a specificity of 92.3-100% (EpiSkin™ SIT: 96. 0%, SkinEthic™ SIT: 96. 0%, EpiDerm™ SIT: 92. 3%, LabCyte SIT: 100%), and an accuracy of 90.3-93.3% (EpiSkin™ SIT: 93. 3%, SkinEthic™ SIT: 93. 3%, EpiDerm™ SIT: 90. 3%, LabCyte SIT: 92. 0%), indicating that the correlation between the human patch test and *in vitro* was higher than that between the primary skin irritation in the rabbits and *in vitro*. Moreover, a PPV of 66.7-100% supported the performance of *in vitro* methods to predict the results of 24-hr human patch test.

The scatter plot in Fig. 3 demonstrates the relationship between the human patch test (S.I.I.) and cellular viability. False negatives were 0.1% Benzalkonium Chloride (#2) (EpiSkin™ SIT, SkinEthic™ SIT), 1.0% Lauryl Betaine (#39) (EpiDerm™ SIT, LabCyte SIT), and 1.0% Sodium Lauryl Sulfate (#38) (LabCyte). In both epidermal models, a false negative result was observed in the surfactants.

Therefore, the examination was conducted to determine the lowest concentration that would produce positive results in the RhE test methods, by increasing the test concentration of test chemicals applied onto RhE models. As a result, 0.2% Benzalkonium Chloride was determined as classified in EpiSkin™ SIT and SkinEthic™ SIT. Similarly, 2.0% Sodium Lauryl Sulfate was classified with LabCyte SIT and 2.0% Lauryl Betaine with EpiDerm™ SIT and LabCyte SIT. Therefore, all false-negative test chemicals using all RhE models were classified at 2-fold higher concentrations (see Table 5).

DISCUSSION

Preparation of primary skin irritation test dataset

Although the test methods adopted in OECD TG 439 were not developed for evaluating skin irritation in Japanese quasi-drugs approval system, the usefulness of the *in vitro* test methods was examined for predicting primary irritation in rabbits and patch tests in humans. To begin with, primary skin irritation test results based on 24-hr exposure in rabbits and 24-hr occlusive human patch test results for quasi-drugs in Japan were prepared as an *in vivo* dataset. Chemicals included in the *in vivo* dataset were not re-tested with animals or humans, and therefore existing historical data were used. The collected *in vivo* dataset was considered appropriate as the *in vivo* dataset for evaluating the usefulness of the alternative methods, as chemicals in the dataset included an approx-
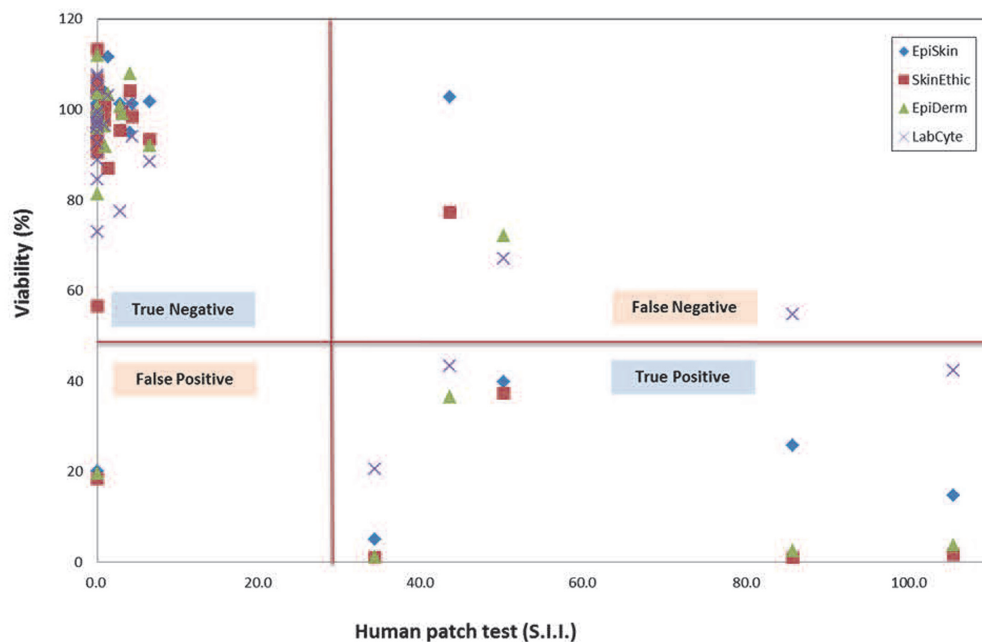
Comparative study of *in vivo* and *in vitro* to predict skin irritation

