ORIGINAL RESEARCH

A method for ranking compounds based on their relative toxicity using neural networking, *C. elegans*, axenic liquid culture, and the COPAS parameters TOF and EXT

Martine Ferguson[1]
Marc Boyer[1]
Robert Sprando[2]

[1]United States Food and Drug Administration, Center for Food Safety and Applied Nutrition, Office of Food Defense Communication and Emergency Response, Division of Public Health and Biostatistics, College Park, MD, USA; [2]United States Food and Drug Administration, Center for Food Safety and Applied Nutrition, Office of Applied Research and Safety Assessment, Division of Toxicology, Laurel, MD, USA

Abstract: *Caenorhabditis elegans* (L1s) were exposed to (in order of decreasing toxicity) sodium arsenite, sodium fluoride, caffeine, valproic acid, sodium borate, or dimethyl sulfoxide in *C. elegans* habitation medium (CeHM) for 72 consecutive hours. At this time point nematode growth and development were assessed using a Complex Object Parametric Analyzer and Sorter (COPAS™). The COPAS generated biomarkers of growth (time of flight [TOF] – a measure of axial length) and development (extinction [EXT] – a measure of optical density) were subsequently utilized to rank compounds according to their relative toxicity, as measured by the rat oral LD-50, using artificial neural network methods. Neural network methods were utilized to analyze this data because of their ability to model nonlinear endpoints and a multilayer perceptron neural network method was used because of its capability to function well in the presence of collinearity. Using a neural network approach we found that the LD-50 was correctly predicted 96% of the time. The present study demonstrates that neural network methods can be utilized to rank compounds according to their relative toxicity using COPAS-generated data (TOF and EXT) obtained from exposing a large number of nematodes to water-soluble compounds in axenic liquid culture.

Keywords: neural network, TOF, EXT, COPAS, *C. elegans*, rat oral LD-50

Introduction

The Complex Object Parametric Analyzer and Sorter (COPAS™; Union Biometrica, MA), an instrument designed to facilitate the rapid, accurate, and efficient analysis, dispensing and/or sorting of a large number of nematodes by measuring nematode axial length (time of flight, TOF), optical density (extinction, EXT), and the intensity of specific fluorescent markers, is commonly utilized in the development of high throughput assays. The quantity of data obtained using COPAS technology can be enormous and statistical methodologies are being developed for the analysis/evaluation of the large datasets obtained when using this technology. In a previous study[1] we assessed the use of COPAS parameters TOF and EXT to gauge growth and development in nematodes exposed to compounds of known toxicity in axenic liquid culture and proposed statistical methods to analyze the large COPAS generated datasets. The objectives of the present investigation are to propose more sophisticated statistical approaches for the analysis of COPAS data and assess the usefulness of our approach for use in future toxicity testing. We will judge the feasibility of using a neural network model to rank water soluble compounds in terms of their acute toxicity (as measured

Correspondence: Martine Ferguson
5100 Paint Branch Parkway, Room 2D022, College Park, MD20740, USA
Tel +1 301 436 2087
Fax +1 301 210 4699
Email martine.ferguson@fda.hhs.gov

139

by their rat oral LD-50s) using the COPAS parameters TOF and EXT. Artificial neural networks were inspired by the design of the biological neural network and are supervised learning techniques commonly used for pattern recognition or data classification. Our intent is to develop a neural network model that will predict relative toxicity of test compounds in the scheme of select training compounds (of known toxicity) previously tested using axenic liquid culture, *Caenorhabditis elegans*, and the COPAS parameters TOF and EXT.

Materials and methods
Test animals (nematode culture)

Bristol N2 (wild-type) worms utilized for dosing were obtained from our axenically raised nematode colony. Adult worms were maintained at 22 + 1°C in 250 mL BD Falcon Tissue Culture Flasks with canted necks and vented tops (BD Falcon, Two Oak Park, Bedford, MA) containing 30 mL of *C. elegans* habitation media (CeHM[2]). Eggs were obtained by treating gravid adults with hypochlorite.[3] Collected eggs, after several washes in sterile distilled water, were transferred into M9 salt solution where they remained until hatching. All worms utilized in the present study were maintained in M9 for 3 days prior to dosing.

