

# Analysis of quantitative high throughput screening data using a robust method for nonlinear mixed effects models

Chorong Park<sup>a</sup>, Jongga Lee<sup>a</sup>, Changwon Lim<sup>1,a</sup>

<sup>a</sup>Department of Applied Statistics, Chung-Ang University, Korea

---

## Abstract

Quantitative high throughput screening (qHTS) assays are used to assess toxicity for many chemicals in a short period by collectively analyzing them at several concentrations. Data are routinely analyzed using nonlinear regression models; however, we propose a new method to analyze qHTS data using a nonlinear mixed effects model. qHTS data are generated by repeating the same experiment several times for each chemical; therefore, they can be viewed as if they are repeated measures data and hence analyzed using a nonlinear mixed effects model which accounts for both intra- and inter-individual variabilities. Furthermore, we apply a one-step approach incorporating robust estimation methods to estimate fixed effect parameters and the variance-covariance structure since outliers or influential observations are not uncommon in qHTS data. The toxicity of chemicals from a qHTS assay is classified based on the significance of a parameter related to the efficacy of the chemicals using the proposed method. We evaluate the performance of the proposed method in terms of power and false discovery rate using simulation studies comparing with one existing method. The proposed method is illustrated using a dataset obtained from the National Toxicology Program.

**Keywords:** nonlinear mixed effects model, robust estimation, quantitative high throughput screening

---

## 1. Introduction

Quantitative high throughput screening (qHTS) assays take advantage of robotic technology and high sensitivity detectors (Michael *et al.*, 2008) to assess the potential toxicity of several chemicals in a short period. Each library chemical is tested at multiple concentrations to generate the assay dataset. For example, the NCATS Chemical Genomics Center (NCGC) developed a qHTS paradigm to generate dose-response curves for more than 60,000 chemicals in a single experiment (Inglese *et al.*, 2006). This development allowed all chemicals to be screened for dose dependent responses that allowed higher accuracy assessment of biological cell activity (Xia *et al.*, 2008). Data used in the current paper are qHTS for BG1 ER luciferase reporter gene assays. The US Tox21 program screened a library of 8,306 chemicals for ER reporter gene cell lines (Huang *et al.*, 2014), with each chemical repeatedly measured 1–48 times at 15 concentrations from 1.1nM–92 $\mu$ M. We are particularly interested in trans-activation activity occurring along estrogen signaling pathways measuring luminescence. Responses were normalized relative to positive and vehicle controls (Huang *et al.*, 2014).

Many previous studies have considered qHTS data to assess potential toxicity. Shockley (2015) mentioned that qHTS data could be analyzed using the Hill model, a nonlinear regression model,

---

Chorong Park and Jongga Lee equally contributed to this work.

<sup>1</sup> Corresponding author: Department of Applied Statistics, Chung-Ang University, 47, Heukseok-ro, Dongjak-Gu, Seoul 06974, Korea. E-mail: [clim@cau.ac.kr](mailto:clim@cau.ac.kr)

Published 30 November 2020 / journal homepage: <http://csam.or.kr>

© 2020 The Korean Statistical Society, and Korean International Statistical Society. All rights reserved.

with significant potential for toxicity testing. He suggested that a robust method be developed to take full advantage of qHTS assays, to overcome large uncertainties in parameter estimation when fitting nonlinear models to qHTS data. Lim *et al.* (2013a) evaluated qHTS assays using the Hill model and made decisions regarding toxicity using parameter estimates from the models. They proposed a preliminary test estimation (PTE) based methodology that was robust to the variance structure as well as any potential outliers and influential observations. Thomas *et al.* (2012) developed a cross-validation model comparison to evaluate the predictive performance of HTS assays and predicted hazards using various statistical classification models, including distance scoring, general linear model, k-nearest neighbor, logistic regression model, and partition tree.

Several methods have been proposed to screen toxic chemicals using qHTS data. Inglese *et al.* (2006) estimated parameters using an OLSE method by fitting a nonlinear regression model to the qHTS data. Fitted dose-response curves were subsequently categorized into classes using the parameter estimates and the toxic chemicals were declared based on the coefficient of determination ( $r^2$ ); number of asymptotes; dose level corresponding to 50% of maximum response change ( $ED_{50}$ ); and efficacy of the chemicals ( $E_{max}$ ). Parham *et al.* (2009) suggested that there could be some variation depending on the location of chemicals being treated on the plate (i.e., the experimental container). To eliminate this variation, they proposed a model to normalize the data, estimating the parameters using maximum likelihood estimation (MLE) by fitting the normalization model, and performed a likelihood ratio test for the parameters. Subsequently, the fitted dose-response curves were divided into several categories and toxic chemicals were classified based on criteria similar to Inglese *et al.* (2006). Shockley (2012) proposed a three-stage algorithm to analyze thousands of chemicals collectively and classify them into three categories.

Repeated measures data are generated by observing a number of individuals repeatedly under differing experimental conditions, where individuals are assumed to constitute a random sample from a population of interest (Lindstrom and Bate, 1990). For the qHTS assays in the current study, each chemical was measured 1–48 times at 15 concentrations. Therefore, for an analysis purposes these qHTS data may be technically treated as repeated measures data taken on multiple subjects, even though they have commonly been evaluated using nonlinear regression models.

Nonlinear mixed effects models are generally used to analyze repeated measures data. For example, Strathe *et al.* (2010) applied a nonlinear mixed effects model to a growth model using repeated measurements of pig bodyweight every two weeks. Parameters for nonlinear mixed effects models are generally estimated using Gaussian quasi maximum likelihood methods. However, these estimates can be significantly inaccurate in the presence of outliers or influential observations. Several robust estimation methods have been proposed to minimize the influences. Mancini *et al.* (2005) proposed a robust M-estimation that assigns a lower weight to outliers than the Gaussian maximum likelihood estimation. Yeap and Davidian (2001) proposed a robust two-stage estimation in a nonlinear mixed effects model to accommodate outliers or influential observations in a subject as measurements. Meza *et al.* (2012) proposed a robust method to infer parameters by assuming heavy tailed elliptical distributions for both random effects and errors. Williams *et al.* (2015) proposed a one-step approach by linearizing Gaussian likelihood for nonlinear mixed effects models.

