Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 67(2) 2013 141-150

Statistical Analysis of Data from Quantitative High Throughput Screening (qHTS) Assays- Methods and Challenges

Shyamal D. Peddada
Biostatistics Branch, NIEHS (NIH), RTP, NC 27709, USA

Received 16 May 2013; Accepted 23 May 2013

SUMMARY

Humans are exposed to thousands of chemicals, some of which are potentially toxic and even carcinogenic. For example, farmers are exposed to pesticides, workers cleaning oil spills are exposed to complex mixtures of compounds, miners are exposed to various chemicals in the dust they inhale and so on. Identification of toxins and carcinogens among such exposures and determination of their effects on human health is a complex process. While epidemiological studies at the population level serve an important purpose to this end, laboratory based toxicological studies play an equally important role. Despite the fact that extrapolation from lower order animals and cell lines to humans is a challenge, a major advantage of laboratory based toxicological studies is that one can control for various confounders when evaluating a chemical. For this reason toxicological studies, *e.g.* the standard two-year rodent cancer bioassay, are widely used for evaluating toxicity and carcinogenicity of various chemicals. Although such assays are considered to be robust and informative, they tend to be slow and expensive. Since not every chemical humans are exposed to is a toxin or a carcinogen, performing a rodent cancer bioassay on every chemical may not be time or cost effective. Consequently, there is considerable interest in conducting high or medium throughput screening assays using cells or lower order animals such as nematodes (*e.g. Caenorhabditiselegans*). Such assays are designed to evaluate several thousands of chemicals in a single experimental run, resulting in substantial reduction in cost and time. There are, however, several statistical issues that need to be considered when designing and analyzing such studies. The focus of this paper is to survey some of the statistical methods used for analyzing data obtained from the rodent cancer bioassay and those obtained from quantitative high throughput screening (qHTS) assays. Some of the statistical challenges will also be described.

Keywords: Carcinogenicity, Epidemiology, Nonlinear regression analysis, Optimal designs, Pesticides, Quantitative high through put screening assays, Toxicology.

1. INTRODUCTION

I thank the Indian Society of Agricultural Statistics, New Delhi, India, for inviting me to deliver the Technical Address at its 66th Annual Conference held in Delhi, India in December 2012. Indeed it is a great honor to be selected to give this talk. This paper elaborates some of the major items discussed in the talk as well as the material described in my discussion of Professor P.K. Sen's paper Sen (2013) appearing in this special issue. Thus some of the ideas mentioned here also appear in the discussion paper.

The green revolution of the 1960's pioneered by the Nobel Prize winning agronomist Dr. Norman Borlaug played a major role in modernizing agriculture and feeding billions around the world. Dr. M.S. Swaminathan, who embraced Dr. Borlaug's modern agricultural practices, brought the green revolution to India which included the use of high yielding variety of crops, different cropping patterns, efficient use of fertilizers and pesticides etc. The success of the green revolution in countries such as India is very evident. However, the impact of long term exposure to

[*] *Technical Address delivered at the 66th Annual Conference of the Indian Society of Agricultural Statistics held at IASRI, New Delhi during 18-20 December, 2012.*
E-mail address: peddada@niehs.nih.gov

chemicals, such as pesticides, on human health is still being investigated by researchers all over the world. Blair and Zhan (1995) discovered that compared to the general population, farmers in their study experienced higher rates of various cancers (*e.g.* leukemia, non-Hodgkin's lymphoma, stomach cancer, brain cancer). They described the need for studies to systematically characterize and evaluate the effects of various exposures occurring on a farm and to various health outcomes. Later in their 1998 paper (Blair and Zhan 1998) they observed that the use of insecticide lindane increased the risk of non-Hodgkin's lymphoma among white men. Scientists at NIEHS, NCI and EPA established the Agricultural Health Study (AHS) in 1993, to understand the effect of living (and working) on a farm on various health outcomes. The study includes about 90,000 people either living or working on a farm. For more details regarding AHS one may refer to NCI's website http://www.cancer.gov/cancertopics/factsheet/Risk/ahs. Numerous publications have resulted from this ongoing project. As noted in my discussion of the paper by Sen (2013) appearing in this special issue, several pesticides have been identified to be associated with various health outcomes. For example, Alavanja *et al*. (2009) noted that herbicides metolachlor and pendimethalin and insecticides chlorpyrifos and diazinon may be associated with the risk of lung cancer. Hopin *et al*. (2009) discovered that high doses of herbicides such as, coumaphos, heptachlor, parathion, 80/20 mix (carbon tetrachloride/carbon disulfide), and ethylene dibromide are associated with allergic asthma, while DDT is associated with non-allergic asthma. Several other diseases have been demonstrated to be associated with the use of pesticides, such as Parkinson's disease (Kamel *et al*. 2007), prostate cancer (Van Maele Fabry and Willems 2004, Meyer *et al*. 2007), etc. Recently Sarkar *et al*. (2012) conducted a comprehensive study of the effects of modern agriculture on public health. This is perhaps the most comprehensive paper written to-date on this subject from India's perspective. More research along these lines will be very important to carry out, especially in developing countries where the environmental stressors and their effects on various health outcomes have not been fully understood.