Exposure period: dose selection and test compounds

Cohorts of L1 stage larvae were exposed to water soluble compounds sodium arsenite, sodium fluoride, sodium borate, caffeine, valproic acid at five different doses (300, 600, 1,200, 2,500, and 10,000 ppm) and dimethyl sulfoxide at doses 1,200, 2,500, and 10,000 ppm) for 72 h. Test compounds were selected to represent a range of toxicities. The doses utilized in this study were established based on preliminary dose range finding studies previously conducted internally (unpublished data). Controls were exposed to sterilized double distilled water and control groups were included for each compound tested. The pH range of each control (water plus CeHM) and dosing solution (compound plus CeHM) was recorded. The six test compounds and their corresponding rat oral LD-50 values are presented in Table 1.

Table 1 The LD-50 for each test compound

Compound	Rat oral LD 50 (in mg/kg)
Sodium arsenite	41
Sodium fluoride	52
Caffeine	192
Valproic acid	670
Sodium borate	4,500
Dimethyl sulfoxide	14,500

Preparation of dosing solutions and animal exposure

Methods for the preparation of the dosing solutions and animal exposure have been previously reported.[1] Briefly, the stock dosing solution was prepared at twice (2X) the desired concentration in sterile distilled water and filtered using a 0.22 μ vacuum filter assembly (Corning Non Pyrogenic, Polystyrene Vacuum Filter Assembly; Corning Life Sciences, New York, NY). Individual dosing solutions were prepared by serially diluting the stock solution with filtered distilled water. The individual dosing solutions were prepared at twice the desired concentration. CeHM culture medium is composed of three component parts. These include ultra-high temperature pasteurized milk, the media proper consisting of salts, amino acids, vitamins etc, and sterile distilled water in a ratio of 2:3:5, respectively. To prepare the final dosing solution, two parts of ultrahigh temperature pasteurized milk to three parts of the media proper and five parts of the desired dosing solution prepared at twice the desired concentration in sterile distilled water were combined in a 25 mL BD Falcon Tissue Culture Flask with canted necks and vented top (BD Falcon). This combination of milk/growth media and the dosing solution will reduce the concentration of the dosing solution by ½.

Nematodes (n = 12,000; day 3 post-hatching L1's) were transferred into the freshly prepared growth media proper containing ultrahigh temperature pasteurized milk and either the dosing or control solutions. This time point was considered 0 h post-feeding (PF). Experimental animals were exposed to their respective dosing solutions for 72 h. Controls were exposed to double distilled water for the same period of time. During this time the growth media was not replaced. The tissue culture flasks containing either control or treated worms were placed on an Innova 2000 platform shaker (Shaker setting: 62 revolutions/min; New Brunswick Scientific, Edison, NJ) in an Ambi Hi Lo Chamber incubator (Lab Line Instruments Inc, Melrose Park, IL) whose temperature was set at 22 + 1°C. Tissue culture flasks were examined after approximately 1 h of incubation to assess worm viability.

COPAS analysis
Calibration

The COPAS was calibrated using, as per manufacturers specifications, GP Control Particles (HI-Fluorescence; Union Biometrica, Holiston, MA). Briefly, 20 mL of the control particles were placed into the primary sample cup and parameters were adjusted to process the fluorescent control particles at

a rate of 5–10 events/s until 1000 events are collected. After calibrating the instrument the sample chamber was rinsed with distilled water. After the final rinse, distilled water was flushed through the sample delivery tubing before sample analysis was initiated.

