International Journal of Public Health Science (IJPHS)
Vol. 11, No. 3, September 2022, pp. 770~778
ISSN: 2252-8806, DOI: 10.11591/ijphs.v11i3.21282

Statistical model for IC_{50} determination of acetylcholinesterase enzyme for Alzheimer's disease

Anwar Fitrianto[1], Siau Man Mah[2], Siau Hui Mah[3]
[1]Department of Statistics, Faculty of Mathematics and Natural Sciences, IPB University, Bogor, Indonesia
[2]Pin Hwa High School, Klang, Malaysia
[3]School of Biosciences, Faculty of Health and Medical Sciences, Taylor's University, Subang Jaya, Malaysia

Article Info

Article history:

Received Sep 2, 2021
Revised Apr 29, 2022
Accepted Jun 20, 2022

Keywords:

Alzheimer's disease
Anticholinesterase
Regression
Synthetic compound

ABSTRACT

This study aimed to formulate a suitable statistical model to determine acetylcholinesterase enzyme's half-maximal inhibitory (IC_{50}) by a series of synthetic compounds. It was done with the same core structure for acetylcholinesterase inhibition for anti-Alzheimer's disease (AD). The IC_{50} of eighteen synthesized compounds on anticholinesterase activities was obtained and statistical methods were applied. Regression models were fitted to the dose-response curve to look for their IC_{50}. Simple linear regression is the simplest model for the dose-response curve. However, polynomial regression models or non-linear regression models fit the data more accurately. The adjusted coefficient of determination (R^2_{adj}) was used to determine the best model among the linear models, while the root mean square error (RMSE) is more suitable in determining the goodness of fit between linear and non-linear model. Four-parameter logistic (4-PLR) regression often fits the dose-response data closely. Based on the RMSE value, a polynomial regression fitted better than 4-PLR with the IC_{50} of 245.52.

Corresponding Author:

Anwar Fitrianto
Department of Statistics, Faculty of Mathematics and Natural Sciences, IPB University
Meranti Wing 22 Level 4, Dramaga, Bogor, 16680 West Java, Indonesia
Email: anwarstat@gmail.com

1. INTRODUCTION

Alzheimer's disease (AD) is a general term for dementia that denotes severe memory loss, language and other intellectual abilities. The exact cause is unknown but it is very likely related to environmental and genetic factors. This disease is 6th leading cause of death in the world [1], [2]. The majority of patients were 65 years of age or older. Symptoms develop slowly and get worse over time. Alzheimer's patients cannot be cured. Instead, treatment can be used to reduce the acceleration of the effects, which involves the treatment of cholinesterase inhibitors [3] to improve cognitive function.

Synthesis of a series of compounds with the same core skeleton is commonly used to improve their biological activities. Thus, it is good to discover compounds with good anti-acetylcholinesterase activities for Alzheimer's disease [2]. Multi-well plates are often used in many kinds of biological assays in pharmacological research. The parameter of interest is usually the concentration of drug dose corresponding to the median response. A measure of a drug's potency is the half-maximal inhibitory concentration (IC50), which is the amount of drug needed to inhibit a biological process by 50%, [4]. The lower the IC50 value, the lower the drug concentration is required [5]. A lower drug dose is preferred because a higher drug concentration may harm our health.

In fitting the dose-response curve, simple linear regression (SLR) is the most straightforward statistical tool to estimate the relationships among the variables. It studies the functional dependencies between factors. The term functional dependencies imply that the independent variable determines the value of the dependent variable. In some situations, simple linear regression does not hold. Hence, polynomial regression is introduced in which the relationship of independent variables and the dependent variable is modeled by nth degree polynomial. It fits a linear relationship between the variable *x* and *y*.

Non-linear regression is another powerful model that can fit virtually any curve. For example, logistic regression fits dose-response data and estimates the expected responses at different doses. A simple 2 parameter logistic regression model adequately summarizes how a binary response relates to dose. The dose-response curve has a lower asymptote of zero and an upper asymptote of one, the limits of the expected range of response probabilities. However, when the response variable is continuous, three or more parameters would be more appropriate [6]. For generalizing t2-parameter logistic regression model, two more parameters need to be added. The upper response asymptote may be less than one, and the lower response may be more than zero. The resulting four-parameter logistic (4-PLR) model increases flexibility at a higher optimization cost. The 4-PLR model [5] is illustrated as (1):

$$y = d + \frac{a-d}{1+\left(\frac{x}{c}\right)^b} \quad (1)$$

where *a* is the minimum value of the response variable, *b* is the hill slope, *c* is the inflection point, d is the maximum value of the response variable, *y* is the dependent variable and *x* is the concentration variable used. The hill slope refers to the rate at which the response decreases concerning the dose. The inflection point is the dose corresponding to the median response. The 4-PLR has several characteristics [7]. For instance, it is perfect symmetry for the sigmoidal curve (S-curve) at its inflection points. It assumes the response is symmetric about the median response and it is monotonic with either constantly increasing or decreasing with the *x* variable. The research objective is to formulate a suitable statistical model to determine the IC50 of a series of synthetic compounds with the same core skeleton for Alzheimer's disease.