Outliers often occur in qHTS assay data; therefore, this paper analyzes qHTS assay data by introducing the one-step approach proposed by Williams *et al.* (2015) for nonlinear mixed effects models. We assessed potential toxicity of various chemicals based on statistical significance. The methods proposed by Inglese *et al.* (2006) and Parham *et al.* (2009) considered only estimated parameter values, rather than the statistical significance. Even if a parameter such as chemical efficacy is large, it should be considered as 0 if it is not statistically significant. Therefore, we classified chemical toxicity

based on the significance of estimated parameters.

The remainder of this paper is structured as follows. Section 2 describes the basic nonlinear mixed effects model, current methods, one-step approach and the proposed method to analyze qHTS data. The performance of the proposed method is compared with that of an existing method on the basis of simulation studies described in Section 3. Section 4 summarizes the results of applying the proposed method to qHTS data. We conclude the paper and discuss potential future research topics in Section 5.

## 2. Methods

### 2.1. Assay description

The qHTS assay considered in this study is the “TOX21\_ERa\_LUC\_BG1\_Agonist” assay from the Tox21 library. The assay identifies small molecule agonists of the estrogen receptor alpha (ER-alpha) signaling pathway using the BG1 cell line. Chemical interactions with estrogen receptor (ER) disrupts normal endocrine function. Therefore, it is important to understand the effect of environmental chemicals on the ER signaling pathway to effectively manage endocrine disrupting chemicals (EDCs). The BG1-Luc-4E2 cell line was employed in the TOX21\_ERa\_LUC\_BG1\_Agonist assay to identify ER agonists, and has been used to screen the Tox21 10K compound library. BG1-Luc-4E2 cell line endogenously expresses full length ER-alpha and is stably transfected with a plasmid containing four estrogen responsive elements (ERE) upstream of a luciferase reporter gene (<https://tripod.nih.gov/tox/apps/assays/>).

### 2.2. Hill model

The Hill model is a nonlinear model primarily used to describe the shape of dose-response curves in biology, chemistry, and pharmacology. We apply the Hill model for each chemical since qHTS data are generated by repeating the same experiment at various concentrations for thousands of chemicals. The Hill model can be expressed by

$$f(x, \beta_0, \beta_1, \beta_2, \beta_3) = \beta_0 + \frac{\beta_1}{1 + \exp\left(\frac{\beta_3 - \log(x)}{\beta_2}\right)}. \quad (2.1)$$

Each parameter in Hill model has a specific bio-chemical meaning (Figure 1).  $\beta_0$  is the lower asymptote of the dose-response curve;  $\beta_1$  is the difference between the upper and lower asymptote of the curve, also known as efficacy of the chemical;  $\beta_3$  is the logarithm of the dose corresponding to 50% of maximum response change, also known as  $\log(\text{ED}_{50})$ ; and  $\beta_2$  represents the steepness of the curve which is mathematically the slope of the tangent line at  $\log(x) = \beta_3$ . If  $\beta_2$  is positive, the response curve increases with increasing dose, whereas if negative, the response curve decreases with increasing dose.

### 2.3. Nonlinear mixed effects model

Repeated measures data are analyzed using a nonlinear mixed effects model, which allows two types of correlation: correlation among random effects within each subject, and correlation in the within-subject variance-covariance matrix of errors (Williams *et al.*, 2015). The nonlinear mixed effects model can be expressed as:

$$y_{ij} = f(x_{ij}, \boldsymbol{\beta}, \mathbf{u}_i) + \epsilon_{ij}, \quad i = 1, \dots, s, \quad j = 1, \dots, n_i, \quad (2.2)$$

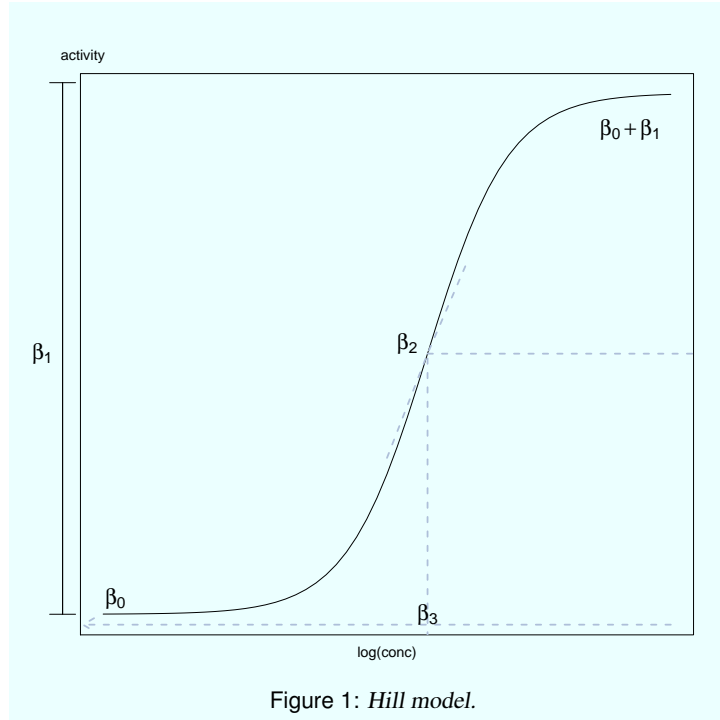


Figure 1: Hill model.

where  $x_{ij}$  is independent variable (dose) for the  $j^{\text{th}}$  response of the  $i^{\text{th}}$  subject;  $s$  is the number of subjects;  $n_i$  is the number of responses in the  $i^{\text{th}}$  subject;  $\boldsymbol{\beta}$  is the  $p \times 1$  vector of fixed effects;  $\mathbf{u}_i = (u_{i1}, u_{i2}, \dots, u_{iq})'$  is the  $q \times 1$  vector of random effects for the  $i^{\text{th}}$  subject; and  $q$  is the number of random effects in subject  $i$ .

