NIH Public Access

Author Manuscript

Technometrics. Author manuscript; available in PMC 2013 July 30.

Published in final edited form as:

Technometrics. 2013 May 1; 55(2): 150–160. doi:10.1080/00401706.2012.749166.

Robust Analysis of High Throughput Screening (HTS) Assay Data

Changwon Lim[1,†], **Pranab K. Sen**[2,3,‡], and **Shyamal D. Peddada**[4,*]

[1]Department of Mathematics and Statistics, Loyola University Chicago, 1032 W Sheridan Rd, Chicago, IL 60660

[2]Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, 338 Hanes Hall, CB#3260, Chapel Hill, NC 27599

[3]Department of Biostatistics, University of North Carolina at Chapel Hill, 3101 McGavran-Greenberg, CB#7420, Chapel Hill, NC 27599

[4]Biostatistics Branch, NIEHS, NIH, 111 T. W. Alexander Dr, RTP, NC 27709

Abstract

Quantitative high throughput screening (qHTS) assays use cells or tissues to screen thousands of compounds in a short period of time. Data generated from qHTS assays are then evaluated using nonlinear regression models, such as the Hill model, and decisions regarding toxicity are made using the estimates of the parameters of the model. For any given compound, the variability in the observed response may either be constant across dose groups (homoscedasticity) or vary with dose (heteroscedasticity). Since thousands of compounds are simultaneously evaluated in a qHTS assay, it is not practically feasible for an investigator to perform residual analysis to determine the variance structure before performing statistical inferences on each compound. Since it is well-known that the variance structure plays an important role in the analysis of linear and nonlinear regression models it is therefore important to have practically useful and easy to interpret methodology which is robust to the variance structure. Furthermore, given the number of chemicals that are investigated in the qHTS assay, outliers and influential observations are not uncommon. In this article we describe preliminary test estimation (PTE) based methodology which is robust to the variance structure as well as any potential outliers and influential observations. Performance of the proposed methodology is evaluated in terms of false discovery rate (FDR) and power using a simulation study mimicking a real qHTS data. Of the two methods currently in use, our simulations studies suggest that one is extremely conservative with very small power in comparison to the proposed PTE based method whereas the other method is very liberal. In contrast, the proposed PTE based methodology achieves a better control of FDR while maintaining good power. The proposed methodology is illustrated using a data set obtained from the National Toxicology Program (NTP). Additional information, simulation results, data and computer code are available online as supplementary materials.

[*](corresponding author) peddada@niehs.nih.gov.

[†]clim2@luc.edu

[‡]pksen@bios.unc.edu

SUPPLEMENTARY MATERIALS

- Additional results: HTS supplement.pdf
- Computer code: Software written in C++ implementing the methods described in the paper, and associated data (zip file)

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Keywords

Dose-response study; False discovery rate (FDR); Heteroscedasticity; Hill model; M-estimation procedure; Nonlinear regression model; Power; Toxicology

1 INTRODUCTION AND MOTIVATION

The classical rodent cancer bioassay used to detect toxicity or carcinogenicity of a chemical is often a slow and expensive process. For instance a typical carcinogenicity bioassay conducted by the National Toxicology Program (NTP) takes more than 2 years and costs several million dollars. It is widely acknowledged that humans are exposed to thousands of chemicals and yet less than 600 chemicals have been investigated by the NTP using the 2-year rodent bioassay. To speed up the process of screening chemicals, the NTP and other agencies as well as some chemical and pharmaceutical industries have begun conducting quantitative high throughput screening (qHTS) assays. Rather than using higher order animals such as rodents, qHTS assays typically treat cells or tissues to various doses of each chemical to determine toxicity of a chemical. Typically these assays are completed in a short period of time which results in substantial reduction of costs and time. Accordingly, in recent years, the design and analysis of qHTS data has been an active area of research (Zhang 2007; Michael et al. 2008; Qu 2010).

Typically, thousands of chemicals are processed at the same time in a qHTS assay and the resulting data are usually analyzed by fitting dose-response curves using the following function, known as the Hill function shown in Figure 1:

$$f(x,\theta)=\theta_0+\theta_1\frac{\theta_3^{\theta_2}}{x^{\theta_2}+\theta_3^{\theta_2}}, \quad (1)$$

where x denotes the dose of a chemical, θ_0 is the lower asymptote, θ_1 is the diference between the mean response at baseline and the lower asymptote (also known as the efficacy of the chemical), θ_2 is the slope or shape parameter of the curve and θ_3 is the dose corresponding to 50% response to the maximum change from baseline (also known as ED_{50}). Throughout this paper $x_{\min}$ and $x_{\max}$ denote the smallest and the largest doses used in the study. In qHTS assays a possible response variable of interest is cell viability measured by intracellular adenosine triphosphate (ATP) levels (Xia et al. 2008). Often such responses are normalized relative to a positive control (−100%) and the vehicle control (0%) as follows:

$$y=\frac{y^0-N}{I-N}\times(-100)$$

where y is percent activity, y^0 is raw data value, N is the median of the vehicle control, I is the median of a positive control (Inglese et al. 2006). Hence, the response is expressed as percentage and can be positive or negative.

Two common methods for analyzing qHTS data are the methods by Xia et al. (2008) and Parham et al. (2009). The former method, which will be referred to as the NCGC method, was developed by researchers at National Institute of Health Chemical Genomic Center (NCGC) and is widely used by researchers in the field. For each chemical, the NCGC methodology fits the Hill model using ordinary least squares and classifies the chemical into

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

various curve classes on the basis of the ordinary least squares estimates (OLSE) $\hat{\theta}$ as follows:

- Class 1: If $\hat{\theta}_1 > 30$, $\hat{\theta}_3 \in (x_{\min}, x_{\max})$ and the curve does not have the lower asymptote.
- Class 2: If $\hat{\theta}_1 > 30$, $\hat{\theta}_3 \in (x_{\min}, x_{\max})$ and the curve has the lower asymptote.
- Class 3 (Inconclusive): If $\hat{\theta}_1 > 30$ and $\hat{\theta}_3 \notin (x_{\min}, x_{\max})$.
- Class 4 (Inactive): If $\hat{\theta}_1 < 30$ then a chemical is classified as inactive.

Chemicals in Class 1 are declared to be active if the multiple correlation coefficient $R^2 > 0.9$. Chemicals in Class 2 are declared to be active if $R^2 > 0.9$ and $\hat{\theta}_1 > 80$. Chemicals that are neither classified to be active nor inactive are classified as inconclusive (Class 3). To define lower asymptote we quote Inglese et al. (2006) "To determine if the curve contained a lower bound asymptote, two points, representing the highest concentration tested and a half-log lower, were plotted according to the curve's Hill parameters. If the slope of these points was > −10 and if the plotted concentrations were past the inflection, the curve was determined to contain a lower asymptote."

A problem with the above strategy is that it ignores uncertainty associated with $\hat{\theta}$. To deal with some of these issues, Parham et al. (2009) first test $H_0 : \theta_1 = 0$ against $H_a : \theta_1 \neq 0$ using a likelihood ratio test (LRT) (chi-squared with 3 degrees of freedom) at $\alpha = 0.05$ with Bonferroni correction for multiple testing. They then classify compounds as follows:

- Active: If H_0 is rejected with $\hat{\theta}_2 > 0$, $\hat{\theta}_3 < x_{\max}$ and $|y_{x_{\max}}| > 10$ where $y_{x_{\max}}$ is the response at $x = x_{\max}$.
- Inactive: If either H_0 is not rejected or $\hat{\theta}_2 < 0$.
- Marginal: If a chemical is neither classified as active nor as inactive.

Throughout this paper we shall refer to the above methodology as "Parham methodology". Although they use a formal procedure for testing θ_1, they too ignore uncertainty associated with the estimates of remaining parameters. Furthermore, they ignore the underlying variance structure by using OLSE even if the variances are heteroscedastic.

In practice, since qHTS assays consist of data on thousands of chemicals, it may not be realistic to assume that all data are homoscedastic. Williams et al. (2007) suggested the use of iterated weighted least squares (IWLS) methodology by modeling the variance as a function of dose. Although such a methodology is practical, it is well-known that when the error variance is nearly homoscedastic then IWLS may not perform well. To illustrate this point, we consider qHTS data on two chemicals from the NTP library of 1,408 chemicals (Tice et al. 2007). We label them as Chemical A and Chemical B. Since outliers are common in qHTS data, throughout this paper we use M-estimators (with Huber-score function) in place of least squares. Thus we use Ordinary M-estimator (OME) and Weighted M-estimator (WME), both of which will be defined precisely later in the paper. The following illustration would work equally well for OLSE and IWLSE.