Fig. 3. Correlation of cell viability % data of the 4 adopted OECD TG 439 test methods with human patch test (S.I.I.). EpiSkin: EpiSkin™ SIT, SkinEthic: SkinEthic™ RHE SIT, EpiDerm: EpiDerm™ SIT (EPI-200), LabCyte: LabCyte EPI-MODEL24 SIT. *In vitro* method using OECD TG439; Positive (Classified): Viability ≤ 50%, Negative (Not Classified): Viability > 50%. Human patch test; Positive (Irritant): S.I.I. (Sugai's Irritation Index) ≥ 30, Negative (Non-Irritant): S.I.I. < 30.
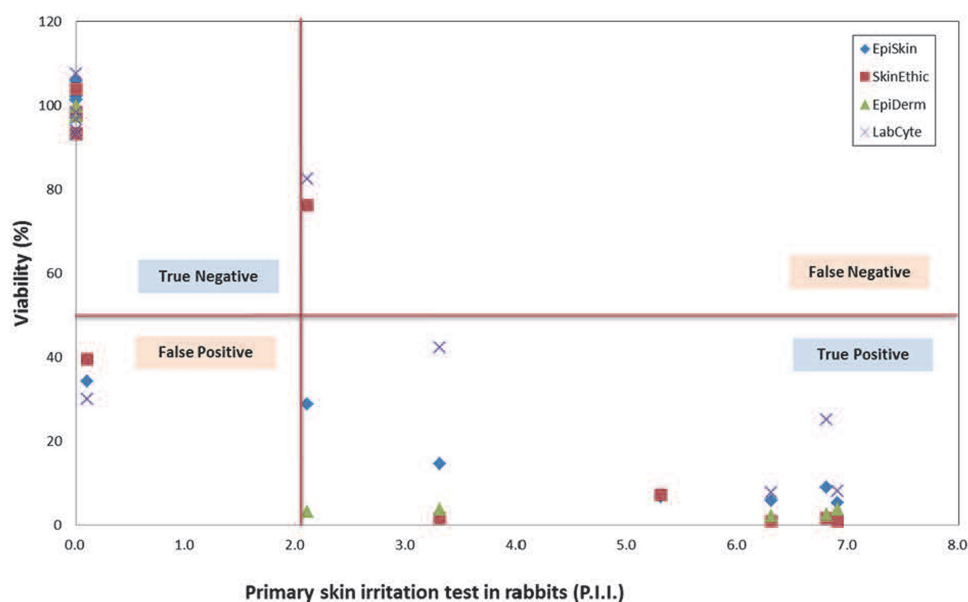
Fig. 4. Correlation of cell viability % data of the 4 adopted OECD TG439 test methods with primary irritation test in rabbits (P.I.I.) (test chemicals: Log Kow ≤ 3.5). EpiSkin: EpiSkin™ SIT, SkinEthic: SkinEthic™ RHE SIT, EpiDerm: EpiDerm™ SIT (EPI-200), LabCyte: LabCyte EPI-MODEL24 SIT. *In vitro* method using OECD TG439; Positive (Classified): Viability ≤ 50%, Negative (Not Classified): Viability > 50%. Primary skin irritation test in rabbits; Positive (Irritant): P.I.I. (Primary Irritation Index) > 2.0, Negative (Non-Irritant): P.I.I. ≤ 2.0.

M. Sugiyama *et al.*

imately equal number of irritants and non-irritants, along with a wide range of chemicals such as surfactants, oils, polymers, acid, and powders.

As human patch tests are usually performed to confirm negative results, it is difficult to obtain data on quasi-drugs that are strongly irritating in animal studies. The lack of positive data from human patch tests is a challenge to developing an alternative method.

