Proceedings of the 5th International Conference on Statistics: Theory and Applications (ICSTA'23)
Brunel University, London, United Kingdom – August 03-05, 2023
Paper No. 10.11159/icsta23.108
DOI: 10.11159/icsta23.108

An Overview of Statistical Models in High Throughput Screening (HTS) for Toxicity Data

Mehdi Razzaghi
Commonwealth University of Pennsylvania
Bloomsburg, Pennsylvania, USA

Abstract **-** There has been an increasing amount of attention focused on using high-throughput *in vitro* screening for identification of hazardous chemicals in the environment over the last decade. The advantage of such methods to identify chemical agents that have potential to cause adverse effects in humans is the fact that it allows the assessment of several chemicals in a short period of time. Although nonlinear regression is routinely used to analyze data generated from the HTS assays, several models have been introduced for classifying and identifying the mechanism of action and developing predictive models for assessment of toxicity. Since HTS screening data are generated by repeating experiments several times, approaches vary in considering mixed effect models to repeated measures and hierarchical models. In this paper, we provide an overview of these approaches and make a comparative analysis of the models. Relative advantages as well as limitations are discussed.

Keywords: Statistical Models, Toxicity Data, Quantitative Risk Assessment, High Throughput Screening.

1. Introduction

Humans are routinely exposed to a multitude of chemicals through drugs, food additives, pollution, industrial waste, and other sources. The industrialization of the modern world has witnessed a steady growth in the consumption of chemical agents in many different settings. Regulatory agencies, pharmaceutical companies, and food industries are continually faced with the challenge of developing safe levels of chemicals and avoiding possible hazards as the result of human exposure to potential toxic substances. Toxicity is defined as any undesirable or adverse effect of exogenous substances on humans, animals, and the environment. Often risk assessment is utilized to determine possible hazards and for achieving safety. This is a process for evaluating and deriving a quantitative or qualitative estimate of the potential health risk to humans or other organisms. It represents the likelihood of an adverse effect or the absence of a beneficial effect. Risk is generally defined as the probability of an adverse effect and risk estimation can potentially present several problems due to scarcity and uncertainty in the data.

Toxicity assessment of chemical agents have for a long time used laboratory animals in bioassay experiments to perform *in vivo* studies and to determine the adverse effects of toxicants. Typically, depending on the toxicological endpoint of interest, e.g. carcinogenicity or developmental effects, specific experiments were designed and laboratory animals, mostly mice and rats, were exposed to a few doses of a toxicant in controlled environments. Several statistical models were developed, especially in the 80's and 90's to describe the toxicological processes mathematically and to facilitate risk assessment. Hoel [1] refers to this era as an "exciting time" because of the attention given by statisticians to the problem of estimating the human health risk due to environmental and occupational exposures. For a thorough description of these models, we refer to Razzaghi [2].

Unfortunately, the traditional bioassay approach to detect toxicity in chemicals proved to be slow, inefficient, and very costly. For example, to determine the cariogenic effect of a chemical, estimation of cancer risk was on the basis of long-term bioassay experiments and exposure generally occurred for a period of 200 days to two years and would cost several million dollars. Besides, those experiments considered one chemical at a time, while in reality, humans are exposed to several chemicals simultaneously. Addressing these inefficiencies, after the publication of the National Research Council [NRC, 3], there was a major shift in the paradigm of toxicity testing