Data for Chemical A, along with OME and WME fitted curves are presented in Figure 2(a). Similarly, plots for Chemical B are presented in Figure 2(b). The individual point estimates are summarized in Table 1. There are a total of 14 doses in the study which are 0.00059, 0.00294, 0.0147, 0.0328, 0.0734, 0.164, 0.367, 0.821, 1.835, 4.103, 9.175, 20.52, 45.87, and 91.74 with 3 replicates per dose. From Figure 2, it seems reasonable to assume that Chemical A data are possibly heteroscedastic as the variance seems to increase with dose, whereas Chemical B data are approximately homoscedastic. We performed a simple linear

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

regression by regressing log of sample variance on the dose and found that the slope parameter for Chemical A is highly significant ($p = 0.004$) while the slope parameter for Chemical B is not ($p = 0.33$). The log linear model for the sample variance seems to be a simple parsimonious model to describe variance as a function of dose. Hence it is used throughout this paper.

As seen from Figure 2, the fits based on OME and WME seem to be equally good. However, parameter estimates and their standard errors seem to differ substantially. Thus this example demonstrates that OME and WME (and their standard errors) can be drastically different from each other depending upon the underlying variance structure. In practice, one never knows *a priori* if the data are homoscedastic or heteroscedastic. Standard diagnostic tools are practically impossible to implement since thousands of models are to be fitted in an automated manner. Thus there is a need for a methodology which automatically chooses between OME and WME.

Recently, in Lim et al. (2012) we proposed the preliminary test estimation (PTE) procedure for possibly heteroscedastic nonlinear models. The basic idea is to select either OME or WME on the basis of a simple preliminary test for heteroscedasticity. Depending upon the outcome of the test, PTE uses either OME or WME. Motivated by the performance of PTE methodology in Section 2 we develop PTE based likelihood ratio type methodology to evaluate if a chemical is active or inactive. In addition to testing, we also propose PTE based confidence intervals for estimating various parameters of the Hill model. We derive suitable critical values for the PTE methodology. Extensive simulation studies are conducted in Section 3 to investigate the performance of the proposed methodology in terms of the false discovery rate (FDR), the power and coverage probabilities of confidence intervals.

It is important to note that unlike linear models, where statistical inference is based on exact distribution theory (under suitable model assumptions), in the case of nonlinear models one relies on the asymptotic theory. The asymptotic approximations are generally reasonable for moderately large tail probabilities. Unfortunately, however, the asymptotic approximations are not good for very small tail probabilities, which are of primary interest in multiple testing problems. This is particularly the case when the data are heteroscedastic. A possible alternative to asymptotic theory based methodology is to use a resampling based method such as bootstrap/permutation methodology. Unfortunately, in the case on nonlinear models, this is a computationally challenging process. Even in the absence of bootstrap, fitting nonlinear models requires a very large number of starting values to avoid convergence to local solutions. It is critical to get a solution as close to the optimum as possible – not only because one wants to estimate the model parameters as accurately as possible, but the uncertainty estimates depend upon the unknown parameters (unlike linear models). This has a downstream effect on statistical inference. Secondly, some of the model parameters can be very unstable/sensitive to the data (such as ED_{50} and slope parameter). Estimates of these parameters are critical for toxicologists. Thus, even in the absence of bootstrap/permutation, the computation time to fit each model can be substantial. Since we are interested in far right tail probabilities (of the order 0.0001 or less), the number of bootstraps/permutations needed are of the order 100,000 or more to get an accurate estimate of small p-values. The current and future qHTS assays process 10,000 to 100,000 chemicals if not more. Thus the bootstrap/permutation based methods are practically infeasible for these assays. For this reason, it is more practical to use asymptotic theory based p-values although they are potentially inaccurate for small tail probabilities, especially when the data are heteroscedastic. As a consequence the FDR cannot be controlled by the Benjamini-Hochberg procedure (Benjamini and Hochberg 1995). Since p-values are inaccurate, we need to empirically determine which “black box” method could control FDR in this case. In this paper we concentrate on using a Bonferroni corrected threshold at level 0.05, which in

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

our simulations below works quite well at assuring FDR control at the same level. The Bonferroni adjusted p-values obtained in this paper can be thought of as “scores” for each chemical, with small scores corresponding to a compound that is more likely to be active. Thus, although the methodology described here is inspired by large sample theory, the inability to obtain accurate p-values for small to moderate sample sizes renders this methodology to be a bit of a “black box”. We note that this problem is not unique to the qHTS context, but may arise in other settings where p-values are not exact but are based on asymptotic methods.

The proposed methodology is illustrated in Section 4 using an assay conducted on NTP’s library of 1,408 compounds. We conclude this paper in Section 5 by providing discussion and open problems. Notations and Proofs are provided in the Appendix. The regularity conditions and results of some additional simulation studies are provided in the online Supplementary Materials.

2 METHODOLOGY

The standard nonlinear regression model may be expressed as

$$y_{ij}=f(x_i,\theta)+\sigma_i\varepsilon_{ij}, \quad i=1,\ldots,k, \quad j=1,\ldots,n_i \qquad (2)$$

where y_{ij} is an observable response variable, x_i is a known constant, $f(x_i, \boldsymbol{\theta})$ and $\boldsymbol{\theta}$ are defined in (1), σ_i is an error variance at x_i, and ε_{ij} is an unobservable random error assumed to be *iid* $N(0, 1)$. The total sample size $n=\sum_{i=1}^{k} n_i$.

Our experience with a sample of qHTS data suggests that σ_i depends upon x_i. To keep the variance model parsimonious, we use the log-linear model log σ_i = log $\sigma(x_i, \boldsymbol{\tau}) = \tau_0 + \tau_1 x_i$, where $\boldsymbol{\tau} = (\tau_0, \tau_1)^t$ is a vector of variance parameters. However, the proposed methodology can be easily modified if a more complex model is justified.

An underlying assumption made by researchers in this field is that if a chemical is active then its mean response can be modeled by the Hill model. On the other hand, if the chemical is not active then it is assumed to have a constant mean response across dose groups. Therefore, one may formulate the statistical problem as a test of the following hypotheses:

$$H_0: E(y)=\beta \ \text{(unknown)} \quad \text{vs.} \quad H_1: E(y)=f(x,\theta). \qquad (3)$$

Assuming that the residuals are homoscedastic and normally distributed, the null distribution of the likelihood ratio test (LRT) for the above hypotheses can be approximated by central F-distribution (cf., Gallant 1987). More precisely, suppose $\hat{\boldsymbol{\theta}}$ is the OLSE of $\boldsymbol{\theta}$ under H_1 with $\mathrm{SSE}_1(\widehat{\theta})=\sum_{i,j}\{y_{ij} - f(x_i,\widehat{\theta})\}^2$. Similarly, suppose $\hat{\beta}$ is the sample mean under the null hypothesis and $\mathrm{SSE}_0(\widehat{\beta})=\sum_{i,j}(y_{ij} - \widehat{\beta})^2$. Then the LRT $L_{\mathrm{OLSE}} = \{(\mathrm{SSE}_0(\hat{\beta}) - \mathrm{SSE}_1(\hat{\boldsymbol{\theta}}))/3\}/\{\mathrm{SSE}_1(\hat{\boldsymbol{\theta}})/(n-4)\}$ is approximately central F distributed with $(3, n-4)$ degrees of freedom. The above test statistic can be approximated as $L_{\mathrm{OLSE}} = \{\hat{\boldsymbol{\eta}}^t(\hat{\mathbf{H}}-\mathbf{H}^0)\hat{\boldsymbol{\eta}}/3\}/\{\hat{\boldsymbol{\eta}}^t(\mathbf{I}_n - \hat{\mathrm{H}})\hat{\boldsymbol{\eta}}/(n-4)\}+o_P(1) = L^{\mathrm{OLSE}}+o_P(1)$, where $\hat{\boldsymbol{\eta}} = \mathbf{Y} - \hat{\beta}\mathbf{1}$, $\mathbf{Y} = (y_{11},\ldots, y_{k,n_k})^t$, $\hat{\beta} = \bar{y}$, $\mathbf{1} = (1,\ldots, 1)^t$, $\hat{\mathbf{H}} = \hat{\mathbf{F}}(\hat{\mathbf{F}}^t\hat{\mathbf{F}})^{-1}\hat{\mathbf{F}}^t$, $\hat{\mathbf{F}} = \mathbf{f}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}) = \{\partial f(x_i, \boldsymbol{\theta})/\partial\theta_j|_{\theta=\hat{\theta}}\}$, and $\mathbf{H}^0 = \mathbf{1}(\mathbf{1}^t\mathbf{1})^{-1}\mathbf{1}^t$. Since L^{OLSE} is not robust to outliers and influential observations, in this paper we make it robust by replacing OLSE by OME $\tilde{\boldsymbol{\theta}}_n$, in the above calculations, where OME is defined as (Lim et al. 2012): $\tilde{\boldsymbol{\theta}}_n = \mathit{Argmin}\,\{\Sigma_{i,j}\, h^2\,(y_{ij} - f(x_i, \boldsymbol{\theta})) : \boldsymbol{\theta} \in \Re^p\}$ where h is taken to be the Huber-score function. For a pre-specified positive constant k_0, $h(u)=u/\sqrt{2}$, if $|u| < k_0$, otherwise $h(u) = \{k_0(|u|$