Correlations between *in vivo* data (rabbits and humans)

In the application for quasi-drugs, primary skin irritation is explained by data from the primary skin irritation test using rabbits and the 24-hr patch test using humans. There are no false negative chemicals in the correlation between the results of the rabbit and human studies (Table 4-1, Fig. 1) confirming the usefulness of the animal-based studies for safe conduct of human patch test. This result was consistent with previous studies in which skin reactivity was significantly higher in animals such as rabbits than in humans after 24 hr of exposure (Kästner, 1977; Motoyoshi *et al.*, 1979; Tsuchiya *et al.*, 1980). The performance of predictability, the sensitivity was 100% whereas PPV was 57.1%, suggesting that primary skin irritation test in rabbits is over predicted against the human patch test.

In addition, the skin reactivity in rabbits with 4-hr exposure was higher than skin reactivity in humans, and it has been reported that the irritation of chemicals is sometimes overestimated in animals. (Basketter *et al.*, 2004) (Jírová *et al.*, 2010). Overall, the usefulness of the animal-based studies for identifying a potential effect in human should be counterbalanced with the low PPV observed both with 4-hr and 24-hr exposure. A chemical being negative in the primary skin irritation test using rabbits has a high probably to be also negative in humans whereas the relevance of a positive effect observed in the primary skin irritation test should be further confirmed.

Differences in evaluation results by reconstructed human epidermal models

Epidermal test methods gave different results for 9 chemicals [0.1% Benzalkonium Chloride (#2), 20 and 100% Lauric Acid (#19), 10% TEA-Laureth Sulfate (#25), 100% Rosemary Oil (#26), 100% Lauryl Alcohol (#29), 100% Dicaprylyl Ether (#32), 100% Orange Oil (#37), 1% Sodium Lauryl Sulfate (#38), and 1% Lauryl Betaine (#39)]. Four of these were surfactants (#2, #25, #38, and #39), and five were lipophilic chemicals (#19, #26, #37, #29, and #32) classified as higher fatty acids, essential oils, higher alcohols, or ethers. The difference of

the evaluation results by the epidermal models used here could not be explained by the physical properties of the chemicals. In addition, in this study, the test chemicals evaluated in the 4 test methods were purchased from various suppliers and therefore were not necessarily obtained from the same chemical providers for the *in vivo* studies. Therefore, differences in reagent sources could affect the test results.

Among the reconstructed epidermal models listed in OECD TG 439, Kano *et al.* (2010) examined the skin permeability coefficients of LabCyte EPI-MODEL and EpiSkinS (EpiSkin) and reported that the permeability of chemicals differed from model to model. There were no differences in cumulative amount of chemicals up to 1 hr after application permeated through RhE due to models (EpiDerm, LabCyte EPI-MODEL, EpiSkin) (Todo, 2016). In the protocol on *in vitro* methods in TG 439, the application period of the test chemical is as short as 15-60 min. The application times were adjusted model-by-model to detect chemicals which are equal to or more irritant than those classified in UN GHS Category 2. Therefore, it is considered that the differences in chemical permeability by the epidermis model did not affect the results significantly. As a result, the predictive ability (sensitivity, specificity, accuracy, PPV, and NPV) of *in vivo* (rabbits, humans) results did not differ clearly between reconstructed human models.

Correlations between *in vitro* test methods (OECD TG 439) with *in vivo* primary skin test (rabbits) results

Imai *et al.* (in press) used the same test chemicals as in this study and reported on the results of the EpiDerm™ SIT and the concordance of the *in vivo* (rabbits) results and speculated that oil-soluble chemicals that do not penetrate the epidermis may be less toxic because of wetting of the epidermal modeled surfaces. On the other hand, surfactant that quickly penetrates the epidermis model was considered to have little influence of the wetting surface of the EpiDerm™ model. Of the 9 surfactants classified as irritants in this study, 7 were judged to be irritants; therefore, the irritancy of the surfactants was easily detectable in the test method. We selected chemicals other than oil-soluble chemicals by consideration for Log Kow above 3.5 and classifications, and examined the correlation of *in vivo-in vitro* for these chemicals (Table 4-4). The number of test chemicals for which oil-soluble chemicals were excluded was 11 for EpiDerm™ SIT and Lab-Cyte SIT and 12 for EpiSkin™ SIT and SkinEthic™ SIT. The correspondence between *in vivo* (rabbits) and *in vitro* results was 80.0-100% for sensitivity, 83.3% for specifici-

Table 5. Lowest concentration judged as Classified.