2. LITERATURE REVIEW

2.1. Previous research on Alzheimer's disease

Many statistical methods have been widely used in the AD research field. An analytic method using non-parametric smoothing has proved the scores from several measures can be combined and related to a time-index estimation of dementia severity based on disease time course or duration [8]. Mini mental state examination (MMSE) has been used extensively to measure the severity of dementia in AD. Research by [9] proposed a method based on "time-index" and modeled the change in the MMSE as a function of time and showed that AD progression over time could be modeled by cubic or a logarithmic function of MMSE score.

Statistical parametric mapping (SPM) algorithm in diagnosing the early stage of AD was discussed in [10]. A few years later, [11] applied the multivariate analysis based on the principal component analysis to evaluate the correlation covariance of activation across brain regions rather than proceeding on a voxel-by-voxel basis. Furthermore, research on edge detection, which is an important initial step in the image processing of the analysis of CT scan image, which is similar to [10], was done by [12] using Weibull. The advantage of this approach over the Gaussian distribution is that it has a limitation to only symmetric shape but Weibull distribution has a symmetric and asymmetric shape.

A study about an acetylcholinesterase inhibitory activity on some herbal medicine to bring in a new management for AD by applying Student's t-test to compare the mean ± SD of the herbal medicine has been carried out by [13]. In the same year, [14] showed that the long-run longitudinal study and meta-analysis examining the broad factors and specific personality facts are associated with the increased risk of AD. Another statistical analysis which is known as multivariate analysis of covariance (MANCOVA), is used to discuss that active smoking is related to a significantly increased risk of AD by [15].

Multi-resolution statistical analysis of brain connectivity graphs in preclinical AD and showed significant connective difference between AD individuals [16]. The proposed algorithm helps identify the potentially subtle differences between patient groups. On the other hand, meta-analysis is applied to investigate whether chronic exposure to aluminum is associated with an increased risk of AD [17]. In the same year, the Q test and I2 statistics were used to examine the heterogeneity between selected studies. In the same year, [18] proposed a delayed-start analysis approach for the first time implemented in the field of AD. They suggested that solanezumab has demonstrated that the expedition program is consistent with a disease-modifying effect in mild AD patients. Recently, Bayesian statistics enhanced the efficiency in analyzing the study data of a disease-modifying drug in AD [19]. Besides, a systematic review and meta-analysis carried out by [20] showed that insomnia is associated with an increased risk for dementia.

2.2. Polynomial regression

A linear regression analysis was used to compute the IC50 for reducing prostaglandin E_2 content by different test substance concentrations [21]. In order to compare between regressions models, coefficient of determination, R^2 can be used and it can be the goodness-of-fit test of models [22]. Besides, the second degree of a polynomial was used to compute the IC50 of a drug by [23]. Meanwhile, [24] proposed a nonparametric model to describe the dose-response relationship and fit the curve by polynomial regression.

Shapiro-Wilk and Anderson Darling tests were adequate for normality when the sample size equals 20 in simple linear regression [25]. When the normality assumption of a simple linear model is not achieved, transformation can be applied [26]. For convenience in inference and improved efficiency in estimating simple linear regression, analyses often incur additional assumptions other than the normally distributed error [27].

2.3. Four-parameter logistic regression (4-PLR)

The 4-PLR model-fitting approach has been proven to increase variance along with the mean response [7]. They illustrated it by radioimmunoassay (RIA) and enzyme-linked immunosorbent (ELISA). The extended 4-PLR model in multianalyte analysis (MELISA) is discussed and compared with a more simplistic approach in ELISA [28]. An add-on package for R-language that simultaneously fits several non-linear regression models focusing on analyzing dose-response curves had been introduced [29]. The functionality applies to arbitrary non-linear regression models such as the 4-PLR model.

