Regulatory Toxicology and Pharmacology 56 (2010) 352–356

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

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

Statistical evaluation of the Local Lymph Node Assay

Ludwig A. Hothorn [a,c,*], Hans-Werner Vohr [b]

[a] Institute of Biostatistics, Leibniz University Hannover, Germany
[b] Bayer HealthCare AG Wuppertal, Germany
[c] Scripps Genomic Medicine, The Scripps Research Institute, 3344 North Torrey Pines Court, La Jolla, CA 92037, United States

ARTICLE INFO

Article history:
Received 6 June 2009
Available online 14 November 2009

Keywords:
LLNA
Statistical evaluation
Proof of hazard
Proof of safety
Inference for ratios
Software

ABSTRACT

In the Local Lymph Node Assay measured endpoints for each animal, such as cell proliferation, cell counts and/or lymph node weight should be evaluated separately. The primary criterion for a positive response is when the estimated stimulation index is larger than a specified relative threshold that is endpoint- and strain-specific. When the lower confidence limit for ratio-to-control comparisons is larger than a relevance threshold, a biologically relevant increase can be concluded according to the proof of hazard. Alternatively, when the upper confidence limit for ratio-to-control comparisons is smaller than a tolerable margin, *harmlessness* can be concluded according to a proof of safety.

1. The problem

The Local Lymph Node Assay (LLNA) provides an alternative method for identifying skin sensitizing chemicals, standardized by OECD Guideline for Testing of Chemicals: Skin Sensitisation: Local Lymph Node Assay OECD/OCDE 429 (2002) and Haneke et al. (2001). In comparison to the traditional guinea pig tests, the LLNA seriously reduces animal distress.

In a randomized one-way layout, at least three doses of the test substance D_i, a negative control C and a positive control C^+, if appropriate, is used with at least five animals per treatment group. Doses are commonly selected from a concentration series, i.e., 100%, 50%, 25%. The primary endpoint is the β-scintillation count of the incorporated 3H-methyl thymidine as disintegrations per minute per node pairs (DPM). A positive response is determined when the stimulation index $SI = \frac{\bar{\text{DPM}}_i}{\bar{\text{DPM}}_C}$ is equal or larger than 3 in any of the doses D_i, where the lymph nodes are pooled for each treatment group (Gerberick et al., 2007) without a variance estimator. Thus, a relative change is used as the relevance criterion, analogously to the k-fold rule in the Ames assays (Cariello and Piegorsch, 1996). However, this relevance threshold is formulated for ratio of mean values only, whereas from the view point of statistical inference a confidence limit should be used estimated from on individual values, i.e. taken the variance and sample sizes also into account. This recommendation follows the ICCVAM Peer Review Panel that agreed that the *decision process to identify a positive response should include a SI greater than or equal to 3, statistical significance, and dose response information* (The Murine Local Lymph Node Assay: A Test Method for Assessing the Allergic Contact Dermatitis Potential of Chemicals/Compounds, 1999).

Several alternative endpoints were proposed for the LLNA, such as cellularity or lymph node weight, to assess cell proliferation without radioactive labeling of lymph node cells (Ehling et al., 2005). In the derivation of the underlying statistical methods, we assume that the endpoint is any continuous endpoint, measured on each animal j, i.e. a univariate endpoint y_{ij} in a design including a negative control and some doses D_i is performed. Based on the relative k-fold relevance criteria, we propose a statistical approach using confidence limits for ratio-to-control comparisons with and without a monotonic dose–response assumption (trend alternative). Based on an inter-laboratory study (Ehling et al., 2005) the k-fold criteria was determined to be $k = 1.5$ for cellularity in BALB/c mice and $k = 1.4$ in NMRI mice.

In the OECD 429 guideline two decision strategies are formulated: (i) the proof of hazard as *identifying skin sensitizing chemicals*, and (ii) the proof of safety (harmlessness) as *confirming that chemicals lack a significant potential to cause skin sensitization*. For both strategies, related statistical approaches are described using confidence intervals respectively.

A notable assumption is the *a priori* definition of the direction of a possible positive effect, i.e. only increases of the cellularity or the lymph node weight are of toxicological interest. Therefore, one-sided hypotheses and one-sided confidence limits for both the proof of hazard and proof of safety are used throughout.

* Corresponding author. Address: Institute of Biostatistics, Leibniz University Hannover, Herrenhauser Str. 2, D-30419 Hannover, Germany.
E-mail address: hothorn@biostat.uni-hannover.de (L.A. Hothorn).

0273-2300/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.yrtph.2009.10.007

2. An example

For two mice strains (NMRI and BALB/c) the two endpoints cellularity (cell) and lymph node weight (lnw) were measured in a control and three doses $(D_{low}, D_{med}, D_{high})$ for six animals each (Vohr, 2007), see the raw data in Table 1.

The box-plots in Fig. 1 indicate some dose-related effects in the BALB mice for both endpoints and for NRMI mice weak effects, outliers, minor variance heterogeneity and approximate symmetric distributions (notice the relative small sample sizes of $n_i = 6$ for appropriate interpretation of the distribution). We selected as an example a substance leading to negative response in the NRMI mice and a positive response in the BALB/c mice.