INCI	Conc.	Human 24 hr Patch Test[1]			*in vitro* evaluation by OECD TG439[2]			
		Positive rate	S.I.I.	Class	EpiSkin™ SIT	SkinEthic™RHE SIT	EpiDerm™SIT (EPI-200)	LabCyte EPI-MODEL24 SIT
Benzalkonium Chloride	0.01%				NC 105.8 ± 3.5%	NC 93.2 ± 4.3%	NC	NC 96.6 ± 8.2%
	0.1%	28.9%	43.4	I	NC 102.8 ± 3.2%	NC 77.4 ± 5.5%	C 36.6 ± 7.6%	C 43.4 ± 11.9%
	0.2%				C 42.0±2.9%	C 12.7 ± 0.4%		C 15.9 ± 1.1%
	0.5%					C 1.5 ± 0.1%		C 15.6 ± 3.6%
	1.0%				C 20.1 ± 4.1%	C 1.0 ± 0.0%		C 10.7 ± 2.7%
Sodium Lauryl Sulfate	1.0%	44.7%	85.5	I	C 25.8 ± 14.1%	C 1.1 ± 0.2%	C 2.5 ± 0.07%	NC 54.7 ± 11.2%
	2.0%				C 4.7 ± 1.1%			C 10.3 ± 2.6%
	5.0%				C 5.8 ± 1.3%	C 0.9 ± 0.1%	C 2.1 ± 0.04%	C 7.8 ± 1.4%
Lauryl Betaine	1.0%	39.5%	50.0	I	C 39.9 ± 0.6%	C 37.3 ± 3.6%	NC 72.2 ± 20.7%	NC 67.2 ± 13.7% (C 47.7 ± 1.1%)
	2.0%				C 11.8 ± 0.4%		C 7.3 ± 7.6%	C 11.1 ± 0.8%
	5.0%					C 1.0 ± 0.3%	C 2.8 ± 0.2%	C 12.1 ± 2.8%
	10.0%				C 7.7 ± 1.9%		C 2.4 ± 0.2%	C 10.2 ± 0.6%
	100.0%				C 9.1 ± 12.8%	C 1.7 ± 0.7%	C 2.6 ± 0.3%	C 25.2 ± 1.3%

1) I (Irritant): S.I.I. (Sugai's Irritation Index) ≥ 30
   NI (Non-Irritant): S.I.I. < 30
2) C (Classified): Viability ≤ 50%
   NC (Not Classified): Viability > 50 %
   INCI, International Nomenclature of Cosmetic Ingredients; OECD TG, Organisation for Economic Co-operation and Development Test Guideline

ty and 81.8-91.7% for accuracy; therefore, the increased sensitivity and accuracy were obtained. In addition, the correlation between the primary skin irritation index (P.I.I.) and cellular viability were investigated after the exclusion of oil-soluble chemicals (Fig. 4).

Although there was one chemical that was judged as false negative [TEA-Laureth Sulfate (#25)], when viability was more than 50%, primary skin irritation index was mostly equal to or less than 2.0.

Therefore, if the chemical is not oil soluble it was suggested that in the *in vitro* method using TG 439 the non-irritant chemicals could correctly predict irritancy in rabbits with 24-hr exposure.

Correlations between *in vivo* human and *in vitro* (TG 439) results

Chemicals classified as oils-soluble showed no irritation in human patch tests but were judged to be irritating with 100% Lauric acid (#19) and 10% Rosemary Oil (#26) by *in vitro* methods (false positive).

