



Evaluation of radioisotopic and non-radioisotopic versions of local lymph node assays for subcategorization of skin sensitizers compliant to UN GHS rev 4

Soojin Ha ^{a,1}, Il Young Ahn ^{b,1}, Da-eun Kim ^a, Jong Kwon Lee ^b, Soojung Sohn ^b,
Mi-Sook Jung ^c, Yong Heo ^d, Takashi Omori ^e, Seungjin Bae ^{a,**}, Kyung-Min Lim ^{a,*}

^a College of Pharmacy, Ewha Womans University, Seoul, Republic of Korea

^b National Institute of Food and Drug Safety Evaluation, Ministry of Food and Drug Safety, Osong, Republic of Korea

^c Pharmacology Efficacy Team, Biototech Co., Ltd, O-Chang, Republic of Korea

^d Department of Occupational Health, College of Natural Sciences, Catholic University of Daegu, Republic of Korea

^e Division of Biostatistics, Department of Social/Community Medicine and Health Science, Kobe University School of Medicine, Kobe, Japan

ARTICLE INFO

Article history:

Received 20 October 2016

Received in revised form

4 February 2017

Accepted 5 February 2017

Available online 10 February 2017

Keywords:

Local lymph node assay

Non-radioisotopic LLNAs

UN Global Harmonization System

GHS

Skin sensitization

Alternative to animal test

ABSTRACT

Recently UN GHS has introduced the sub-categorization of skin sensitizers for which ECT (concentration estimated to induce stimulation index above threshold) of the murine local lymph node assay (LLNA) is used as criteria. Non-radioisotopic variants of LLNA, LLNA: DA, LLNA: BrdU-ELISA, LNCC and LLNA: BrdU-FCM were developed yet their utilities for potency sub-categorization are not established. Here we assessed the agreement of LLNA variants with LLNA or human data in potency sub-categorization for 22 reference substances of OECD TG429. Concordance of sub-categorization with LLNA was highest for LLNA: BrdU-FCM(91%, $\kappa = 0.833$, weighted kappa) followed by LLNA: BrdU-ELISA (82%, $\kappa = 0.744$) and LLNA: DA (73%, $\kappa = 0.656$) whereas LNCC only showed a modest association (64%, $\kappa = 0.441$). With human data, LLNA agreed best (77%) followed by LLNA: DA and LLNA: BrdU-FCM(73%), LLNA: BrdU-ELISA (68%) and LNCC(55%). Bland-Altman plot revealed that ECT's of LLNA variants largely agreed with LLNA where most values fell within 95% limit of agreement. Correlation between ECT's of LLNA and LLNA variants were high except for LNCC(pair-wise with LLNA, LLNA: DA, $r = 0.848$, LLNA: BrdU-ELISA, $r = 0.744$, LLNA: BrdU-FCM, $r = 0.786$, and LNCC, $r = 0.561$ by Pearson). Collectively, these results demonstrated that LLNA variants exhibit performance comparable to LLNA in the potency sub-categorization although additional substances shall be analyzed in the future.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

Local lymph node assay (LLNA), OECD TG 429 (OECD, 2010a), and its non-radioisotopic versions, LLNA:DA (TG 442A) (OECD, 2010a), LLNA: BrdU-ELISA (TG 442B) (OECD, 2010b), LNCC (Basketter et al., 2012) and LLNA: BrdU-FCM (Jung et al., 2010, 2012; Kim et al., 2016; Yang et al., 2015), have been developed as a stand-alone method to replace *in vivo* skin sensitization tests employing guinea pigs, OECD TG 406 (OECD, 1992; Ryu et al., 2016), with an

aim to reflect '3R' principles and to improve animal welfare. Indeed, the employment of LLNAs has considerably reduced the sacrifice of animals (from 45 to less than 25), improved animal welfare by avoiding the use of painful adjuvant, and substantially reduced time and cost compared with conventional guinea pig tests. However, LLNAs are still "*in vivo*" and with the recent advent of true *in vitro* and *in chemico* assays that include KertinoSens™ (OECD, 2015b), hCLAT and Direct Peptide Reactivity Assay (DPRA) (OECD, 2015a), there is a strong voice to phase out LLNAs.

However, *in vitro* and *in chemico* alternatives described above have not fully overcome their intrinsic limitations yet, namely, their applicability has not exceeded the realm of hazard identification. Ultimately, skin sensitization tests must be able to provide potency information that can contribute to the exposure-based quantitative risk assessment/management and subsequent prioritization or

* Corresponding author. College of Pharmacy, Ewha Womans University, 52 Ewhayeodae-gil, Seodaemun-gu, Seoul 03760, Republic of Korea.

** Corresponding author.

E-mail addresses: sjbae@ewha.ac.kr (S. Bae), kmlim@ewha.ac.kr (K.-M. Lim).

¹ These authors contributed equally to this work.

classification for the regulation of chemicals. LLNAs can give information concerning the potency of skin sensitizers as ECt values (mathematically estimated concentration of chemical required to induce a threshold stimulation index (SI)). In this regard, LLNAs would be “currently in use” for substantial length of time as a gold standard for the assessment of sensitization.

United Nation Globally Harmonized System of Classification and Labelling of Chemicals (UN GHS) employs a classification system to designate categories to substances by types or level of hazard (United Nations, 2003). UN GHS has been revised to introduce subcategory for the skin sensitizers (United Nations, 2009). New system recommends the sub-categorization of skin sensitizers into Cat 1A and 1B depending on their potency. Criteria for the sub-categorization into Cat 1A is established as the substance with the EC3 of LLNA $\leq 2\%$, or $\geq 30\%$ responding at $\leq 0.1\%$ intradermal induction dose in guinea pig maximization test (or positive at $\leq 500 \mu\text{g}/\text{cm}^2$ in Human Repeated Insult Patch Test (HRIPT)). Indeed, in the risk assessment of cosmetics, “No Expected Sensitization Induction Level (NESIL)” concept has been recently taken up, which is based on ECt produced by LLNA (Api et al., 2008), supporting the utility of LLNA for quantitative risk assessment.

