Toxicology Letters 234 (2015) 172–179

Contents lists available at ScienceDirect

Toxicology Letters

journal homepage: www.elsevier.com/locate/toxlet

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

Hyeri Yang [a], Jihye Na [a], Won-Hee Jang [b], Mi-Sook Jung [c], Jun-Young Jeon [c], Yong Heo [d], Kyung-Wook Yeo [d], Ji-Hoon Jo [d], Kyung-Min Lim [a,*], SeungJin Bae [a,**]

[a] College of Pharmacy, Ewha Womans University, Seoul 120-808, Republic of Korea
[b] Medical Beauty Research Division, Amorepacific Corp. R&D Center, Republic of Korea
[c] Pharmacology Efficacy Team, Biotoxtech Co., Ltd., Republic of Korea
[d] Department of Occupational Health, College of Natural Sciences, Catholic University of Daegu, Daegu, 712-702, Republic of Korea

HIGHLIGHTS

- Here, we assessed within- and between-laboratory reproducibility of LLNA:BrdU-FCM, a non-radioisotopic analog of LLNA.
- We compared the criteria given by OECD TG429 and formal statistical methodologies.
- We found that reproducibility may be assessed more rigorously through the application of statistical methods.

ARTICLE INFO

Article history:
Received 31 December 2014
Received in revised form 15 February 2015
Accepted 18 February 2015
Available online 27 February 2015

Keywords:
Local lymph node assay
LLNA:BrdU-FCM
Within and between laboratory reproducibility
OECD TG performance standard
Descriptive and inferential statistics

ABSTRACT

Mouse local lymph node assay (LLNA, OECD TG429) is an alternative test replacing conventional guinea pig tests (OECD TG406) for the skin sensitization test but the use of a radioisotopic agent, ^{3}H-thymidine, deters its active dissemination. New non-radioisotopic LLNA, LLNA:BrdU-FCM employs a non-radioisotopic analog, 5-bromo-2′-deoxyuridine (BrdU) and flow cytometry. For an analogous method, OECD TG429 performance standard (PS) advises that two reference compounds be tested repeatedly and ECt(threshold) values obtained must fall within acceptable ranges to prove within- and between-laboratory reproducibility. However, this criteria is somewhat arbitrary and sample size of ECt is less than 5, raising concerns about insufficient reliability. Here, we explored various statistical methods to evaluate the reproducibility of LLNA:BrdU-FCM with stimulation index (SI), the raw data for ECt calculation, produced from 3 laboratories. Descriptive statistics along with graphical representation of SI was presented. For inferential statistics, parametric and non-parametric methods were applied to test the reproducibility of SI of a concurrent positive control and the robustness of results were investigated. Descriptive statistics and graphical representation of SI alone could illustrate the within- and between-laboratory reproducibility. Inferential statistics employing parametric and nonparametric methods drew similar conclusion. While all labs passed within- and between-laboratory reproducibility criteria given by OECD TG429 PS based on ECt values, statistical evaluation based on SI values showed that only two labs succeeded in achieving within-laboratory reproducibility. For those two labs that satisfied the within-lab reproducibility, between-laboratory reproducibility could be also attained based on inferential as well as descriptive statistics.

1. Introduction

Murine local lymph node assay (LLNA) is an OECD-endorsed alternative test method to evaluate the skin sensitization (OECD, 2010a), replacing conventional guinea pig tests (OECD, 1992).

* Corresponding author at: College of Pharmacy, Ewha Womans University, 52 Ewhayeodae-gil, Seodaemun-gu, Seoul 120-808, Republic of Korea. Tel.: +82 2 3277 3055; fax: +82 2 3277 3760.
** Corresponding author. Tel.: +82 2 3277 3056; fax:+82 2 3277 2851.
E-mail addresses: kmlim@ewha.ac.kr (K.-M. Lim), sjbae@ewha.ac.kr (S. Bae).

http://dx.doi.org/10.1016/j.toxlet.2015.02.014
0378-4274/

Although LLNA is still "*in vivo*" and should be an interim measure to be replaced by true *in vitro* alternatives in the future, *in vitro* alternatives currently available or to be available in the foreseeable future as exemplified by hCLAT, direct peptide reactivity assay (DPRA) or KeratinoSens®, are not expected to replace LLNA for a substantial period of time since they do not overcome the intrinsic limitation, *i.e.*, will not be able to transcend the realm of hazard identification. Ultimately, skin sensitization test method must be able to produce relative potency data that can contribute to the risk assessment/management (Kimber et al., 2001). In this context, LLNA has an unmatched merit in its capability to produce dose-response relationship data that can provide invaluable information for risk assessment.

Original LLNA method employs the radioisotopic ^{3}H-thymidine to quantitate proliferating lymph node cells upon exposure to test substances. Due to the *in vivo* use of a radioisotope with a long physical half-life, however, its wide diffusion is hampered since many countries are implementing strict regulations regarding the disposal of radioactive wastes. To circumvent this issue, newly developed LLNA:BrdU-FCM uses a non-radioisotopic thymidine analog, 5-bromo-2′-deoxyuridine (BrdU) and detects the BrdU-incorporated lymph node cells through antibody-assisted flow cytometric method (Jung et al., 2012, 2010).