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

$-k_0/2)\}^{1/2}$. As commonly done, throughout this paper we take k_0 to be 1.5. We shall denote the resulting statistic by L^{OME}. Since OME and OLSE are asymptotically equivalent and OME is consistent and asymptotically normally distributed (Lim et al. 2012), the null distribution of L^{OME} can also be approximated by central F-distribution with (3, $n - 4$) degrees of freedom. Throughout this paper we shall refer to this modified LRT as the OME based methodology.

As often done, for heteroscedastic data we use the weighted version of L_{OLSE} by weighting the least squares with the estimated variances under the log-linear model as follows. First we estimate $\boldsymbol{\tau}_0$ and $\boldsymbol{\tau}_1$ in the log-linear model log $\boldsymbol{\sigma}_i = \boldsymbol{\tau}_0 + \boldsymbol{\tau}_1 x_i$ by performing simple linear regression of log of residuals on dose using OLSE. Using these estimates we obtain WLSE of $\boldsymbol{\theta}$, denoted as $\hat{\boldsymbol{\theta}}$ under H_1 and WLSE of β, denoted $\hat{\beta}$ under H_0. Let $L_0 = \Sigma_{i,j}(y_{ij} - \hat{\beta})^2/\exp(2\hat{\boldsymbol{\tau}}_1 x_i)$, $L_1 = \Sigma_{i,j}(y_{ij} - f(x_i, \hat{\boldsymbol{\theta}}))^2/\exp(2\hat{\boldsymbol{\tau}}_1 x_i)$ and $\hat{\beta} = \Sigma_{i,j} y_{ij} \exp(-2\hat{\boldsymbol{\tau}}_1 x_i)/\Sigma_i n_i \exp(-2\hat{\boldsymbol{\tau}}_1 x_i)$. Then the weighted version of L_{OLSE} is given by $L^{\text{WLSE}} = \{(L_0 - L_1)/3\}/\{L_1/(n-6)\}$. As done in the case of homoscedastic data, to robustify against outliers and influential observations, we replace the WLSE by WME in the above test statistic, where WME $(\widehat{\theta}_n^t, \widehat{\boldsymbol{\tau}}_n^t)^t$, is defined as (Lim et al. 2012) $(\widehat{\theta}_n^t, \widehat{\boldsymbol{\tau}}_n^t)^t = Argmin\left[\sum_{i,j}\{h^2((y_{ij} - f(x_i, \theta))/\sigma(\mathbf{z}_i, \boldsymbol{\tau})) + \log\sigma(\mathbf{z}_i, \boldsymbol{\tau})\} : \theta \in \Re^p, \boldsymbol{\tau} \in \Re^q\right]$. Under suitable regularity conditions, the asymptotic normality and consistency of WME is established in Lim et al. (2012). The WME version of L^{WLSE} is denoted as L^{WME}. Again due to asymptotic equivalence of WME and WLSE and the asymptotic equivalence of L^{WLSE} and L^{WME}, the null distribution of L^{WME} can be approximated by the approximate null distribution of L^{WLSE} (see Theorem 1).

As noted earlier, in practice one does not know if the data are homoscedastic or heteroscedastic. For this reason we now describe the PTE methodology.

As in Lim et al. (2012), we test for heteroscedasticity under the log-linear model log $\boldsymbol{\sigma}_i = \boldsymbol{\tau}_0 + \boldsymbol{\tau}_1 x_i$, by testing $H_0 : \boldsymbol{\tau}_1 = 0$ vs. $H_1 : \boldsymbol{\tau}_1 \neq 0$ using $T_n = \widehat{\tau}_{1n}/\sqrt{\text{Var}(\widehat{\tau}_{1n})}$, which is approximately central t distributed with $n - 2$ degrees of freedom. Throughout this paper we perform this preliminary test at $\boldsymbol{\alpha} = 0.5$. Thus the PTE $\widehat{\theta}_n^{\text{PT}}$ is defined as $\widehat{\theta}_n^{\text{PT}} = \tilde{\theta}_n I(|T_n| \leq t_{\alpha/2,n-2}) + \widehat{\theta}_n I(|T_n| > t_{\alpha/2,n-2})$, where $I(\cdot)$ is the usual indicator function. There is no special reason for choosing an $\boldsymbol{\alpha}$ of 0.50. Our proposed methodology is flexible enough that one could use any level of significance in the pre-test. We use the pre-test as a model selection procedure and not a formal test. It mimics classical model selection procedures such as forward selection/stepwise regression etc. in the linear/logistic/Cox regression model literature. Most standard software packages (and applied researchers) typically set the default value to be larger than 0.05. For example, in its FORWARD selection method implemented in PROC REG, SAS sets it at 0.50 as the default value (http://support.sas.com/onlinedoc/913/docMainpage.jsp). Unlike in most instances, in the present context of pretesting for homoscedasticity vs. heteroscedasticity, the Type II error is more important than the Type I error. This is because, using a homoscedastic method for heteroscedastic data usually results in an inflated false positive rate. On the other hand, if we use a heteroscedastic method for homoscedastic data, then the damage done is potentially small because the estimate of the slope parameter of the log-linear variance model (corresponding to the dose) would be small enough that the dose may not contribute much to the variance. Thus the presumed heteroscedastic model would be “close to” homoscedastic case. Hence, we arbitrarily chose $\boldsymbol{\alpha}$ to be 0.50.

The asymptotic covariance matrix of $\widehat{\theta}_n^{\text{PT}}$, derived in Lim et al. (2012), is as follows:

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$\begin{aligned}\mathrm{Var}(\widehat{\theta}_n^{\mathrm{PT}}) &= \alpha^*\mathrm{Var}(\tilde{\theta}_n)+(1-\alpha^*)\mathrm{Var}(\widehat{\theta}_n)\\ &= \alpha^*\left(\tfrac{1}{n}\mathbf{\Gamma}_{4n}(\theta)\right)^{-1}\ \left(\tfrac{1}{n}\mathbf{\Gamma}_{33n}(\theta)\right)\ \left(\tfrac{1}{n}\mathbf{\Gamma}_{4n}(\theta)\right)^{-1}\ +(1-\alpha^*)\left(\tfrac{1}{n}\mathbf{\Gamma}_{1n}(\theta,\boldsymbol{\tau})\right)^{-1}\ \left(\tfrac{1}{n}\mathbf{\Gamma}_{31n}(\theta,\boldsymbol{\tau})\right)\ \left(\tfrac{1}{n}\mathbf{\Gamma}_{1n}(\theta,\boldsymbol{\tau})\right)^{-1},\end{aligned}$$

where $\alpha^*=F_t\left(t_{\alpha/2,n-2}-\tau_1/\sqrt{\mathrm{Var}\,(\widehat{\tau}_{1n})}\right)-F_t\left(-t_{\alpha/2,n-2}-\tau_1/\sqrt{\mathrm{Var}\,(\widehat{\tau}_{1n})}\right)$, F_t is the cumulative distribution function of the t-distribution with $n-2$ degrees of freedom, and $\mathbf{\Gamma}_{1n}\,(\boldsymbol{\theta},\ \boldsymbol{\tau})$, $\Gamma_{31n}\,(\boldsymbol{\theta},\ \boldsymbol{\tau})$, $\Gamma_{33n}\,(\boldsymbol{\theta})$ and $\Gamma_{4n}\,(\boldsymbol{\theta})$ are defined in Appendix A.