Sample analysis

Nematode growth and development were monitored after 72 h of continuous incubation/exposure in CeHM. After the exposure period, treated and control worms were each collected by centrifugation (800 rpm/45 seconds), washed three times in 5 mL of M9 then transferred into the primary sample chamber of the COPAS. The sample tube was rinsed two times with M9 into the primary sample chamber (volume of worms and M9 in sample chamber = 15 mL). The primary sample chamber was then filled with 15 mL of M9 bringing the total volume of the primary sample chamber to 30 mL. Initial data observations (first 30 s after the initiation of sample analysis) were discarded in order to prevent data carry over from the previously analyzed sample. Observations (3,000–3,100 observations) were collected from each dose and control groups for each of the compounds tested. The experiment was repeated three times, resulting in three replicate cohorts. For each replicate, the treatment groups and controls were randomized prior to analysis. The primary sample chamber and the sample delivery tubing were rinsed with distilled water between the analysis of each experimental and control sample. The data collected included information on the TOF and extinction. The TOF measures the amount of time the instrument microprocessor was detained in the analysis of a signal. TOF is an indicator of the length of an object. The EXT measured the decrease in laser light when a particle or organism passed through the laser beam. Extinction is an indicator of the size and the internal structure of the object (see Sprando et al[1] for details).

pH measurements

Three pre-exposure pH measurements per replicate were made on all high and low dose and control dosing solutions for each test compound using a Corning pH/ion analyzer 350.

Statistical analysis

Standardization

TOF and EXT were standardized to their respective controls as follows: for observation i in replicate j of compound k, the standardized TOF value (TOF_{ijk}) was:

$TOF_{ijk} = TOF_{ijk} - MEAN\_TOF_{jk0}/STD\_TOF_{jk0}$ where TOF_{ijk} is the raw TOF value for observation i of replicate j of compound k, $MEAN\_TOF_{jk0}$ is the mean TOF value of the control for replicate j of compound k and STD_TOF_{jk0} is the standard deviation of the TOF values of the control for replicate j of compound k. EXT was similarly standardized. The means (and accompanying standard deviations) of the standardized TOF and EXT values of the compounds are presented in Table 2 for each concentration.

Prediction of LD-50

For each compound, the observations were randomly split into a training dataset containing 65% of the observations and a validation dataset containing the remaining 35%. This process

Table 2 The means of the TOF and EXT standardized-to-control values and their standard deviations (SD) at each concentration

Compound	Concentration	EXT mean (SD)	TOF mean (SD)
Sodium borate	0	0.000 (1)	0.000 (1)
Sodium borate	0.03	−0.027 (0.965)	−0.007 (0.978)
Sodium borate	0.06	−0.895 (0.556)	−0.659 (0.645)
Sodium borate	0.12	−1.320 (0.493)	−1.014 (0.492)
Sodium borate	0.25	−1.469 (0.424)	−1.212 (0.454)
Sodium borate	1	−1.691 (0.422)	−1.609 (0.396)
Caffeine	0	0.000 (1)	0.000 (1)
Caffeine	0.03	−0.366 (0.846)	−0.222 (0.919)
Caffeine	0.06	−0.345 (0.852)	−0.193 (0.902)
Caffeine	0.12	−0.938 (0.629)	−0.609 (0.787)
Caffeine	0.25	−1.391 (0.448)	−0.936 (0.617)
Caffeine	1	−2.234 (0.251)	−1.948 (0.287)
Dimethyl sulfoxide	0	0.000 (1)	0.000 (1)
Dimethyl sulfoxide	0.12	−0.087 (1.001)	−0.028 (1.031)
Dimethyl sulfoxide	0.25	0.084 (0.992)	0.034 (0.961)
Dimethyl sulfoxide	1	−0.433 (0.679)	−0.455 (0.652)
Sodium arsenite	0	0.000 (1)	0.000 (1)
Sodium arsenite	0.03	−2.520 (0.186)	−2.078 (0.328)
Sodium arsenite	0.06	−2.513 (0.142)	−2.045 (0.331)
Sodium arsenite	0.12	−2.552 (0.139)	−2.129 (0.334)
Sodium arsenite	0.25	−2.555 (0.131)	−2.146 (0.331)
Sodium arsenite	1	−2.581 (0.138)	−2.239 (0.334)
Sodium fluoride	0	0.000 (1)	0.000 (1)
Sodium fluoride	0.03	−1.714 (0.162)	−1.507 (0.314)
Sodium fluoride	0.06	−1.720 (0.156)	−1.495 (0.325)
Sodium fluoride	0.12	−1.743 (0.18)	−1.558 (0.363)
Sodium fluoride	0.25	−1.749 (0.151)	−1.555 (0.32)
Sodium fluoride	1	−1.791 (0.185)	−1.673 (0.328)
Valproic acid	0	0.000 (1)	0.000 (1)
Valproic acid	0.03	−0.159 (0.934)	−0.045 (0.99)
Valproic acid	0.06	−0.483 (0.805)	−0.292 (0.894)
Valproic acid	0.12	−1.046 (0.663)	−0.764 (0.814)
Valproic acid	0.25	−1.556 (0.438)	−1.310 (0.635)
Valproic acid	1	−1.969 (0.259)	−1.965 (0.309)