A few comparative fit of non-linear models where 4-PLR model is included was carried out on bone marrow endothelial cell lines, replicated at various doses of an inhibitory agent [30]. Meanwhile, [31] developed an EM algorithm suited for maximum likelihood estimation under the 4-PLR model after conceptualizing the problem as a mixture of subpopulations. Some subjects respond regardless of dose, some fail to respond regardless of dose, and some fail regardless of respond with a probability that depends on the dose. The 4-PLR model hill equation is used to derive IC50 in HDAC4 and ENPP2 high-throughput screening data sets and compared with a series of numerical simulations [32].

Moreover, [33] had shown that several approaches, including the 4-PLR model to quantity the inhibition of drug efflux in a cell assay, can result in different IC50 values. In the same year, [34] presented a Solver-based Microsoft Excel template and a stand-alone GUI-based "point and click" program, called HEPB to fit the Hill equation. Both programs estimate Hill equation parameters were identical to GraphPad Prism and programming language R. When the actual curve is asymmetric, [6] showed that sometimes the 4-PLR model offers better performance due to bias-variance trade-off than 5PL. Another approach for dose-response includes the 4-PLR model has been proposed [29]. Although the 4-PLR model is often used in the dose-response curve, from the new algorithm, a portion of cases was proved better described by a multiphasic model than the Hill model [35].

3. RESEARCH METHOD

3.1. Data

The data were obtained from an experiment that undergoes Ellman's colorimetric method. Some minor calculations were needed to determine the inhibitory activities of acetylcholinesterase (AChE) [36]. Electric eel AChE acted as the enzyme source, while acetylcholine iodide (ATCI) acted as substrates of the reaction. Meanwhile, 5,5"-dithio-bis-2nitrobenzoic acid (DTNB) was used to measure the cholinesterase activity.

Sodium phosphate buffer was added to the 96 wells, followed by the synthesized compound solution at different concentrations and AChE solution. The mixture was then incubated for a period at room temperature. After that, the reaction was initiated by adding DTNB and ATCI. The hydrolysis of ACh was determined by monitoring the formation of yellow 5-thio-2-nitrobenzoate anion due to the reaction of DTNB with thiocholine, released by the enzymatic hydrolysis of ACh at a wavelength of 412 nm utilizing a 96-well plate reader. The absorbance reading was recorded 10 times with a 1-minute interval by using a microplate reader. The percentage of inhibition of AChE was determined by Equation 1, where PI was the percentage of inhibition, b_c was the slope of control (without any drug) and b_d was the slope of synthesized compounds or tacrine.

$$PI = \frac{b_c - b_d}{b_c} \times 100 \quad (2)$$

Eighteen synthetic compounds with the same core structures were synthesized. All the compounds labeled 8, 9, 11, 15, 16, 17, and 18 did not show any inhibition within the tested concentration range. Hence,

their results were discarded. The results from the remaining synthetic compounds were kept for further analysis.

3.2. Methodology

The following polynomial regression models were used to fit the PI data set using to determine the IC_{50}. The fitted models of all the equations were then written as (3)-(6):

$$\hat{y} = b_0 + b_1 x \tag{3}$$

$$\hat{y} = b_0 + b_1 x + b_2 x^2 \tag{4}$$

$$\hat{y} = b_0 + b_1 x + b_2 x^2 + b_3 x^3 \tag{5}$$

$$\hat{y} = \hat{d} + \frac{\hat{a} - \hat{d}}{1 + \left(\frac{x}{\hat{c}}\right)^{\hat{b}}} \tag{6}$$

where $b_0, b_1, b_2, b_3, \hat{a}, \hat{d},$ and $\hat{c}$ are the estimated parameters.

Since there are several models can be used to fit the PI data set, adjusted coefficient of determination (R^2_{adj}) was used to determine the best fitted linear regression models among SLR, quadratic regression (QR), and cubic regression (CR) [37]. The R^2_{adj} is a modified version of coefficient of determination (R^2) adjusted for the number of predictors in the model. It increases when the new term improves the model and decreases when the predictors improve it less than expected. High R^2_{adj} indicates that the model is fitted adequately. Hence, the best fit of linear regression always goes to the one with a higher R^2_{adj}.

The R^2_{adj} is suitable for comparing the linear regressions with a different number of terms. Unfortunately, using non-linear models such as 4-parameter logistic regression [32]. The root means square error (RMSE) is more appropriate in determining the goodness of fit between non-linear and linear models [38]. It measures the differences between the predicted values and observed values. In contrast with R^2_{adj}, the lower RMSE indicates a better model fit. The RMSE is written as (7):

$$RMSE = \sqrt{\frac{(\hat{y}_i - y_i)^2}{df_e}} \tag{7}$$

where $\hat{y}_i$= *i*th fitted response, y_i= *i*th observed response value, and df_e= degree of freedom of the error term.

