Regulatory Toxicology and Pharmacology 59 (2011) 191–196

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

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

Safety assessment of allergic contact dermatitis hazards: An analysis supporting reduced animal use for the murine local lymph node assay

Joseph K. Haseman [a], Judy Strickland [b], David Allen [b], Eleni Salicru [b], Michael Paris [b], Raymond R. Tice [c], William S. Stokes [d,*]

[a] J.K. Haseman Consulting, Raleigh, NC 27614, USA
[b] Integrated Laboratory Systems, Inc., Research Triangle Park, NC 27709, USA
[c] The National Institute of Environmental Health Sciences, The National Toxicology Program, Biomolecular Screening Branch, Research Triangle Park, NC 27709, USA
[d] The National Institute of Environmental Health Sciences, The National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods, Research Triangle Park, NC 27709, USA

ARTICLE INFO

Article history:
Received 25 August 2010
Available online 23 October 2010

Keywords:
Local lymph node assay
Skin sensitization
Alternative test method
Animal reduction
Sample size
OECD test guideline 429

ABSTRACT

The original Organisation for Economic Co-operation and Development Test Guideline 429 (OECD TG 429) for the murine local lymph node assay (LLNA) required five mice/group if mice were processed individually. We used data from 83 LLNA tests (275 treated groups) to determine the impact on the LLNA outcome of reducing the group size from five to four. From DPM measurements, we formed all possible four- and five-mice combinations for the treated and control groups. Stimulation index (SI) values from each four-mice combination were compared with those from five-mice combinations, and agreement (both SI <3 or both SI $\geqslant$3) determined. Average agreement between group sizes was 97.5% for the 275 treated groups. Compared test-by-test, 90% (75/83) of the tests had 100% agreement; agreement was 83% for the remaining eight tests. Disagreement was due primarily to variability in animal responses and closeness of the SI to three (positive response threshold) rather than to group size reduction. We conclude that using four rather than five mice per group would reduce animal use by 20% without adversely impacting LLNA performance. This analysis supported the recent update to OECD TG 429 allowing a minimum of four mice/group when each mouse is processed individually.

Published by Elsevier Inc.

1. Introduction

The murine local lymph node assay (LLNA[1]) (Dean et al., 2001; Haneke et al., 2001; ICCVAM, 1999; Sailstad et al., 2001) is an alternative skin sensitization test method that requires fewer animals and less time than currently accepted guinea pig tests (e.g., the guinea pig maximization test and the Buehler test) and represents a significant reduction in animal pain and distress. The LLNA is based on the principle that sensitizing chemicals induce lymphocyte proliferation in the lymph nodes draining the test substance application site. Cell proliferation is determined by analyzing the extent of incorporation of a radioactive marker into newly synthesized DNA. Under appropriate test conditions, proliferation is proportional to the dose applied, and provides a means of obtaining an objective, quantitative measurement of sensitization (EPA, 2003; ICCVAM, 2009; OECD, 2010). The LLNA was the first alternative test method evaluated and recommended by the US Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) for consideration by US regulatory agencies (ICCVAM, 1999). Since 2002, regulatory authorities internationally have recognized the LLNA as an acceptable alternative to guinea pig tests for most testing situations (Stokes and Schechtman, 2008).

The current US and original international test guidelines for the LLNA, US Environmental Protection Agency (EPA) Health Effects Test Guidelines on Skin Sensitization OPPTS 870.2600 (EPA, 2003) and Organisation for Economic Co-operation and Development (OECD) Test Guideline (TG) 429 for Skin Sensitisation: LLNA (OECD, 2002) are based on similar LLNA protocols. Both guidelines include a comparative assessment of lymph node cell proliferation in treated and control groups of mice by measuring the incorporation of ^{3}H-thymidine or ^{125}I-iododeoxyuridine (measured as disintegrations per minute [DPM]) into the DNA of draining auricular lymph nodes. The stimulation index (SI) is the ratio of the

* Corresponding author. Address: 530 Davis Drive, P.O. Box 12233, K2-16, Research Triangle Park, NC 27709, USA. Fax: +1 919 541 0947.
E-mail address: stokes@niehs.nih.gov (W.S. Stokes).

[1] *Abbreviations used:* DPM, disintegrations per minute; EPA, US Environmental Protection Agency; ICCVAM, Interagency Coordinating Committee on the Validation of Alternative Methods; LLNA, murine local lymph node assay; NICEATM, National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods; OECD, Organisation for Economic Co-operation and Development; OPPTS 870.2600, US Environmental Protection Agency Office of Prevention, Pesticides and Toxic Substances Health Effects Test Guidelines on Skin Sensitization 870.2600; SI, stimulation index; TG, test guideline; TG 429, OECD TG 429 – Skin Sensitisation: Local Lymph Node Assay.

0273-2300/$ - see front matter Published by Elsevier Inc.
doi:10.1016/j.yrtph.2010.10.004

incorporated radioactivity, in DPM, of the treated group to that of the vehicle control group. If the SI $\geqslant$3 the substance is classified as a skin sensitizer. If the SI <3 the substance is classified as a nonsensitizer.

The current EPA and the original OECD test guidelines have specific differences however. EPA OPPTS 870.2600 requires at least five mice per group and the collection of individual animal data so that interanimal variability can be assessed. The original version of OECD TG 429 allowed for as few as four mice per dose group when the lymph nodes of the mice in each dose group were pooled. When individual animal data were collected, consistent with the EPA test guideline, OECD TG 429 required at least five mice per dose group. Because many international animal care and use regulations require that the minimum number of mice necessary be used for testing, many laboratories opted to collect pooled data from only four mice per dose group.