Yet, it is to be demonstrated whether the non-radioisotopic variants of LLNA, (i.e., LLNA: DA, LLNA: BrdU-ELISA, LNCC and LLNA: BrdU-FCM) are compatible to the original LLNA for quantitative risk assessment. Since LLNA has a critical drawback, namely, *in vivo* use of radioisotope, ^3H -labeled thymidine during the experimental procedure, which seriously deters its use in some countries, proof of the utility of non-radioisotopic variants in the potency subcategorization may be helpful in replacing further conventional guinea pig tests. The main purpose of this study was to examine agreement between LLNA and its non-radioisotopic variants, LLNA: DA, LLNA: BrdU-ELISA, LNCC, and LLNA: BrdU-FCM, for classification of 22 reference substances enlisted in the performance standards of OECD TG429. As the secondary purpose, we compared the ECt values of LLNA and its variants.

2. Materials & methods

2.1. ECthreshold data for radioisotopic and non-radioisotopic LLNAs

Currently, one radioisotopic LLNA and two non-radioisotopic LLNAs, (LLNA: BrdU-ELISA and LLNA: DA) have been officially approved by OECD as test guidelines for the test of skin sensitization. In case of radioisotopic LLNA (TG 429), stimulation index (SI), which represents the measured extent of proliferation of lymph node cells (LNCs), and ECt, a hypothetical intra-polated concentration of a test chemical that induces SI beyond the threshold, or cutoff to determine as a sensitizer were published for 22 reference substances in OECD TG 429 performance standards (OECD, 2010a). These data are the basis for the sub-categorization of sensitizers into Category 1A and 1B compliant to UN GHS 4th revision. For LLNA:DA and LLNA: BrdU-ELISA, different threshold values (1.8 for DA and 1.6 for BrdU-ELISA) are adopted and EC_t for reference substances was published in the ICCVAM 2010 review report (ICCVAM, 2010a; ICCVAM, 2010b). Data for some substances unavailable in the report, were provided by JaCVAM (xylene for LLNA: DA and methyl methacrylate, chlorobenzene, xylene and nickel chloride for LLNA: BrdU-ELISA). In addition, data for LLNA: BrdU-FCM and LNCC were published with which ECt values were derived (Ahn et al., 2016; Basketter et al., 2012; Kim et al., 2016) and for LLNA: BrdU-FCM, optional 4 substances (Sodium lauryl sulfate (Sigma-Aldrich, St. Louis, MO), ethylene glycol dimethacrylate, xylene, Nickel chloride) were tested again to conform to the current protocol version 1.3. LLNA: BrdU-FCM test was done in Biototech (Ochang, Korea) in a coded fashion and the experimental procedure

was well-described in the recent paper (Ahn et al., 2016). When multiple values were known, the lowest values among them were used to sub-categorize the potency.

2.2. United nation global harmonization system (UN GHS) classification

Based on the collected ECt values, the potency of sensitizers were subcategorized into 3 levels, namely, No category, Category 1B and 1A based on EC_t. UN GHS classification system was revised to incorporate the potency concept into labeling recently (rev 4.). And this is based on EC3 value of radioisotopic LLNA. When the EC3 of the substance is measured to be $\leq 2\%$ in LLNA, it is categorized to Category 1A, if EC3 $> 2\%$ then, Category 1B and when EC3 could not be determined (when maximum SI values did not exceed threshold SI value for respective LLNA variant), then No category. In a same vein, the categorization with other LLNAs were done according to respective ECt value.

2.3. Statistical analysis

2.3.1. UN GHS classification

The proportion of concordance of LLNA variants with LLNA or human data for binary decision (sensitizer vs non-sensitizer) and for the UN GHS subcategorization was assessed. The concordance in the UN GHS subcategorization (Cat 1A, Cat 1B, or No Category) was further assessed with weighted kappa since the UN GHS is an ordinal variable. Concordance is generally interpreted as “almost perfect agreement” when Kappa value is 0.81–0.99, “substantial agreement” at 0.61–0.8, “moderate agreement” at 0.41–0.6, “fair agreement” at 0.21–0.40 and “slight agreement” at 0.01–0.20 (Hunt, 1986; Landis and Koch, 1977).

2.3.2. ECt values

ECt value is a continuous data ranging between 0 and 100%. First, Bland-Altman analysis was employed to describe the agreement of LLNA variants with LLNA through drawing plot and 95% limit of agreement lines (Altman and Bland, 1983; Bland and Altman, 1986). In addition, Pearson correlation analysis (Jang et al., 2015) was done to quantify the pairwise correlation of ECt values between two LLNAs. Correlation coefficient is between -1 and 1 , and the closer to 1 the absolute value is, the stronger the correlation becomes. Generally, strong correlation is when $0.7 < r < 1$ and moderate, $0.3 < r < 0.7$ and weak when $r < 0.3$. For analysis, when ECt value could not be determined (non-sensitizer), then ECt value was assumed to be 100. Statistical analysis was done with SAS version 9.3 (SAS Institute Inc, Cary, NC, USA).

3. Result

ECt values for 22 reference substances with LLNA and its non-radioisotopic variants were presented in Table 1. These substances cover the various chemical classes, physicochemical properties and wide range of potencies, and were enlisted in OECD TG 429 performance standards for the evaluation of newly developed variants of LLNA. With the obtained ECt, substances were classified and sub-categorized according to UN GHS rev 4.0 (i.e., Cat 1A, 1B or No category when ECt is $\leq 2\%$, $2\% < \text{ECt} < 100\%$ or $= 100\%$, respectively) as shown in Table 2.

When compared with the results of traditional LLNA as a gold standard, LLNA: DA produced 2 false positives (chlorobenzene, salicylic acid) and 1 false negative (xylene) while LLNA: BrdU-ELISA had 2 false positives (chlorobenzene, lactic acid). LLNA: BrdU-FCM also produced 2 false negatives (2-mercaptobenzothiazole, methyl methacrylate) while LNCC determined 4 among 6 non-sensitizers

as positives (namely, false positives for chlorobenzene, methyl salicylate, salicylic acid and nickel chloride) and 2 false negatives (2-mercaptobenzothiazole, methyl methacrylate). Overall concordance for binary decision (sensitizer vs non-sensitizer) for 22 reference substances versus LLNA was 91% for LLNA: BrdU-ELISA and LLNA: BrdU-FCM, 86% for LLNA: DA and 73% for LNCC.