As with the case of original vs. generic drugs, OECD test guideline recommends that a new analogous or me-too test method should be evaluated for its equivalence to the original test method in reproducibility and predictive capacity based on the pre-determined criteria described in performance standard (PS) (as appended to the original test guideline). To evaluate the within-laboratory reproducibility of a new test method to LLNA, OECD TG429 PS advises that one reference (positive) compound (hexylcinnamaldehyde, HCA) shall be tested repeatedly four times and ECt values, an estimate of the test material concentration required to produce a stimulation index (SI) of threshold or cutoff for determination of sensitizers, must fall within pre-determined acceptable range (5–20%). To evaluate the between-laboratory reproducibility, two positive reference compounds (HCA and 2,4-dinitrochlorobenzene, DNCB) shall be tested by 3 independent laboratories (that attained within laboratory reproducibility) and ECt values obtained must fall within acceptable range (5–20% for HCA and 0.025–0.01% for DNCB). Up to now two analogous LLNA methods, LLNA:DA(OECD, 2010a) and LLNA:BrdU-ELISA (OECD, 2010b), have been approved by OECD after demonstrating that they satisfy the criteria provided by PS. However, this criteria is somewhat arbitrary and the sample size of ECt is less than 5, raising some concerns about insufficient reliability and statistical power.

Previously, the consistency in ECt as expressed in coefficients of variation has been used as an index for the estimation of within- or between-laboratory reproducibility (Dean et al., 2001). Actually, ECt is a figure summarizing a line constructed by the regression of 3 concentration-SI points (Loveless et al., 1996). Each concentration-SI point is again obtained from 4 to 5 animals, reflecting that an ECt value represents at least 12 SI values. Considering that one point of a concurrent positive control (generally, 25% HCA) is included with 3 concentration points of test article, numerous SI values are generated during each LLNA trial. Actually, SI is an important index by providing a cut-off for the classification of sensitizer (Basketter et al., 1999) and Ehling et al., tried to provide a statistical rationale for determining sensitizers in LLNA employing SI values (Ehling et al., 2005a,b).

Several studies attempted to demonstrate the reproducibility of LLNA or LLNA:BrdU-ELISA through the graphical representation of mean and standard deviation (SD) or 95% confidence interval of SI for each concentration point (Kojima et al., 2011; Omori et al., 2008; Omori and Sozu, 2007). To obtain representative SI values produced from multiple laboratories, the variance component, τ^2, which is commonly used in meta-analysis and estimated based on random-effect model for the log-transformed SI, was used to appraise the between-lab variation. (Kojima et al., 2011; Omori et al., 2008). Haneke et al. (2001) estimated the within- and between-laboratory reproducibility (or reliability) by calculating the consistency statistics (k and h, respectively) with SI. Kimber et al. (1991) examined between-laboratory reproducibility by applying analysis of covariance with the dose-dependent SI values produced from 4 laboratories (Kimber et al., 1991). Another important and commonly used methodology for estimating reproducibility is to evaluate the consistency in binary decision (non-sensitizer vs. sensitizer) between trials in single laboratory or trials of multiple laboratories (Idehara et al., 2008; Kimber et al., 1998; Scholes et al., 1992) which can be further analyzed by kappa-statistics (Viera and Garrett, 2005).

Unlike inferential statistics which tests the null hypothesis based on the level of significance and the p-value thus rejects or fails to reject the null hypothesis, descriptive statistics only "describe" the magnitude of reproducibility. Incidentally, SI values for a concurrent positive control have limited sample size (N = 4–5 in one LLNA trial), common sense indicates that normality assumption may fail; yet considering that baseline characteristics of experimental studies are relatively homogenous (Hothorn, 2014; Na et al., 2014) as contrasted with those of clinical studies, small sample sizes of SI do not necessarily imply that the classification of sensitizer based on SI may be statistically flawed (Basketter et al., 2009) or that parametric method should not be used. However, unlike clinical studies which emphasize the importance of sufficiently large sample sizes and the use of the appropriate statistical methods (Pagano and Gauvreau, 2000), few experimental studies have fully discussed the appropriate application of the statistical methods (Na et al., 2014).

In this study, using the data produced by 3 laboratories, we first investigated the within- and between-laboratory reproducibility of LLNA:BrdU-FCM based on ECt values of HCA and DNCB, according to criteria given by PS of OECD TG429. To further examine the reproducibility, we analyzed the SI data employing descriptive statistics and inferential statistics. Briefly, the mean and SD of all SI values was graphically presented, and the reproducibility in SI values of a concurrent positive control, HCA 25% (Dearman et al., 2001) was investigated based on inferential statistics. We employed both parametric (one-way ANOVA and student t-test) and non-parametric (Kruskal–Wallis and Wilcoxon rank sum test) methods to evaluate the within- and between-laboratory reproducibility along with examining assumptions behind parametric approach (such as test of normality and equal variance assumption), and results obtained from both approaches were compared and discussed.

2. Materials and methods

2.1. Chemicals and reagents

2,4-Dinitrochlorobenzene (DNCB), hexylcinnamaldehyde (HCA) and 5-bromo-2_-deoxyuridine (BrdU) were obtained from Sigma–Aldrich (San Diego, CA., USA). DNCB and HCA were dissolved in acetone:olive oil (AOO; 4:1). BrdU was dissolved in phosphate-buffered saline (PBS) at a concentration of 20 mg/mL.

2.2. LLNA:BrdU-FCM

Both the animal care and study protocol employed were in accordance with Institutional Animal Care and Use Committee (IACUC) of each participating laboratory. LLNA:BrdU-FCM assays were conducted according to previous reports (Jung et al., 2012;