Blackwell Publishing

Asian J Androl 2005; 7 (1): 107–108
DOI: 10.1111/j.1745-7262.2005.00012.x

AJA

·Letters to the Editor·

Statistical testing and distribution for lead chloride toxicity

John H. Lange

Envirosafe Training and Consultants, P.O. Box 114022, Pittsburgh, PA 15239, USA
Tel: +412-780-5075; E-mail: johnhlange@worldnet.att.net

Received 2004-08-04 Accepted 2004-11-12

Dear Sir,

Graca *et al.* [1] provided an interesting investigation on the toxicity of lead chloride and sperm development in mice. However, I would like to make a comment on the statistical analysis presented. Table 1 and its results suggest that a comparison of treated (experiment) and control mice were undertaken using the *t*-test. The authors indicate that they used the *t*-test along with complementation of ANOVA analysis. It appears that the *t*-test was used for analysis in Table 1 and ANOVA, as indicated, for Table 2. Use of *t*-test for comparing three of more groups is not appropriate since this may result in a multiple comparison problem (increasing the type I error rate) [2, 3]. Multiple comparisons can result in the reporting of a *P* value that is significant (or of lower value) when in actuality it is not. It is better to use, for example, a one-way ANOVA followed by a post-test (post-hoc) which can take into account all comparisons. Other statistical testing can also be employed to control the overall type I error, such as Tukey-HSD (honest significant difference), Scheffe's and Bonferroni-Dunn methods [2].

The authors confuse readers by indicating the use of ANOVA while also mentioning the *t*-test as if they were used together. There should be a clear indication of statistical analysis for experiments reported without the necessity of readers making assumptions (i.e. Table 1 using the *t*-test). Some may interpret that the ANOVA was a post-hoc test for the *t*-test, which is inappropriate. Post-hoc tests are performed after finding significance from the ANOVA [3].

In Table 2, the authors report statistical comparison of percent values. It should be noted that it would be more appropriate to evaluate the actual numbers rather than percent. It is likely that the authors did conduct comparisons on actual numbers, but this is not made clear in the paper.

It should also be noted that using mice or animals from the same litter may result in a relatedness problem. If this is of concern to the investigator, methods that allow for correlated outcomes can be used, such as generalized estimating equations [4].

Finally, it should be noted that in many toxicology studies results (data) are not normally distributed and require use of non-parametric methods, such as the Kruskal-Wallis test [3]. This is particularly important when a small number of animals are employed. Although, this is not the case for body and organ weights [3], analysis of distribution or reference to others performing this type of evaluation is important for understanding appropriate statistical analysis. Distribution analysis can be conducted using a number of different tests [5], with one of the simplest being the Shapiro-Wilk test [6].

Although my comments are critical on the reporting and use of statistics, these issues are not uncommon in the literature [6]. It should be noted that my comments do not infer that this is a poor quality study, but illustrate the importance of proper statistical analysis and its use in interpretation. It is possible that the issue of a multiple comparison problem, specifically as suggested for Table 1, does not alter the basic interpretation presented, although similarity of test and control values (e.g. testis weight) that are reported as significant must be questioned. I would like to suggest that the study's authors

consider providing the actual *P* value so readers can better judge significance, although reporting a comparison as < 0.05, for example, is acceptable. Thus, proper use of statistics will enhance interpretative capacity of studies.

References

1 Graca A, Ramalho-Santos J, de Lourdes Pereira M. Effect of lead chloride on spermatogenesis and sperm parameters in mice. Asian J Androl 2004; 6: 237–41.

2 Sheskin DJ. Handbook of parametric and nonparametric statistical procedures. Boca Raton: CRC Press; 1997.

3 Gad SC. Statistics and experimental design for toxicologists. Boca Raton: CRC Press; 1999.

4 Hardin JW, Hibe JM. Generalized estimation equations. Boca Raton: Chapmen and Hall/CRC; 2003.

5 Sokal RR, Rohlf FJ. Biometry. New York: WH Freeman and Company; 1995.

6 Lange JH. Statistical analysis of exposure data – formaldehyde in occupied and unoccupied caravans in Australia. Indoor Built Environ 2001; 10: 116–7.

Reply to 'Statistical testing and distribution for lead chloride toxicity'

Maria de Lourdes Pereira[1], J. Ramalho-Santos[2]

[1]Department of Biology, University of Aveiro, 3810 Aveiro, Portugal
[2]Department of Zoology, University of Coimbra, Coimbra, Portugal
Tel: +351-234-370-770, Fax: +351-234-426-408; E-mail: lpereira@bio.ua.pt

Received 2004-11-12 Accepted 2004-11-12

Dear Sir,

We are very grateful for the letter written by Dr Lange, and indeed apologize for the mistakes noted in the wording of our text regarding statistical analysis. This was due to changes carried out while revising the manuscript at the request of reviewers, whom we thank for, pointing out several issues that were actually similar to those noted by Dr. Lange. Unfortunately, we were unable to describe and discuss our findings properly in the context of the revision.

To clarify: in Table 2 we compared controls and treatments (columns 1 and 2, or 3 and 4) using the *t*-test (i.e. two groups at a time, as noted by Dr. Lange). We also compared controls for lead injection and recovery for the same variable (columns 1 and 3). The *P* values presented in the table as significant are from those comparisons. At the request of reviewers, we indeed performed further analysis on our data using one-way ANOVA, to compare all columns, with post-tests taking into account all comparisons (Bonferroni and Tukey). These tests showed that the only significant differences were the ones already discussed in the first version of the Manuscript. However, this was swiftly and inappropriately worded in the final version of the manuscript, and, we now realize, was not clear at all to readers.

We have since also performed the Kruskal-Wallis test (for reasons Dr. Lange clearly explains in his letter) with Dunn's post-test, to similar results.

A few short clarifications: The mice were bought on several different occasions from a supplier, in numbers that imply several litters, and assigned randomly to different experimental groups. Although we cannot be sure, we believe there should be no relatedness issue in this case. In regard of semen analysis, following the WHO guidelines for human semen [1], we evaluated motility, sperm abnormalities and acrosome reaction in terms of percentage.

This is also not at all uncommon when evaluating sperm from animal models, given that different procedures (treatments, genetic manipulations, *etc*.) may have an effect on sperm count, thus rendering total number comparisons misleading.

References

1 World Health Organization. WHO Laboratory Manual for the Examination of Human Semen and Sperm-Cervical Mucus Interaction (Fourth Edition). Cambridge: Cambridge University Press; 1999.