DATA MINING POTENCY ESTIMATORS FROM TOXICOLOGICAL DATABASES

Piegorsch, Walter W.
Department of Statistics, University of South Carolina

Simmons, Susan J.
Department of Mathematics and Statistics, University of North Carolina at Wilmington

Zeiger, Errol
Errol Zeiger Consulting

https://doi.org/10.5109/12578

出版情報：Bulletin of informatics and cybernetics. 36, pp.51-62, 2004-12. Research Association of Statistical Sciences
バージョン：
権利関係：

DATA MINING POTENCY ESTIMATORS FROM TOXICOLOGICAL DATABASES

by

Walter W. Piegorsch, **Susan J. Simmons** and **Errol Zeiger**

Reprinted from the Bulletin of Informatics and Cybernetics
Research Association of Statistical Sciences, Vol.36

FUKUOKA, JAPAN
2004

Bulletin of Informatics and Cybernetics, Vol. 36, 2004

DATA MINING POTENCY ESTIMATORS FROM TOXICOLOGICAL DATABASES

By

Walter W. Piegorsch[*] **Susan J. Simmons**[†] and **Errol Zeiger**[‡]

Abstract

We discuss use of data mining techniques to study estimators of toxic potency (activity) in toxicological databases. The methods are slight variations on the standard data mining motif, but fit fully within the larger context of knowledge discovery in databases. An example illustrates the general theme of the approach, using results from a U.S. National Toxicology Program *Salmonella* mutagenicity database.

Key Words and Phrases: Ames test, bioinformatics, feature extraction, knowledge discovery in databases, mutagenic potency, *Salmonella* assay.

1. Introduction

A major goal in modern toxicology is identification of the damaging effects from exposure to potentially hazardous agents. Bioassays on small mammals or other biological systems, including in vitro systems, are typically used. A critical quantitative component of such studies is statistical characterization of the dose-response to the agent, and whether this response is indicative of a significant toxic effect in the system under study. Summary measures describing the potency of the agent's toxicological activity are of interest, and the development of such measures is an area of active statistical research (Bailer and Piegorsch, 2000).

In recent years, rapid technological and scientific developments have enhanced our ability to generate large amounts of toxicological data on a variety of endpoints, even when a laboratory's resources are limited. As a result, a wealth of archived toxicity data has emerged; see, e.g., Gold and Zeiger (1996), Haseman (2000), Keenan *et al.* (2002), Mattingly *et al.* (2003), or the online collections referred to in Helma *et al.* (2000), http://helma.informatik.uni-freiburg.de/db_links/, and Schmidt (2003), see http://ehp.niehs.nih.gov/txg/docs/2003/111-6/focus-abs.html. Databases such as these represent excellent sources for studying the statistical characteristics of newly-developed (and some older) estimators of toxic potency. In effect, this represents a form of *data mining*, where information buried in large collections of data is quarried in a systematic fashion to uncover data characteristics - here, the statistical attributes of a potency estimator - that can lead to new knowledge discovery (Hand *et al.*, 2000).

[*] Department of Statistics, University of South Carolina, Columbia, SC 29208, USA. piegorsc@stat.sc.edu

[†] Department of Mathematics and Statistics, University of North Carolina at Wilmington, Wilmington, NC 28403, USA.

[‡] Errol Zeiger Consulting, 800 Indian Springs Rd., Chapel Hill, NC 27514

As used in toxicology, data mining is often directed at identifying/manipulating quantitative structure-activity relationships (QSARs) of hazardous chemicals (Wang *et al.*, 1999), and there are strong overlaps between the use of data mining for potency estimation and that for QSAR identification (Klopman and Rosenkranz, 1991; Cunningham *et al.*, 1998). One can also use toxicological databases to study quantitative activity-activity relationships (QAARs), where, say, the mutagenic activity of a set of hazardous chemicals is compared for predictive purposes against their carcinogenic activity (Piegorsch and Hoel, 1988; Fetterman *et al.*, 1997). See the review by Helma *et al.* (2000) for more on QSAR/QAAR knowledge discovery in toxicology. Data mining is also growing in allied applications such as pharmacology and drug discovery (Abt *et al.*, 2001), ecotoxicology (Yuan *et al.*, 2000), and in the burgeoning studies of gene expression (Creighton and Hanash, 2003), microarrays (Zweiger, 1999; Valafar, 2002), and toxicogenomics (Aardema and MacGregor, 2002).

In this paper, we discuss bioinformatic methods for applying data mining to identify statistical features of toxicological potency estimators. We start in Section 2 with a short review of potency estimation, then move in Section 3 to how basic principles of data mining can be applied to understand the statistical characteristics of such estimators. Section 4 ends with a short example using mutagenic potency in the bacterium *Salmonella typhimurium*.