The PTE based test statistic, which is robust to heteroscedasticity as well as outliers and influential observations, is given by $L^{\mathrm{PT}} = L^{\mathrm{OME}}\ I(|T_n| \le t_{\alpha/2,n-2}) + L^{\mathrm{WME}}\ I(|T_n| > t_{\alpha/2,n-2})$. Using the fact that OME and WME are consistent and asymptotically normally distributed under the regularity conditions (Lim et al. 2012), we obtain approximate critical values for L^{WME} and L^{PT} in the following theorem.

Theorem 1

Suppose $\varepsilon_{ij} \sim^{\mathit{independent}} N(0, \exp\{2(\tau_0 + \tau_1 x_i)\})$. *Then,*

a. *Under the conditions* [S1] – [S9] in the supplementary material, *the distribution of* L^{WME} *under the null hypothesis* H_0 *can be approximated by the central F-distribution with* $(3,\ n-6)$ *degrees of freedom.*

b. *Under the conditions* [S1] – [S11] in the supplementary material, $\limsup_{n\to\infty} P(L^{PT} > f_{\alpha_1,3,n-6}|H_0) \le \alpha_1$ *where* $f_{\alpha_1,3,n-6}$ *is the upper* α_1 *percentile of F-distribution with* $(3,\ n-6)$ *degrees of freedom and* α_1 *is the significance level.*

The proof of the theorem is provided in Appendix B.

Our proposed methodology using OME, WME or PTE for screening chemicals in qHTS assays can be summarized as follows. For each chemical test (3) using L^T (i.e., L^{OME}, L^{WME} or L^{PT}) with Bonferroni correction for multiple testing. Chemicals that are significant based on this test are declared to be active while the remaining ones are declared to be inactive.

Once a chemical is declared to be active, researchers are interested in estimating individual parameters of the Hill model along with their confidence intervals. Standard errors and the critical values for the confidence intervals using OME and WME methodologies are available from Lim et al. (2012). However, Lim et al. (2012) did not derive the critical values for the confidence intervals centered at PTE, which are provided in the following theorem.

Theorem 2

For $i = 1,\ldots, 4$ *define* $c_{\alpha_1}=t_{\alpha_1/2,n-6}\max\{SE(\tilde{\theta}_i),SE(\widehat{\theta}_i)\}/SE(\widehat{\theta}_i^{PT})$. *Then, under the conditions* [S1] – [S11] in the supplementary material, $\liminf_{n\to\infty} P(|\widehat{\theta}_i^{PT}-\theta_i|/SE(\widehat{\theta}_i^{PT})\le c_{\alpha_1})\ge 1-\alpha_1$.

The proof of the theorem is provided in Appendix B.

Note that we are constructing confidence intervals for parameters of those models that are selected by the above testing process. Consequently, one needs to be concerned about the overall coverage probability along the lines of Benjamini and Yekutieli (2005). Unfortunately, as noted earlier, the asymptotic p-values, especially those corresponding to far right tail of the distribution, are likely to be incorrect for small to moderate sample sizes.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Thus, under the null hypothesis, the p-values are not necessarily uniformly distributed. As a consequence, procedures such as those of Benjamini and Yekutieli (2005) are not directly applicable. This presents an interesting problem for future research.

3 SIMULATION STUDIES

In this simulation study we compare the proposed PTE based methodology with the two existing methods, namely, NCGC and Parham methodologies as well as compare it with OME and WME based methodologies. Note that the WME and OME methods are similar to the PTE method. In contrast, NCGC and Parham are completely different methods. Our primary criteria of comparison are the estimated false discovery rate (FDR) and the power. As usual, for a given method, the estimated FDR is the proportion of true null hypotheses rejected among all rejected hypotheses and the estimated power is the proportion of cases in which the null hypothesis is correctly rejected among all cases where the null is false.

Since investigators are also interested in estimating parameters of the Hill model for chemicals that are considered to be active, in this paper we also compare the performance of the estimators, OME, WME and PTE in terms of coverage probability.

3.1 Study design

The design of our simulation study was modeled after a real qHTS data set obtained from the NTP. Thus our simulation experiment consisted of 14 doses (0.59 nM, 2.94 nM, 14.7 nM, 32.8 nM, 73.4 nM, 0.164 μM, 0.367 μM, 0.821 μM, 1.835 μM, 4.103 μM, 9.175 μM, 20.52 μM, 45.87 μM, and 91.74 μM) and 3 observations per dose. We generated 10,000 "chemicals" of which $\gamma \times 100\%$ corresponded to "active" (i.e., none of the Hill parameters are zero) and the remaining $(1 - \gamma) \times 100\%$ were true nulls which corresponded to no change in mean response across the dose groups. We now describe the selection of θ for the non-null data in our simulation study. To keep the parameters of our simulation study consistent with the NTP's data on 1,408 chemicals, we first fitted a Hill model to each of these 1,408 compounds. From these we selected 100 patterns of curves that displayed various shapes of dose-response to get a reasonably broad selection of patterns. Since the observed responses of the NTP's qHTS data were normalized using the positive control (set to be −100%) and the vehicle control (set to be 0%) obtained using dimethylsulfoxide (DMSO) (the solvent used for compound transfer) only, and since NCGC uses 30% as the minimum response to be active, θ_1 was chosen in the interval (32, 112). Motivated by Parham et al. (2009), θ_2 was chosen to range from 0.8 to 4.9 and $\theta_3 < x_{\max}$. Hence the 100 parameter sets so chosen include a wide range of patterns of shapes of Hill curves (see online supplementary materials for the values). The descriptive statistics of the parameters are presented in Table 2. We generated the null data by taking the mean response to be zero across all doses.

The patterns of variances considered in our simulation study were also deduced from the 1,408 chemicals data. For the selected 100 compounds, we tested for homoscedasticity and estimated the standard deviation for homoscedastic data sets. The range of the estimated standard deviation is between 2.9 and 48.6 and the median is 5.3. Thus, we considered five patterns of homoscedastic errors with standard deviation, $\sigma = 4$, 6, 8, 15 and 30, respectively.

To obtain patterns of variances to model heteroscedasticity, we fitted log-linear model for the variance (as described in the previous section) to the heteroscedastic data sets among the selected 100 compounds. The range of the estimated slope $\hat{\tau}_1$ of the log standard deviation is between −0.034 and 0.027 and the median is 0.007. Then, we arrived at four patterns of parameters, namely, $(\tau_0, \tau_1) = (1.5, -0.015)$, $(1, 0.01)$, $(1, 0.02)$ and $(3, 0.01)$. The ranges of

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

σ_i with the choice of these parameters were (1.132, 4.482), (2.718, 6.803), (2.718, 17.027) and (20.086, 50.270), respectively.

In practice one generally does not know *a priori* what proportion of chemicals are heteroscedastic in a given qHTS assay. To keep our simulations realistic, we considered three patterns of proportion of heteroscedastic chemicals in a given run, namely, 10%, 25% or 50% heteroscedastic chemicals. The remaining are homoscedastic.

In summary we have 100 patterns of parameters with 9 patterns of variances for non-null data and 9 patterns of variances for null data. Thus we created a universe of 90,000 non-null data by generating 100 data sets for each of the 900 combinations of the patterns, and 9,000 null data by generating 1,000 data sets for each of the 9 patterns. From these 99,000 data sets, we obtained a random sample of 10,000 “chemicals” of which $\gamma \times 10,000$ were non-null and the remaining were null patterns. We repeated the simulation study 100 times and estimated the FDR and the power. We considered two different patterns of $\gamma = 0.05$ and 0.10 and three different patterns of the proportion of heteroscedastic data (as stated above).

For comparing OME, WME and PTE in terms of the coverage probability and the length of confidence interval, we used non-null data only. From 90,000 non-null data sets, we obtained a random sample of 10,000 non-null chemicals with three different patterns of proportions of heteroscedastic chemicals, which were 0.10, 0.25 or 0.50. Thus, either 10% or 25% or 50% of the non-null chemicals were heteroscedastic.

3.2 Results

The estimated FDR and power for the methods when $\gamma = 0.10$ are summarized in Figure 3. The results for the case of $\gamma = 0.05$ are provided in the supplementary material. The standard error of all FDR and power estimates provided in these figures was less than 0.005. In all the cases, the Parham method had a very high FDR compared to other methods. For example, when the proportion of non-null data is 0.10 and the proportion of heteroscedastic data is 0.50, the overall FDR of the Parham method is 0.421 and its FDR for heteroscedastic data is 0.475. On the other end of the spectrum, the NCGC method produced zero FDR in all cases. Accordingly it has a very low power. Additionally, we see from Table 3, both NCGC and Parham methods tend to declare a large proportion of chemicals to be inconclusive or marginal, thus requiring additional testing and resources to evaluate such chemicals.