We can generally assume that the number of random effects and the variance structure of random effects are the same for all subjects, respectively (Williams *et al.*, 2015). Therefore, the responses for each subject can be stacked, and Equation (2.2) can be expressed as:

$$\mathbf{Y} = f(\mathbf{X}, \boldsymbol{\beta}, \mathbf{u}) + \boldsymbol{\epsilon}, \quad (2.3)$$

where

$$f(\mathbf{X}, \boldsymbol{\beta}, \mathbf{u}) = \begin{pmatrix} f(X_1, \boldsymbol{\beta}, \mathbf{u}_1) \\ f(X_2, \boldsymbol{\beta}, \mathbf{u}_2) \\ \vdots \\ f(X_s, \boldsymbol{\beta}, \mathbf{u}_s) \end{pmatrix},$$

$\mathbf{X}_i = (x_{i1}, x_{i2}, \dots, x_{in_i})'$ ,  $\mathbf{Y} = (y_{11}, y_{12}, \dots, y_{sn_s})'$ , and  $\boldsymbol{\epsilon} = (\epsilon_{11}, \epsilon_{12}, \dots, \epsilon_{sn_s})'$ . We assume that

$$\boldsymbol{\epsilon} \sim N_N(\mathbf{0}, \mathbf{R}),$$

where  $\mathbf{R} = \text{blockdiag}(\mathbf{R}_i)$ ,  $\mathbf{R}_i$  is the matrix whose elements are variances and covariances of the error vector conditional on the random effects, and  $N = \sum n_i$ . We also assume that

$$\mathbf{u} \sim N_Q(\mathbf{0}, \mathbf{G}),$$

where  $\mathbf{G} = \text{blockdiag}(\mathbf{G}_i)$ ,  $\mathbf{G}_i$  is the matrix whose elements are variances and covariances of the random effects, and  $\mathbf{Q} = sq$ .

Various methods have been proposed to estimate parameters for mixed effects models; however, we restrict this paper to maximum likelihood (ML) methods.

#### 2.4. Existing robust estimation of the nonlinear mixed effects model

This section introduces one of robust estimation methods, the one-step approach (Williams *et al.*, 2015) using linearized Gaussian likelihood, extending the method of linear mixed effects models (Gill, 2000).

##### 2.4.1. Linearization of nonlinear mixed effects model

We apply the multivariate first-order Taylor expansion to approximate  $f(\mathbf{X}, \boldsymbol{\beta}, \mathbf{u})$  (Williams *et al.*, 2015).

$$f(\mathbf{X}, \boldsymbol{\beta}, \mathbf{u}) \approx f(\mathbf{X}, \hat{\boldsymbol{\beta}}, \hat{\mathbf{u}}) + \mathbf{D}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + \mathbf{Z}(\mathbf{u} - \hat{\mathbf{u}}), \quad (2.4)$$

where

$$\mathbf{D} = \left( \frac{\partial}{\partial \boldsymbol{\beta}'} \right) f(\mathbf{X}, \boldsymbol{\beta}, \mathbf{u}), \quad \text{and} \quad \mathbf{Z} = \left( \frac{\partial}{\partial \mathbf{u}'} \right) f(\mathbf{X}, \boldsymbol{\beta}, \mathbf{u}).$$

Substituting Equation (2.4) into Equation (2.3) gives Equations (2.5)–(2.6).

$$\mathbf{Y} \approx f(\mathbf{X}, \hat{\boldsymbol{\beta}}, \hat{\mathbf{u}}) + \mathbf{D}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) + \mathbf{Z}(\mathbf{u} - \hat{\mathbf{u}}) + \boldsymbol{\epsilon}, \quad (2.5)$$

$$\mathbf{Y} - f(\mathbf{X}, \hat{\boldsymbol{\beta}}, \hat{\mathbf{u}}) + \mathbf{D}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\mathbf{u}} \approx \mathbf{D}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}. \quad (2.6)$$

If the left side of Equation (2.6) is replaced by  $\tilde{\mathbf{Y}}$ , then  $\tilde{\mathbf{Y}}$  become the pseudo response  $\tilde{\mathbf{Y}} \approx \mathbf{D}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}$ , and  $\tilde{\mathbf{Y}}$  has multivariate normal distribution with mean vector  $\mathbf{D}\boldsymbol{\beta}$  and covariance matrix  $\mathbf{V}(\boldsymbol{\theta})$ .

$$\mathbf{V}(\boldsymbol{\theta}) = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}, \quad (2.7)$$

where  $\boldsymbol{\theta} = (\boldsymbol{\theta}_G', \boldsymbol{\theta}_R')$ ;  $\boldsymbol{\theta}_R'$  is the variance parameter of fixed effects, and  $\boldsymbol{\theta}_G'$  is the variance parameter of random effects.