Recently, OECD TG 429 was updated (OECD, 2010) to allow for a minimum of four mice per dose group whether the lymph nodes are processed individually for each mouse or whether the lymph nodes of the mice in each dose group are pooled. This update to OECD TG 429 was based on the analysis contained herein and conducted by the National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) in support of an ICCVAM evaluation to determine if the LLNA would continue to support the same level of public health protection if the number of animals in each LLNA dose group is reduced from five to four.

1.1. Objectives

Collecting lymph nodes from individual mice has several advantages over pooling lymph nodes. Interanimal variability can be assessed, which allows for a statistical comparison of differences between test substance and vehicle control groups, along with an opportunity to identify outlier responses using statistical tests such as Dixon's test (Dixon and Massey, 1983). Identifying outlier responses may prevent false results for substances that produce responses near threshold values. Substances that normally would induce an SI value just above or below three might be incorrectly classified due to a low or high outlier value, respectively, if the outlier is not identified and excluded.

The purpose of this analysis was to determine whether the requirement of five mice per dose group for individual animal data collection in the original OECD TG 429 and current EPA 870.2600 protocols could be reduced to four mice per dose group without adversely affecting the accuracy of the LLNA. Because the "true" underlying sensitizer status for individual chemicals may not be known, this investigation focuses primarily on the degree of agreement between outcomes with groups of four or five mice rather than on which observed outcome was the "correct" one.

Although the SI value is the primary determinant of the LLNA outcome, a statistical test might be used in addition to the SI decision criterion. In fact, the EPA test guideline includes a requirement that investigators also submit LLNA data for statistical comparisons of the mean DPM values for treated and vehicle control groups (EPA, 2003). For this reason, we also used a Student's *t*-test based on log-transformed DPM data to compare each dosed group with its concurrent vehicle control. We compared the frequency of LLNA outcomes with either four- or five-mice group sizes using SI $\geqslant$3 or statistical significance to classify substances as positive.

2. Methods

This retrospective evaluation used individual animal data from LLNA tests submitted to NICEATM. These data were submitted by six laboratories that used inbred CBA mice, the strain recommended in LLNA test guidelines by OECD (2002) and EPA (2003). The 78 substances tested include individual chemicals and proprietary formulations from 83 LLNA tests. There were two tests for formaldehyde and five tests for hexyl cinnamic aldehyde. Each test consisted of three or four dose groups and a vehicle control group.

Of the 83 test results, 50 tests yielded positive results (i.e., maximum SI $\geqslant$3) and 33 tests yielded negative results (i.e., all SI <3). Among the 277 dose groups and the 67 control groups, the number of mice per group ranged from two to nine (Table 1). Two dose groups, one with two mice and one with three, contained too few mice for the comparison of LLNA outcomes and were excluded from SI $\geqslant$3 criterion analyses. LLNA test results were evaluated on a dose-by-dose basis as well as on a test-by-test basis, recognizing that the dose groups within a test used a common vehicle control group. Also, in certain laboratories, a common vehicle control group was used for multiple chemicals.

For each LLNA test that used five mice per dose group, SI values were calculated for all possible four-mice combinations in both the treated and vehicle control groups (25 possible combinations per test). The SI value of each of these combinations was compared with the SI value determined from all five mice. The proportion of outcomes with four mice that agreed with the outcome based on five mice was determined. The outcomes agreed if (1) both protocols produced SI <3 or (2) both produced SI $\geqslant$3.

For each LLNA test that had more than five mice per group, a similar procedure was applied. In these cases, however, it was necessary to form all possible four- and five-mice combinations from the full dataset. This resulted in significantly more possible combinations of samples (e.g., 8100 possible combinations for tests with six animals per dose group compared to 25 possible combinations for tests with five mice per dose group).

For those tests with more than five mice per dose group, we examined the relative impact of animal-to-animal variability and sample size reduction on the disagreement in study outcome. That is, we compared the disagreement related to reducing the sample size from five to four mice per dose group to the disagreement that would result from simply taking a second sample of five mice per dose group.

In addition to the SI $\geqslant$3 criterion, formal statistical testing was also considered. All data were log-transformed prior to statistical analyses to normalize the frequency distribution. A Student's *t*-test was used to compare each dose group with its concurrent vehicle control, and statistically significant differences ($p < 0.05$) between treated and vehicle control groups were regarded as positive test results (i.e., sensitizers). All other results ($p > 0.05$) were regarded as negative (i.e., nonsensitizers). Power calculations based on a two-sided Student's *t*-test were also conducted using a Web-based statistics program (DanielSoper.com Statistics Calculators version 2.0 [http://www.danielsoper.com/statcalc/calc49.aspx]) to determine the impact of reducing the sample size from five to four mice per group.

3. Results

3.1. Use of the SI to identify sensitizers

Table 2 shows the frequency of the various SI values among the 275 dose groups, together with the average agreement between LLNA outcomes with four or five mice per group. Only 12% (34/275) of the dose groups had less than 100% agreement between four- or five-mice outcomes. Disagreement was limited to those SI values from 2.1 to 4.7, but some dose groups in this range produced 100% agreement (see Table 2 and Supplementary Tables 1–6). Note also that, as expected, the degree of disagreement was greatest at SI values close to three (Table 2). The overall average