While epidemiological studies, such as AHS, are extremely important as they provide direct information regarding exposures and public health, they can be expensive and may potentially be difficult to interpret sometimes due to possible unmeasured confounders. To this end, laboratory based toxicological studies may provide an effective complement since they are controlled experiments. The National Toxicology Program (NTP) uses rodents (rats and mice) in its 2-year cancer bioassay to evaluate the toxicity and carcinogenicity of a chemical. These studies are very meticulously conducted and the resulting data are considered to be the "gold standard" by toxicologists world-wide. Although these data are considered to be pristine and reliable, they are very expensive and time consuming to obtain. As a result, the NTP and other agencies in the US, such as the NIH Chemical Genomics Center (NCGC) (http://www.ncats.nih.gov/research/reengineering/ncgc/ncgc.html) and the US Environmental Protection Agency (EPA), started to conduct quantitative high throughput screening (qHTS) assays. In a given run, a qHTS assay may evaluate as many as 10,000 chemicals. Using the data derived from these assays researchers may prioritize chemicals for further testing and hope to eventually be able to classify a chemical as toxic or not toxic.

The focus of this paper is to review some of the statistical methods currently available for analyzing data obtained from rodent cancer bioassay as well as the qHTS assay. The design and analysis of NTP's 2-year cancer bioassay data will be described in Section 2, whereas the design and analysis of qHTS assay will be described in Section 3. Several challenges and open research problems are also presented in each of these sections.

2. NTP's 2-YEAR RODENT CANCER BIOASSAY

2.1 Experimental Design and the Data

For each chemical (*e.g.* a pesticide) that needs to be evaluated for toxicity and carcinogenicity, the National Toxicology Program conducts a 2-year bioassay that consists of exposing male and female rats and mice to different doses of chemical. Typically there are four dose groups, namely, *control, low-dose, medium-dose and high-dose* and $n = 50$ animals (of roughly same age) are randomly assigned to each dose group. The two rodent species and the two sexes are treated as separate experiments. Thus, typically, there are 4 dose-response experiments and data from each experiment is analyzed separately.

The *route of exposure* to chemical is often matched with typical route of exposure humans experience for that chemical. For example, if humans are exposed to the chemical under study (i.e. test substance) through skin contact then animals in the experiment are also typically dermally exposed to the test substance. If the typical route of exposure for humans is via drinking water then the animals are also exposed to the chemical by mixing the chemical in drinking water. The route of exposure is very critical since tumor incidences for a given organ may vary substantially by the route of exposure.

Similar to the route of exposure, the *vehicle* used to deliver the test substance to the animal is an important factor. Suppose animals are exposed to the chemical via inhalation. That is, the test substance is mixed with ordinary air and animals are put in inhalation chambers where they breathe the test substance along with regular air. Then the vehicle in this case is ordinary air. Thus the group of animals not receiving the test substance (*i.e.* animals in the control group) is called *vehicle control* group.

Once the route of exposure and the vehicle are determined, animals in each dose group are exposed to the test substance for a two year period using the route of exposure and the vehicle. Since the natural life span of rats and mice is approximately 2 years, the duration of the study is typically set to 2 years (t_{sac} ~ 730 days). At the end of 2 years all surviving animals are sacrificed. All animals in the study are dissected and all tissues are inspected for the presence or absence of tumors.

2.2 Statistical Analysis

For the j^{th} tissue in i^{th}animal in d^{th} dose group, let T_{id} denote number of days the animal survived and let X_{ijd} denote the binary outcome which takes a value of 1 if tumor is present otherwise it takes a value of 0. Let $Y_{jd} = \sum_{i=1}^{n} X_{ijd}$ denote the number of animals in the d^{th} dose group that have tumor in the t^{th} tissue. For animals in the d^{th} dose group and j^{th} tissue, let $\beta_{jd}(t)$ denote the death rate of tumor free animals with survival function, $S_{\beta jd}$, let the tumor incidence rate be denoted by$\lambda_{jd}(t)$, with the corresponding survival function given by $S_{\lambda jd}$ and let $\gamma_{jd}(t)$ denote the death rate of animals with tumors. Then the three states an animal can be during the two year period is described in Fig. 1 (see Bailer and Portier 1988). Some other useful references in this regard are Dinse (1985, 1988a, 1988b, 1991, 1994, 1998), Lindsey and Ryan (1993) and references therein.

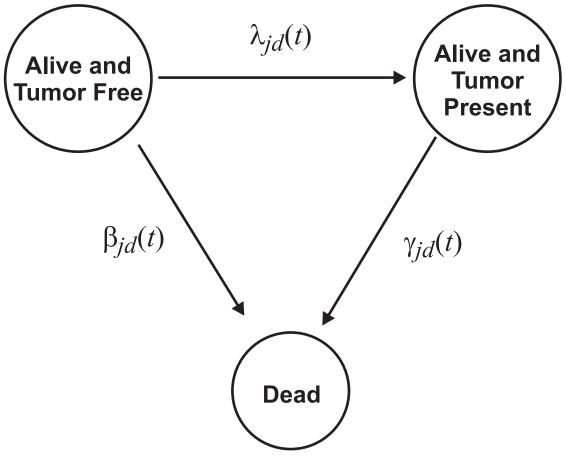

Fig. 1. Three-state stochastic model

Suppose for animals in the d^{th} dose group and j^{th} tissue π_{jd} denote the life-time risk of developing tumor. Making a simplifying assumption that the tumors are non-lethal (*i.e.* time to death is independent of the presence or absence of tumor) we obtain the following relationship between life time risk of developing tumors and tumor incidence rate (see Bailer and Portier 1988).