##### 2.4.2. Robust maximum likelihood estimation

Since  $\tilde{\mathbf{Y}}$  has multivariate normal distribution with mean vector  $\mathbf{D}\boldsymbol{\beta}$  and covariance matrix  $\mathbf{V}(\boldsymbol{\theta})$ , from Section 2.4.1, the likelihood function of  $\tilde{\mathbf{Y}}$  can be expressed by

$$l(\boldsymbol{\beta}, \boldsymbol{\theta} | \tilde{\mathbf{y}}) = -\frac{1}{2}N \log(2\pi) - \frac{1}{2} \log |\mathbf{V}(\boldsymbol{\theta})| - \mathbf{r}'\mathbf{r}, \quad (2.8)$$

where  $\mathbf{r} = \mathbf{V}(\boldsymbol{\theta})^{-1/2}(\tilde{\mathbf{y}} - \mathbf{D}\boldsymbol{\beta})$  is the studentized pseudo-residual vector. The likelihood function is robustified by replacing  $\mathbf{r}'\mathbf{r}$  with the sum of the Huber functions,  $\sum \rho(r_i)$ . This provides robust estimation using ML methods. The Huber  $\rho$  function (Huber and Ronchetti, 2009) is defined as:

$$\rho(r_i) = \begin{cases} \frac{1}{2}r_i^2, & |r_i| \leq c, \\ c|r_i| - \frac{1}{2}c^2, & \text{otherwise,} \end{cases} \quad (2.9)$$

where  $r_i$  is  $i^{\text{th}}$  element of the studentized pseudo-residual vector,  $\mathbf{r}$ ; and  $c$  is a tuning constant that indicates the degree of robustness to outliers or influential observations. Thus, the robustified likelihood function is described as:

$$l(\boldsymbol{\beta}, \boldsymbol{\theta}|\tilde{\mathbf{y}}) = -\frac{k}{2}N \log(2\pi) - \frac{k}{2} \log |V(\boldsymbol{\theta})| - \sum_{i=1}^N \rho(r_i), \quad (2.10)$$

where  $k = P(|r_i| \leq c)$ . The tuning constant  $c$  can be arbitrarily set, but normally  $c \in [1, 2]$  (Huber and Ronchetti, 2009). We calculate the robust estimation using Newton-Raphson algorithm (Williams *et al.*, 2015) iteratively.

$$\boldsymbol{\beta}^{(h+1)} = \boldsymbol{\beta}^{(h)} + \left(\mathbf{H}_{\boldsymbol{\beta}\boldsymbol{\beta}'}^{(h)}\right)^{-1} \frac{\partial}{\partial \boldsymbol{\beta}} l^{(h)}(\boldsymbol{\beta}, \boldsymbol{\theta}|\tilde{\mathbf{y}}), \quad (2.11)$$

and

$$\boldsymbol{\theta}^{(h+1)} = \boldsymbol{\theta}^{(h)} + \left(\mathbf{H}_{\boldsymbol{\theta}\boldsymbol{\theta}'}^{(h)}\right)^{-1} \frac{\partial}{\partial \boldsymbol{\theta}} l^{(h)}(\boldsymbol{\beta}, \boldsymbol{\theta}|\tilde{\mathbf{y}}), \quad (2.12)$$

where  $\mathbf{H}_{\boldsymbol{\beta}\boldsymbol{\beta}'}$  and  $\mathbf{H}_{\boldsymbol{\theta}\boldsymbol{\theta}'}$  are matrices of second derivatives of the robustified likelihood function with respect to  $\boldsymbol{\beta}$  and  $\boldsymbol{\theta}$ , respectively.

## 2.5. Proposed qHTS data analysis process

We propose a method to identify toxic chemicals using a series of qHTS data analyses, as shown in Figure 2 and are detailed as:

Step 1. Perform one sample  $t$ -test for each chemical

- a.  $H_0 : E(y) = \mu = 0$  vs.  $H_1 : E(y) = \mu \neq 0$
- b. Classify chemicals that fail to reject the null hypothesis as ‘Inactive’.
- c. If chemical rejects the null hypothesis, go to next Step.

Step 2. Fit a nonlinear regression model to each chemical and identify the sign of estimates  $\beta_2$ , i.e., slope of dose-response curve.

- a. Classify chemicals with negative estimates ( $\hat{\beta}_2 < 0$ ) as ‘Inconclusive’.
- b. If estimates are positive ( $\hat{\beta}_2 > 0$ ), go to next Step.

Step 3. Fit different models according to replication count for each chemical.

- a. Fit a nonlinear regression model to chemicals that were only measured once.
- b. Fit a nonlinear mixed effects model to chemicals that have at least two observations at each concentration

Step 4. After fitting a model to each chemical in Step 3, identify toxic chemicals based on the significance of  $\hat{\beta}_1$ .

- a. If  $\hat{\beta}_1$  is significant, classify as ‘Active’.

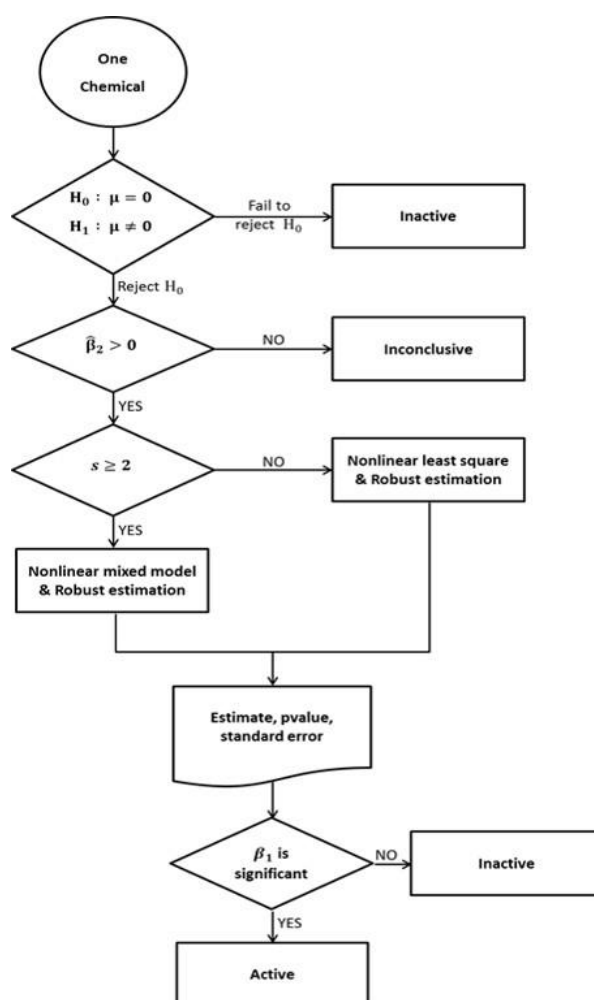


Figure 2: Proposed qHTS data analysis process.

b. If  $\hat{\beta}_1$  is not significant, classify as 'Inactive'.