In the following sections, both a proof of hazard and a proof of safety approach for ratio-to-control comparisons and related one-sided confidence limits are provided.

Table 1
Cellularity raw data.

	Group	cellNRMI	lnwNRMI	cellBALB	lnwBALB
1	Control	12.0	9.6	6.8	6.6
2	Control	13.1	10.9	5.0	4.9
3	Control	15.3	9.6	7.0	6.2
4	Control	12.1	10.1	5.4	5.8
5	Control	20.2	10.5	4.7	5.5
6	Control	10.3	7.4	9.8	8.2
7	Dlow	6.4	5.5	9.0	4.9
8	Dlow	15.5	9.9	8.1	5.9
9	Dlow	18.4	11.3	10.1	6.4
10	Dlow	15.0	11.1	11.7	7.3
11	Dlow	16.3	10.6	6.1	5.7
12	Dlow	19.1	10.5	7.7	5.1
13	Dmed	23.6	13.1	12.6	5.6
14	Dmed	11.3	10.9	11.2	6.0
15	Dmed	7.2	7.5	8.9	6.1
16	Dmed	17.2	11.7	11.5	7.8
17	Dmed	13.0	11.2	8.7	6.6
18	Dmed	12.8	10.6	8.3	6.5
19	Dhigh	15.6	10.9	19.1	10.0
20	Dhigh	13.2	8.8	18.6	10.9
21	Dhigh	16.2	9.4	24.5	11.2
22	Dhigh	23.7	14.5	17.4	10.6
23	Dhigh	11.8	8.6	19.2	11.6
24	Dhigh	12.8	10.4	16.1	11.1

3. A proof of hazard approach for ratio-to-control comparisons

The comparisons of several doses versus a control in a design $C, D_1, \ldots, D_k$ are commonly performed using Dunnett (1955) procedure which belongs to the most cited statistical approaches (Ryan and Woodall, 2005) mainly used in pharmacology and toxicology. Consequently, the evaluation of the LLNA by means of Dunnett's procedure was described by Takeyoshi et al. (2003), where stars in the figures and tables indicate significant difference compared to control. Although Dunnett's procedure is a reasonable starting point for data analysis, three aspects require to reconsider its usefulness for routine evaluations.

Firstly, Dunnett's procedure compares mean differences between doses and a control, whereas the relevance criterion (defined relatively by $SI_i = \frac{\bar{y}_i}{\bar{y}_C}$) considers ratios of the means at doses to the control mean. Therefore, simultaneous (joint) confidence limits for ratios-to-control according to Dilba et al. (2004) present a better suited approach. This approach estimates directly ratios of the means at doses to the control mean as effect sizes and their simultaneous confidence intervals, whereas simultaneous means multiplicity-adjusted for the several comparisons. The common variance estimator and common degree of freedom of an one-way layout are used, and the correlations between the comparisons depending on the sample sizes, the contrast coefficients for the control versus doses comparisons and the estimated mean ratios itself are used for estimating the confidence limits.

Secondly, in a design including a control and several dose groups, a monotonicity assumption of the effects is appropriate, i.e. reasoning of a dose–response trend is intended. Williams (1971) procedure can be seen as an enhancement of the Dunnett procedure comparing doses versus control under the assumption

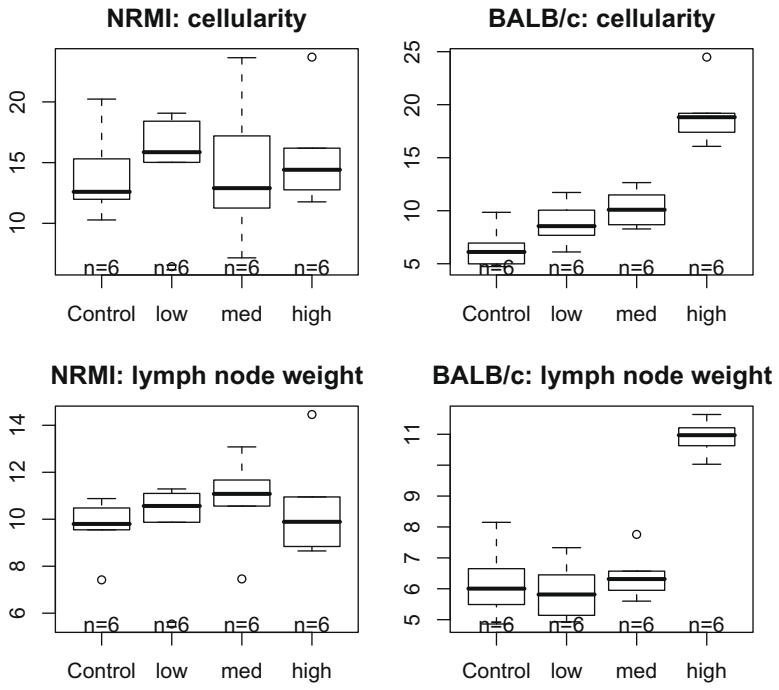

Fig. 1. Box-plots of the example data.