$$\pi_{jd} = \int_{0}^{\infty} I_{[0,t_{sac}]}(u)\lambda_{jd}(u)S_{\lambda jd}(u)S_{\beta jd}(u)du.$$

As a consequence of the above relationship we notice that $\lambda_{jd}(t) = c_j(t)$ [*i.e.* not dependent on dose] is not equivalent to $\pi_{jd} = \pi_j$ [*i.e.* not dependent on dose], unless the mortality rates are constant across dose groups. For each tissue, toxicologists are typically interested in testing the null hypothesis that there is no difference among the *D* dose groups in terms of tumor incidence against the alternative that the tumor incidence monotonically increases with dose. That is

$$\begin{aligned} &H_0 : \lambda_{j1}(t) = \lambda_{j2}(t) = \ldots = \lambda_{jD}(t), \quad \forall t \\ &H_\alpha : \lambda_{j1}(t) \le \lambda_{j2}(t) \le \ldots \le \lambda_{jD}(t) \end{aligned} \tag{1}$$

with at least one strict inequality. Unfortunately, since scheduled interim sacrifice of animals is prohibitively expensive, therefore they are not commonly done and hence tumor incidence cannot be properly estimated and consequently the above hypotheses is not testable directly. On the other hand, although lifetime risk of developing tumor π_{jd} can be estimated using the standard binomial proportion, *i.e.*, $\hat{\pi}_{jd} = \frac{Y_{jd}}{n}$, since π_{jd} is confounded by the survival of an animal it is not appropriate to test (1) using $\hat{\pi}_{jd}$. Bailer and Portier (1988) suggested a modification that accounted for the number of days an animal survived. The resulting estimator, known as the Poly-3 estimator, is obtained as follows. For the i^{th} animal in d^{th} dose group, if $T_{id} \le t_{sac}$ and the animal did not develop tumor in the j^{th} tissue then let $m_{ijd} = \left(\frac{T_{id}}{t_{sac}}\right)^3$ otherwise $m_{ijd} = 1$. Then Poly-3 estimator of π_{jd}, derived in Bailer and Portier (1988), which accounts for animals that died without tumor but are subject to risk of developing tumor in the j^{th} tissue had they lived their full life of $t_{sac} = 730$ days, is given by, $\hat{\pi}^*_{jd} = \frac{Y_{jd}}{n^*_{jd}}$, where. $n^*_{jd} = \sum_{i=1}^{n} m_{ijd}$. Bailer and Portier (1988) suggested using the well-known Cochran–Armitage (CA) trend test (Cochran 1954 and Armitage 1955) for testing hypotheses (1) using the survival adjusted binomial proportions $\hat{\pi}^*_{jd}$ rather than using the standard binomial proportions $\hat{\pi}_{jd}$. It is important to note that n^*_{jd} is a random variable. Consequently, one needs to account for variability in n^*_{jd} when computing the standard errors of the estimate of the regression parameter. Bieller and Willaims (1993) introduced a jackknife based methodology that accounts for the variability in n^*_{jd}. The resulting CA trend test using Poly-3 survival adjusted binomial proportions of Bailer and Portier (1988) with jackknife variance estimator of Bieller and Willaims (1993) is currently in use by various researchers, including the NTP, for analyzing the 2-year rodent cancer bioassay. The NTP calls this test as the "Poly-3 trend test".

The CA test (and hence the Poly-3 trend test) is ideal when testing for a linear trend in the dose response against the null hypothesis that the slope parameter is zero. However, it is well-known that dose-response in many toxicological studies are not necessarily linear, although monotonic (Peddada *et al*. 2005 and Peddada and Kissling 2006). As noted in Peddada *et al*. (2005), Peddada and Kissling (2006), in such situations, the above mentioned tests can be severely under powered in comparison to methods based on order-restricted inference. In a series of papers, Peddada *et al*. (2001), Peddada *et al*. (2005), Peddada and Kissling (2006), the use of methods based on order restricted inference to analyze such data was discussed. Specifically, in Peddada and Kissling (2006), the authors introduced a hybrid test that attempts to be as powerful as the Poly-3 trend test when the dose-response is linear and attempts to be as powerful as the isotonic trend test of Peddada *et al*. (2005) when the dose-response is monotonic but not linear. The resulting test is called the "max-Iso-Poly-3" trend test. It takes the maximum of isotonic regression trend test and the Poly-3 trend test and the asymptotic p-values are derived by simulating data from standard normal distribution.

2.3 Use of Historical Control Information

Each time a laboratory conducts a dose-response study to evaluate the toxicity and/or carcinogenicity of a chemical, it obtains data on the vehicle control group. Over time, it accrues data on control groups. The database consisting of all such historical data is often called the historical control database. Most agencies and pharmaceuticals maintain their own internal database of historical control data. When evaluating the data for a given chemical, the historical control database is used very judiciously by matching species, sex, route of exposure and vehicle. This is because there are differences in tumor incidence among control animals belonging to different species, sex, route of exposure and vehicle. Furthermore, to avoid any bias due to genetic drift over time, the NTP typically uses studies conducted within a 5 year window of the time the chemical under consideration was tested.