2. Toxicological potency estimation

When quantifying a hazardous agent's toxic effects, one often generates data that describe the response of the test organisms to various exposure doses of the agent, called a *dose-response experiment*. From these data, statistical estimators of potency represent summary measures of the observed dose response. When no significant response is evidenced, the potency is set equal to some boundary value that indicates no potent response. (Zero is common, although other boundary values are possible, depending on the measure used to describe the toxic activity.) For data exhibiting significant responses to the agent, some quantity departing from this boundary value is calculated.

Statistical issues dealing with potency estimation have a long history, ranging back to bioassays of pesticides, herbicides, and other poisons (Bliss, 1935; Finney, 1952). The simplest concern is how to measure or define the potency in an unambiguous manner: a number of quantitative solutions exists to answer this question, each based on a slightly different feature of the observed dose response. We consider a few of the more general forms in this section, taken in large part from the discussion in Piegorsch and Bailer (1997, §7.3).

2.1. Median Effective Dose

A basic measure used to summarize dose response is the *median effective dose*, denoted as ED_{50}. If the environmental agent is given as a concentration, we use the term *median effective concentration*, or EC_{50}. In either case, this is defined as the amount of agent required to produce a response at the median (50%) level (Trevan, 1927). For example, suppose we measure quantal response data as the proportion of times organisms respond to a toxic stimulus. Then, since the proportions are bounded between 0 and 1, ED_{50} is the dose at which the response equals 0.50. Note that ED_{50} is an inverse measure of potency: a higher ED_{50} suggests a weaker agent, since a greater

dose is required to produce the same median result.

Obviously, the ED_{50} depends upon the outcome variable under study, and the terminology has developed to indicate this. For instance, when simple lethality is the outcome the ED_{50} is more specifically the *median lethal dose*, or LD_{50}.

Traditionally, ED_{50} has been used as a summary measure of the dose effect due to its central location on the dose-response curve. Many other response levels are possible, however, such as ED_{01}, ED_{10}, ED_{90}, etc. Depending on the application, any of these may be appropriate for summary use.

2.2. NOAEL/LOAEL and Regression-Slope Potency Measures

Many other measures have been proposed to describe the potency of a toxic agent. Perhaps the simplest ones are known as observed effect levels. Specifically, the no-observed-adverse-effect level (NOAEL), is the largest level of dose where no adverse effect is observed. If the stimulus is recorded as a concentration, we often write NOAEC. By extension, the lowest-observed-adverse-effect level (LOAEL) is the lowest level of dose where a significant adverse effect is observed. If a significant increase over background occurs at the lowest dose, then the NOAEL will not exist, while the LOAEL will correspond to that lowest tested dose level. For both NOAEL and LOAEL, estimates are determined statistically by performing a series of pairwise comparisons at each non-zero dose level with the response at control or background levels. Clearly, correction for multiple testing is required, using some form of multiplicity adjustment (Yanagawa *et al.*, 1994). Numerous concerns have been raised regarding the NOAEL/LOAEL approach, since it suffers from a variety of instabilities. For instance, a NOAEL or LOAEL must by definition correspond to the actual levels of the agent under study. Thus, spacing of the dose levels is critical to these summary values. Dose selection must be performed carefully, and include a dense enough grid of dose levels to derive a LOAEL and/or NOAEL with sufficient accuracy. If the experiment is badly designed or poorly controlled, excessively high variability can result. This typically raises the values of NOAEL and LOAEL so that high variability in the response will translate into assertions of low potency (high NOAELs), regardless of the true response effect. Also, a design that allocates too few experimental units to the dose levels may overestimate or even fail to identify a NOAEL or LOAEL. Indeed, it is unclear which parameter(s) of the dose-response relationship these quantities actually estimate. These difficulties have led to concerns that observed-effect levels such as NOAEL are poor summary statistics for dose-response data (Chapman *et al.*, 1996; Yanagawa and Kikuchi, 2001). One feature that both the ED50 and the NOAEL share is that they are increasing measures of potency; i.e., an agent with a smaller ED50 or NOAEL is more potent than an agent with a larger ED50 or NOAEL. Some authors find this sort of "inverse potency" to be counterintuitive, calling instead for transformations that recover direct proportionality, or to development of new, directly proportional measures. For instance, suppose the form of the dose-response function is known and that an increasing response indicates a detrimental effect. Since in this case a sharper increase indicates a more potent agent, a direct measure of potency is the rate of increase or slope of the dose-response curve. Crawford and Wilson (1996) argue that the incremental potency of the agent is then equivalent to the slope or tangent line of the dose-response curve. Unfortunately, for most nonlinear models the slope itself varies and so some specific dose must be chosen at which to measure the curve's slope. A common solution is to use the incremental

rate of change in the dose response just past $d = 0$. That is, given a dose-response function $\mu(d)$, define the potency of an environmental agent as the dose-response slope at $d = 0$: this is typically the first derivative at zero, $\mu'(0)$. This 'slope-at-zero' measure is motivated from low-dose estimation arguments: human exposures to a hazardous agent often approximate the lower dose levels of a laboratory assay, so dose-response behavior observed at low doses in the animals may better mimic human responses to the agent. As such, incremental change in dose response is of greatest interest near $d = 0$ (Bogen, 1995).