UN GHS classification by LLNA and its variants were largely concordant where 4 of 6 Cat1A substances were all categorized into Cat 1A. Isoeugenol, a Cat1A sensitizer in LLNA was correctly sub-categorized by LLNA: DA and LLNA: BrdU-FCM, but other variants determined it as Cat1B. Decision for 2-mercaptobenzothiazole, a Cat1A sensitizer in LLNA was substantially diverged; LLNA: DA and LLNA: BrdU-ELISA determined it as Cat1B while LLNA: BrdU-FCM and LNCC as NC. For 6 of 10 Cat1B sensitizers, decision was identical in all 5 variants. However, decision was diverged again among

LLNA variants for phenyl benzoate, methyl methacrylate and sodium lauryl sulfate. While LLNA: BrdU-ELISA resulted in Cat1B completely conforming to LLNA, LLNA: DA over-predicted phenyl benzoate and sodium lauryl sulfate as Cat1A, and LNCC over-predicted sodium lauryl sulfate as Cat1A. Methyl methacrylate was a false negative for LNCC and LLNA: BrdU-FCM while LLNA: DA and LLNA: BrdU-ELISA correctly determined it as Cat1B. The decision for NC was strikingly different among LLNA variants, for which LLNA: BrdU-FCM produced completely concordant results with LLNA while LNCC falsely determined 4 of 6 NCs (chlorobenzene, methyl salicylate, salicylic acid, and nickel chloride) as Cat1B. LLNA: DA incorrectly determined chlorobenzene and salicylic acid as Cat1B while LLNA: BrdU-ELISA produced false positive results for chlorobenzene and lactic acid.

Comparison with human data with LLNA or LLNA variants

Table 1

ECt values for 22 reference substances with traditional LLNA and its non-radioisotopic variants.

No.	Substance	Abb.	CAS No.	LLNA ^a	LLNA:DA	LLNA: BrdU- ELISA	LNCC ^d	LLNA: BrdU- FCM	HMT/HRIPT ^e (human result)
				EC3	EC1.8	EC1.6	EC1.5	EC2.7	
1	5-Chloro-2-methyl-4-isothiazolin-3-one (CMI)/2-methyl-4-isothiazolin-3-one (MI)	CMI	26172-55-4/ 2682-20-4	0.009	0.009 ^b	0.065 ^b	0.011	*1.062	DSA ₀₅ HRIPT = 4.3 (Cardin et al., 1986)
2	DNCB	DNCB	97-00-7	0.049	0.032 ^b	0.032 ^b	<0.025	*0.016	DSA ₀₅ HRIPT = 8.8 (Basketter et al., 2014)
3	4-Phenylenediamine	PPD	106-50-3	0.11	0.04 ^b	0.29 ^b	0.1	0.101	DSA ₀₅ HRIPT = 10 (Basketter et al., 2014)
4	Cobalt chloride	CC	7646-79-9	0.6	0.9 ^b	0.3 ^b	<0.25	0.199	DSA ₀₅ HRIPT = 172–453 (Basketter et al., 2014)
5	Isoeugenol	IE	97-54-1	1.5	1.5 ^b	5.2 ^b	2.7	*1.2	DSA (LOEL) HRIPT = 69 (Griem et al., 2003)
6	2-Mercaptobenzothiazole	MBT	149-30-4	1.7	8.0 ^b	12.1 ^b	NA (7.5)	NA (25)	DSA (LOEL)HMT = 5217 DSA ₀₅ HMT = 2269 (Kligman, 1966b)
7	Citral	Citral	5392-40-5	9.2	2.1 ^b	7.1 ^b	9.2	13.1	DSA ₀₅ HRIPT = 775
8	HCA	HCA	101-86-0	9.7	6.3 ^b	12.9 ^b	16.2	15.1	DSA (NOEL)HRIPT = 23622 (ICCVAM, 2011)
9	Eugenol	EUG	97-53-0	10.1	2.6 ^b	8.9 ^b	8.7	16.5	DSA ₀₅ HRIPT = 1938 (Basketter et al., 2014)
10	Phenyl benzoate	PB	93-99-2	13.6	0.7 ^b	17.0 ^b	4	*5.5	DSA ₀₅ HRIPT = 9448 (Basketter et al., 2014)
11	Cinnamic alcohol	CA	104-54-1	21	5.2 ^b	24.1 ^b	26	44.3	DSA ₀₅ HRIPT = 3000 (Basketter et al., 2014)
12	Imidazolidinyl urea	IU	39236-46-9	24	6.3 ^b	49.5 ^b	21.5	32.0	DSA ₀₅ HRIPT = 2000 (Basketter et al., 2014)
13	Methyl methacrylate	MMA	80-62-6	90	99 ^b	79.1 ^c	NA (100)	NA (100)	<u>Insufficient data but+human clinical data</u>
14	Chlorobenzene	CB	108-90-7	NA (25)	17.9 ^b	21.4 ^c	79	NA (50)	<u>Insufficient data</u>
15	Isopropanol	IS	67-63-0	NA (50)	NA (50) ^b	NA (100) ^b	NA (100)	NA (100)	DSA (LOEL) HMT = 6897 (Basketter et al., 1998)
16	Lactic acid	LA	50-21-5	NA (25)	NA (50) ^b	15.2 ^b	NA (25)	NA (50)	<u>Insufficient data</u>
17	Methyl salicylate	MS	119-36-8	NA (20)	NA (25) ^b	NA (50) ^b	49	NA (50)	DSA (LOEL) HMT = 5517 (Opdyke, 1979)
18	Salicylic acid	SA	69-72-7	NA (25)	17.7 ^b	NA (25) ^b	15.8	NA (5)	DSA (LOEL) HMT = 13793 (Kligman, 1966b)
19	Sodium lauryl sulfate	SLS	151-21-3	8.1	1.6 ^b	13.3 ^b	1.6	5.2	DSA (LOEL) HMT = 6897 (Kligman, 1966b)
20	Ethylene glycol dimethacrylate	EGDM	97-90-5	28	19.2 ^b	31.8 ^b	38	97.20	<u>+ human from</u> (Basketter et al., 1999)
21	Xylene	Xylene	1330-20-7	95.8	NA (100) ^c	15.9 ^c	28.2	*34.14	DSA (LOEL) HMT = 68966 (Kligman, 1966b)
22	Nickel chloride	NC	7718-54-9	NA (5)	NA (10) ^b	NA (5) ^b	3.6	NA (2.5)	DSA ₀₅ HMT = 28 (Kligman, 1966a) (but data expressed as nickel salt)

Data source, a: OECD TG 429 b: ICCVAM, 2010a, c: JaCVAM, d: Basketter et al., 2012, e: ICCVAM LLNA performance standards, 2009/ICCVAM test method evaluation report, 2011

* indicates the case when EC2.7 was obtained assuming y-intercept is 1.0. NA: not assessed due to weak sensitization potency and ECt was assumed to be 100% for statistical analysis. Value in the bracket is the maximum concentration tested.