and High Throughput Screening (HTS) techniques which had gained popularity since the early 1990's found favor in evaluating chemical toxicities. Such experiments allowed for *in vitro* and rapid evaluation of bioactivity of chemicals and enabled the simultaneous assessment of large numbers of chemical compounds. The goal of HTS is to identify and compare chemicals that are likely to perturb normal biological processes that can lead to human or environmental adverse effects. Three main quantities are used to compare chemicals [4]. These are the probability of an active response, the potency or the concentration that caused the biological change, and the magnitude of the response called the efficacy of the chemical. In the application of HTS, assays were usually conducted at a single test concentration and suffered from a large false positive error. Therefore, the methodology was not compatible with toxicity testing, which requires evaluation of compounds with weak activity. Introduction of Quantitative High Throughput Screening (qHTS) method introduced by Iglese et al. [5] yielded ground for the HTS methods and provided the opportunity to test a chemical agent at several concentration levels while significantly reducing the false positive rate. The advantage of the qHTS method is that it can generate concentration-response curves for several thousands of compounds in a single experiment. Development of qHTS method has clearly given way to new approach methodologies as much more efficient techniques in comparison to the old-fashioned low throughput toxicity testing. The qHTS approach can produce results for a multitude of chemicals in a matter of days. The methodology has become so popular that in 2019, the US Environmental Protection Agency announced that they plan to eliminate all mammalian toxicity testing by the year 2035. For this reason, statistical methods are being developed to analyze the data from qHTS experiments and here, we introduce some of these models and methods. In the next section, we provide a brief introduction to a formal risk assessment methodology and discuss the popular Benchmark Dose (BMD) approach. In the subsequent three sections, we discuss methods based on non-linear regression, robust regression and the hierarchical and Bayesian methods. We close with some concluding remarks and possible paths for future research.

2. Risk Assessment Methodology

Risk assessment is the process of evaluating and deriving a quantitative or qualitative estimation of the potential health risk to humans or other organisms from a defined source of hazard. The National Research Council in the United States formalized the process of risk assessment in a publication [6] which later became known as the "Red Book." Accordingly, the process pf risk assessment was defined as having four distinct stages. The first stage is hazard identification which consists of qualitatively identifying the source of hazard. The second stage is Exposure Assessment which entails the identification of the possible routes of exposure. Third, is the Dose-Response Assessment, which is the critical stage of quantification of the relationship that may exist between the risk and the exposure level. Often, a sigmoid-shaped mathematical function is used to describe the effect as a function of exposure level. The fourth and the final stage of the risk assessment process is Risk Characterization which involves the integration of the information gathered form the previous three stages in order to quantitatively estimate the risk. Traditionally, to estimate the risk, no-observed-adverse-effect-level (NOAEL) method was used. But, that method has been abandoned for the most part and replaced by the more favorable benchmark dose (BMD) methodology introduced by Crump [7,8]. The main advantage of this method is that, unlike the NOAEL, it utilizes the entire shape of the concentration-response relationship and is not limited to only the experimental dose levels. After fitting a statistical model to the data, BMD is determined as the concentration level that causes a fixed preset change in response, called the benchmark response (BMR) usually fixed at 5% or 10%. The statistical lower confidence (often 95%) limit called BMDL is then determined and used as the so-called point of departure (PoD) which is the starting point for calculating an acceptable exposure level. For qHTS data, an alternative approach to derive the PoD was introduced by Sands et al [9]. Their method is based on the signal-to-noise crossover dose (SNCD) defined as the "dose at which the ratio between the additional effect and the difference between upper and lower bounds of the two- sided 90% confidence interval on absolute effect correspond to some critical value (e.g. 0.67)". To estimate the human exposure levels, the PoD is then divided by a series of adjustment factors

(AFs) or uncertainty factors (UFs) to account for inter-species extrapolation, inter-individual differences and other uncertainties involved in toxicological data interpretation. Haber et al. [10] give a thorough discussion of the BMD methodology and its properties. For qHTS data, Anderson and Krewski [11] suggest an additional AF due to in vitro to in vivo extrapolation to allow for the calibration of cellular test concentration system used with corresponding human doses using pharmacokinetic models.

With the emergence and growth of the new data on toxicity pathway assays using high-throughput tests, the stages of risk assessment described above are also going through major revisions. The US Environmental Agency initiated the NextGen project for developing the paradigm in risk assessment for high-throughput data. The specific goal of the project was making risk assessment faster, less expensive, and more scientifically robust [12]. The process of risk Assessment for the next generation of risk science is well delineated in [13].