was repeated 50 times to obtain 50 training and validation datasets. For each training dataset the mean standardized TOF and EXT values at the six concentrations (excluding control), for each of the three replicates, were fed into a feed forward multilayer perceptron neural network with the output layer containing the ordinal LD-50 values. Linear combination functions were used in both the hidden and output layers. The hyperbolic tangent activation function was used in the hidden layer and the logistic activation function was used in the output layer. Because the target variable was ordinal, the multiple Bernoulli error function was employed. Minimization of the misclassification rate, via Levenberg–Marquardt Optimization, was the criterion used to select the model. To avoid over-fitting, the neural network models were cross-validated by classifying the compounds in each validation set using the neural network trained from the corresponding training set. Any missing values (including the values for dimethyl sulfoxide at 300 and 600 ppm) were interpolated with cubic splines using the remaining non missing portion of the data vector prior to input to the neural network. The neural network routines were implemented using the Neural Network Node of SAS® Enterprise Miner™ 4.3 (Copyright© 1999–2003; SAS Institute Inc, Cary, NC).[4]

Results

pH measurements

The mean pH from each of the six control groups (range: 6.14 ± 0.004 to 6.19 ± 0.002) was not significantly different from the mean pH of the culture media for the high (range: 6.12 ± 0.004 to 7.76 ± 0.001) and low dose groups (range: 6.14 ± 0.003 to 6.32 ± 0.002) for each of the compounds tested. The mean pH calculated for the high borax dose group (7.76 ± 0.001) was slightly higher than the mean pH for the control group (6.14 ± 0.004) however the mean pH of the high dose borax treatment group was close to neutral. These measurements excluded marked changes in pH as possible confounding variables in producing the effects on growth and development observed in the present study.

Prediction of LD-50

Multiple neural network architectures were tested with the eventual choice balancing out the simplicity of the chosen network with the relatively low misclassification rate of the validation datasets. The ultimate neural network had one hidden layer with one node containing three units. Please refer to Figure 1 for a pictorial description of this ultimate neural network with equations specified at each layer. We do not propose that this architecture is optimal for all future