Once the historical control data of interest are identified, a longstanding question has been how to use these data and develop a formal statistical procedure to evaluate the tumor incidence data obtained from the current study. Often toxicologists use the range of tumor incidence obtained from the historical control data. Although there are no established criteria, a common strategy is to see if the tumor incidence in the current control (often called "concurrent control") is within the range of historical controls. If it is in the range and if the tumor incidence in one or more dose groups are outside the range of historical controls and if the Poly-3 trend test finds a significant dose-related trend in tumor incidence in the current study then toxicologists may declare a significant chemical effect. Recently there has been a detailed discussion on this topic in Keenan *et al.* (2009). It is important to note that as the number of studies increases, so does the range of historical controls. Furthermore, when a new strain of animal is introduced there may not be sufficient studies in the historical control database to estimate the range efficiently.

Use of historical control range as a criterion to evaluate data from current study appears to be ad-hoc and unsatisfactory. Consequently, several alternate statistical methods have been proposed in the literature. These methods range from using hierarchical models, such as the beta-binomial model (Tarone 1982), a variety of Bayesian methods (Ibrahim *et al*. 1996, 1998, Dunson and Dinse 2001) and most recently frequentist's methods using order restricted inference (Peddada *et al.* 2007). Comparison of historical control with the current control group was discussed in Dinse and Peddada (2011). An important feature of Peddada *et al.* (2007) and Dinse and Peddada (2011) is that rather than making any complicated modeling assumptions regarding the population of controls in the historical control database, they take a nonparametric approach by assuming that all controls in the historical control data under consideration come from a common population with common mean proportion parameter π_j (for the j^{th} tissue) and variance of the form $\sigma_j^2\pi_j(1-\pi_j)$, thus allowing for extra-binomial variation without requiring the distribution to be beta-binomial. Extensive simulation studies conducted in Peddada *et al*. (2007) and Dinse and Peddada (2011) suggest that these new procedures provided a good control of the Type I error at the desired nominal level.

2.4 Multivariate Methods

It is well-known to toxicologists that some of the tumors may co-exist, or alternatively, the occurrence of one type of tumor increases the probability of occurrence of another particular type. For example, pituitary gland tumors in female rats are known to trigger the increased incidence of mammary gland tumors through the prolactin pathway (McComb *et al*. 1984). Despite such associations, toxicologists typically analyze the rodent bioassay data for each tumor type separately using univariate methods such as those described above. This practice is perhaps largely due to the fact that multivariate analogs of the methods described in the above have not been developed in the literature until recently.

Davidov and Peddada (2012) developed methodology for testing for multivariate stochastic order among binary random variables. For a randomly chosen animal from the d^{th} dose group, X_d is a p-dimensional binary random vector, where p is the number of tissues, and each component of X_d takes a value of 0 or 1 depending upon whether the corresponding tissue is either tumor free or has tumor.

Definition (e.g. Shaked and Shantikumar 2007): For two $p\times 1$ random vectors X and Y, X is said to be multivariate stochastically smaller than Y (denoted by $X \leq Y$) if for all upper sets $\bigcup\in \mathcal{R}^p$, $P(X\in\bigcup)\leq P(Y\in\bigcup)$ with a strict inequality for at least one upper set.

Recall that a set $\bigcup\in \mathcal{R}^p$, is said to be an upper set, if for all u, $v\in \mathcal{R}^p$, with $u\leq v$ (inequality is defined component wise), if $u\in\bigcup$ then $v\in\bigcup$.

For dose-response studies such as those conducted in the rodent cancer bioassay, Davidov and Peddada (2012) developed a general framework to test multivariate stochastic order for binary random vectors. Assuming that the vector of responses is multivariate stochastically non-decreasing in dose level (a reasonable assumption in toxicological studies), they developed a non-parametric test for testing the following hypotheses:

$$H_0 : X_1 =^d X_2 =^d \ldots =^d X_D$$
$$H_\alpha : X_1 \leq X_2 \leq \ldots \leq X_D \text{(with at least one strict inequality).}$$

The resulting methodology is a simple and powerful nonparametric procedure that can be widely used for studying multivariate dose-related trends for not only toxicology data such as the NTP's 2-year cancer bioassay data but other data as well.

2.5 Concluding Remarks

The rodent cancer bioassay, such as those conducted by the NTP, is a well-planned bioassay with over three decades of history. The NTP alone has evaluated hundreds of chemicals for toxicity and carcinogenicity. Despite an outstanding track record, these assays take years to complete and are expensive to conduct. Furthermore, since not every chemical (*e.g.* pesticide) is likely to be a toxin at doses humans are exposed to, it may not be cost efficient to conduct rodent cancer bioassay on every chemical. Many of the chemicals can be potentially filtered away using other assays, such as those based on cell-lines and lower order animals such as nematodes [*e.g.Caenorhabditiselegans (c. elegans)*] and zebra fish etc. This idea about lower cost pre-screening for many chemicals led to the design and analysis of high throughput screening assays which are briefly described in the next section of this article.

3. QUANTITATIVE HIGH THROUGHPUT SCREENING (qHTS) ASSAYS

3.1 Experimental Design and the Data

Unlike the rodent cancer bioassay, where only one chemical is tested at a time using 3 dose groups and a vehicle control, the qHTS assay tests several thousands of chemicals at a time using a large number of concentrations ranging broadly from picomolar to millimolar levels. The experimental design consists of several plates, with each plate containing several hundred wells. Each plate usually corresponds to a single concentration of a chemical. Furthermore, on each plate several positive and negative controls are loaded.