When actually fitting the curve on the real data, the method above may not be ensured. For example, there are some cases where the model fit fails to converge. One possible reason for the failure might be that the plateau of the Hill curve is not specified by the data. Nevertheless, we see that the Hill function can still be used to represent the data. Thus, we have set a maximum iteration number to be 50 in all steps. If not converged, the result of the last iteration is returned. We checked for individual cases and confirmed that the number of cases is small. Most have converged successfully. Another remedy to address this issue for our proposed method is to have the first step where we conduct a simple  $t$ -test for the hypothesis of a constant mean response. We fit the Hill function only to the data for cases which pass the first step.

Another possible problem is that the covariance matrix of the estimated parameters is not positive definite. We observed that there are a few cases having the problem. To resolve this issue, we applied

a ridge regression type approach similar to the method proposed by Lim (2015). When a covariance matrix was not positive definite, we added a very small value to the diagonal elements of the Hessian matrix and re-calculated the covariance matrix, which turned to be positive definite.

Last, we know that starting values are very important for a nonlinear model to be converged and bad initial points lead the optimization procedure to converge to local minima. Therefore, it is often recommended to use a grid of starting values to ensure to fit the model correctly. However, from our long-term experience of the Hill function, we know how to select very good starting values. Using them showed that the model fit converged in most cases. Therefore, we did not utilize the grid method for starting values in this paper.

### 3. Simulation studies

In this simulation study we compare the proposed method with one existing method proposed by Parham *et al.* (2009). It uses a likelihood ratio test to test the hypothesis  $H_0 : \beta_1 = 0$  against  $H_0 : \beta_1 \neq 0$  and classifies chemicals into 3 categories as:

- 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  $H_0$  is not rejected or  $\hat{\theta}_2 < 0$ .
- Marginal: Neither classified as active nor inactive.

Our criteria of comparison are the power and the false discovery rate (FDR). Power is the ratio of predicted actives among true actives and FDR is the ratio of true inactives among predicted actives. High power with small FDR is preferred, while high power with high FDR is not.

#### 3.1. Study design

The design of the simulations was modeled after the real qHTS data considered in this paper. Logarithmic value of the concentration  $\log(x)$  in the real data ranges from approximately  $-23$  to  $-6$ , and experimented on 15 levels. Therefore, our simulation experiments have 15 values of  $x$  ranged from  $-22$  to  $-8$  equally separated in logarithmic scale. Since we consider a nonlinear mixed effects model, we generate data such that the number of repeated measurements is selected randomly among 1, 2, 3, 5, and 7. Accordingly, the sample size for each chemical is between 15 and 105. We generate 10,000 chemicals in total,  $\gamma \times 100\%$  of which are active and the others are inactive. To evaluate the methods for various situations having different number of active chemicals, we consider four scenarios of  $\gamma = 0.05, 0.1, 0.15$ , and  $0.2$ .

Active data are expected to follow the Hill model in Equation (2.1), and we generate data for active chemicals using the nonlinear mixed effects model Equation (2.2). We assume that  $\epsilon_{ij}$  are independent and identically distributed with  $N(0, \sigma^2)$ . The selection of  $\beta$  and  $\sigma$  emulates ones estimated in the real data in Section 4. Since there are cases where some estimated values are too large or too small, we considered these values as outliers and trimmed them out. The values of  $\beta$  are sampled from a normal distribution with mean vector  $(-2.2, 36, 1.1, -13.5)'$  and its covariance matrix  $\text{diag}(2, 120, 1, 2)$ . We set the covariance matrix to be diagonal since  $\beta_0, \beta_1, \beta_2$  and  $\beta_3$  are not significantly correlated ( $< 0.2$ ) for the real data. There is a chance that a sampled value is extreme since we sample from a normal distribution. Thus, the sampled values should be controlled so that they are placed within a certain range. For example, we set  $\beta_1$  to be greater than 30. Likewise, we make restrictions such that  $\beta_3$  is within



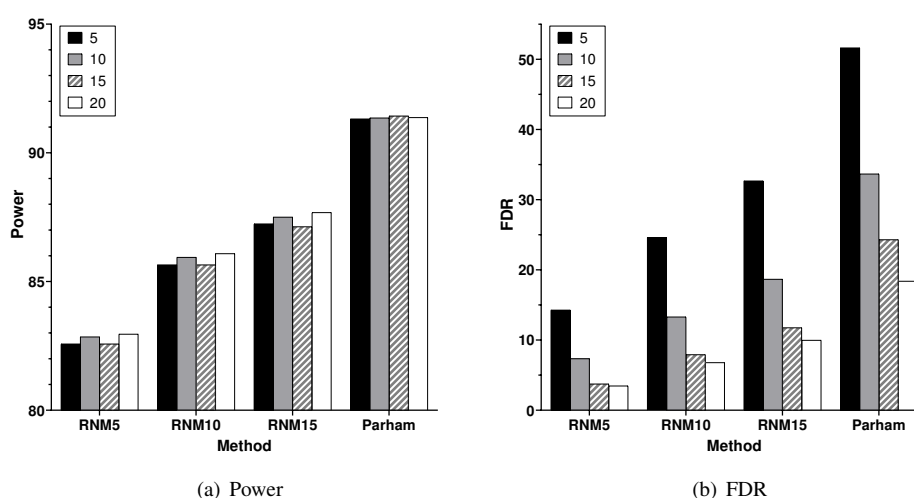


Figure 3: Estimated power and FDR for the proposed methods and Parham method.