Once the best model was identified, estimated equations were formed with its estimated coefficient. Then, substitute the value of 50% at the response variable *y*. The value of *x* is the IC_{50} in the unit of microgram per millimeter ($\mu g/mL$). At *y* equal to 50%, the QR and CR curves could have more than one value of *x*. Hence, the lowest positive distinct point was chosen as the $IC_{50,}$ and discarded the other [23].

4. RESULTS AND DISCUSSION

4.1. Comparing the linear models

Table 1 displays the results of fitting the linear regressions for compound 1 at each replication. The fitted equation was $\hat{y} = 21.03 + 0.09339x$ with the R^2_{adj} =91.8. It indicated that 91.8% of the inhibition of compound 1 could be explained by the concentration used. This result is in line with the research conducted by [39], who found that the inhibition of acetylcholinesterase is an essential enzyme for Alzheimer's disease. Moreover, [40] and [41] stated that inhibitor concentration significantly influences activity inhibition.

Then, the R^2_{adj}increased to 94.3 and 94.2 when QR and CR were fitted to the data, respectively. The increment in R^2_{adj}suggested that the QR and CR were better than SLR. Since there were minor differences in the value R^2_{adj}, the QR was selected as the best-fitted model due to its simplicity of lower degree of the polynomial model. As a result, the selected model for the first replication was $\hat{y} = 17.01 + 0.1545x - 0.000082x^2$.

At the second experiment run, the R^2_{adj} for all three fitted models were lower than the R^2_{adj} of the first experiment. However, it had similar results with the first run. The cubic regression (CR) and quadratic regression (QR) fit better than SLR, but QR and CR were very close to each other. The QR was selected again as the best model due to its simplicity. The $R^2_{adj}$for the three fitted models for *n* =3 were not similar to

each other. Hence, the best-fitted model went to the SLR with the highest R^2_{adj}at 90.8 and the fitted equation was $\hat{y} = 30.10 + 0.07475x$. In summary, the selected model for compound 1 at each replication was listed in Table 1. An additional column is the IC_{50}, which was calculated based on the corresponding selected fitted regression.

Table 2 displays the IC50 of compound 1 for each experimental unit. The IC_{50} for data at the first experiment run was 245.52 μg/mL which indicated that 245.52 μg/mL of compound 1 was needed to perform a 50% inhibition. Compared with *n*=2 and *n*=3, the amount needed to have a 50% inhibitory was 216.61μg/mL and 266.22 μg/mL, respectively. At each replication, the IC_{50} obtained was slightly different from each other, but they represent the same inhibitory ability of the compound. Hence, the study of the mean of IC_{50} for each compound gave more informative results. The mean of these IC_{50} values was 242.78 μg/mL. Overall, compound 1 needed an average value of 242.78 μg/mL to get the 50% of inhibitory [5] with the electric eel acetylcholinesterase under Ellman's assay [36]. The rest of the synthetic compounds were analyzed using the same methods and the result of the IC_{50} was shown in Table 3. Overall, the lowest mean of IC_{50} exhibited by compound 3 (31.84 μg/mL) indicated its most substantial inhibition among the compounds, followed by compounds 6 (50.82 μg/mL) and 10 (57.15 μg/mL). However, the mean of IC_{50} obtained for these compounds showed that their inhibition effect was moderate if compared to the common standard drug tacrine, used in the Ellman's assay [42].

Table 1. Fitted simple, quadratic and cubic regression for compound 1 for each replication

*i*th experiment run, *n*	Type of regression	R^2_{adj}	Fitted equations
1	Simple Linear Regression	91.8	$\hat{y} = 21.03 + 0.09339x$
	Quadratic Regression	94.3	$\hat{y} = 17.01 + 0.1545x - 0.000082x^2$
	Cubic Regression	94.2	$\hat{y} = 19.17 + 0.08546 + 0.000162x^2 - 0.0000002056x^3$
2	Simple Linear Regression	79.5	$\hat{y} = 26.54 + 0.07338x$
	Quadratic Regression	87.0	$\hat{y} = 20.89 + 0.1593x - 0.000115x^2$
	Cubic Regression	87.4	$\hat{y} = 23.83 + 0.06503x + 0.000218x^2 - 0.0000002806x^3$
3	Simple Linear Regression	90.8	$\hat{y} = 30.10 + 0.07475x$
	Quadratic Regression	89.0	$\hat{y} = 29.83 + 0.07892 - 0.000006x^2$
	Cubic Regression	86.7	$\hat{y} = 30.82 + 0.04721x + 0.000106x^2 - 0.0000000944x^3$