For clarity of exposition, the typical study design is illustrated using an example from Xie *et al*. (2008).

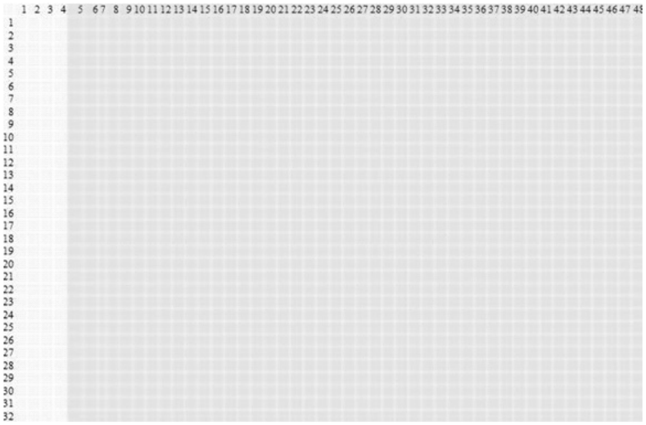

Fig. 2. Arrangement of chemicals on a 1536 well plate consisting of 32 rows and 48 columns. The 1408 test chemicals are placed in the shaded grey area with each cell corresponding to one chemical. The first 4 columns are used for positive control and vehicle control DMSO (*i.e.* negative control).

In this study the cytotoxicity of 1408 chemicals are evaluated. The design consists of 18 plates with 1536 wells per plate arranged in a matrix consisting of 32 rows and 48 columns. The vehicle control (*i.e.* the negative control) used in the study was Dimethyl Sulfoxide (denoted as DMSO). Each chemical that is to be tested in this assay is dissolved in this vehicle. The plates were numbered 1 through 18 with plates 1, 2 and 17, 18 containing DMSO, whereas plates 3 through 16 contained various doses of 1408 chemicals with concentration increasing with plate number. Typical arrangement of chemicals on each plate is as shown in Fig. 2. The 14 doses were as follows:

0.59 nM (nano molar), 2.95 nM, 14.8 nM, 33nM, 74 nM, 0.165 μM, 0.369 μM, 0.824 μM, 1.84 μM, 4.12 μM, 9.22 μM, 20.6 μM, 46 μM, and 92 μM (micro molar).

3.2 Statistical Analysis

Analysis of qHTS data generally relies on fitting the following nonlinear function called the "Hill function" (Fig. 3):

$$f(x,\theta)=\theta_0+\theta_1\frac{\theta_3^{\theta_2}}{\theta_3^{\theta_2}+x^{\theta_2}}, \qquad (2)$$

where, for a monotonically decreasing dose-response, $\theta_0 + \theta_1$ represents the mean response at baseline (control group, dose $x = 0$), θ_0 represents the minimum mean response (*i.e* as $x \to \infty$), θ_3 represents dose corresponding to a mean response halfway from baseline to minimum mean response. The slope of the Hill curve is described by θ_2. If the dose-response relationship is monotonically increasing then the above function can be suitably reparametrized. Several methods have been proposed in the literature for analyzing qHTS data using the Hill function.

In most instances the basic idea is to fit the Hill function for each chemical and then based on the estimates of various parameters different methods arrive at different decisions regarding each chemical. For example, researchers at the National Institute of Health Chemical Genomic Center (NCGC) (Xia *et al.* 2008) used a heuristic approach based on the ordinary least squares estimators (OLSE) $\hat{\theta}=(\hat{\theta}_0,\hat{\theta}_1,\hat{\theta}_2,\hat{\theta}_3)$ of $\theta=(\theta_0,\theta_1,\theta_2,\theta_3)$ to classify if a chemical is active (*i.e.* potentially toxic), inactive (*i.e.* non-toxic) or

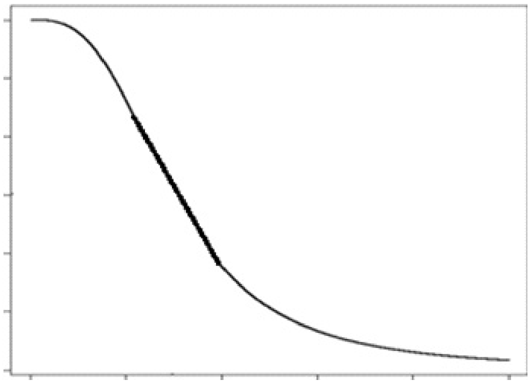

Fig. 3. Shape of a Hill function.

inconclusive. Their decision is based on the values of the point estimates and does not take into account underlying variability in the data, and hence in the variability in the estimates. As a consequence, their decision rule is not designed to control any notion of statistical errors, let alone errors due to multiple testing. Parham *et al*. (2009) developed a decision rule that included performing a likelihood ratio test on θ_1 and then use the estimated values of the remaining parameters without accounting for statistical uncertainties associated with those estimates (*i.e*. not taking into account their standard errors). Hence, as with the NCGC method, the classification rule obtained in Parham *et al*. (2009) is not designed to control any notion of statistical errors. Based on an exhaustive simulation study conducted in Lim *et al*. (2013), the authors noted that the NCGC method may have a false discovery rate (FDR) as low as 0, which results in very small power compared to other methods. On the other hand, the methodology of Parham *et al*. (2009) has an inflated FDR, as high as 0.6 in some instances.