3. Nonlinear Regression Methods

A critical step for the analysis of qHTS data, is the establishment of the mathematical model that characterizes the concentration-response relationship. Although there are many mathematical functions that represent a sigmoid-shaped curve, the Hill model is widely accepted and used for qHTS data. As pointed out in Shockly [14], the Hill model has a long and merited reputation for describing the concentration-response in many biological experiments, partly because the model parameters "have convenient biological interpretations and direct comparison of its parameter estimates from different experiments provides a convenient way to compare concentration-response profiles.". Mathematically, the Hill function can be defined in different forms. Some authors (e.g.[15]) define the Hill function as

$$f(x, \boldsymbol{\theta}) = \theta_{\mathbf{0}} + \frac{\theta_1}{1 + \left(\frac{x}{\theta_3}\right)^{\theta_2}} \tag{1}$$

where x is the dose of a chemical, $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2, \theta_3)^T$ is the vector of parameters with the following interpretations:

θ_0: Lower asymptote.

θ_1: Difference between the mean response baseline (zero concentration) and the lower asymptote, also referred to as the chemical efficacy.

θ_2: Slope or shape parameter.

θ_3: Is the median effective dose known as ED_{50}, which is the dose corresponding to 50% maximal response.

Alternatively, other authors (e.g. [16]) define the Hill function in its logistic format as

$$f(x, \boldsymbol{\theta}) = \theta_{\mathbf{0}} + \frac{\theta_1}{1 + e^{\left[\frac{\log(\theta_3) - \log(x)}{\theta_2'}\right]}} \tag{2}$$

where $\theta_2' = -\frac{1}{\theta_2}$ and represents the steepness of the curve, being the slope of the tangent line at $\log(ED_{50})$.

Suppose now that an experiment consists of k concentrations $x_1, \ldots, x_k$ and let $n_i$be the number of replications at concentration x_i $i = 1, \ , k$. The observed responses are first normalized using one of several standard methods [17]. The nonlinear regression approach to analyzing the qHTS data, generally begins by fitting the Hill model and assuming that

$$y_{ij} = f(x_i, \boldsymbol{\theta}) + \varepsilon_{\boldsymbol{ij}} \qquad i = 1, \ldots, k, j = 1, \ldots, n_i \tag{3}$$

where y_{ij} denotes the normalized response of the jth assay at the ith concentration and ε_{ij} is the unobserved random error assumed to have a normal distribution with zero mean and an unknown variance σ_i^2 $i = 1, \ldots, k$. An optimization technique is used to fit the above model. Using the Ordinary Least Squares method, [18] developed a classification algorithm whereby after fitting the Hill model, and estimating the model parameters, each chemical was classified into one of four classes based on the estimated values of θ_1 and θ_3. For chemicals in classes 1 and

2, the procedure then declared a chemical active if in addition, the regression coefficient of determination R^2 exceeded 0.9. Noting that this methodology ignores the uncertainty associated with the parameter estimates, Parham et al. [19] first used a likelihood ratio test with Bonferroni correction to test the null hypothesis that θ_1 is zero and if the null hypothesis was rejected, the chemical was then declared active based on some additional constraints on the estimates of θ_2 and θ_3. A three-stage algorithm based on statistical testing for classification of chemicals was also developed in [20]. The method of maximum likelihood for estimating the parameters of the Hill model was used by [21]. Assuming the following structure for the model variance

$$\sigma_i^2 = \sigma_0^2 y_{ij}^{2\lambda} \qquad \lambda \geq 0 \tag{4}$$

where σ_0 and λ are constants to be determined, an optimal design for the qHTS assay is developed using the BMD as the PoD. Note that the in the above variance structure, λ=0 and λ=1 correspond respectively to constant variance and constant coefficient of variation cases.

4. Application of Robust Regression

As pointed out in [14], the parameter estimates of the Hill model usually suffer from high uncertainties, which can arise if the range of tested concentrations fail to include at least one of the two asymptotes. This means that the responses are non-homoscedastic and concentration spacing is not optimal. It is concluded that the problem with qHTS data cannot be resolved as long as we rely on parameter estimates derived from naïve curve-fitting procedures. The problem does not lie in the estimation method, rather in the application of nonlinear regression to study designs that lack suitable concentration spacing and sufficient replications. Various robust techniques have been proposed to improve the situation with the parameter estimation. Modifying the nonlinear regression equation (3) as