(0.2, 1.5) and  $\beta_4$  within  $(-22, -8)$ . The range of  $\beta_3$  is motivated by Parham *et al.* (2009). Standard deviation  $\sigma$  is sampled among five values,  $\sigma = 4, 6, 8, 15, 20$  with probability (0.2, 0.3, 0.3, 0.15, 0.05). Finally, in order to sample the random effects, we need to set its covariance matrix, which is represented by  $\mathbf{G} = \text{blockdiag}(\mathbf{G}_i)$  in this paper. We first assume that  $\mathbf{G}_i = \text{diag}(\sigma_{u_0}^2, \sigma_{u_1}^2, \sigma_{u_2}^2, \sigma_{u_3}^2)$ ,  $i = 1, \dots, s$ . Then, again in reflection of real data, we make the variances either 10% or 20% of their corresponding fixed effects with probability 0.3 and 0.7, respectively. For example, if  $\beta_1$  is 60, then the variance of its corresponding random effect is either 6 or 12.

Inactive data do not follow the Hill model and scatter around 0, with no trends. They are sampled from normal distribution with mean 0 and  $\sigma = 4, 6, 8, 15, 20$  with probability (0.2, 0.3, 0.3, 0.15, 0.05).

At Step 1 of our method, we set the significance level of  $t$ -test to be 0.05, 0.1 and 0.15, denoted by RNM5, RNM10, RNM15, respectively, since the significance level may have an effect on power and FDR. The RNM stands for the robust nonlinear mixed effects model. The existing method that we compare our method with is denoted by Parham. We repeat the simulation 100 times and estimate the power and FDR for each of the four scenarios of  $\gamma = 0.05, 0.1, 0.15$ , and 0.2.

### 3.2. Simulation result

The estimated power and FDR are summarized in Figure 3. The standard errors of all power and FDR estimates provided in this figure are less than 2% and 21% of the estimates, respectively. In all cases, the Parham method has a very high FDR compared to our proposed methods. For example, when the proportion of active chemicals is 0.05, the FDR of the Parham method is 51.6%. However, those of RNM5, RNM10 and RNM15 are 14.3%, 24.6%, and 32.7%, respectively. As the proportion of active chemicals increases, the FDR decrease for all methods. However, even when the proportion is 0.2, the FDR of the Parham method is 18.4%, still quite high. On the other hand, those of RNM5, RNM10 and RNM15 are 3.47%, 6.80%, and 9.98%, respectively, which are much lower than that of the Parham method. The proposed methods outperform the Parham method in terms of the FDR. The RNM5 performs best.

The power of the Parham method is around 91% in all cases. On the other hand, those of RNM5,

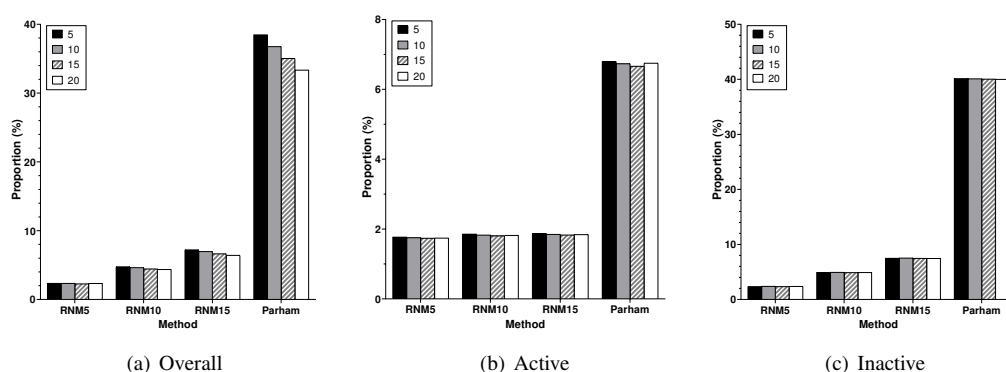


Figure 4: Estimated proportion of chemicals declared as inconclusive (marginal) among (a) overall (b) active (b) inactive chemicals based on the proposed methods and Parham method.

RNM10 and RNM15 are around 82%, 85% and 87%, respectively, in all cases. Although the proposed methods have lower power than the Parham method, they maintain reasonably good power in all cases. However, the Parham method has a much larger FDR than the proposed methods and hence it is not recommended. As described, the three variants of the proposed method are based on the significance level of the  $t$ -test at Step 1. Smaller significance level classifies more inactive chemicals at Step 1 that resulting in less false positive cases. Among the three, we recommend the RNM5 since it is the most conservative with good power.

Additionally, Figure 4 shows that the Parham method declare a much larger proportion of chemicals to be marginal than the proposed methods. For example, when the proportion of active chemicals is 0.10, the Parham method declared 40.10% of inactive chemicals as marginal whereas the RNM5 declared only 2.39% as inconclusive.

In summary, our simulation studies suggest that the proposed method outperforms the method proposed by Parham *et al.* (2009) by providing a better control of FDR while maintaining good power. The proposed method performs the best at a significance level  $\alpha = 0.05$  at Step 1. In this case, the Parham method has a five times higher FDR. Even though the Parham method has 10% higher power, the FDR is unacceptably high.

## 4. Data analysis

### 4.1. qHTS data

The US Tox21 program established a library of approximately 10,000 chemicals comprising commercial origin substances by the US Environmental Protection Agency, National Toxicology Program, and NCGC (Huang *et al.*, 2014).