The three alternative methods discussed in this paper, namely, OME, WME and PTE, had a better control of FDR in comparison to Parham method while maintaining reasonably good power in all the cases considered here. They were uniformly more powerful than NCGC method. The power of OME, WME and PTE based methods were almost same in every case but they differed in terms of FDR. As expected, the OME has the smallest FDR when the data are homoscedastic and generally has the largest FDR when the data are heteroscedastic. Overall, it has a larger FDR than WME and PTE based methods and hence we do not recommend its use for qHTS data. Interestingly, WME and PTE perform very similarly. This is largely due to the fact that for homoscedastic data the estimate of τ_1 used in the WME methodology is small enough for WME to perform similar to OME (and hence similar to PTE). However, when estimating confidence intervals for active chemicals, our simulation study reveals that PTE methodology outperforms WME (as well as OME) in terms of coverage probability (Figure 4). The coverage probabilities based on PTE are closer to the nominal level (0.95) than those based on OME and WME. One exception is the generally elevated θ_2 levels, in which PTE can be slightly higher and WME is close to nominal. The OME is subject to severe under-coverage especially for heteroscedastic data. For example, the lowest coverage probability based on OME for θ_0 is as low as 0.78 when 25% of the data are heteroscedastic. The WME is also subject to under-coverage in several

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

cases, with the coverage probability as low as 0.90. On the other hand, the lowest coverage probability based on PTE is 0.93.

In summary, our simulation studies suggest that the proposed methodology based on PTE outperforms the existing methods of NCGC and Parham et al. (2009) by providing a better control of FDR while maintaining good power.

4 ANALYSIS OF HTS ASSAYS DATA

The NTP library of 1,408 compounds was established for evaluation in qHTS assays (Smith et al. 2007; Tice et al. 2007). These include solvents, preservatives, flavoring agents, therapeutic agents, inorganic and organic pollutants, pesticides, natural products, etc. Among 1,408 compounds, 1,353 are unique compounds and 55 are tested twice to evaluate the reproducibility of the assay. For details, we refer the reader to Xia et al. (2008).

We illustrate the proposed methodology using the HepG2 cell triplicate data obtained from the above experiment. For each compound, 14 different concentrations, listed previously in Section 3.1 and ranging from 0.00059 to 91.74 μM, were used. There were 3 replicates at each concentration thus resulting in a total sample size $n = 42$ observations.

The results of the analysis using all the methods discussed in this paper are summarized in Figure 5. According to our preliminary test 782 out of 1,408 compounds (56%) potentially have a heteroscedastic variance structure and the remaining 626 (44%) have a homoscedastic variance structure.

According to the NCGC method 5% of the 1,408 chemicals are active, while 26% are active according to the method of Parham et al. (2009). On the other hand, using the proposed OME, WME and PTE methodologies we discovered 19%, 17% and 17% of the 1,408 chemicals to be active, respectively. The NCGC method declared 88% chemicals as inconclusive while the Parham method declared 41% chemicals as marginal. For homoscedastic data, 5%, 27%, 18%, 16% and 15% of 626 chemicals are active while for heteroscedastic data, 4%, 25%, 21%, 18% and 18% of 782 chemicals are active according to the NCGC, Parham, OME, WME and PTE methods, respectively.

Venn diagrams of active chemicals declared by the various methods are provided in in Figure 6. From Figure 6(a), 205 chemicals were found to be active by OME, WME and PTE methods. Figure 6(b) shows that 66 chemicals were discovered by all three methods in the diagram.

5 DISCUSSION AND OPEN PROBELMS

With the advent of high throughput screening (HTS) assays and the need for fitting thousands of nonlinear regression models to classify chemicals into various toxicity categories, it is important that statistical methods are developed which are robust to various assumptions commonly made in classical regression analysis. This paper takes the first step in this direction. Based on the simulation studies and the example of real qHTS data (Section 4), it appears that the proposed methodology performs better than two currently available methods in terms of the false discovery rate and the power, which is desirable for the analysis of qHTS assays data. Although the PTE and WME performed equally well in terms of FDR and power, the PTE as an estimator outperforms both OME and WME in terms of the coverage probability of confidence intervals.

In the course of this investigation we identified several important research problems for future research. As noted in the introduction, the asymptotic p-values derived in the context

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

of nonlinear models are not necessarily accurate for small to moderate sample sizes, especially those corresponding to small tail areas which are important for multiple testing problems. This problem is not unique to the present context, but may arise in other such multiple testing problems where asymptotic p-values are used. Unfortunately, resampling based procedures such as bootstrap/permutation methods are not practical in such situations for computational reasons.

A related problem is, for some data, especially the null data, the condition number of the information matrix can be extremely large. Note that, unlike linear models, in nonlinear models the information matrix is a function of the unknown parameters of the model. Toxicologists seem to recognize this issue and hence tend to discount/ignore data with either large slopes or large ED_{50} values since they can't trust those values (e.g., Parham et al. 2009). Analogous to SAM methodology for microarray data, in such cases we believe that the testing procedure can perhaps be modified by considering a shrinkage or a ridge type M-estimator in place of the regular M-estimator considered in this paper.

Although the theory of optimal designs is well developed in the case of linear and nonlinear models, it is has not been very well developed for high throughput screening assays with the exception of Qu (2010). Optimal designs developed in Qu (2010) are useful when the investigator is interested in making comparisons across chemicals. However, in the context described in this paper the problem of interest is not necessarily the comparison among thousands of chemicals but to screen chemicals that are potentially toxic. Hence the design issues discussed in Qu (2010) is not directly applicable here. Since the information matrix depends upon the unknown parameters of the model, the optimal design is not only a function of the dose but it is also a function of the unknown parameters of the model. The problem is exacerbated by the fact that the study involves not one nonlinear model but thousands of nonlinear models. Given that qHTS assays are being routinely conducted, it is important to derive suitable optimal designs. Since the condition number of the information matrix plays an important role, perhaps one may explore optimal designs by taking the objective function to be the expected value of the condition number of the information matrix. The expectation may be taken over a suitable prior distribution of $\boldsymbol{\theta}$, representing a wide range of chemicals.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported, in part, by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences [Z01 ES101744-04]. This work was initiated while Dr. Lim was a post-doctoral research fellow at NIEHS under the supervision of Dr. Peddada. We thank Drs Kissling, Shockely, the editor, the associate editor and the two anonymous referees for several helpful comments that led to substantial improvement in the content and presentation of the manuscript.

APPENDIX A: NOTATIONS

In this appendix we define the notations used in the paper.

i. $\boldsymbol{\Gamma}_{1n}(\theta,\boldsymbol{\tau})=\gamma_2\sum_{i=1}^{n}k^2(\mathbf{z}_i,\boldsymbol{\tau})\mathbf{f}_\theta(x_i,\theta)\mathbf{f}_\theta^t(x_i,\theta)$, where $\gamma_2 = E\psi'(\varepsilon)(\neq 0)$, $k(\mathbf{z}_i, \boldsymbol{\tau}) = 1/\boldsymbol{\sigma}(\mathbf{z}_i, \boldsymbol{\tau})$, and $\mathbf{f}_{\boldsymbol{\theta}}(x, \boldsymbol{\theta}) = (\partial/\partial\boldsymbol{\theta})\, f(x, \boldsymbol{\theta})$.

ii. $\boldsymbol{\Gamma}_{31n}(\theta,\boldsymbol{\tau})=\sigma_{\psi 1}^2\sum_{i=1}^{n}k^2(\mathbf{z}_i,\boldsymbol{\tau})\mathbf{f}_\theta(x_i,\theta)\mathbf{f}_\theta^t(x_i,\theta)$, where $\sigma_{\psi 1}^2=E\psi^2(\varepsilon)(<\infty)$.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

iii. $\boldsymbol{\Gamma}_{33n}(\theta)=\sigma^2_{\psi 3}\sum_{i=1}^{k} n_i w_1(x_i)\mathbf{f}_\theta(x_i,\theta)\mathbf{f}^t_\theta(x_i,\theta)$, where $\sigma^2_{\psi 3}w_1(x)=E\psi^2(\sigma(\mathbf{z},\boldsymbol{\tau})\varepsilon)(<\infty)$.

iv. $\boldsymbol{\Gamma}_{4n}(\theta)=\gamma_4\sum_{i=1}^{k} n_i\mathbf{f}_\theta(x_i,\theta)\mathbf{f}^t_\theta(x_i,\theta)$, where $\gamma_4 = E\psi'(\boldsymbol{\sigma}(\mathbf{z},\ \boldsymbol{\tau})\varepsilon)(\neq 0)$.