$$y_{ij} = f(x_i, \boldsymbol{\theta}) + \sigma_{\boldsymbol{i}}\, \varepsilon_{\boldsymbol{ij}} \qquad i = 1, \ldots, k, j = 1, \ldots, \tag{5}$$

where σ_i the error standard deviation at x_i, Lim et al [22] describe three procedures based on M-estimates for the nonlinear regression model. The first method is the ordinary M-estimator (OME) for $\boldsymbol{\theta}$ defined as the solution of the optimization problem

$$\widehat{\boldsymbol{\theta}} = argmin\left[\sum_{i,j} h^2\left(y_{ij} - f\left(x_{i,}, \boldsymbol{\theta}\right)\right)\right] \tag{6}$$

where h is a suitable Huber score function. The second method is the weighted M-estimator (WME) where the argument of the score function is weighted by σ_i^{-1} and a penalty function of $\log(\sigma_i)$ is added to the objective function (6). The third method is the so-called preliminary test estimate (PTE) where after fitting the OME for each x_i, the residuals are tested for heteroscedasticity under a loglinear model using a t test and either the OME or the WME is selected based on a cut-off value of the t score. Using the dose response data for two compounds from the US National Toxicology Program library of 1408 compounds that were evaluated using qHTS assay, they show that all three methods produce estimates that are robust to outliers and influential observations. However, the PTE method is also robust to error variance and may be more useful in practice. The performance of the PTE method is evaluated by simulation using the false discovery rate (FDR) and power in Lim et al [15]. Properties of the robust ridge regression was also explored in [23] and [24]. An ordinary ridge M-estimator is the solution of the following minimization problem

$$\widehat{\boldsymbol{\theta}}_{\boldsymbol{R}}(k) = argmin\left[\sum_{i,j} h^2\left(y_{ij} - f\left(x_{i,}, \boldsymbol{\theta}\right)\right) + k\, \boldsymbol{\theta}^T \boldsymbol{\theta}\right] \tag{7}$$

where, as before, k is the number of dose levels and $\boldsymbol{\theta}^T$ is the transpose of the unknown parameter vector $\boldsymbol{\theta}$. Similarly, for the weighted ridge regression the first term in objective function (7) is weighted by σ_i^{-1} and the penalty function $\log(\sigma_i)$ is also added. Note that in (7), when the score function h is the identity function, the

ordinary ridge regression estimator obtained from ordinary least squares estimator is resulted. The asymptotic properties of the estimators are derived and the through simulation, the ridge regression estimators are compared with other methods. Specifically, it is shown that the biases of the ridge regression estimators are much smaller than the corresponding standard estimators in many cases.

5. Hierarchical and Bayesian Approaches

For qHTS assays, one is estimating the response of a large number of chemicals across a variety of screening assays with sparse dose-response data for each chemical and assay combination. Fitting a dose-response model to each assay for each chemical requires the estimation of a large number of parameters and often leads to poor variance estimates. Noting this, [4] points out that there are currently several semi-parametric Bayesian methods for monotone regression for a single curve. In addition, other ad-hoc methods such as piecewise linear spline models, Bernstein polynomial model are not adequate because in qHTS assays, we are estimating the response of several chemicals on multiple assays. Hence, "HTS requires a dose-response method that is robust to the sparsity of the data for each chemical-assay combination, takes advantage of the larger number of chemicals and assays, and accurately estimates the efficacy, potency, and probability of an active response." The authors propose a Bayesian hierarchical model for dose-response that is specifically tailored to the high -dimensional, sparse data setting. Their model is called the zero-inflated piecewise log -logistic model. The dose response function is defined as a mixture of non-active response and an active response. Thus, the model depends on a latent variable which is equal to 1 if the response is active and 0 otherwise. Using a hierarchical structure based on normal distribution on the model parameters and the latent variables, they apply a MCMC algorithm which is a hybrid Gibbs and Metropolis sampler to derive the posterior distribution. Using simulation, they show that their hierarchical approach to analyzing qHTS data outperforms method that treat each curve as independent and ignores correlation between assays.