The qHTS experiment in this paper automated screening in the 1,536-well plate format and tested each library chemical at multiple concentrations. The qHTS test vessel is the plate, incorporating a grid of 1,536 small open divots (Figure 5). The first and second rows of divots were positive controls, and third and fourth rows were vehicle controls. The remaining rows were treated with chemicals and measurements were normalized relative to the controls,

$$y = \frac{y^{\text{raw}} - V_{\text{vehicle}}}{V_{\text{positive}} - V_{\text{vehicle}}} \times 100, \quad (4.1)$$

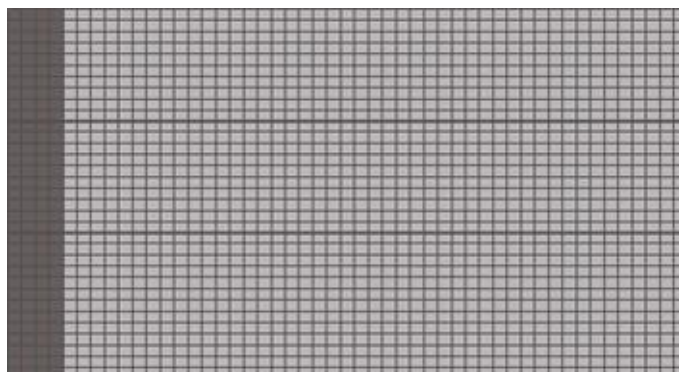


Figure 5: 1,536 well plate used for qHTS assays.

Table 1: Screening results for 11064 chemicals of BG1-Luc-4E2 cell agonist data using the proposed method (RMN5) and Parham method

Class	Count (%)	
	RMN5	Parham
Active	1,600 (14.5%)	1,978 (17.9%)
Inactive	6,711 (60.7%)	5,782 (53.3%)
Inconclusive (Marginal)	2,753 (24.8%)	3,304 (29.8%)
Total	11,064 (100%)	11,064 (100%)

where  $y^{\text{raw}}$  is response variable of interest, cell viability measuring protease activities within live cells,  $V_{\text{positive}}$  is median of the positive control wells, and  $V_{\text{vehicle}}$  is median of the vehicle control wells (Huang *et al.*, 2014).

The qHTS data included 8,306 chemicals and a variable that indicated the cluster by CASANOVA (within each chemical) a profile belonged to. CASANOVA refers to Cluster Analysis by Subgroups using ANOVA (Shockley *et al.*, 2019). For the same chemical, the cluster was classified according to the location where the profile was done. The cluster could take one of NA, 1, 2, 3, 4, or 5; each of which denotes where the profile belongs. If a chemical has a single cluster (i.e., cluster = 1 for all profiles), then the assay worked well; whereas if a chemical has a cluster = NA (i.e., there was no cluster available, hence the profile is within a noise band) then the measured data inadequately expresses the Hill model. We regarded the same chemical as being different depending on the cluster value. Therefore, the analysis was conducted for approximately 11,064 independent chemicals. Each chemical was tested at 15 concentrations from 1.1nM–92 $\mu$ M. We expect an increasing response with an increasing concentration since this is agonist data (Huang *et al.*, 2014).

Tests were run various times at same concentration, i.e., the number of replication was in (1, 48), hence sample size for each chemical was in (15, 720). The data set for each chemical was analyzed using the one-step approach for the nonlinear mixed effects model (Williams *et al.*, 2015).

## 4.2. Results

Table 1 summarizes the results of the analysis using the proposed method (RMN5) and Parham method. In Step 1 of the proposed method, among 11,064 chemicals, 5,270 chemicals classified as ‘Inactive’ because they failed to reject the null hypothesis in the one sample  $t$ -test. In Step 2, among 5,794 chemicals, 2,753 (24.8%) chemicals are classified as ‘Inconclusive’ because  $\hat{\beta}_2 < 0$ .

We expect an increasing response with an increasing dose since qHTS assays in current study was agonist data. Therefore, we classified chemicals with  $\hat{\beta}_2 < 0$  as 'Inconclusive'. In Step 3, among 3,041 chemicals, 557 chemicals replicated once at each dose and 2,464 replicated more than twice at each same dose. In Step 4, 272 chemicals among 577 were classified as 'Active' due to  $\hat{\beta}_1$  being significant, and 1,328 chemicals among 2,464 were classified as 'Active' due to  $\hat{\beta}_1$  being significant. The remaining were classified as 'Inactive'. Therefore, using the proposed method, we discovered 1,600 (14.5%) and 6,711 (60.7%) of the 11,064 chemicals to be active and inactive, respectively. On the other hand, according the Parham method, 1,978 (17.9%), 5,782 (53.3%), and 3,304 (29.8%) of the 11,064 chemicals are active, inactive, and marginal, respectively.

## 5. Conclusion

In this paper we analyzed quantitative high throughput screening (qHTS) data using a robust one-step approach, and assessed potential toxicity for a number of chemicals. The ToxCast pipeline (TCPL) package (Filer *et al.*, 2017) is the standard data analysis pipeline at the EPA. It fits all the data for a substance (typically 3 replicates at each concentration) into a single model. Instead, we used nonlinear mixed effects model, whose assumption might be more appropriate in qHTS data. Our method also allows for statistical significance, making our decision more reasonable. Even where a parameter has large value, the estimates may not differ from 0 unless it is statistically significant. Therefore, we proposed a method to classify chemical toxicity considering relevant parameter statistical significances.

We first perform a one sample *t*-test for each chemical and classify chemicals as 'Inactive' if they fail to reject the null hypothesis. We then fit nonlinear regression models to each chemical and classify chemicals with negative slope estimates as 'Inconclusive'. We fit different models depending on the replication level. For chemicals tested once, we fit nonlinear regression models, whereas we fit nonlinear mixed effects models for chemicals tested more than once at the same dose. Finally, we assess potential chemical toxicity from the parameter estimate,  $\hat{\beta}_1$ , i.e., chemical efficacy by statistical testing of the significance. Based on the simulation studies and the example of real qHTS data, it appears that the proposed method performs better than one existing method in terms of power and FDR.