2.3. DUI

For curvilinear dose-response models, Margolin and Risko (1988) defined a potency estimator in terms of the dose needed to increase the response by one 'unit.' Called the *Dose per Unit Increase* (or *DUI*), their approach was similar in spirit to finding a slope estimate from a nonlinear model. Assume that the dose-response curve can be well-approximated by a second-order polynomial of the form $\beta_0 + \beta_1 d + \beta_2 d^2$, and that the variances of the observations are constant. Then, point estimates, b_j, of each β_j are determined using ordinary least squares. The dose that incurs a single unit increase over the response at $d = 0$ is estimated by solving for d in the relationship

$$b_0 + 1 = b_0 + b_1 d + b_2 d^2$$

and setting the result equal to the DUI. From the quadratic formula, this is the smallest positive solution of $\{-b_1 \pm \sqrt{b_1^2 + 4b_2}\}/2b_2$. When $(b_1^2 + 4b_2) < 0$, DUI is set to infinity. If $b_2 = 0$ and $b_1 > 0$, then DUI simplifies to $1/b_1$, while if $b_2 = 0$ and $b_1 < 0$, or if $b_2 < 0$ and $b_1 < 0$, there is no positive solution and the DUI is again set to infinity. For all other cases, the solution for DUI results in the unique estimator

$$\text{DUI} = \frac{-b_1 + \sqrt{b_1^2 + 4b_2}}{2b_1}. \tag{1}$$

This can be shown as follows, using the case $b_1 > 0$ and $b_2 > 0$ as a guide. If $b_1 > 0$ and $b_2 > 0$, the quadratic equation that produces the solution has exactly one sign change in its coefficients ($b_2 > 0$, $b_1 > 0$, and $-1 < 0$). From Descartes' famous Rule of Signs (Descartes, 1637; see Borwein and Erdélyi, 1995, §3.2), the number of positive real roots must then be no larger than one. But since $b_1 > 0$ and $b_2 > 0$, the numerator in (1) and hence the entire expression is clearly positive. Since (1) is also clearly larger than its competitor $-b_1 - \sqrt{b_1^2 + 4b_2}/2b_2$ when $b_1 > 0$, it must therefore be a unique positive real root. As such, we employ it as the DUI. Similar sorts of arguments show that (1) is the unique positive real root for $b_1 > 0$ and $b_2 < 0$ and for $b_1 < 0$ and $b_2 > 0$. Notice again that DUI is an inverse potency measure. To recover direct proportionality, Simmons (2002) suggests use of measures such as $1/\text{DUI}$ or $1/\log(1 + \text{DUI})$.

2.4. Maximum Pairwise Slope

For cases where specification of a particular form for the dose-response function is undesirable, Fetterman *et al.* (1997) and Margolin *et al.* (1994) proposed use of a potency estimator based on the simple two-point slopes between the control-level/background data and the various exposed-level data. That is, suppose the toxic outcomes are ob-

served to have sample means $\overline{U}_k$ at doses d_k $(k = 0, \ldots, K)$, and that the control/background-level dose is d_0. Then, the pairwise slope from the origin at each non-control dose is $(\overline{U}_k - \overline{U}_0)/(d_k - d_0)$. From these, calculate the maximum observed pairwise slope

$$b_x = \max_{k=2,\ldots,K} \left\{ \frac{\overline{U}_k - \overline{U}_0}{d_k - d_0} \right\}$$

and use this as an estimate of potency. (Here again, this assumes that increasing response is indicative of greater potency; if instead decreasing response indicates higher potency then use the minimum pairwise slope.)

3. Data mining potency estimators

For the bioinformatic exercise we have in mind, the overall goal is one of building new models or validating existing forms that describe the statistical nature of a potency estimator in a specific context. Efforts to build such descriptive models are a common theme in data mining (Hand *et al.*, 2001, Ch. 9). Crucial to our approach is the use of *feature extraction* to select and efficiently analyze pertinent information from the database, and to gain a better understanding of the potency estimator(s) (Hand *et al.*, 2001, §6.6). In addition, we can use these extracted features to generalize to new data that may arise where potency calculations/inferences are required.