DSA₀₅: induction dose per skin area, in µg/cm², in a human repeat-insult patch test or human maximization test that produces a positive response in 5% of the tested population.

Table 2
Potency sub-categorization to UN GHS-compliant based on EC₀₁ values.

No.	Substance	Human (DSA ₀₅)	LLNA	LLNA: DA	LLNA: BrdU-ELISA	LNCC	LLNA: BrdU-FCM
1	5-Chloro-2-methyl-4-isothiazolin-3-one(CMI)/2-methyl-4-isothiazolin-3-one (MI)	Cat1A	Cat1A	Cat1A	Cat1A	Cat1A	Cat1A
2	DNCB	Cat1A	Cat1A	Cat1A	Cat1A	Cat1A	Cat1A
3	4-Phenylenediamine	Cat1A	Cat1A	Cat1A	Cat1A	Cat1A	Cat1A
4	Cobalt chloride	Cat1A	Cat1A	Cat1A	Cat1A	Cat1A	Cat1A
5	Isoeugenol	Cat1A	Cat1A	Cat1A	Cat1B	Cat1B	Cat1A
6	2-Mercaptobenzothiazole	Cat1B	Cat1A	Cat1B	Cat1B	NC	NC
7	Citral	Cat1B	Cat1B	Cat1B	Cat1B	Cat1B	Cat1B
8	HCA	Cat1B	Cat1B	Cat1B	Cat1B	Cat1B	Cat1B
9	Eugenol	Cat1B	Cat1B	Cat1B	Cat1B	Cat1B	Cat1B
10	Phenyl benzoate	Cat1B	Cat1B	Cat1A	Cat1B	Cat1B	Cat1B
11	Cinnamic alcohol	Cat1B	Cat1B	Cat1B	Cat1B	Cat1B	Cat1B
12	Imidazolidinyl urea	Cat1B	Cat1B	Cat1B	Cat1B	Cat1B	Cat1B
13	Methyl methacrylate ^a	Cat1B	Cat1B	Cat1B	Cat1B	NC	NC
14	Chlorobenzene ^a	NC	NC	Cat1B	Cat1B	Cat1B	NC
15	Isopropanol	Cat1B	NC	NC	NC	NC	NC
16	Lactic acid ^a	NC	NC	NC	Cat1B	NC	NC
17	Methyl salicylate	NC	NC	NC	NC	Cat1B	NC
18	Salicylic acid	NC	NC	Cat1B	NC	Cat1B	NC
19	Sodium lauryl sulfate	NC	Cat1B	Cat1A	Cat1B	Cat1A	Cat1B
20	Ethylene glycol dimethacrylate ^a	Cat1B	Cat1B	Cat1B	Cat1B	Cat1B	Cat1B
21	Xylene	NC	Cat1B	NC	Cat1B	Cat1B	Cat1B
22	Nickel chloride	Cat1A	NC	NC	NC	Cat1B	NC
		Yes or No		86%(19/22)	91%(20/22)	73%(16/22)	91%(20/22)
Concordance vs LLNA		Cat1A		83%(5/6)	67%(4/6)	67%(4/6)	83%(5/6)
		Cat1B		70%(7/10)	100%(10/10)	80%(8/10)	90%(9/10)
		NC		67%(4/6)	67%(4/6)	33%(2/6)	100%(6/6)
		Total		73% (16/22)	82%(18/22)	64%(14/22)	91%(20/22)
		Yes or No	82%(18/22)	77%(17/22)	73%(16/22)	64%(14/22)	73%(16/22)
Concordance vs Human		Cat1A	83%(5/6)	83%(5/6)	67%(4/6)	67%(4/6)	83%(5/6)
		Cat1B	80%(8/10)	80%(8/10)	90%(9/10)	70%(7/10)	70%(7/10)
		NC	67%(4/6)	50%(3/6)	33%(2/6)	17%(1/6)	67%(4/6)
		Total	77%(17/22)	73%(16/22)	68%(15/22)	55%(12/22)	73%(16/22)

NC: No Category.

^aNo HRIPT data but based on clinical case report, Shaded: discordance between LLNA and LLNA variants. Red text: discordance between Human Result and LLNA variants.

suggests that the predictive capacity of LLNA and LLNA variants is around 64–82% for yes or no decision (namely sensitizer or non-sensitizer), and 55–77% for UN GHS classification, where LLNA was the best predictive of human responses followed by LLNA: DA, LLNA: BrdU-FCM, LLNA: BrdU-ELISA and LNCC (Table 2). However, since the high variability and non-standardized protocol of human patch tests (HRIPT), it shall be taken with caution.

Agreement in UN GHS classification was further analyzed with weighted kappa statistics (Table 3). LLNA: BrdU-FCM showed most concordant results with LLNA for UN GHS sub-categorization ($\kappa = 0.833$). LLNA: DA and LLNA: BrdU-ELISA also exhibited substantial agreements ($\kappa = 0.656$ and 0.744 , respectively). LNCC showed only moderate agreement with LLNA ($\kappa = 0.441$). The analysis for the correlations of LLNA variants vs. Human in the sub-categorization results suggest that LLNA matches best with human data ($\kappa = 0.588$), followed by LLNA: BrdU-FCM, LLNA: BrdU-ELISA, LLNA: DA and LNCC (Table 3).

Bland-Altman analysis to compare the correlations of LLNA

variants with LLNA as shown in Fig. 1 suggests that all LLNA variants exhibited similar mean difference with respect to LLNA, ranging from −6.9 (LLNA: BrdU-FCM) to +10.2 (LLNA: DA), with LLNA: BrdU-FCM showed the lowest absolute difference, thus smallest

Table 3
Level of agreement among test methods, quantified by weighted kappa.