endeavors. We rather suggest that the researcher use similar cross-validation approaches to obtain the architecture that best fits their data.
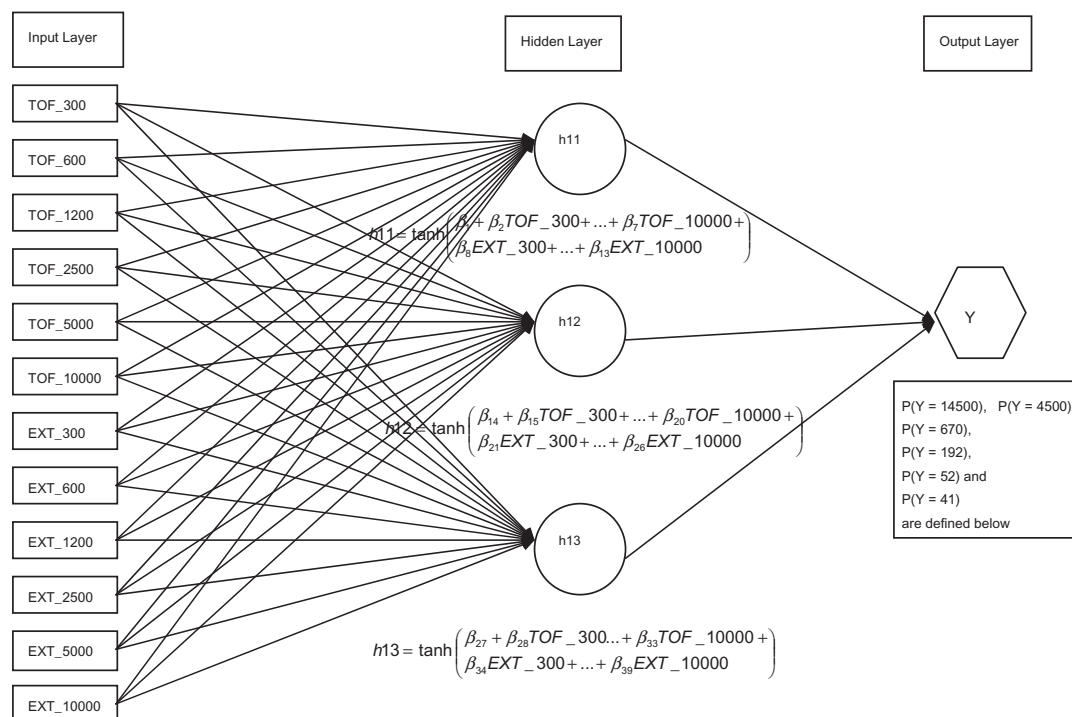
The mean misclassification rates averaged across all fifty validation sets for each compound were as follows: 12% for caffeine, 9.3% for valproic acid, 2.7% for sodium borate, and 0% for dimethyl sulfoxide, sodium arsenite, and sodium fluoride. Averaged across all six compounds, the mean misclassification rate was therefore 4.0%.

Examining the misclassifications more closely, we found that: 1) sodium borate was misclassified as more toxic than its LD-50 indicates for all its misclassifications with a predicted LD-50 of 670; 2) valproic acid was misclassified as more toxic than its LD-50 indicates with a predicted LD-50 of 192 for 21% of its misclassifications and less toxic, with a predicted LD-50 of 4500, for the remaining misclassifications; 3) caffeine was misclassified as more toxic than its LD-50 indicates with a predicted LD-50 of 52 for 61% of its misclassifications and less toxic with a predicted LD-50 of 670 for the remaining 39% of the misclassifications.

Once the neural network is properly trained and the architecture that best fits both the training and validation data is obtained, that same network can be used to determine the LD-50s of future compounds. The compounds dimethyl sulfoxide and caffeine were run in duplicate, allowing us to set aside these duplicate runs to mimic this process and further validate the prediction abilities of our neural network. The parameters for our network with one hidden layer containing one 3-unit node were re-estimated on the full dataset comprising all the training compounds. Recalling that all compounds were run in triplicates, dimethyl sulfoxide's LD-50 was correctly predicted for all three replicates. The LD-50 for caffeine was correctly predicted for two of its replicates, with the LD-50 for the third replicate predicted to be 670.

Discussion

Neural networks are data mining techniques that can be taught to identify complex relationships in relatively large datasets. "Their advantage over classical statistical models used to analyze data, such as regression and ANOVA, is that they can fit data where the relationship between independent and dependent variables is nonlinear and where the specific form of the nonlinear relationship is unknown".[5] In other words, they are very effective in modeling highly nonlinear patterns. Since our main goal was pattern recognition and prediction and not explicitly identifying the exact nature of the relationship between the LD-50 and TOF and EXT, artificial neural networks were selected for data analysis.

