Application of public-domain statistical analysis software for evaluation and comparison of comet assay data

Pablo E. Verde [a], Laura A. Geracitano [b], Lílian L. Amado [b,c], Carlos E. Rosa [b,d], Adalto Bianchini [b,c,d], José M. Monserrat [b,c,d,*]

[a] Coordination Centre for Clinical Trials and Department of Social Medicine, Heinrich Heine Universitaet Duesseldorf, Moorenstr. 5, D-40225 Duesseldorf, Germany
[b] Departamento de Ciências Fisiológicas, Fundação Universidade Federal do Rio Grande (FURG), R. Eng. Alfredo Huch 475 (96201-900), Rio Grande, RS, Brazil
[c] Programa de Pós-Graduação em Oceanografia Biológica, FURG, Brazil
[d] Programa de Pós-Graduação em Ciências Fisiológicas – Fisiologia Animal Comparada, FURG, Brazil

Abstract

A novel approach for statistical analysis of comet assay data (i.e.: tail moment) is proposed, employing public-domain statistical software, the R system. The analytical strategy takes into account that the distribution of comet assay data, like the tail moment, is usually skewed and do not follow a normal distribution. Probability distributions used to model comet assay data included: the Weibull, the exponential, the logistic, the normal, the log normal and log-logistic distribution. In this approach it was also considered that heterogeneity observed among experimental units is a random feature of the comet assay data. This statistical model can be characterized with a location parameter m_{ij}, a scale parameter r and a between experimental units variability parameter θ. In the logarithmic scale, the parameter m_{ij} depends additively on treatment and random effects, as follows:

$$\log(m_{ij}) = a_0 + a_1 x_{ij} + b_i,$$

where $\exp(a_0)$ represents approximately the mean value of the control group, $\exp(a_1)$ can be interpreted as the relative risk of damage with respect to the control group, x_{ij} is an indicator of experimental group and $\exp(b_i)$ is the individual risk effects assume to follows a Gamma distribution with mean 1 and variance θ. Model selection is based on Akaike's information criteria (AIC). Real data coming from comet analysis of blood samples taken from the flounder *Paralichtys orbignyanus* (Teleostei: Paralichtyidae) and from samples of cells suspension obtained from the estuarine polychaeta *Laeonereis acuta* (Nereididae) were employed. This statistical approach showed that the comet assay data should be analyzed under a modeling framework that take into account the important features of these measurements. Model selection and heterogeneity between experimental units play central points in the analysis of these data.
© 2006 Elsevier B.V. All rights reserved.

Keywords: Comet assay; Tail moment; Statistical modeling; Model selection; Akaike information criteria; Estuarine organisms

1. Introduction

The comet assay is a simple, rapid and sensitive tool, first introduced by Östling and Johanson in 1984 as a microelectrophoretic technique for direct visualization

* Corresponding author. Tel.: +55 53 233 6856;
fax: +55 53 233 6850.
E-mail address: jose@octopus.furg.br (J.M. Monserrat).

of DNA damage in individual cells [1]. Since its initial proposal, a number of laboratories have been working in the improvement of this technique for a wide range of research domains [2]. In aquatic toxicology, the comet assay has been employed successfully in different invertebrate and vertebrate species, reflecting the pollutants effects in field and also in laboratory experiments [3]. The basis of the comet assay involves embedding a suspension of single cells in low melting point (LMP) agarose on a microscope slide. The cells are lysed, the DNA is electrophoresed and stained with a fluorescent DNA binding dye. The extent of DNA strand breakage can be assessed microscopically based on the appearance of damage nuclei (visual scoring) or by using image analysis systems that provide more detailed information, such as comet tail length, % of DNA in the comet tail and tail moment. This last parameter can be calculated as the product of comet tail length and % of DNA in the comet tail length [4].

Up to date, no clear statistical procedures have been proposed for analysis of comet assay data. The Comet Assay Interest Group (http://www.cometassay.com/) has included in their recommendations the use of the median as a position parameter, taking into account that statistical distribution of comet assay data are usually skewed and do not follow a normal distribution. In this way, a sequence of 50 tail moments (or the selected parameter employed to quantify DNA damage) is summarized in a single value, the median. Under this context, it is expected that comet assay data coming from control groups should present higher frequent values on the left tail of the distribution (low damage) than those from exposed groups or positive controls treated with oxidants like hydrogen peroxide. In order to perform statistical analysis, the Comet Assay Interest Group suggests calculating a mean response for each treatment, using the median response of each replicate as data. In this way, a priori, it can be compared the mean of medians of each treatment through conventional analysis such as ANOVA. Duez et al. [5] have summarized some of the usually employed statistical methods, including χ^2 applied after visual scoring of comets, parametric tests (*t*-test or ANOVA) without any transformation or after log transformation. The effects of applying mathematical transformation have been previously analyzed after performing simulation studies to evaluate the statistical behavior of DNA damage estimates like tail length and tail moment [6]. For tail moment it was established that the log transformation clearly augmented the power of statistical analysis using a three-way nested linear model.

The analysis of comet assay data deserves several important considerations. First, as previously recognized

by Lovell et al. [7], observations are nested within experimental units (i.e., animals), generating an intra-individual variability and, as result, observations within individuals might be auto-correlated. Second, the number of observations within experimental units is usually different (unbalanced design), which generates variability changes between the median response of different experimental units [8]. Ignoring these data collection issues may lead to inappropriate application of ANOVA procedures [9].

It is clear that comet assay methodology generates great number of data and it seems more interesting to analyze the data generating process for a parameter like tail length and then assess the difference of experimental groups (for example, control and treated one). As mentioned previously, data such as tail moment present skewed distribution, so some authors [10,11] have proposed the Weibull distribution or even the sum of two Gaussian curves in order to fit this kind of data. However, other statistical models, like the log-logistic, can also be employed to fit tail moment data. In general, the analysis of positive measurement values with asymmetric distributions has a long history in statistical analysis, with special application in survival analysis or failure data [12].

Taking into account the facts cited above, the aims of this paper were two folds:

1. To introduce a two levels hierarchical model as a statistical modeling framework to analyze tail moment data. The first level corresponds to tail moment measurements within individuals and the second level corresponds to variability between individuals. This approach makes possible to study different distributional assumptions in both levels and to assess treatment effects.
2. To employ R, a public domain and open source statistical analysis software, to analyze tail moment measurements [13]. The main R computational tool involved in the analysis presented here is the package "survival" [14]. Full documented worked examples of this package can be found in Ref. [15].

2. Material and methods

2.1. Samples preparation

In order to get real data to compare different statistical approaches, comet assays were performed on blood samples of the flounder *Paralichtys orbignyanus* (Teleostei: Paralichtyidae) and from samples of cells suspension obtained from the estuarine polychaeta *Laeonereis acuta* (Nereididae). Flounders