APPENDIX B: PROOFS OF THEOREMS

Proof of Theorem 1

(a) It is enough to show that the distribution of L^{WLSE} under H_0 is approximately central F-distribution with (3, $n - 6$) degrees of freedom. To begin with we assume $\boldsymbol{\tau}_1(\neq 0)$ is known. Let $y^*_{ij}=y_{ij}/\exp(\tau_1 x_i)$, $g(x_i, \boldsymbol{\theta}) = f(x_i, \boldsymbol{\theta})/\exp(\boldsymbol{\tau}_1\ x_i)$ and $\boldsymbol{\sigma}_0 = \exp(\boldsymbol{\tau}_0)$. Then, the nonlinear regression model (2) can be expressed by $y^*_{ij}=g(x_i,\theta)+\sigma_0\varepsilon_{ij}$. Then, $L^*_1=\sum_{i,j}(y^*_{ij}-g(x_i,\widehat{\theta}))^2=\eta^t_1(\mathbf{I}_n-\mathbf{H}^*)\eta_1+o_P(1)$, where $\boldsymbol{\eta}_1 = \mathbf{Y}^* - g(\boldsymbol{\theta})$, $\mathbf{Y}^*=(y^*_{11},\ldots,y^*_{k,n_k})^t$, $g(\boldsymbol{\theta}) = (g(x_1, \boldsymbol{\theta}), \ldots, g(x_k, \boldsymbol{\theta}))^t$ ($n \times 1$ vector), $\mathbf{H}^* = \mathbf{G}(\mathbf{G}^t\mathbf{G})^{-1}\mathbf{G}^t$ and $\mathbf{G} = \mathbf{g}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = \{\partial g(x_i, \boldsymbol{\theta})/\partial\theta_j\}$. Under H_0 in (3) the above regression model is replaced by $y^*_{ij}=\beta d_i+\sigma_0\varepsilon_{ij}$, where $d_i = 1/\exp(\boldsymbol{\tau}_1\ x_i)$. Then, $L^*_0=\sum_{i,j}(y^*_{ij}-\widehat{\beta}d_i)^2=\eta^t_0(\mathbf{I}_n-\mathbf{G}^0)\eta_0$, where $\widehat{\beta}=\sum_{i,j}d_i\,y^*_{ij}/\sum_{i,j}n_i d^2_i$, $\boldsymbol{\eta}_0 = \mathbf{Y}^* - \mathbf{D}\beta$, $\mathbf{D} = (d_1, \ldots, d_k)^t$ ($n \times 1$ vector) and $\mathbf{G}^0 = \mathbf{D}(\mathbf{D}^t\mathbf{D})^{-1}\mathbf{D}^t$. If we apply the standard theory of the LRT in nonlinear regression models, under H_0 $L^*=\{(L^*_0-L^*_1)/3\}/\{L^*_1/(n-6)\}$ is approximately following the central F-distribution with (3, $n - 6$) degrees of freedom.

Since $\boldsymbol{\tau}_1$ is unknown we replace it by its estimator $\hat{\boldsymbol{\tau}}_{1n}$. Noting that $\hat{\boldsymbol{\tau}}_{1n} = \boldsymbol{\tau}_1 + o_P(1)$ and hence $\exp(\hat{\boldsymbol{\tau}}_{1n}\ x_i) = \exp(\boldsymbol{\tau}_1 x_i) + o_P(1)$, we therefore have $L_1=L^*_1+o_P(1)$ and $L_0=L^*_0+o_P(1)$. Consequently, $L^{\mathrm{WLSE}} = L^* + o_P(1)$. Therefore, under the null hypothesis H_0, L^{WLSE} is approximately distributed as central F-distribution with (3, $n - 6$) degrees of freedom.

(b) We show that the Type I error for the test statistic L^{PT} is bounded by $\boldsymbol{a}_1$ in the following equation. Here the independence between T_n and (L^{WME}, L^{OME}) is needed, which can be deduced from the asymptotic joint normality of ($\tilde{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\tau}}_{1n}$) (Theorem 2, Lim et al. 2012).

$$\begin{gathered}
P(L^{\mathrm{PT}}>f_{\alpha_1,3,n-6}|H_0)\\
=P(L^{\mathrm{WME}}>f_{\alpha_1,3,n-6},|T_n|>t_{\alpha/2,n-2}|H_0)+P(L^{\mathrm{OME}}>f_{\alpha_1,3,n-6},|T_n|\leq t_{\alpha/2,n-2}|H_0)\\
=P(L^{\mathrm{WME}}>f_{\alpha_1,3,n-6}|H_0)P(|T_n|>t_{\alpha/2,n-2}|H_0)+P(L^{\mathrm{OME}}>f_{\alpha_1,3,n-6}|H_0)P(|T_n|\leq t_{\alpha/2,n-2}|H_0)\\
\leq P(L^{\mathrm{WME}}>f_{\alpha_1,3,n-6}|H_0)P(|T_n|>t_{\alpha/2,n-2}|H_0)+P(L^{\mathrm{OME}}>f_{\alpha_1,3,n-4}|H_0)P(|T_n|\leq t_{\alpha/2,n-2}|H_0)\\
=\alpha_1.
\end{gathered}$$

Proof of Theorem 2

For $i = 1, \ldots, 4$, from the independence between T_n and ($\tilde{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\theta}}$) (Theorem 2, Lim et al. 2012),

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

$$
\begin{aligned}
&P\left(\frac{|\widehat{\theta}_i^{PT}-\theta_i|}{SE(\widehat{\theta}_i^{PT})}\le c_{\alpha_1}\right)\\
&=P\left(|\widehat{\theta}_i-\theta_i|\le c_{\alpha_1}SE(\widehat{\theta}_i^{PT}),|T_n|>t_{\alpha/2,n-2}\right)+P\left(|\widetilde{\theta}_i-\theta_i|\le c_{\alpha_1}SE(\widehat{\theta}_i^{PT}),|T_n|\le t_{\alpha/2,n-2}\right)\\
&=(1-\alpha^*)P\left(|\widehat{\theta}_i-\theta_i|\le c_{\alpha_1}SE(\widehat{\theta}_i^{PT})\right)+\alpha^*P\left(|\tilde{\theta}_i-\theta_i|\le c_{\alpha_1}SE(\widehat{\theta}_i^{PT})\right)\\
&=(1-\alpha^*)P\left(|U|\le c_{\alpha_1}\frac{SE(\widehat{\theta}_i^{PT})}{SE(\widehat{\theta}_i)}\right)+\alpha^*P\left(|V|\le c_{\alpha_1}\frac{SE(\widehat{\theta}_i^{PT})}{SE(\widetilde{\theta}_i)}\right)\\
&\ge(1-\alpha^*)P\left(|U|\le c_{\alpha_1}\frac{SE(\widehat{\theta}_i^{PT})}{\max\{SE(\tilde{\theta}_i),SE(\widehat{\theta}_i)\}}\right)+\alpha^*P\left(|V|\le c_{\alpha_1}\frac{SE(\widehat{\theta}_i^{PT})}{\max\{SE(\tilde{\theta}_i),SE(\widehat{\theta}_i)\}}\right)\\
&=(1-\alpha^*)P\left(|U|\le t_{\alpha_1/2,n-6}\right)+\alpha^*P\left(|V|\le t_{\alpha_1/2,n-6}\right)\\
&\ge(1-\alpha^*)P\left(|U|\le t_{\alpha_1/2,n-6}\right)+\alpha^*P\left(|V|\le t_{\alpha_1/2,n-4}\right)=1-\alpha_1,
\end{aligned}
$$

where U and V are random variables following t-distributions with $n-6$ degrees of freedom and $n-4$ degrees of freedom, respectively.