Recently, Shockley (2012) developed a systematic three-stage algorithm to classify qHTS response profiles as “active”, “inactive” or “inconclusive”. His methodology used both unweighted and weighted least squares to fit the Hill model and used an F-test to make decisions regarding each chemical. Various decision rules were developed that used information obtained from unweighted and weighted least squares methodologies. However, he acknowledged that overall Type I error (associated with each chemical) and the multiple testing issues need to be addressed for his methodology.

Lim *et al*. (2013) took a more principled approach to the problem by developing an M-estimation based procedure that is not only robust to outliers and influential observations, which are common to qHTS assay, but their procedure is robust to heteroscedasticity, which is also a serious problem to consider when dealing with thousands of chemicals. For each chemical, Lim *et al*. (2013) perform a pre-test to determine if the data are heteroscedastic. If the null hypothesis is rejected then their methodology uses a weighted M-estimator (WME) that accounts for heteroscedasticity otherwise it uses ordinary M-estimator (OME). Finally, their method attempts to control the FDR. Based on their simulation studies, it appears that their method does not sacrifice too much power in comparison to the method of Parham *et al.* (2009) while the FDR remains close to the desired nominal level of 0.05.

3.3 Challenges and Future Research

As observed in Lim *et al*. (2013) as well as in my discussion of Sen (2013) appearing in this special volume, there are several issues with the design and analysis of qHTS assay that need further evaluation by statisticians. Issues range from numerical computations, statistical methods of analysis and experimental designs.

3.3.1 Numerical computations and statistical errors

It is important to note that, unlike in linear models, the information matrix in a nonlinear model is function of unknown parameters. It is therefore important to estimate θ as accurately as possible since the estimated information matrix is a function of the estimate of θ. The challenge therefore is not only the statistical accuracy but also numerical accuracy of the estimator of θ. Essentially, the point estimator of θ has two sources or components of bias and variance. One is statistical which can be dealt with if one is able to increase the sample size. The second is numerical, which requires large number of starting points whatever be the numerical algorithm. The minimization problem for computing the point estimator of θ, whether weighted for heteroscedasticity or unweighted for homoscedastic errors, is computationally intensive because one needs to explore a very large number of starting values for the optimization problem. Without such an exhaustive search, there is a danger to converge to local solutions that could result poor statistical inference. This issue is often overlooked or minimized in the context of nonlinear regression models. Although it may not be a serious issue in most other contexts, but in the context qHTS data one needs to be cautious since thousands of nonlinear models are fitted and the sample sizes are not very large.

3.3.2 Estimation of p-values

Again, in the case of standard linear models, if the model assumptions are valid, then the p-values obtained using the F-test are accurate. However in the case of nonlinear models, even if the model assumptions are valid, the p-values are approximate. As noted in Lim *et al*. (2013), in the context of qHTS assay, since several thousand models are fitted and accordingly several thousand inferences are being drawn, one needs to perform multiple testing corrections to the p-values. Thus, whether one uses Bonferroni corrections or Benjamini-Hochberg procedure (Benjamini and Hochberg 1995) the p-value threshold for significance is often extremely small. The F-test approximation used in nonlinear models provides a reasonable approximation to the true tail probabilities as long as the tails are of moderate size (*e.g*. tails corresponding to 5% level of significance). However, the p-values derived from F-distribution may not be very accurate for far right tail probabilities, where significance is often determined in the context of high dimensional data (*e.g*. tail corresponding to small levels of significance such as 0.000005). Since the sample sizes used in qHTS assay are typically small, the asymptotic approximations for such small tail probabilities are very likely to be poor. This is particularly true under heteroscedasticity because the F approximation in that case may not be accurate. It is the classical Beherens-Fisher problem and is exacerbated in nonlinear regression models. As an alternative to the asymptotic tests, one could possibly consider some resampling procedure. As discussed in detail by Lim *et al*. (2013), such a strategy is numerically infeasible. A possible alternative is to modify the denominator of the test statistic derived in Lim *et al*. (2013) by exploring shrinkage estimator of the variance, similar to the SAM methodology of Tusher *et al*. (2001).

3.3.3 Optimal designs for qHTS assay

Theory of statistical designs for qHTS assay is almost non-existent at present, despite the fact that a lot of resources are spent on these assays. The only effort in this area is due to Qu (2010) who discussed optimal designs for qHTS assay when the goal is to compare various chemicals. However, often the goal of qHTS assay is to identify chemicals that are "active" (or toxic) and those that are "inactive" (or non-toxic) and not necessarily to making comparisons among chemicals. There are several issues that need to be considered in designing these experiments.

3.3.3(a) Dose spacing

Fitting a nonlinear model such as the Hill model requires proper dose-spacing such that different parts of the curve are captured. For example, it will be ideal to have doses spaced so that, for each chemical, there are data points available at the upper asymptote, the linear part of the curve and the lower asymptote. Unfortunately, however, this is not an easy problem since several thousands of chemicals are being processed at the same time and each chemical may have a different value of θ, which is always unknown, and hence need different dose spacing pattern. Dose spacing that is good for one chemical profile may not be good for another. Perhaps one way to handle the problem is to take a Bayesian approach so that the dose spacing is ideal for an average chemical. Or minimize the maximum of the objective function over a high probability region of the prior support. In linear models, objective functions typically used for deriving optimal designs include, trace or generalized variance or largest eigenvalue of the information matrix. In the present context one may take a similar strategy except that the information matrix is replaced by the expected value of the information matrix where the expectation is taken over the prior distribution of θ. One can perhaps consider other strategies including a sequential design where intermediate doses are predicted to optimize the objective function. This is a wide open research area that requires a careful thought.