Application of the Bayesian regression tree was explored in Low-Cam et al [25]. It is argued that the advantage of this approach is that unlike the classical approaches that rely on data summaries, the Bayesian tree structure solves the problem of low sample size by combining all measurements from a general exposure experiment across doses, time of exposure, and replicates. Wheeler [26] notes that for qHTS data, the observed dose-response curves are cross sections of a surface defined by a chemical's structural properties. Thus, he proposes a model to characterize this surface as a sum of learned basis functions formed as the tensor product of lower dimensional functions. Wheeler's model expresses the dose response function as the product of spline functions defined over the space of observed vector $\mathcal{X}$ and the space of the dose levels $\mathcal{D}$. Then, each spline function is assumed to be the span of spline basis. He then defines a basis over the product space of $\mathcal{X} \times \mathcal{D}$ where each basis function is the tensor product of two surfaces defines over $\mathcal{X}$ and $\mathcal{D}$. He shows that his approach is computationally more effective.

6. Concluding Remarks

There is a large number of chemicals used in industry and commerce. Hagiwara [21] states that only a fraction the approximately 140,000 chemicals have been evaluated in depth. Noting that the traditional animal bioassay toxicity testing has proved to be extremely inefficient, taking a long time and being very expensive, the National Research Council (NRC) in the United States [3] has delineated a long-term vision for toxicity testing. Their proposal has received an overwhelming support nationally and internationally, and has created a major turnaround in toxicity testing. More and more the toxicity testing paradigm is rapidly moving towards the high-throughput *in vitro* screening (HTS). Although there is a major change in the methodologies for toxicity, they are still compatible with the risk assessment paradigm [5] established by the NRC earlier [10]. A wide range of statistical models have been introduced and utilized to manage and analyze HTS data. Here, an attempt has been made to provide an overview of these models and their properties. Although many of these models are mathematically elegant and statistically powerful, the problems relating to analyzing HTS data is far from over. First, the problem of optimal

design of such experiments is an issue that requires attention. In addition, the problem of developmental toxicity requires more in-depth study. Sipes et al [27] assumes that the developmental toxicity of *in vivo* animal studies guideline correlate with cell-based and cell-free *in vitro* HTS data. However, it is not clear how the cell-based *in vitro* data would create a mechanistic relationship in identifying chemicals with potential to cause developmental toxicity. Although in the *in vivo* animal studies, these relationships have been largely established, a thorough study of these problems for HTS data would be enlightening. Models that can predict pathways linked to specific developmental effects could greatly improve our understanding of developmental complexities.

References

[1] D.G. Hoel, "Quantitative risk assessment in the 1970s: A personal remembrance. *Dose-Response*, vol. 16, 1559325818803230.PMID: 303302069, 2018.

[2] M. Razzaghi, Statistical Models in Toxicology. Boca Raton, FL: CRC Press, 2020.

[3] National Research Council, Toxicity Testing in the 21st Century: A vision and a strategy. Washington DC: The National Academic Press, Washington, D.C, 2007.

[4] A. Wilson, D. M. Reif, and B. J. Reich, Hierarchcal dose-response modeling for high-throughput toxicity screening of environmental chemicals. Biometrics, vol. 70, 237-246, 2014.

[5] J. Inglese, D.S. Auld, A. Jadhav, R.L. Johnson, A. Simeonov, A. Yasgar, W. Zhang, and C.P. Austin, "Quantitative high-throughput screening: A titration-based approach that efficiently identifies biological activities in large chemical libraries," Proc. Natl. Acad. Sci. USA., vol 103, no. 31, 11473-11478, 2006.

[6] National Research Council. Risk assessment in the federal government: Managing the process. The National Academy Press, Washington, D.C., 1983.

[7] K. Crump, A new method for determining allowable daily intakes. Fundamental and Applied Toxicology, vol 4, no. 5, 854-871, 1984.

[8] K. Crump, Calculation of benchmark dose for continuous data. Risk Analysis, vol.15, no. 1, 79 -89.

[9] S. Sand, C. J. Portier, D. Krewski. A signal-noise-crossover dose as the point of departure for health risk assessment, Environmental Health Perspectives, vol. 119, 1766-1774, 2011.