At the core of any data mining algorithm is a set of four basic components: (1) determination of model structures and other important data patterns, (2) quantifying the quality of a fitted model via a pre-defined 'score' function, (3) search/optimization using the score function, and (4) large-scale data management (Hand *et al.*, 2001, §1.5). For examining the statistical characteristics of potency estimators we can apply these concepts in various ways, depending on the specific database and the specific estimator(s) being studied. For example, when using a potency estimator to make statistical inferences, the statistical distribution of the estimator must be well understood. Typically, we assume that the estimator has some form of normal distribution, at least asymptotically, but this is not guaranteed. (Indeed, many of the forms described in §2 can exhibit decidedly non-normal variation, even in large samples.) Thus a common problem with many potency estimators is proper determination of their distributional features. We can use large toxicological databases to study these features and determine if a particular potency estimator or some natural transformation thereof exhibits some desired distributional characteristic.

Within the context of the four data mining components mentioned above, this sort of exercise can be viewed as (1) descriptive model determination, using (2) a measure to score an estimator's departure from normality, via (3) a search strategy that eliminates those estimators whose scores indicate strong departure, by (4) managing and manipulating the database to facilitate this operation. For example, suppose a toxicity database exists on a series of N chemical agents, and it is of interest to study the characteristics of $Q \geq 1$ putative potency estimators. Suppose also that the database already contains summary QSAR values for each chemical, taken from a known source such as Klopman's (1985) Computer-automated Structure Evaluation (CASE) technique. Viewing these summary values as pairs $(x_i, y_{qi}) = (\mathrm{QSAR}_i, \mathrm{potency}_{qi})$, $q = 1, \ldots, Q$, $i = 1, \ldots, N$, we might fit some form of linear or nonlinear regression model to each qth potency via least squares. From this fit, we find the residuals $e_{qi} = y_{qi} - \hat{y}_{qi}$, and study them for

pertinent patterns of interest. In particular, the goal might be to determine if any of the potency estimators appear approximately normal within the context of the fitted regression model. If so, we could score each estimator via a P-value from a goodness-of-fit test of normality on the e_{qi}s. Potency estimators with small P-values would be screened out from further consideration. Or, if resources permit interactive study of the residuals, graphical diagnostics such as histograms, kernel density plots, or normal probability plots (Hand *et al.*, 2001, §3.3; Fernandez, 2002, §3.2.2) could be considered. [This can be viewed as a form of *retrieval by content*; cf. Hand *et al.* (2001, §14.1).] The overall result would be a list of potential potency estimators that appear to satisfy the desired feature of approximate normality.

One could of course imagine multiple variations on this theme; for instance, if the chemical agents spanned a wide variety of different chemical classes, as indicated by the QSAR data or by some other outside source, we might stratify our operations across those chemical classes (Benigni and Richard, 1996). Or, if the QSAR information were not available, we could still compute the Q competing potency estimators for each chemical agent to gain an indication of their ranges and scales of variation. Using these, we might perform a sensitivity analysis via computer, using Monte Carlo methods to study how each estimator performs under specific model conditions and identifying pertinent features such as instabilities in calculation, distributional characteristics of the estimator(s), sensitivity to outliers, etc. We give in the next section a short example of such a study, using a database of mutagenic outcomes in the bacterium *Salmonella typhimurium*.

Alert readers may note that the effort to collect together large amounts of disparate information on a single targeted endpoint, and then to employ that information in making a combined inference about the endpoint, shares many features with what is known as meta-analysis in the biomedical and social sciences (Sutton *et al.*, 2000; Hunter and Schmidt, 2004). To our knowledge, however, little has appeared on ways to formally coordinate a meta-analysis within a data mining framework. [Two exceptions: Kübler and Weihrauch (2002) describe a novel meta-analysis for managing the large amount of mined information on patient drug use in a clinical trial, while Sohn (1999) discusses a meta-analytic model for classifying pattern recognition algorithms.] Of course, not every data mining and knowledge discovery exercise is designed for making formal statistical inferences, and so the advanced features of a meta-analysis may not be required. For cases when they might be warranted, however, the ability to connect features that the two methodological constructs share may enhance the overall data analytic effort; efforts to develop such enhancements may be a fruitful avenue of future research.