	Kappa value, κ (95% C-I)
vs LLNA	
LLNA: DA	0.656 (0.409–0.904)
LLNA: BrdU-ELISA	0.744 (0.514–0.974)
LNCC	0.441 (0.119–0.763)
LLNA: BrdU-FCM	0.833 (0.604–1.000)
vs Human	
LLNA	0.588 (0.288–0.889)
LLNA: DA	0.473 (0.141–0.806)
LLNA: BrdU-ELISA	0.533 (0.203–0.864)
LNCC	0.357 (0.025–0.689)
LLNA: BrdU-FCM	0.539 (0.233–0.846)

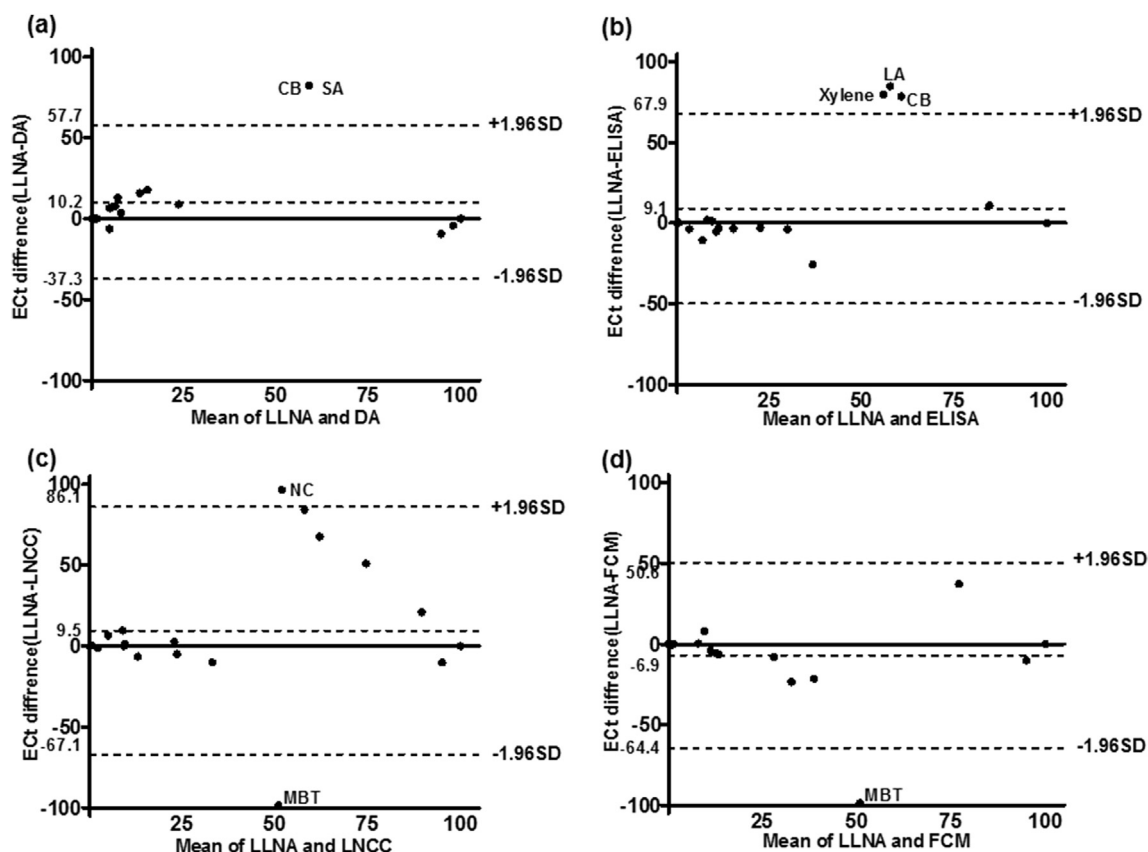


Fig. 1. Bland-Altman plots of Ect values of LLNA variants with LLNA. Differences in Ect values of LLNA variants from LLNA (tLLNA) are plotted with the lines representing confidence interval for agreement (95% CI). X-axis indicates mean of Ect of two tests for a substance while Y-axis, the difference of Ect. Substances fell outside the CI was indicated with abbreviation described in Table 1. (a) LLNA: DA vs LLNA, (b) LLNA: BrdU-ELISA vs LLNA, (c) LNCC vs LLNA and (d) LLNAL BrdU-FCM vs LLNA.

systematic bias compared with LLNA. Regarding the limits of agreement, 1 to 3 substances fell outside the range (chlorobenzene and salicylic acid for LLNA: DA; xylene, lactic acid, and chlorobenzene for LLNA: BrdU-ELISA; nickel chloride and 2-mercaptobenzothiazole for LNCC, and 2-mercaptobenzothiazole for LLNA: BrdU-FCM), with LLNA: BrdU-FCM showing the fewest outliers. Fig. 2 illustrates the correlation of Ect values of LLNA variants with LLNA as scatter plots. Most of LLNA variants showed strongly correlated results as shown by the results of Pearson correlation analysis (Table 4) except for LNCC in which relatively weak linearity was shown (Fig. 2C). There were 2–3 substances that show uncorrelated results in LLNA: DA, LLNA: BrdU-ELISA and LLNA: BrdU-FCM.

4. Discussion

Here we examined the utility of non-radioisotopic variants of LLNA for the potency classification of skin sensitizers in comparison with LLNA and human data employing 22 reference substances in OECD TG 429 performance standards. Overall, LLNA variants exhibited the performance comparable to LLNA in the potency sub-categorization. More than 67% (4/6) Cat1A sensitizers as defined by LLNA were correctly classified by all LLNA variants while remaining 33% (2/6) was under-predicted as Cat1B or NC. More than 60% (6/10) of Cat1B sensitizers were correctly classified by LLNA variants. Interestingly, LLNA: BrdU-ELISA classified Cat1B sensitizers with a perfect concordance with LLNA.