142

Open Access Bioinformatics 2010:2

Input Layer | Hidden Layer | Output Layer

TOF_300
TOF_600
TOF_1200
TOF_2500
TOF_5000
TOF_10000
EXT_300
EXT_600
EXT_1200
EXT_2500
EXT_5000
EXT_10000

$h11 = \tanh\left(\beta_1 + \beta_2 TOF\_300 + ... + \beta_7 TOF\_10000 + \beta_8 EXT\_300 + ... + \beta_{13} EXT\_10000\right)$

$h12 = \tanh\left(\begin{matrix}\beta_{14} + \beta_{15} TOF\_300 + ... + \beta_{20} TOF\_10000 + \\ \beta_{21} EXT\_300 + ... + \beta_{26} EXT\_10000\end{matrix}\right)$

$h13 = \tanh\left(\begin{matrix}\beta_{27} + \beta_{28} TOF\_300... + \beta_{33} TOF\_10000 + \\ \beta_{34} EXT\_300 + ... + \beta_{39} EXT\_10000\end{matrix}\right)$

Y

P(Y = 14500),   P(Y = 4500),
P(Y = 670),
P(Y = 192),
P(Y = 52) and
P(Y = 41)
are defined below

The probabilities of the output layer (Y) are:

$$P1 = \frac{1}{1 + \exp\left(-\left(\beta_{40}h11 + \beta_{41}h12 + \beta_{42}h13 + \beta_{43}\right)\right)}$$

$$P2 = \frac{1}{1 + \exp\left(-\left(\beta_{40}h11 + \beta_{41}h12 + \beta_{42}h13 + \beta_{44}\right)\right)}$$

$$P3 = \frac{1}{1 + \exp\left(-\left(\beta_{40}h11 + \beta_{41}h12 + \beta_{42}h13 + \beta_{45}\right)\right)}$$

$$P4 = \frac{1}{1 + \exp\left(-\left(\beta_{40}h11 + \beta_{41}h12 + \beta_{42}h13 + \beta_{46}\right)\right)}$$

$$P5 = \frac{1}{1 + \exp\left(-\left(\beta_{40}h11 + \beta_{41}h12 + \beta_{42}h13 + \beta_{47}\right)\right)}$$

The probabilities of observing the various LD-50s are then defined as follows:

P(Y = 41) = 1 - P5
P(Y = 52) = P5 - P4
P(Y = 192) = P4 - P3
P(Y = 670) = P3 - P2
P(Y = 4500) = P2 - P1
P(Y = 14500) = P1

Y is then predicted to be the LD-50 which corresponds to the maximum of these probabilities.

The network was trained on each of the 50 training datasets, to obtain a set of 50 parametes $\vec{\beta}$ estimate vectors, which were then used to predict the Y for the 50 validation datasets, concluding the cross-validation cycle.

Figure 1 Architecture of the neural network ultimately chosen: For each of the training datasets, the inputs into the neural network were the EXT and TOF mean standardized values for each compound at the 6 different concentrations excluding control (the first concentration of 0 ppm was the standardization control). For each compound, the inputs therefore were: EXT_C, the mean of the (standardized-to-control) EXT values at the Cth concentration, and TOF_C, the mean of the (standardized-to-control) TOF values at the Cth concentration, where C = 300, 600, 1,200, 2,500, 5,000, and 10,000 ppm.

Multilayer perceptron neural networks were furthermore chosen for their ability to cope with collinearity. Both TOF and EXT are growth endpoints and are therefore highly correlated. Moreover, as expected, the mean TOF and EXT values exhibit a high degree of autocorrelation, with the correlation decreasing as the difference between the doses increases. Due to the high multicollinearity exhibited by the data vectors, multilayer perceptron neural networks were chosen since they are well equipped to handle collinear data. As Omdivar and Elliott[6] point out, the neural network model derived from collinear data is valid only when the correlation holds. Therefore it is important that the approximate structure and degree of correlation be preserved in any validation and test datasets that the neural network will be tested on. We explored lessening the degree of collinearity by reducing the input data vector to a small number

of principal components but found that neural networks run on these principal components did not have improved predictive abilities.