4. Example: mutagenic potency estimation in *Salmonella typhimurium*

To illustrate the approach we outline above, consider the problem of estimating toxic potency from the perspective of mutation induction. Mutagenicity databases often provide abundant sources of data, and exploration of them can yield important insights into fundamental mechanisms of genotoxic response (Liu *et al.*, 1996; Bacha *et al.*, 2002). For instance, the U.S. National Toxicology Program (NTP) routinely conducts mutagenesis assays of environmental chemicals in strains of the bacterium *Salmonella typhimurium*. This microbe has been engineered to allow for rapid, sensitive mutagenicity testing (Maron and Ames, 1983), and the *Salmonella* assay - also known as the *Ames test* - is a mainstay of most national and international regulatory genetic toxicity testing

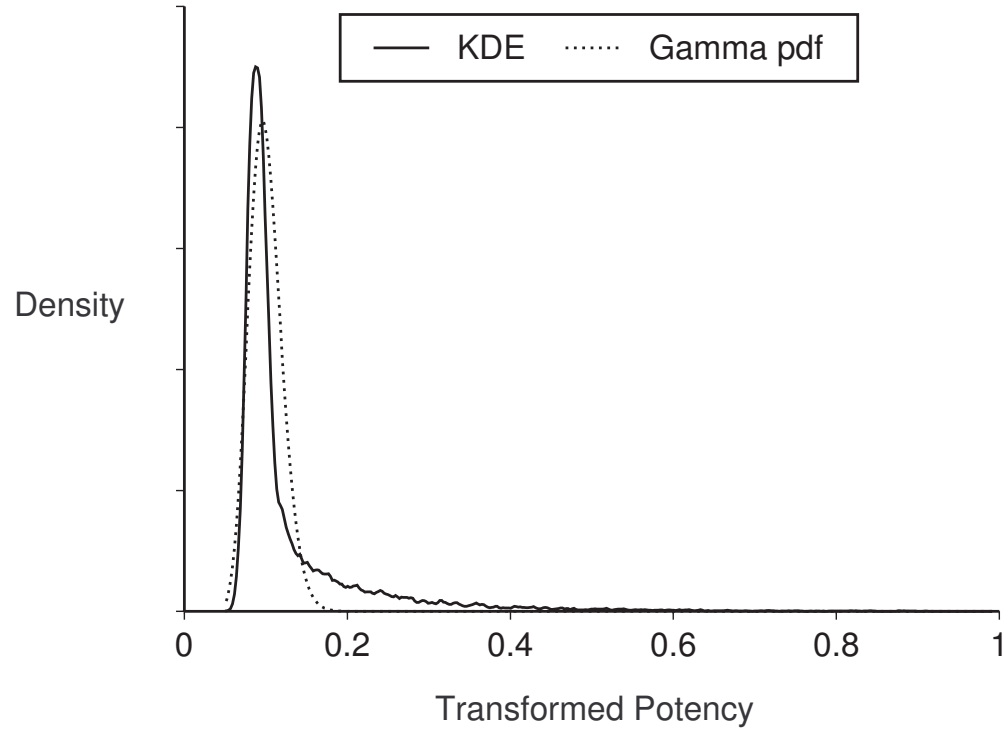
programs. As a result, a large *Salmonella* mutagenicity database has developed (Zeiger, 1996), on which a wide variety of data mining exercises may be performed.

The *Salmonella* assay generates counts of mutant colonies at exposure doses, d, of the test agent. Observing an increase in the counts over increasing dose indicates mutagenic activity of the agent, for which a variety of potency estimators have been proposed. These range in complexity from some of the simpler methods described above in §2, to complex estimators based on nonlinear regression models for the assay's dose response (Margolin *et al.*, 1981; Bernstein *et al.*, 1982; Snee and Irr, 1984). Each brings a different quality or characteristic to the potency estimation problem. Fetterman *et al.* (1997) describe many useful forms; however, to our knowledge no in-depth study has ever been performed of their distributional characteristics. We chose to undertake such a study by mining the NTP Salmonella database. For simplicity, we operated with a subset of the database, comprising the most recent (as of January 2002) assay results on 4,378 experiments representing 184 different chemicals. This included a subset of 115 chemicals also studied by Fetterman *et al.* in their 1997 article. Preprocessing was required to remove data entries that had only control group data (from which no potency estimator could be calculated), replicated experimental entries, or entries that contained contaminated data. (The contamination was due to external sources of toxicity that invalidated the observed assay count. This occurred during the original experiment, not during the data entry or accumulation phases.) In total, we worked with a final subset of 2,356 different experiments. Further details on the database is given by Simmons (2002).