For surfactants, there were five irritants in human patch tests, of which there were false negatives in the RhE models for 0.1% Benzalkonium Chloride (#2), 1% Sodium Lauryl Sulfate (#38), and 1% Lauryl Betaine (#39). Of these chemicals, 0.1% Benzalkonium Chloride (#2) has been reported to be less irritating when applied to humans for 4-hr (Kanto *et al.*, 2013). As rabbits are more sensitive than humans, it was considered that a 4-hr patch test could be performed safely if the test was judged to be Not Classified in the test method adopted in OECD TG 439. These chemicals, #2, #38 and #39, were presumed to be chemicals with enhanced skin reactions when the exposure time was extended from 4 to 24 hr.

In conclusion, the correlation between the human patch test and the *in vitro* test methods were better than that of the primary skin irritation test with rabbits exposed for 24 hr. To avoid potential false negative outcomes in the case of surfactants, it was suggested that using the *in vitro* test methods listed in OECD TG 439 while setting the margin to 2 is an useful approach before conducting a 24-hr occlusive patch test safely in Japan for quasi-drugs. This study also revealed that the *in vitro* test using recon-

structed human epidermis model adopted in OECD TG 439 was useful for evaluating human skin irritation.

Conflict of interest---- The authors declare that there is no conflict of interest.

REFERENCES

Balls, M., Blaauboer, B., Brusick, D., Frazier, J., Lamb, D., Pemberton, M., Reinhardt, C., Roberfroid, M., Rosenkranz, H., Schmid, B., Spielmann, H., Stammati, A. and Walum, E. (1990): Report and recommendations of the CAAT/ERGATT workshop on the validation of toxicity test procedures. Altern. Lab. Anim., **18**, 313-337.

Basketter, D.A., York, M., McFadden, J.P. and Robinson, M.K. (2004): Determination of skin irritation potential in the human 4-h patch test. Contact Dermat., **51**, 1-4.

Cosmetic and quasi-drugs manufacturing and selling guidebook study group (2017): Guide to Marketing and Manufacturing of Quasi-drug and cosmetic Regulations in Japan 2017, pp.154-157, Yakuji Nippo, Ltd., Tokyo. (in Japanese)

Cruzan, G., Dalbey, W.E., D'Aleo, C.J. and Singer, E.J. (1986): A composite model for multiple assays of skin irritation. Toxicol. Ind. Health, **2**, 309-320.

Draize, J.H., Woodard, G. and Calvery, H.O. (1944): Methods for the study of irritation and toxicity of substances applied topically to the skin and mucous membranes. J. Pharmacol. Exp. Ther., **82**, 377-390.

Draize, J.H. (1959): Dermal toxicity. In: Appraisal of the safety of chemicals in foods, drugs and cosmetics, pp.46-48, The Association of Food and Drug Officials of the United States, Austin.

EpiDerm™ (2009): DB-ALM EpiDerm™ Skin Irritation Test Protocol. Available at: http://ecvam-dbalm.jrc.ec.europa.eu (accessed 11. July 2018).

EpiSkin™ (2018): DB-ALM EpiSkin™ Skin Irritation Test Protocol. Available at: http://ecvam-dbalm.jrc.ec.europa.eu (accessed 10. July 2018).

Gilman, M.R., Evans, R.A. and De Salva, S.J. (1978): The influence of concentration, exposure duration, and patch occlusivity upon rabbit primary dermal irritation indices. Drug Chem. Toxicol., **1**, 391-400.

Hamada, T., Mizutani, H., Abe, T., Ogawa, T., Nagura, T. and Kuramoto, M. (1984): Rabbit skin closed patch test. Hifu, **26**, 1084-1091.

Imai, N., Nomura, S., Goto, Y., Masunaga, T. and Nakade, M. (2018): Predicting *in vitro* skin irritation: use of the reconstructed human epidermis test method (EpiDerm™ SIT) in a 24-h exposure test for skin irritation. Altern. Animal Test. Exp., **23**, 1-8.

Japan Cosmetic Industry Association (2015): Guidance for the safety evaluation of cosmetics 2015, pp.111-112, 153-154, Yakuji Nippo, Ltd., Tokyo.

Jírová, D., Basketter, D., Liebsch, M., Bendová, H., Kejlová, K., Marriott, M. and Kandárová, H. (2010): Comparison of human skin irritation patch test data with *in vitro* skin irritation assays and animal data. Contact Dermat., **62**, 109-116.