[10] L.T. Haber, M.L. Dourson, B.C. Allen, R.C. Herzberg, A. Parker, M.J. Vincent, A. Maier, and A.R. Boobies. Benchmark dose modeling: Current practice, issues, and challenges. Critical Review in Toxicology, vol. 48, 387-425, 2018.

[11] M.E. Anderson and D. Krewski, Toxicity testing in the 21st century: Bringing the vision to life, Toxicological Science, vol. 107, 324-330, 2009.

[12] I. Cote, P.T. Anastas, L. S. Birnbaum, , R. M. Clark, D. J. Dix, and S. W. Edwards, Advancing the next generation of risk assessment. Environmental Health Perspectives vol. 120, 1499-1502, 2012.

[13] D. Krewski, M. Westphal, M.E. Anderson, G.M. Paoli, W. A. Chiu, Al-Zoughool, M. C. Croteau, L.D. Burgoon, and i. Cote, A framework for the next generation of risk science, Environmental Health Perspectives, vol. 122, no. 8, 2014.

[14] K. R. Shockley, Quantitative high throughput screening data analysis, Drug discovery Today, vol. 20, no. 3, 2976-300, 2015.

[15] C. Lim, P. K. Sen, and S.D. Peddada, Robust analysis of high throughput screening (HTS) assay data, Technometrics vol. 55 no. 2, 150- 160, 2013.

[16] C. Park, J. Lee, and C. Lim, Analysis of high throughput screening data using robust method for nonlinear mixed effect, Communications for Statistical Applications and Methods vol 27, no. 6, 701-714, 2020.

[17] N. Malo, J. A. Hanley, S. Cerquozzi, J. Pelletier, and R. Nadon, Statistical practice in high throughput screening data analysis, Nature Technology, vol. 24, no. 2, 167-175, 2006.

[18] M. Xia, R. Huang, K. Witt, N. Southall, J. Foster, M.-H. Cho, A. Jadhav, C.S. Smith, C.S. Iglese, C.J. Portier, R.R. Tice, and C.P. Austin, Compound cytotoxicity profiling using quantitative high throughput screening , Environmental Health Perspectives, vol. 116, 284-291, 2008.

[19] F. Parham, C. Austin, N. Southall, R. Huang, R. Tice, and C. Portier, Dose-Response modeling of high throughput screening data, Journal of Biomolecular Screening vol. 14, 1216-1227, 2009.

[20] K. R. Shockley. A three-stage algorithm to make technologically relevant toxicity calls from quantitative high throughput screening data, Environmental Health Peraspectives, vol. 120, no. 8, 1107-1115, 2012.

[21] S. Higawara. Dose-response modeling and optimization of quantitative high throughput screening assays, PhD dissertation, Carleton University, 2020.
[22] C. Lim, P. K. Sen, and S.D. Peddada. Robust nonlinear regression in applications, Journal of the Indian Agricultural Statistics vol. 67, no. 2, 215-234, 2013.
[23] C. Lim, P. K. Sen, and S. D. Peddada, Accounting for uncertainty in heteroscedasticity in nonlinear regression. Journal opf Statistical Planning and Inference, vol. 142, 1047-1062.
[24] C. Lim. Robust ridge regression estimators for nonlinear models with applications to high throughput screening assay data, Statistics in Medicine, vol. 34, 1185-1198, 2014.
[25] C. Low-Cam, D. Ji. Telesca, H. Zhang, T. Xia, and J. I. Zink. A Bayesian regression tree approach to identify the effect of nanoparticles properties on toxicity profiles. The Annals of applied Statistics, vol. 9, 383 – 401, 2015.
[26] M. W. Wheeler. Bayesian additive adaptive basis tensor product models for modeling high dimensional surfaces: An application to high-throughput toxicity testing. Biometrics vol. 75, 193-201, 2018.
[27] N. S. Sipes, M. T. Martin, D. M. Reif, N. C. Kleinstreuer, R. S. Judson, A. V. Singh, K. J. Chandler, D.J. Dix, R. J. Kavlock, and T.B. Knudson, Predictive models of prenatal developmental toxicity from ToxCast high-through-put screening data. Toxicological Sciences vol. 124, no. 1, 109-127, 2011.