The potency estimators were applied to each of the 2,356 experiments in order to determine if they exhibited any computational instability - some of the estimators require iterative or recursive calculation, and they do not guarantee a final potency for every data set - and to gauge their range of possible values and variation. Of those we found to exhibit the greatest computational stability four stood out, producing valid potencies in over 99% of the database's experiments. These were: (i) the maximum pairwise slope from §2.4, (ii) the DUI estimator from §2.3 (*Salmonella* assay data can be affected by competing, nonmutagenic, toxic effects that lead to an unusual, nonmonotone dose response; this can in some cases be approximated by a parabola, roughly validating the DUI assumptions), (iii) a modified DUI, $1/log(1 + \text{DUI})$, that allows for directly proportional potencies, and (iv) a maximum quasi-likelihood (MQL) estimator from fitting the nonlinear model $\mu(d) = (\beta_0+\beta_1 d)\exp\{-\beta_2 d^2\}$ to the observed mutant counts, where $\mu(d)$ is the mean mutagenic response at dose d (Krewski *et al.*, 1993; Leroux and Krewski, 1993). In the latter case, MQL estimation also requires specification of a variance function; following Leroux and Krewski (1993) we used a variance function proportional to $[\mu(d)]^{3/2}$. Then, appealing to the regression-slope potency argument in §2.2 we find $\mu'(0) = \beta_1$, so the MQL estimate b1 was taken as the potency estimator.

Via examination of the experimental database we determined that the various dose-response patterns in the NTP *Salmonella* database could be represented by roughly six different forms: a flat form (no dose response), a roughly linear increase, and four non-monotone forms that increase and then decrease as d gets large [sharp increase/sustained decrease, moderate increase/moderate decrease, sharp increase/moderate decrease, and moderate increase/sharp decrease; these are described in more detail by Simmons (2002, Fig. 5.1).] We then constructed six statistical models to represent each of these patterns, based on a biomechanistic formulation for *Salmonella* response data proposed by

Margolin *et al.* (1981). We used these models as the basis of a Monte Carlo simulation study to examine the distributional characteristics of the four potency estimators listed above. Full details of our Monte Carlo calculations are given by Simmons (2002); here, for expediency's sake, we simply summarize the results.



We found that of the four stable potency estimators, all deviated from normality, generally skewing far to the right. This was a particularly interesting observation with the MQL estimator, since in large samples it should converge to a normal distribution. Sample sizes in *Salmonella* experiments range typically between $n = 15$ and $n = 30$ Petri dishes that each provide an observed count, and we infer from this result that such values are not large enough to validate the asymptotics with this MQL model.

We also studied a variety of transformations of the four stable estimators, and again found that none could consistently be manipulated to induce normal distribution characteristics. We did find, however, that the logistic-type transformation $e^{10b_1}/(10+e^{10b_1})$ appeared to yield a very stable pattern of variation across all models: the variation was still heavily right-skewed, but also well-approximated with a gamma distribution. Figure 1 shows a typical kernel density estimate (here, under the sharp increase/sustained decrease model, which was one of the more common structures seen in the database) for the empirical distribution of e10b1/(10 + e10b1) found via our Monte Carlo study, overlaid on a gamma density. [The density estimate was constructed using a Gaussian kernel (Hand *et al.*, 2001, §3.3) in the SAS procedure PROC KDE (SAS Institute Inc., 2000).] From these results, we can report that certain potency estimators for *Salmonella* mutagenicity experiments can exhibit stable statistical characteristics. More generally, we find that use of even simple data mining practices such as those we undertook herein

can lead to important knowledge discovery in toxicological databases. We believe that with proper implementation they can improve the use of potency estimators in a wide variety of toxicity studies, and we strongly recommend their further development and use.

Acknowledgement

Thanks are due Dr. Brian G. Leroux for his support with the MQL estimator, and to an anonymous reviewer for helpful comments on the manuscript. This work was funded in part under grant #R01-CA76031 from the U.S. National Cancer Institute. Its contents are solely the responsibility of the authors and do not necessarily reflect the official views of the National Cancer Institute

References

Aardema, M. J. and MacGregor, J. T. (2002). Toxicology and genetic toxicology in the new era of "toxicogenomics:" impact of "-omics" technologies. *Mutation Research* **499**, 13–25.

Abt, M., Lim, Y., Sacks, J., Xie, M., and Young, S. S. (2001). A sequential approach for identifying lead compounds in large chemical databases. *Statistical Science* **16**, 154–168.

Bacha, P. A., Gruver, H. S., Den Hartog, B. K., Tamura, S. Y., and Nutt, R. F. (2002). Rule extraction from a mutagenicity data set using adaptively grown phylogenetic-like trees. *Journal of Chemical Information and Computer Sciences* **42**, 1104–1111.

Bailer, A. J., and Piegorsch, W. W. (2000). Quantitative potency estimation to measure risk with bio-environmental hazards. In *Handbook of Statistics Volume 18: Bioenvironmental and Public Health Statistics*, Sen, P. K. and Rao, C. R. (eds.), 441-463. Amsterdam, North-Holland/Elsevier.