LLNA was originally designed to give 'yes' or 'no' decision to identify sensitizing potential of a chemical. However, with a recent

revision of UN GHS classification for skin sensitizers, Ect values from LLNA are being employed as a criteria to differentiate the potency of Cat 1 substances along with guinea pig test methods and HRIPT. Since most of guinea pig tests have been replaced by LLNA for 3R, and HRIPT is being criticized for ethical reasons due to potential risk of causing allergic contact dermatitis in the human subjects, LLNA is virtually a single test available for this purpose now. ICCVAM and many papers have reviewed the utility of LLNA for the purpose of potency classification recently (ICCVAM, 2011; Loveless et al., 2010) and came to a conclusion that LLNA can be employed for this purpose although some limitation exists. Indeed, LLNA has a strong level of agreement with HRIPT results (LLNA and HRIPT data (number of substances = 23; $r = 0.77$)) as shown previously (Schneider and Akkan, 2004). However, according to ICCVAM review, only 52% (14/27) of Cat 1A sensitizers in human were correctly identified in LLNA and 48% (13/27) of Cat 1A was under-classified. In addition, as much as 71% (35/49) of Cat 1B sensitizers and 42% (25/60) of NC were correctly identified by LLNA. This finding is well reproduced in the substances employed in our study. 1 of 6 (16%) Cat 1A in LLNA was actually known to be Cat 1B in human and 2 of 10 Cat 1B (20%), NC and 2 of 6 (40%) NC were classified substances, suggesting that applicability of LLNA for the sub-categorization relevant to human responses shall be done cautiously.

Of 4 LLNA variants, LLNA: BrdU-FCM produced the sub-categorization best concordant to LLNA (91%, 20/22) followed by LLNA: BrdU-ELISA (82%, 18/22). It is noteworthy that BrdU methods showed the highest agreement with LLNA. This may be from the same experimental principle employed in these variants for the

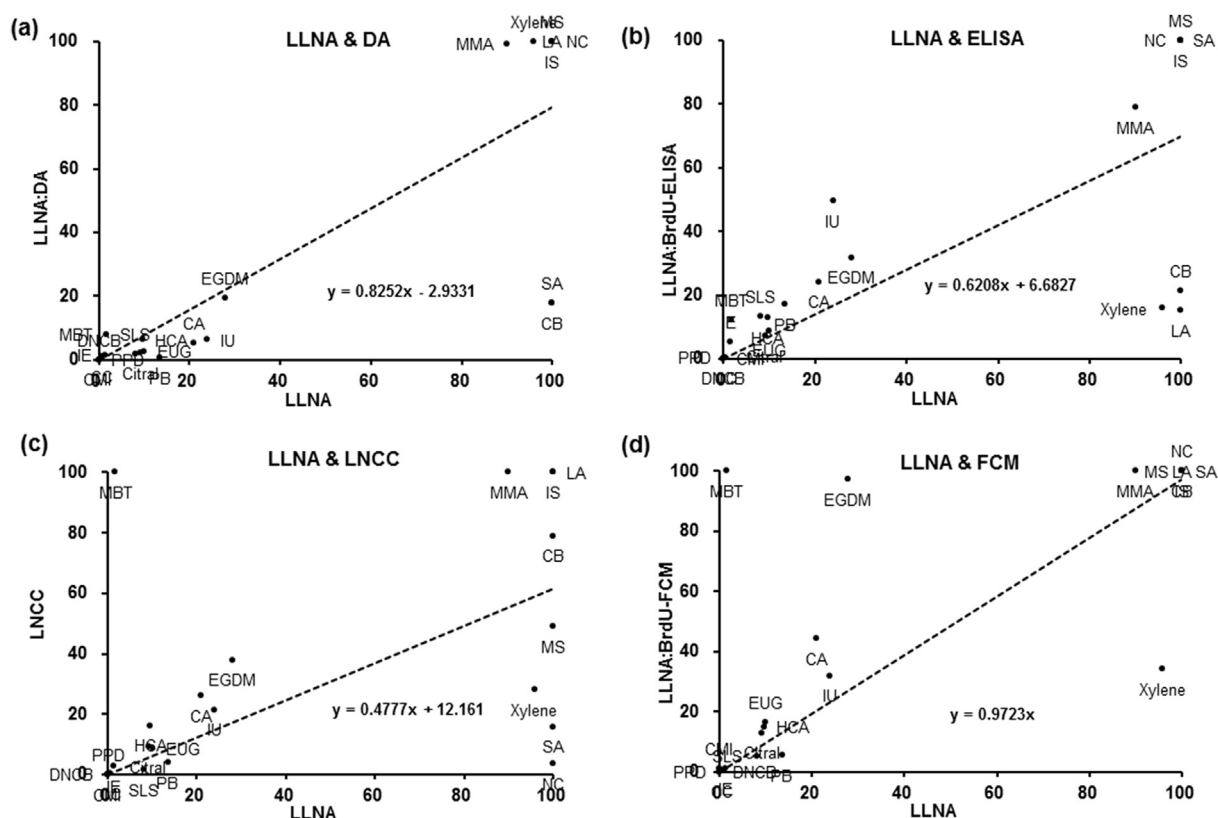


Fig. 2. Correlation of Ect values of LLNA variants with LLNA. Ect values of LLNA variants were drawn against those of LLNA to evaluate the correlation. (a) LLNA: DA vs LLNA, (b) LLNA: BrdU-ELISA vs LLNA, (c) LNCC vs LLNA and (d) LLNA: BrdU-FCM vs LLNA. Equations were produced from the trend line obtained by the least square method.

Table 4

Pearson correlation analysis of Ect values of LLNA variants with LLNA.

	LLNA	LLNA:DA	LLNA: BrdU-ELISA	LNCC	LLNA: BrdU-FCM
LLNA		0.848	0.744	0.561	0.786
LLNA:DA	-		0.652	0.568	0.653
LLNA: BrdU-ELISA	-	-		0.342	0.712
LNCC	-	-	-		0.771
LLNA: BrdU-FCM	-	-	-	-	

measurement of lymph node cell proliferation, i.e., using a nucleoside analog to quantitate DNA synthesis. In contrast, LNCC showed discordance with LLNA, only accomplishing 64% accuracy. This may be from the distinct technology used in LNCC, namely CASY cell counting technology which capitalizes the specific electrical resistant properties of viable cells (Haslam et al., 2000). Although the technology may be specific and selective to viable cells, the signal to noise ratio appears much smaller than the uptake of ^3H -thymidine as can be indicated by lower SI cutoff value (1.5 vs 3.0 in LLNA). Interestingly, LLNA: DA showed a tendency to over-classify the sensitizing potency compared to LLNA. 4 of 6 wrong predictions were the over-classification, which is in a clear contrast with LLNA: BrdU-FCM, where all 2 incorrect sub-categorization were under-classifications.