References

1. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, Series B. 1995; 57(1):289–300.
2. Benjamini Y, Yekutieli Y. False discovery rate controlling confidence intervals for selected parameters. Journal of the American Statistical Association. 2005; 100(469):71–80.
3. Gallant, AR. Nonlinear Statistical Models. Wiley; New York: 1987.
4. Inglese J, Auld DS, Jadhav A, Johnson RL, Simeonov A, Yasgar A, Zheng W, Austin CP. Quantitative high-throughput screening: a titration-based approach that efficiently identifies biological activities in large chemical libraries. Proc Natl Acad Sci. 2006; 103:11473–11478. [PubMed: 16864780]
5. Lim C, Sen PK, Peddada SD. Accounting for uncertainty in heteroscedasticity in nonlinear regression. Journal of Statistical Planning and Inference. 2012; 142:1047–1062. [PubMed: 22345900]
6. Michael S, Auld D, Klumpp C, Jadhav A, Zheng W, Thorne N, Austin CP, Inglese J, Simeonov A. A robotic platform for quantitative high-throughput screening. ASSAY and Drug Dev Tech. 2008; 6(5):637–657.
7. Parham F, Austin C, Southall N, Huang R, Tice R, Portier C. Dose-response modeling of high-throughput screening data. J Biomolec Screening. 2009; 14(10):1216–1227.
8. Qu X. Optimal row-column designs in high-throughput screening experiments. Technometrics. 2010; 52(4):409–420.
9. Smith CS, Bucher J, Dearry A, Portier C, Tice RR, Witt K, et al. Chemical selection for NTP's high throughput screening initiative (Abstract). Toxicologist. 2007; 46:247.
10. Tice RR, Fostel J, Smith CS, Witt K, Freedman JH, Portier CJ, et al. The National Toxicology Program high throughput screening initiative: current status and future directions (Abstract). Toxicologist. 2007; 46:246.
11. Williams JD, Birch JB, Woodall WH, Ferry NM. Statistical monitoring of heteroscedastic dose-response profiles from high-throughput screening. J Agric Bio and Envir Statist. 2007; 12(2):216–235.
12. Xia M, Huang R, Witt KL, Southall N, Fostel J, Cho M-H, Jadhav A, Smith CS, Inglese J, Portier CJ, Tice RR, Austin CP. Compound cytotoxicity profiling using quantitative high-throughput screening. Env Health Persp. 2008; 116(3):284–291.
13. Zhang XD. 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 Biomolec Screening. 2007; 12(5):645–655.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

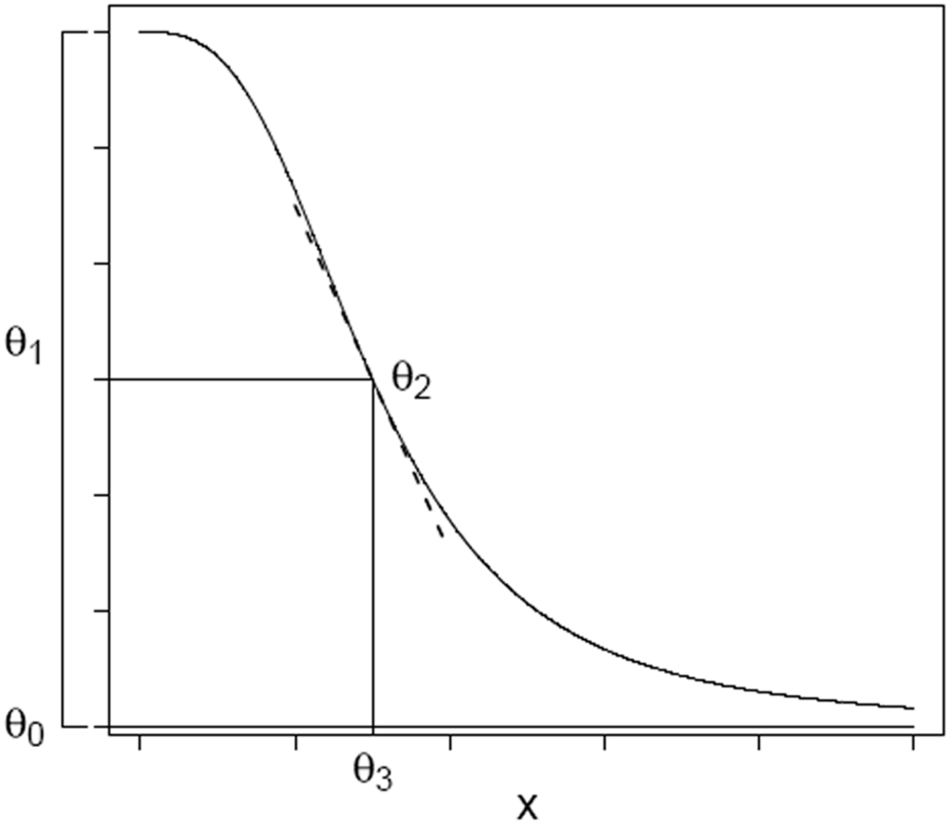
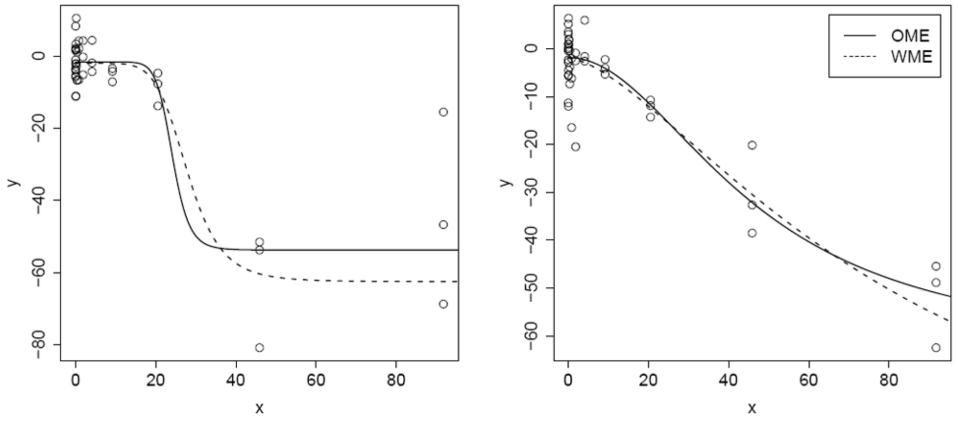


Figure 1.
Hill function

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript



(a) Heteroscedastic data (Chemical A)

(b) Homoscedastic data (Chemical B)

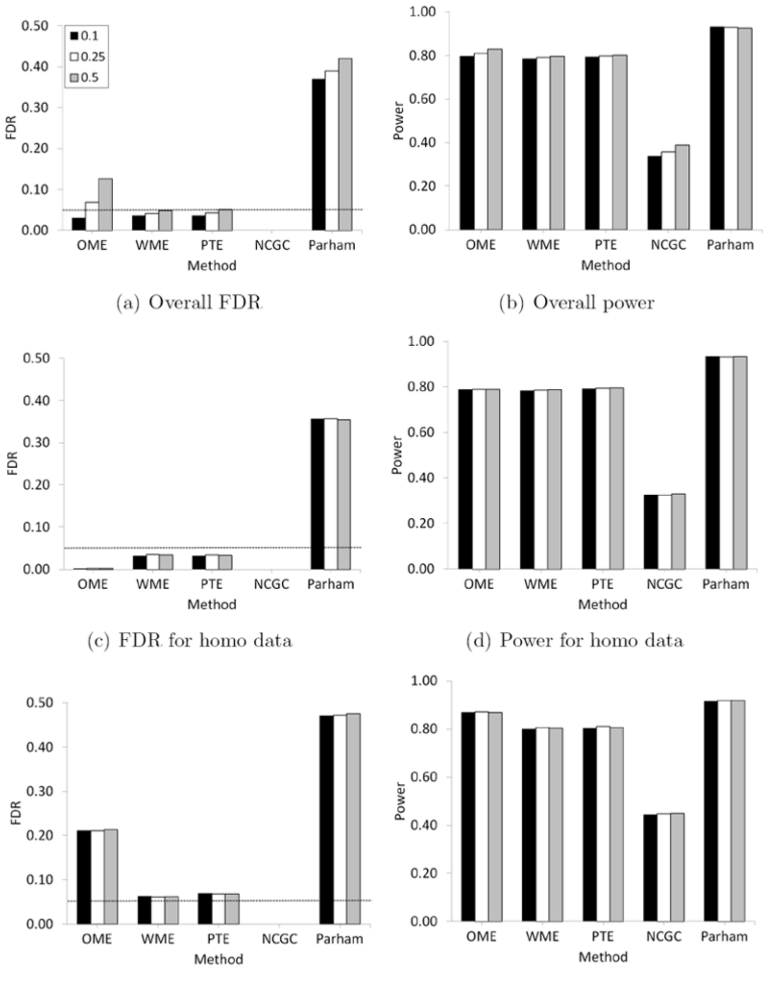
Figure 2.