3.3.3(b) Spatial effects of locations on plates

As noted earlier, each plate (*e.g*. 1536 well-plate) in the assay corresponds to a particular dose. Arrangement of the chemicals does not change from plate to plate (*i.e*. dose to dose). Thus, a chemical "C" is located at the same location (*i, j*), *i.e*. i^{th} row and j^{th} column, over all plates in the experiment. While this is very convenient operationally for the robot, it can be unsatisfactory since it can introduce potential bias in the data. For example, as noted in Parham *et al*. (2009)

there are potential spatial effects. Due to the technology, depending upon the location on the plate, some wells may intrinsically exhibit high background intensities than others. It is therefore important to deal with the spatial effects either by modifying the experimental designs suitably, which may be a non-trivial task from a practical stand point due to the robotics, or account for spatial effects in the nonlinear model and analyze the data suitably. In any case, it is an important issue that needs to be addressed in order to obtain unbiased experimental data.

In summary, given the importance of qHTS assays, I believe that there are numerous statistical issues, both design as well as analysis, that need to be carefully addressed for a better analysis and interpretation of qHTS data.

ACKNOWLEDGEMENT

The research was supported [in part] by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences (Z01 ES101744). I thank Dr. David Umbach and Dr. Keith Shockley for careful reading and numerous comments that improved the presentation of this paper substantially.

REFERENCES

Alavanja, M.C., Dosemeci, M., Samanic, C., Lubin, J., Lynch C.F., Knott, C., Barker, J., Hoppin, J., Sandler, D., Coble, J., Thomas, K. and Blair, A. (2004). Pesticides and lung cancer risk in the Agricultural Health Study Cohort. *Amer. J. Epidemiol.*, **160**, 876-885.

Armitage, P. (1955). Tests for linear trends in proportions and frequencies. *Biometrics*, **11**, 375-386.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc.,* **B57**, 289-300.

Benjamini, Y. and Yekutieli, Y. (2005). False discovery rate-adjusted multiple confidence intervals for selected parameters. *J. Amer. Statist. Assoc*., **100**, 71-80.

Bieler, G. and Williams, R. (1993). Ratio estimates, the delta method, and quantal response tests for increased carcinogenicity. *Biometrics*, **49**, 793-801.

Blair, A. and Zahm, S.H. (1995). Agricultural exposures and cancer. *Environ. Health Perspectives,* **103**, 205-208.

Bailer, A. and Portier, C. (1988). Effects of treatment-induced mortality and tumor-induced mortality on tests for carcinogenicity in small samples. *Biometrics*, **44**, 417-431.

Cochran, W. (1954). Some Methods for Strengthening the Common χ^2 Tests. *Biometrics*, **10**, 417-451.

Davidov, O. and Peddada, S.D. (2011). Order restricted inference for multivariate binary data with application to toxicology. *J. Amer. Statist. Assoc*., **106,** 1394-1404.

Dinse, G.E. (1985). Testing for a trend in tumor prevalence rates: I. Nonlethal tumors. *Biometrics*, **41**, 751-770.

Dinse, G.E. (1988a). A prevalence analysis that adjusts for survival and tumor lethality. *J. Roy. Statist. Soc.,* **C37,** 435-445.

Dinse, G.E. (1988b). Simple parametric analysis of animal tumorigenicity data. *J. Amer. Statist. Assoc*., **83**, 638-649.

Dinse, G. (1991). Constant risk differences in the analysis of animal tumorigenicity data. *Biometrics*, **47**, 681-700.

Dinse, G. (1994). A comparison of tumour incidence analyses applicable in single-sacrifice animal experiments. *Statistics Medicine*, **13**, 689-708.

Dinse, G. (1998).Tumour incidence experiments. In: *Encyclopedia of Biostatistics*, vol. 6 (eds P. Armitage and T. Colton), 4597-4609. Wiley, Chichester.

Dinse, G. and Peddada, S.D. (2011). Comparing tumor rates in current and historical control groups in rodent cancer bioassays. *Stat. Biopharmac. Res.*, **3**, 97-105.

Dunson, D. and Dinse, G. (2001). Bayesian incidence analysis of animal tumorigenicity data. *Appl. Stat.*, **50**, 125-141.

Hoppin, J.A., Umbach, D.M., London, S.J., Henneberger, P.K., Kullman, G.J., Coble, J., Alavanja, M.C., Beane Freeman, L.E. and Sandler, D.P. (2009). Pesticide use and adult-onset asthma among male farmers in the Agricultural Health Study. *European Respiratory J.*, **34**, 1296-1303.

Ibrahim, J.G. and Ryan, L.M. (1996). Use of Historical Controls in Time-Adjusted Trend Tests for Carcinogenicity. *Biometrics*, **52**, 1478-1485.