Benigni, R., and Richard, A. M. (1996). QSARS of mutagens and carcinogens: Two case studies illustrating problems in the construction of models for noncongeneric chemicals. *Mutation Research* **371**, 29–46.

Bernstein, L., Kaldor, J., McCann, J., and Pike, M. C. (1982). An empirical approach to the statistical analysis of mutagenesis data from the Salmonella test. *Mutation Research* **97**, 267–281.

Bliss, C. I. (1935). The calculation of dosage-mortality curves. *Annals of Applied Biology* **22**, 134–167.

Bogen, K. T. (1995). Improved prediction of carcinogenic potencies from mutagenic potencies for chemicals positive in rodents and the Ames test. *Environmental and Molecular Mutagenesis* **25**, 37–49.

Borwein, P., and Erdélyi, T. (1995). *Polynomials and Polynomial Inequalities.* New York, Springer-Verlag

Chapman, P. M., Caldwell, R. S., and Chapman, P. F. (1996). A warning: NOECs are inappropriate for regulatory use. *Environmental Toxicology and Chemistry* **15**, 77–79.

Crawford, M., and Wilson, R. (1996). Low-dose linearity: the rule or the exception? *Human and Ecological Risk Assessment* **2**, 305–330.

Creighton, C., and Hanash, S. (2003). Mining gene expression databases for association rules. *Bioinformatics* **19**, 79–86.

Cunningham, A. R., Rosenkranz, H. S., Zhang, Y. P., and Klopman, G. (1998). Identification of 'genotoxic' and 'non-genotoxic' alerts for cancer in mice: the carcinogenic potency database. *Mutation Research* **398**, 1–17.

Descartes, R. (1637). *La Géométrie*, Nouvelle Edn. Paris, A. Hermann, Librairie Scientifique.

Fernandez, G. (2002). *Data Mining Using SAS Applications.* Boca Raton, FL, Chapman & Hall/CRC.

Fetterman, B. A., Kim, B. S., Margolin, B. H., Schildcrout, J. S., Smith, M. G., Wagner, S. M., and Zeiger, E. (1997). Predicting rodent carcinogenicity from mutagenic potency measured in the Ames Salmonella assay. *Environmental and Molecular Mutagenesis* **29**, 312–322.

Finney, D. J. (1952). *Statistical Method in Biological Assay.* London, Chas. Griffin & Co.

Gold, L. S., and Zeiger, E. (1996). *Handbook of Carcinogenic Potency and Genotoxicity Databases.* Boca Raton, FL, CRC Press.

Hand, D., Mannila, H., and Smyth, P. (2002). *Principles of Data Mining.* Cambridge, MA, MIT Press

Hand, D. J., Blunt, G., Kelly, M. G., and Adams, M. N. (2000). Data mining for fun and profit. *Statistical Science* **15**, 111–131.

Haseman, J. K. (2000). Using the NTP database to assess the value of rodent carcinogenicity studies for determining human cancer risk. *Drug Metabolism Reviews* **32**, 169–186.

Helma, C., Gottmann, E., and Kramer, S. (2000). Knowledge discovery and data mining in toxicology. *Statistical Methods in Medical Research* **9**, 329–358.

Hunter, J. E., and Schmidt, F. L. (2004). *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings, 2nd Edn.* Newbury Park, CA, Sage Publications.

Keenan, C., Hughes-Earle, A., Case, M., Stuart, B., Lake, S., Mahrt, C., Halliwell, W., Westhouse, R., Elwell, M., Morton, D., Morawietz, G., Rittinghausen, S., Deschl, U., and Mohr, U. (2002). The North American Control Animal Database: A resource based on standardized nomenclature and diagnostic criteria. *Toxicologic Pathology* **30**, 75-79.

Klopman, G. (1985). Predicting toxicity through a computer automated structure evaluation program. *Environmental Health Perspectives* **61** 269–274.

Klopman, G., and Rosenkranz, H. S. (1991). Structure-activity relations: Maximizing the usefulness of mutagenicity and carcinogenicity databases. *Environmental Health Perspectives* **96**, 67–75.

Krewski, D., Leroux, B. G., Bleuer, S. R., and Broekhoven, L. H. (1993). Modeling the Ames *Salmonella*/microsome assay. *Biometrics* **49**, 499–510.

Kübler, J., and Weihrauch, T. R. (2002). Integrated summaries and meta-analyses in clinical drug development. *Drug Information Journal* **36**, 127–133.

Leroux, B. G., and Krewski, D. (1993). AMESFIT: A microcomputer program for fitting linear-exponential dose-response models in the Ames *Salmonella* assay. *Environmental and Molecular Mutagenesis* **22**, 78–84.