HepG2 cell triplicate data, potentially (a) heteroscedastic and (b) homoscedastic, from qHTS assays; the corresponding fitted curves using OME and WME methods.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript



(a) Overall FDR

(b) Overall power

(c) FDR for homo data

(d) Power for homo data

(e) FDR for hetero data

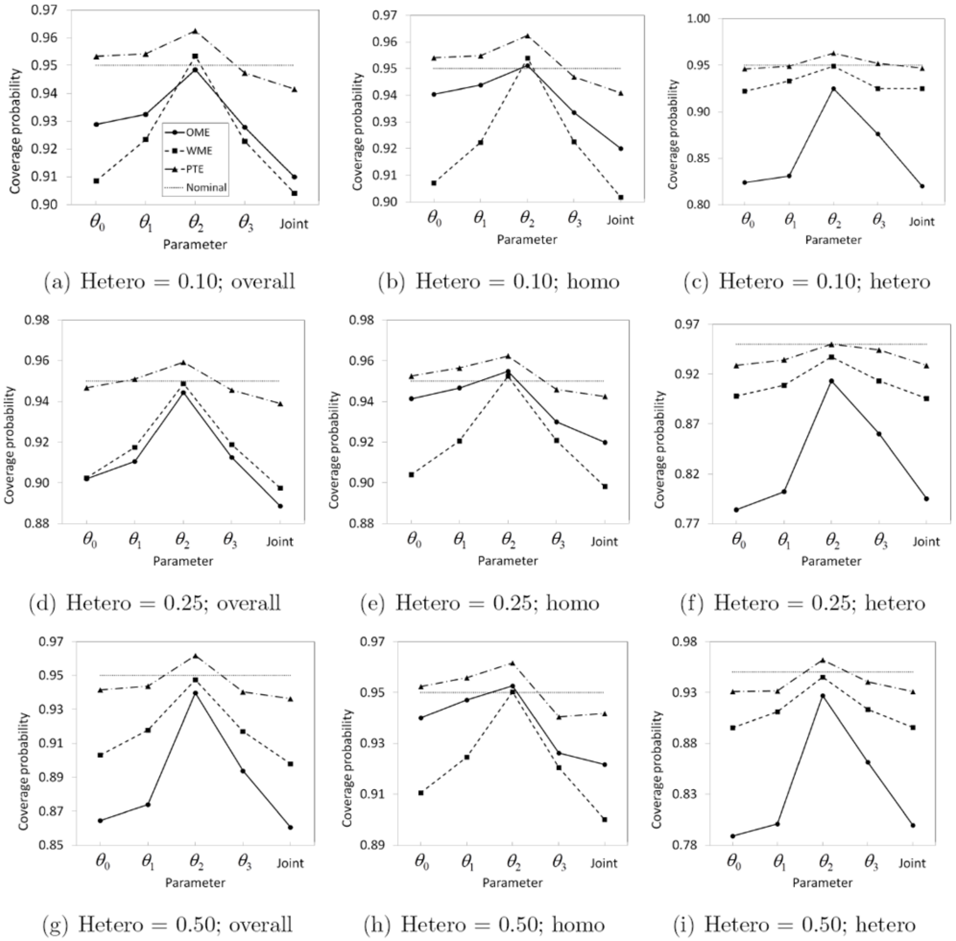
(f) Power for hetero data

Figure 3.
Estimated FDR and power for OME, WME, PTE, NCGC and Parham methods when the proportion of heteroscedastic data is 0.10, 0.25 and 0.50. Here $\gamma = 0.10$ and $\alpha = 0.05/10,000$.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript



(a) Hetero = 0.10; overall

(b) Hetero = 0.10; homo

(c) Hetero = 0.10; hetero

(d) Hetero = 0.25; overall

(e) Hetero = 0.25; homo

(f) Hetero = 0.25; hetero

(g) Hetero = 0.50; overall

(h) Hetero = 0.50; homo

(i) Hetero = 0.50; hetero

Figure 4.

Estimated coverage probability based on OME, WME and PTE methods for non-null (overall, homoscedastic or heteroscedastic) data with $1 - \alpha = 0.95$ when the proportion of heteroscedastic data is 0.10, 0.25 or 0.50.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

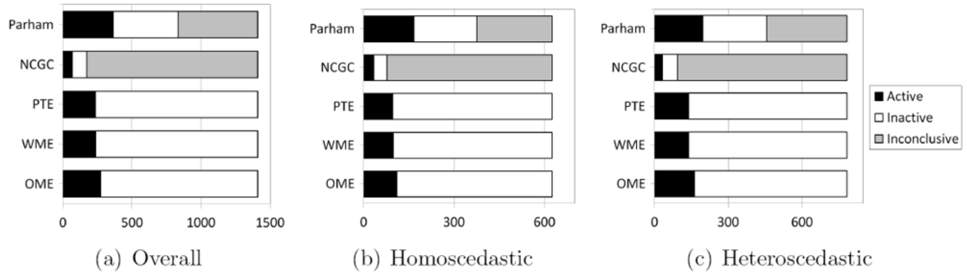


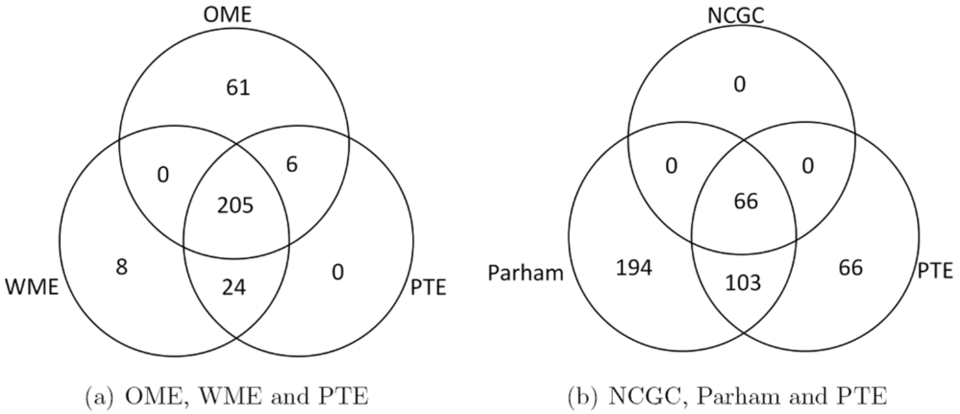
Figure 5.

Screening results for 1,408 chemicals of HepG2 cell triplicate data using the NCGC, Parham, the proposed method based on OME, WME and PTE with $\alpha = 0.05/1408$.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript



(a) OME, WME and PTE

(b) NCGC, Parham and PTE

Figure 6.

Venn diagrams of active chemicals using (a) OME, WME and PTE methods and (b) NCGC, Parham and PTE methods from qHTS assays.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 1

Estimate and Standard Error for parameters of the models for HepG2 cell triplicate data using OME and WME methods.

		Chemical A (possibly heteroscedastic)			Chemical B (possibly homoscedastic)		
PAR	**Method**	**EST**	**SE**	***p*-value**	**EST**	**SE**	***p*-value**
θ_0	OME	-53.7	3.5	0.000	-63.7	15.7	0.000
	WME	-62.5	3.1	0.000	-103.1	74.8	0.088
θ_1	OME	52.1	3.7	0.000	61.8	15.9	0.000
	WME	60.4	3.3	0.000	101.2	75.1	0.093
θ_2	OME	11.6	201.3	0.477	2.0	0.7	0.004
	WME	6.6	2.6	0.008	1.5	0.6	0.009
θ_3	OME	24.1	68.3	0.363	46.8	15.3	0.002
	WME	27.9	3.7	0.000	84.7	82.6	0.156

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 2

Summary statistics of 100 sets of parameters used in the simulation study.

Parameter	Range	Mean	Standard deviation
θ_0	(−117, −31)	−74.420	26.377
θ_1	(32, 112)	71.110	26.237
θ_2	(0.8, 4.9)	2.467	1.060
θ_3	(0.055, 83.49)	30.826	21.467

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 3

Estimated proportion of inconclusive (marginal) data among non-null and null data based on NCGC and Parham methods with $\alpha = 0.05/10, 000$.

γ	Hetero.	Method	Overall		Homoscedastic		Heteroscedastic	
			Non-null	Null	Non-null	Null	Non-null	Null
0.10	0.10	NCGC	0.637	0.951	0.650	0.952	0.524	0.943
		Parham	0.069	0.285	0.068	0.295	0.077	0.198
	0.25	NCGC	0.616	0.950	0.647	0.952	0.521	0.943
		Parham	0.069	0.270	0.067	0.293	0.074	0.201
	0.50	NCGC	0.584	0.948	0.643	0.952	0.525	0.943
		Parham	0.072	0.246	0.066	0.294	0.077	0.199