Ibrahim, J.G., Ryan, L.M. and Chen, M.H. (1998). Using historical controls to adjust for covariates in trend tests for binary data. *J. Amer. Statist. Assoc.*, **93**, 1282-1293.

Kamel, F., Tanner, C.M. and Umbach, D.M., *et al*. (2007). Pesticide exposure and self-reported Parkinson's disease in the agricultural health study. *Am. J. Epidemiology*, **165**, 364-374.

Keenan, C., Elmore, S., Franck-Carroll, S., Kemp, R., Kerlin, R., Peddada, S.D., Pletcher, J., Rinke, M., Schmidt, S., Taylor, I. and Wolf, D. (2009). Best practices for use of historical control data of proliferative rodent lesions. *Toxicologic Pathology*, **37(5)**, 679-693.

Lim, C., Sen, P.K. and Peddada, S.D. (2011). Statistical inference in nonlinear regression under heteroscedasticity. *Sankhya,* **B72**, 202-218.

Lim, C., Sen, P.K. and Peddada, S.D. (2012). Accounting for uncertainty in heteroscedasticity in nonlinear regression. *J. Statist. Plann. Inf.*, **142,** 1047-1062. Epub 2012/02/22. doi: 10.1016/j.jspi.2011.11.003.

Lim, C., Sen, P.K. and Peddada, S.D. (2013). Robust analysis of high throughput screening assays. *Technometrics, in press*.

Lindsey, J.C. and Ryan, L.M. (1993). A Three-state multiplicative model for rodent tumorigenicity experiments. *J. Roy. Statist. Soc.,* **C42,** 283-300.

McComb, D.J., Kovacs, K., Beri, J. and Zak, F. (1984). Pituitary adenomas in old sprague-dawley rats: A histologic, ultrastructural, and immunocytochemical study. *J. National Cancer Instit.*, **73**, 1143-1166.

Meyer, T.E., Coker, A.L., Sanderson, M. and Symanski, E. (2007). A case control study of farming and prostate cancer in African American and Caucasian men. *Occupational Environmental Medicine*, **64**, 155-160.

Parham, F., Austin, C., Southall, N., Huang, R., Tice, R. and Portier, C. (2009). Dose–response modeling of high-throughput screening data. *J. Biomolecular Screening*, **14**, 1216-1227.

Peddada, S.D., Dinse, G. and Kissling, G. (2007). Incorporating historical control data when comparing tumor incidence rates. *J. Amer. Statist. Assoc.*, **102**, 1212-1220.

Peddada, S.D., Dinse, G. and Haseman, J. (2005). A survival-adjusted quantal response test for comparing tumor incidence rates. *J. Roy. Statist. Soc.,* **C54**, 51-61.

Peddada, S.D. and Kissling, G. (2006). A survival-adjusted quantal-response test for analysis of tumor incidence rates in animal carcinogenicity studies. *Environ. Health Perspectives*, **114**, 537-541.

Peddada, S.D., Prescott, K. and Conaway, M. (2001). Tests for order restrictions in binary data. *Biometrics*, **57**, 1219-1227.

Qu, X. (2010). Optimal row–column designs in high-throughput screening experiments. *Technometrics*, **52**, 409-420.

Sen, P.K. (2013). Agricultural environmental epidemiology: Some statistical perspectives. *J. Ind. Soc. Agril. Statist.,* **67(2)**.

Shaked, M. and Shanthikumar, J.G. (2007). *Stochastic Orders*. Springer Series in Statistics, Springer, New York, USA.

Shockley, K.R. (2012). A three-stage algorithm to make toxicologically relevant activity calls from quantitative high throughput screening data. *Environ. Health Perspective*, **120**, 1107-1115.

Smith, C.S., Bucher, J., Dearry, A., Portier, C., Tice, R.R.,Witt, K. and Collins, B. (2007). Chemical selection for NTP's high throughput screening initiative (Abstract), *Toxicologist*, **46**, 247.

Tarone, R.E. (1982). The Use of historical control information in testing for a trend in proportions. *Biometrics*, **38**, 215-220.

Tice, R.R., Fostel, J., Smith, C.S., Witt, K., Freedman, J.H., Portier, C.J., Dearry, A. and Bucher, J. (2007). The national toxicology program high throughput screening initiative: Current status and future directions (Abstract). *Toxicologist*, **46**, 246.

Tusher, V.G., Tibshirani, R. and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl Acad Sci.*, **98**, 5116-5121.

Van Maele Fabry, G. and Willems, J.L. (2004). Prostate cancer among pesticide applicators: a meta analysis. *Int. Arch. Occupational Environmental Health*, **77**, 559-570.

Williams, J.D., Birch, J.B., Woodall, W.H. and Ferry, N.M. (2007). Statistical monitoring of heteroscedastic dose–response profiles from high throughput screening. *J. Agril. Biol. Environ. Statist.*, **12**, 216-235.

Xia, M., Huang, R., Witt, K.L., Southall, N., Fostel, J., Cho. M.H., Jadhav, A., Smith, C.S., Inglese, J., Portier, C.J., Tice, R.R. and Austin, C.P. (2008). Compound cytotoxicity profiling using quantitative high throughput screening. *Environ. Health Perspectives*, **116**, 284-291.

Zhang, X.D. (2007). A new method with flexible and balanced control of false negatives and false positives for hit selection in RNA interference high-throughput screening assays. *J. Biomolecular Screening*, **12**, 645-655.