Although outliers often occur in qHTS data, it is not easy to differentiate between outliers and natural variability in the data. However, we can still say that there may be potential outliers in the data and if there are any, then the standard least squares estimators are affected by them. Our goal is thus not to identify outliers, but to estimate parameters still relatively correctly even though there are outliers by using a robust method such as the M-estimator. It is well known that the M-estimator is robust to outliers or influential observations in linear regression. Recently, the authors introduced the M-estimation methods to nonlinear regression and showed that it is again robust to outliers in nonlinear regression through extensive simulations. For more details, see Lim (2015), Lim *et al.* (2012, 2013a, 2013b) among others.

The proposed qHTS data analysis process will be useful for classifying chemicals with toxicity and rapidly processing the collected 10,000 chemical datasets. The approach could subsequently be applied to other qHTS data libraries. We analyzed qHTS data using the one-step approach (Williams *et al.*, 2015), but several previous analysis methods have been proposed for qHTS data. Therefore, future investigation should consider simulations to compare the proposed method performance with previous methods. We have modeled on only the Hill model. Since not all dose-response curves follow this, we have made use of *t*-test which is expected to treat some non-Hill model cases to some extent. However, there are even non-monotonic curves. In addition, this paper implicitly assumed

the data was homoscedastic. Therefore, future studies should consider heteroscedastic datasets in addition to addressing the other topics mentioned.

## Acknowledgement

This research was supported both by Next-Generation Information Computing Development Program through the National Research Foundation (NRF) of Korea funded by the Ministry of Science, ICT [NRF-2017M3C4A7083281] and the Chung-Ang University Graduate Research Scholarship in 2019.

## References

- Filer DL, Kothiya P, Setzer RW, Judson RS, and Martin MT (2017). tcpl: the ToxCast pipeline for high-throughput screening data, *Bioinformatics*, **33**, 618–620.
- Gill PS (2000). A robust mixed linear model analysis for longitudinal data, *Statistics in Medicine*, **19**, 975–987.
- Huang R, Sakamuru S, Martin MT, *et al.* (2014). Profiling of the Tox21 10K compound library for agonists and antagonists of the estrogen receptor alpha signaling pathway, *Scientific Reports*, **4**, 5664.
- Huber PJ and Ronchetti EM (2009). *Robust Statistics* (2nd Ed), John Wiley & Sons, New York.
- Inglese J, Auld DS, Jadhav A, Johnson RL, Simeonov A, Yasgar A, Zheng W, and Austin CP (2006). Quantitative high-throughput screening: a titration-based approach that efficiently identifies biological activities in large chemical libraries, *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 11473–11478.
- Lim C (2015). Robust ridge regression estimators for nonlinear models with applications to high throughput screening assay data, *Statistics in Medicine*, **34**, 1185–1198.
- Lim C, Sen PK, and Peddada SD (2012). Accounting for uncertainty in heteroscedasticity in nonlinear regression, *Journal of Statistical Planning and Inference*, **142**, 1047–1062.
- Lim C, Sen PK, and Peddada SD (2013a). Robust analysis of high throughput screening (HTS) assay data, *Technometrics*, **55**, 150–160.
- Lim C, Sen PK, and Peddada SD (2013b). Robust nonlinear regression in applications, *Journal of the Indian Society of Agricultural Statistics* **67**, 215–234.
- Lindstrom ML and Bates DM (1990). Nonlinear mixed effects models for repeated measures data, *Biometrics*, **46**, 673–687.
- Mancini L, Ronchetti E, and Trojani F (2005). Optimal conditionally unbiased bounded-influence inference in dynamic location and scale models, *Journal of the American Statistical Association*, **100**, 628–641.
- Meza C, Osorio F, and Cruz RD (2012). Estimation in nonlinear mixed-effects models using heavy-tailed distributions, *Statistics and Computing*, **22**, 121–139.
- Michael S, Auld D, Klumpp C, Jadhav A, Zheng W, Thorne N, Austin CP, Inglese J, and Simeonov A (2008). A robotic platform for quantitative high-throughput screening, *ASSAY and Drug Development Technologies*, **6**, 637–657.
- Parham F, Austin C, Southall N, Huang R, Tice R, and Portier C (2009). Dose-response modeling of high-throughput screening data, *Journal of Biomolecular Screening*, **14**, 1216–1227.
- Shockley KR (2012). A three-stage algorithm to make toxicologically relevant activity calls from quantitative high throughput screening data, *Environmental Health Perspectives*, **120**, 1107–1115.
- Shockley KR (2015). Quantitative high-throughput screening data analysis: challenges and recent

- advances, *Drug Discovery Today*, **20**, 296–300.
- Shockley K, Gupta S, Harris S, Lahiri SN, and Peddada S (2019). Quality control of quantitative high throughput screening data, *Frontiers in Genetics*, **10**, 387.
- Strathe AB, Danfær A, Sørensen H, and Kebreab E (2010). A multilevel nonlinear mixed-effects approach to model growth in pigs, *Journal of Animal Science*, **88**, 638–649.
- Thomas RS, Black MB, Li L, Healy E, Chu TM, Bao W, Andersen MD, and Wolfinger RD (2012). A comprehensive statistical analysis of predicting in vivo hazard using high-throughput in vitro screening, *Toxicological Sciences* **128**, 398–417.
- Williams JD, Birch JB, and Abdel-Salam AS (2015). Outlier robust nonlinear mixed model estimation, *Statistics in Medicine*, **34**, 1304–1316.
- Xia M, Huang R, Witt KL, *et al.* (2008). Compound cytotoxicity profiling using quantitative high-throughput screening, *Environmental Health Perspectives*, **116**, 284–291.
- Yeap BY and Davidian M (2001). Robust two-stage estimation in hierarchical nonlinear models, *Biometrics*, **57**, 266–272.

*Received September 29, 2020; Revised September 24, 2020; Accepted November 1, 2020*