Comparison with human data revealed that LLNA, LLNA: DA and

LLNA: BrdU-FCM showed nearly equivalent levels of predictive capacity in potency sub-categorization. Interestingly, LLNA and LLNA: DA exhibited tendencies to over-predict while LLNA: BrdU-FCM, to under-predict. 3 of 5, and 4 of 6 wrong predictions were over-prediction in LLNA and LLNA: DA, respectively while only 2 of 6 were over-predicted in LLNA: BrdU-ELISA. These results suggest that LLNA variants shall be utilized with caution for the prediction of the potency of skin sensitizers, taking into consideration for these differences.

Safety evaluation methods shall ultimately predict the response of human to test chemicals. Of the substances misclassified in LLNA variants when compared with human data, nickel chloride was a human Cat1A skin sensitizer. It has been demonstrated that murine is not responsive to the skin sensitization potential of nickel chloride due to species difference in toll-like receptor 4 from human

(Schmidt et al., 2010), providing a reasonable excuse for the discordancy. Interestingly, LNCC predicted nickel chloride as Cat1B, suggesting that reliability of LNCC may be questionable. Two wrong predictions made for Cat1B were 2-mercaptobenzothiazole and isopropanol. Of a particular note, prediction made for 2-mercaptobenzothiazole was significantly diverged from 'NC' in LNCC and LLNA: BrdU-FCM to 'Cat1A' in LLNA while LLNA: DA and LLNA: BrdU-ELISA correctly predicted it as 'Cat1B'. Prediction for methyl methacrylate was also different among LLNA variants from NC to Cat1B but ECt's for methyl methacrylate were near 100%, signifying that it is a very weak sensitizer, which conformed to the previous case reports (Basketter et al., 2014). Isopropanol was known to be a sensitizer in human (García-Gavín et al., 2011) but many studies involving mice suggest that it is a non-sensitizer in mice, alluding possible species difference. Overall, the discrepancy shown by LLNA variants in the decision of Cat1B may stem from the weak sensitization potentials of Cat1B and inherently substantial variation of animal experiments (Basketter et al., 2012).

To evaluate the association and consistency among test methods, several statistical methods can be employed. According to the comprehensive review on the correlation analysis in the continuous variables by Zaki et al., Bland-Altman's limits of agreement is the most frequently used method (85%, 178/210 studies) (Zaki et al., 2012). This method establishes the confidence limit for the difference between two methods and visually presents the agreement by plotting the corresponding values from the methods of interests. Therefore, the closer the difference is to 0 and the larger the number of values fall within the confidence limit, the more concordant the comparing methods become (Altman and Bland, 1983; Bland and Altman, 1986). In our study, Bland-Altman plot showed that all LLNA variants are a little bit biased to overestimate ECt compared to LLNA except the BrdU-FCM, yet the BrdU-FCM showed the smallest magnitude of bias.

Bland Altman plot, however, only can illustrates the correlation of two variables, yet it does not quantify the strength of correlation, thus should be augmented by other inferential statistics. To supplement this, we employed Pearson (paired) or Intra-class correlation coefficient (three or more groups) based on the ECt values, that provided with quantitative figures to assess the extent of correlation between LLNAs. Based on the coefficient, we could conclude that LLNA: DA, LLNA: BrdU-ELISA and LLNA: BrdU-FCM are comparable to LLNA with respect to ECt values. The concordance of sub-categorization decision was analyzed by weighted kappa, since we are interested in the agreement of the ordinal data; thus weighted kappa, rather than Spearman Correlation coefficient, was used in our analysis. Weighted kappa considers the agreement and disagreement between subcategorization results for each substance, but also examines the extent of disagreement (Cohen, 1968). For example, 2-mercaptobenzothiazole is categorized as Cat 1A in LLNA, yet being categorized as Cat 1B (one level difference) based on LLNA: DA. In contrast, LNCC categorized 2-mercaptobenzothiazole as No category (two level difference). Although both LLNA: DA and LNCC disagreed with LLNA, weighted kappa counts them differently. Since our study sought to evaluate the consistency of the decision (categorization) by substance-base not by overall rank, weighted kappa should be more heavily considered, which suggested that LLNA: BrdU-FCM has the highest agreement to LLNA (0.833 (95% C.I. 0.604–1.000)) and second best agreement to LLNA with human result (0.539 (95% C.I. 0.233–0.846)) based on the point estimate, yet the 95% confidence intervals overlapped; given that we have limited sample size, further study is needed to evaluate the comparative performance of the tests.

In conclusion, our analysis showed that LLNA variants like LLNA: DA, LLNA: BrdU-ELISA, and LLNA: BrdU-FCM are highly concordant

with LLNA in their subcategorization. The overall quantified measure showed that these LLNA variants are non-inferior to LLNA in terms of the concordance of the classification decision and the correlation of the ECt values. However, this conclusion shall be taken with caution since limited number of substances were considered.

Conflict of interest

None.

Acknowledgements:

This research was supported by grants (13172MFDS987, 15181MFDS457, 16181MFDS386 and Statistical Analysis of Validation of Animal Alternative Test (Skin Sensitisation and Eye Irritation)) from the Ministry of Food & Drug Safety of Korea.

Transparency document

Transparency document related to this article can be found online at <http://dx.doi.org/10.1016/j.yrtph.2017.02.003>.