Liu, M., Sussman, N., Klopman, G., and Rosenkranz, H. S. (1996). Structure-activity and mechanistic relationships: The effect of chemical overlap on structural overlap in data bases of varying size and composition. *Mutation Research* **372**, 79–85.

Margolin, B. H., Kaplan, N., and Zeiger, E. (1981). Statistical analysis of the Ames *Salmonella*/microsome test. *Proceedings of the National Academy of Sciences, USA* **76**, 3779–3783.

Margolin, B. H., Kim, B. S., Smith, M. G., Fetterman, B. A., Piegorsch, W. W., and Zeiger, E. (1994). Some comments on potency measures in mutagenicity research. *Environmental Health Perspectives* **102**, Suppl. 1, 91-94.

Margolin, B. H., and Risko, K. J. (1991). The statistical analysis of in vivo genotoxicity data: Case studies of the rat hepatocyte UDS and mouse bone marrow micronucleus assays. In *In Evaluation of Short-Term Tests for Carcinogens: Report of the International Programme on Chemical Safety's Collaborative Study on In Vivo Assays*, Ashby, J., de Serres, F. J., Shelby, M. D., Margolin, B. H., Ishidate, M. and Becking, G. C. (eds.), 1.29-1.42. Cambridge, Cambridge University Press.

Maron, D. M., and Ames, B. N. (1983). Revised methods for the Salmonella mutagenicity test. *Mutation Research* **113**, 173–215.

Mattingly, C. J., Colby, G. T., Forrest, J. N., and Boyer, J. L. (2003). The Comparative Toxicogenomics Database (CTD). *Environmental Health Perspectives* **111**, 793–795.

Piegorsch, W. W., and Bailer, A. J. (1997). *Statistics for Environmental Biology and Toxicology.* Boca Raton, FL, Chapman & Hall/CRC.

Piegorsch, W. W., and Hoel, D. G. (1988). Exploring relationships between mutagenic and carcinogenic potencies. *Mutation Research* **196**, 161–175.

SAS Institute Inc. (2000). *SAS/STAT ® User's Guide, Version 8, Volumes 1, 2, and 3.* Cary, NC, SAS Institute Inc.

Schmidt, C. W. (2003). Data explosion. Bringing order to chaos with bioinformatics. *Environmental Health Perspectives* **111**, A340–A345.

Simmons, S. J. (2002). *Data Mining for Mutagenic Potency Estimation in Hierarchical Linear Models.* Unpublished Ph.D. Thesis, University of South Carolina, Columbia, SC.

Snee, R. D., and Irr, J. D. (1984). A procedure for the statistical evaluation of Ames *Salmonella* assay results: comparison of results among 4 laboratories. *Mutation Research* **128**, 115–125.

Sohn, S. Y. (1999). Meta analysis of classification algorithms for pattern recognition. *Transactions on Pattern Analysis and Machine Intelligence* **21**, 1137–1144.

Sutton, A. J., Abrams, K. R., Sheldon, T. A., and Song, F. (2000). *Methods for Meta-Analysis in Medical Research.* New York, John Wiley & Sons.

Trevan, J. W. (1927). The error of determination of toxicity. *Proceedings of the Royal Society, Ser. B* **101**, 483–514.

Valafar, F. (2002). Pattern recognition techniques in microarray data analysis. A survey. *Annals of the New York Academy of Sciences* **980**, 41–64.

Wang, J. S., Lai, L. H., and Tang, Y. Q. (1999). Data mining of toxic chemicals: Structure patterns and QSAR. *Journal of Molecular Modeling* **5**, 252–262.

Yanagawa, T., and Kikuchi, Y. (2001). Statistical issues on the determination of the no-observed-adverse-effect levels in toxicology. *Environmetrics* **12**, 319–325.

Yanagawa, T., Kikuchi, Y., and Brown, K. G. (1994). Statistical issues on the no-observed-adverse-effect level in categorical response. *Environmental Health Perspectives* **102**, Suppl. 1, 95–104.

Yuan, B., Wang, X. Z., and Morris, T. (2000). Software analyser design using data mining technology for toxicity prediction of aqueous effluents. *Waste Management* **20**, 677–686.

Zeiger, E. (1996). Genotoxicity database. In *Handbook of Carcinogenic Potency and Genotoxicity Databases*, Gold, L. S. and Zeiger, E. (eds.), 687–729. Boca Raton, FL, CRC Press.

Zweiger, G. (1999). Knowledge discovery in gene-expression-microarray data: mining the information output of the genome. *Trends in Biotechnology* **17**, 429–436.

Received October 10, 2003
Revised June 15, 2004