Kano, S., Todo, H., Sugie, K., Fujimoto, H., Nakada, K., Tokudome, Y., Hashimoto, F. and Sugibayashi, K. (2010): Utilization of Reconstructed Cultured Human Skin Models as an Alternative Skin for Permeation Studies of Chemical Compounds. Altern. Animal Test. Exp., **15**, 61-70.

Kanto, H., Washizaki, K., Ito, M., Matsunaga, K., Akamatsu, H.,

Kawai, K., Katoh, N., Natsuaki, M., Yoshimura, I., Kojima, H., Okamoto, Y., Okuda, M., Kuwahara, H., Sugiyama, M., Kinoshita, S. and Mori, F. (2013): Optimal patch application time in the evaluation of skin irritation. J. Dermatol., **40**, 363-369.

Kästner, W. (1977): Zur Speziesabhängigkeit der Hautverträglichkeit von Kosmetikgrundstoffen. J. Soc. Cosmet. Chem., **28**, 741-754.

Kawamura, T., Sasagawa, S., Masuda, T., Honda, S., Kinoshita, M., Harada, S., Ishizaki, T., Nagai, R., Hirokawa, K., Anzai, T., Anekoji, K., Hidano, A., Kawano, T., Ikegami, I., Sato, S. and Aoyama, T. (1970): Basic Studies on the Standardization of Patch Test. Japanese Journal of Dermatology, **80**, 301-314.

LabCyte. (2011): EPI-MODEL24 SIT SOP, Version 8.3, Skin Irritation Test Using the Reconstructed Human Model "LabCyte EPI-MODEL24". Available at: http://jacvam.jp/ (accessed 8. July 2018).

MHLW Grants System (2011): Final document of committee for study about how to prepare document for assessment of safety in applications for quasi-drugs approval. -Report of skin irritation-FY 2009 Scientific Research of Ministry of Health, Labor and Welfare, Establishment of a safety evaluation system using an alternative method of animal experiments and research on international cooperation, https://mhlw-grants.niph.go.jp/niph/search/NIDD00.do?resrchNum=200940003A (accessed 27. May 2018).

Motoyoshi, K., Toyoshima, Y., Sato, M. and Yoshimura, M. (1979): Comparative studies on the irritancy of oils and the synthetic perfumes to the skin of rabbit, rat, guinea pig, miniature swine and man. Cosmet. Toiletries, **94**, 41-48.

OECD (2015a): Acute Dermal Irritation/Corrosion. Test Guideline 404, Organisation for Economic Co-operation and Development, adopted 28 July 2015, <https://www.oecd-ilibrary.org/environment/test-no-404-acute-dermal-irritation-corrosion_9789264242678-en> (accessed 14. July 2018).

OECD (2015b): In Vitro Skin Irritation: Reconstructed Human Epidermis Test Method. Test Guideline 439, Organisation for Economic Co-operation and Development, adopted 28 July 2015, <https://www.oecd-ilibrary.org/environment/test-no-439-in-vitro-skin-irritation_9789264090958-en> (accessed 14. July 2018).

SkinEthic™ (2018): DB-ALM SkinEthic™ RHE Skin Irritation Test Protocol no. 135. <http://ecvam-dbalm.jrc.ec.europa.eu> (accessed 8. July 2018).

Spielmann, H., Hoffmann, S., Liebsch, M., Botham, P., Fentem, J.H., Eskes, C., Roguet, R., Cotovio, 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. Altern. Lab. Anim., **35**, 559-601.

Sugai, T. (1995): Safety Evaluation of Cosmetic Products. J. Jap. Cosmet. Sci. Soc., **19** (extraordinary supplement), 49-56. (in Japanese)

Todo, H. (2016): Difference in chemical permeation properties through cultured human skin models. Bioindustry, **33**, 3-9. (in Japanese)

Tsuchiya, S., Kondo, M., Okamoto, K. and Takase, Y. (1980): Skin irritancy of cosmetic products and their materials. Hifu, **22**, 373-377.

Young, J.R., How, M.J., Walker, A.P. and Worth, W.M. (1988): Classification as corrosive or irritant to skin of preparations containing acidic or alkaline substances, without testing on animals. Toxicol. In Vitro, **2**, 19-26.