In the present study[1] we have demonstrated how a properly trained neural network may be used to rank order compounds with unknown LD-50s in the scheme of compounds with known LD-50s. It is our hope that with further exploration of this method we may be able to develop a method to rank order compounds of unknown toxicity and in so doing refine, replace, and reduce animal use in toxicity testing.

In a previous study[1] we utilized a multivariate analysis of variance which was run for each replicate for each control to non-control comparison to jointly model the normalized TOF and EXT variables on concentration (control or non-control) and the *P*-values from Hotelling–Lawley's trace F-test were computed on subsamples of nematodes. The percentage of significant *P*-values was calculated for each dose from each replicate and a mean significant *P*-value ratio was calculated for each dose group. This mean *P*-value ratio was then used as a measure of preponderance of evidence against the null hypothesis by determining the lowest concentration at which this ratio was relatively "large". In contrast to the methods in the previous study, the current method is more suited to the inherent nature and size of the data and keeps the focus of the study on classifying the compounds into their respective rat LD-50 values. The nonlinearity of the data is handled seamlessly by the neural network and the high multicollinearity present in the data is efficiently accommodated by the multilayer perceptron based network. Since the amount of data is so vast, some sort of *P*-value scaling or adjustment, as with the ratio method previously utilized, would almost certainly be necessary with hypothesis testing. For this reason we bypassed the hypothesis testing method and approached the problem from a neural-network based prediction perspective.

In conclusion, more training compounds should be evaluated representing different classes of compounds using this modeling system as the more compounds tested the more precisely the model will be able to predict relative toxicity.

If a sufficient number of LD-50s is tested, then the LD-50 target response may better be viewed as a continuous variable and could be predicted using regression for continuous, instead of ordinal, dependent variables. The rat oral LD-50 was chosen as the indicator of acute toxicity because it is widely available and universally considered the standard for expressing and comparing chemical toxicity. Although the testing of this method against other measures of acute toxicity is beyond the scope of this paper, there is no reason this method cannot be extended to approximate other measures of acute toxicity in lieu of rat LD-50s. It would be interesting, for future studies, to measure the ability of the model to approximate other toxicity measures such as those examined in Rosenkranz and Cunningham[7] including minnow lethality, rat maximum tolerated dose, and mouse maximum tolerated dose. Rosenkranz and Cunningham[7] conclude that the relationship of in vitro cell toxicity data to endpoints more relevant to human health than the rat LD-50 deserves investigation.

Disclosure

No conflicts of interest were declared in relation to this paper.

References

1. Sprando R, Olejnik N, Cinar H, Ferguson M. A method to rank order water soluble compounds according to their toxicity using Caenorhabditis elegans, a Complex Object Parametric Analyzer and Sorter, and axenic liquid media. *Food Chem Toxicol*. 2009;47: 722–728.
2. Nass R, Hamza I. The nematode C. elegans as a model to explore toxicology in vivo: solid and axenic growth culture conditions and compound exposure parameters. *Curr Protocols Toxicol*. 2007: 1.9.1–1.9.18.
3. Sulston J, Hodgkin J. Methods. In: The Nematode *Caenorhabditis elegans*. Wood WB, editor and Cold Spring Harbor, NY. 1988. Cold Spring Harbor Laboratory Press; p. 587–606.
4. SAS 9.1.3 Help and documentation. Cary, NC: SAS Institute Inc; 2002–2004.
5. Francis L. Neural networks demystified, casualty actuarial society forum. 2001; p. 253–320.
6. Omidvar O, Elliott D. *Neural Systems for Control*. 1st ed. 1997. Morgan Kaufmann Publishers, Inc. San Francisco, CA. p. 224.
7. Rosenkranz HR, Cunningham AR. Lack of predictivity of the rat lethality (LD50) test for ecological and human health effects. *Alternatives Lab Animals*. 2005;33:9–19.