References

- Ahn, I., et al., 2016. Performance standard-based validation study for local lymph node assay: 5-bromo-2-deoxyuridine-flow cytometry method. *Regul. Toxicol. Pharmacol.* 80, 183–194.
- Altman, D.G., Bland, J.M., 1983. Measurement in medicine: the analysis of method comparison studies. *statistician* 307–317.
- Api, A.M., et al., 2008. Dermal sensitization quantitative risk assessment (QRA) for fragrance ingredients. *Regul. Toxicol. Pharmacol.* 52, 3–23.
- Basketter, D., et al., 2012. Experience with local lymph node assay performance standards using standard radioactivity and nonradioactive cell count measurements. *J. Appl. Toxicol.* 32, 590–596.
- Basketter, D.A., et al., 2014. Categorization of chemicals according to their relative human skin sensitizing potency. *Dermatitis* 25, 11–21.
- Basketter, D.A., et al., 1998. Strategies for identifying false positive responses in predictive skin sensitization tests. *Food Chem. Toxicol.* 36, 327–333.
- Basketter, D.A., Lea, L.J., Cooper, K., Stocks, J., Dickens, A., Pate, I., et al., 1999. Threshold for classification as a skin sensitizer in the local lymph node assay: a statistical evaluation. *Food. Chem. Toxicol.* 37 (12), 1167–1174.
- Bland, J.M., Altman, D., 1986. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 327, 307–310.
- Cardin, C.W., et al., 1986. Dose-response assessments of Kathon biocide. (II). Threshold prophetic patch testing. *Contact Dermat.* 15, 10–16.
- Cohen, J., 1968. Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit. *Psychol. Bull.* 70, 213.
- García-Gavín, J., et al., 2011. Allergic contact dermatitis caused by isopropyl alcohol: a missed allergen? *Contact Dermat.* 65, 101–106.
- Griem, P., et al., 2003. Proposal for a risk assessment methodology for skin sensitization based on sensitization potency data. *Regul. Toxicol. Pharmacol.* 38, 269–290.
- Haslam, G., et al., 2000. Estimating the number of viable animal cells in multi-well cultures based on their lactate dehydrogenase activities. *Cytotechnology* 32, 63–75.
- Hunt, R.J., 1986. Percent agreement, Pearson's correlation, and kappa as measures of inter-examiner reliability. *J. Dent. Res.* 65, 128–130.
- ICCVAM, 2010a. ICCVAM Test Method Evaluation Report on the Murine Local Lymph Node Assay: BrdU-ELISA.
- ICCVAM, 2010b. ICCVAM Test Method Evaluation Report on the Murine Local Lymph Node Assay: DA.
- ICCVAM, 2011. ICCVAM Test method evaluation report: usefulness and limitations of the murine local lymph node assay for potency categorization of chemicals causing allergic contact dermatitis in humans. NIH Publication No. 11–7709. In: N. T. P. I. C. f. t. E. o. A. T. (Ed.), *Methods*. National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709.
- Jang, W.-h., et al., 2015. Evaluation of Eye Irritation Potential of Solid Substance with New 3D Reconstructed Human Cornea Model, MCTT HCETM. *Biomol. Ther. (Seoul)* 23, 379.
- Jung, K.-M., et al., 2012. B cell increases and ex vivo IL-2 production as secondary endpoints for the detection of sensitizers in non-radioisotopic local lymph node assay using flow cytometry. *Toxicol. Lett.* 209, 255–263.
- Jung, K.M., et al., 2010. Comparison of flow cytometry and immunohistochemistry in non-radioisotopic murine lymph node assay using bromodeoxyuridine. *Toxicol. Lett.* 192, 229–237.
- Kim, D.-e., et al., 2016. Predictive capacity of a non-radioisotopic local lymph node

- assay using flow cytometry, LLNA: BrdU–FCM: comparison of a cutoff approach and inferential statistics. *J. Pharmacol. Toxicol. Methods* 78, 76–84.
- Kligman, A.M., 1966. The identification of contact allergens by human assay. 3. The maximization test: a procedure for screening and rating contact sensitizers. *J. Invest. Dermatol* 47, 393–409.
- Kligman, A.M., 1966. The identification of contact allergens by human assay. II. Factors influencing the induction and measurement of allergic contact dermatitis. *J. Invest. Dermatol* 47, 375–392.
- Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174.
- Loveless, S.E., et al., 2010. Potency values from the local lymph node assay: application to classification, labelling and risk assessment. *Regul. Toxicol. Pharmacol.* 56, 54–66.
- OECD, 1992. Test No. 406: skin sensitisation. In: OECD. OECD Publishing, Paris.
- OECD, 2010a. Test No. 429: skin sensitisation: local lymph node assay. In: OECD. OECD Publishing, Paris.
- OECD, 2010b. Test No. 442B: skin sensitization local lymph node assay: BrdU-ELISA. In: OECD. OECD Publishing, Paris.
- OECD, 2015a. Test No. 442D: in chemico skin sensitisation: Direct Peptide reactivity assay (DPRA). In: OECD. OECD Publishing, Paris.
- OECD, 2015b. Test No. 442D: in vitro skin sensitisation: ARE-Nrf2 luciferase test method. In: OECD. OECD Publishing, Paris.
- Opdyke, D.L.J., 1979. Fragrance raw materials monographs. Citral. *Food Chem Toxicol* 17, 259–266.
- Ryu, H.Y., et al., 2016. Oral Toxicity Study and Skin Sensitization Test of a Cricket. *Toxicol. Res.* 32, 159.
- Schmidt, M., et al., 2010. Crucial role for human Toll-like receptor 4 in the development of contact allergy to nickel. *Nat. Immunol.* 11, 814–819.
- Schneider, K., Akkan, Z., 2004. Quantitative relationship between the local lymph node assay and human skin sensitization assays. *Regul. Toxicol. Pharmacol.* 39, 245–255.
- United_Nations, Globally Harmonized System of Classification and Labelling of Chemicals (GHS). In: of, S.-C. o. E. o. t. G. H. S. o. C. a. L., Chemicals, (Eds.). United Nations, New York and GENEVA, 2003.
- United_Nations, Globally Harmonized System of Classification and Labelling of Chemicals (GHS) rev. 3. In: of, S.-C. o. E. o. t. G. H. S. o. C. a. L., Chemicals, (Eds.). United Nations, New York and GENEVA, 2009.
- Yang, H., et al., 2015. Appraisal of within- and between-laboratory reproducibility of non-radioisotopic local lymph node assay using flow cytometry, LLNA: BrdU-FCM: comparison of OECD TG429 performance standard and statistical evaluation. *Toxicol. Lett.* 234, 172–179.
- Zaki, R., et al., 2012. Statistical methods used to test for agreement of medical instruments measuring continuous variables in method comparison studies: a systematic review. *PLoS One* 7, e37908.