Table 2. IC50 of compound 1 for each experimental run

*i*th experiment run, *n*	Selected fitted linear model	IC_{50} (μg/mL)
1	$\hat{y} = 17.01 + 0.1545x - 0.000082x^2$	245.52
2	$\hat{y} = 20.89 + 0.1593x - 0.000115x^2$	216.61
3	$\hat{y} = 30.10 + 0.07475x$	266.22

4.2. Choosing between linear and non-linear models

In the earlier discussion, besides R^2_{adj}, the better approach for choosing a more suitable linear and non-linear model is based on RMSE [43]. The estimated coefficients of 4-PLR for compound 1 are listed in Table 4. For *n*=1 compound 1, the estimated 4-PLR model which is written as:

$$\hat{y} = 147.53 + \frac{17.94 - 147.53}{1 + \left(\frac{x}{658.35}\right)^{1.15}}$$

The RMSE of this estimated equation was 8.19. However, for the polynomial regression models of compound 1, the estimated QR had a RMSE of 6.97 for *n*=1 which is smaller than the RMSE of the estimated 4-PLR. In contrast with R^2_{adj}, small RMSE indicates a better model. For *n*=2 and *n*=3, both RMSE of selected fitted linear models were lower than fitted 4-PLR. Hence the final fitted model for compound 1 was best fitted by CR and SLR, respectively as shown in Table 4.

Table 3. The best fitted model and the IC_{50} for each synthetic compound

Adamantane	*n*	R^2_{adj}	Best fitted linear model	IC_{50}	Mean of IC_{50}
	1	94.3	$\hat{y} = 17.01 + 0.1545x - 0.000082x^2$	245.52	
1	2	87.0	$\hat{y} = 20.89 + 0.1593x - 0.000115x^2$	216.61	242.78
	3	90.8	$\hat{y} = 30.10 + 0.07475x$	266.22	
	1	97.8	$\hat{y} = 17.38 + 0.4085x - 0.001191x^2 + 0.000001x^3$	114.27	
2	2	91.1	$\hat{y} = 16.84 + 0.4279x - 0.001260x^2 + 0.000001x^3$	110.03	107.49
	3	97.1	$\hat{y} = 27.99 + 0.2891x - 0.000759x^2 + 0.000001x^3$	98.16	
	1	96.6	$\hat{y} = 13.85 + 0.4535x - 0.001141x^2 + 0.000001x^3$	104.82	
3	2	96.5	$\hat{y} = 15.59 + 0.3420x - 0.000315x^2$	112.21	104.66
	3	99.3	$\hat{y} = 14.49 + 0.4385x - 0.000842x^2 + 0.000001x^3$	96.95	
	1	95.8	$\hat{y} = -9.846 + 0.6693x - 0.001040x^2$	107.31	
4	2	74.0	$\hat{y} = 26.86 + 0.1740x$	132.99	120.32
	3	98.6	$\hat{y} = 44.82 - 0.4051x + 0.004678x^2 - 0.000008x^3$	120.68	
	1	91.0	$\hat{y} = -23.47 + 3.506x - 0.03957x^2 + 0.000123x^3$	30.40	
5	2	81.7	$\hat{y} = -4.07 + 2.738x - 0.02872x^2 + 0.000086x^3$	26.56	31.84
	3	76.6	$\hat{y} = 19.26 + 0.9188x - 0.003155x^2$	38.56	
	1	97.0	$\hat{y} = 5.238 + 1.171x - 0.004178x^2$	45.67	
6	2	84.3	$\hat{y} = 1.681 + 0.9317x - 0.002862x^2$	64.73	50.82
	3	77.8	$\hat{y} = 17.68 + 0.9112 - 0.003392x^2$	42.05	
	1	58.3	$\hat{y} = 29.06 + 0.2315x$	90.45	
7	2	85.4	$\hat{y} = 6.284 + 0.9319x - 0.003115x^2$	58.25	66.65
	3	76.5	$\hat{y} = 25.64 + 0.5771x - 0.001984x^2$	51.24	
	1	99.7	$\hat{y} = 24.10 + 0.6236x - 0.002381x^2 + 0.000003x^3$	50.73	
10	2	99.4	$\hat{y} = 21.86 + 0.4421x - 0.000659x^2$	71.21	57.15
	3	96.2	$\hat{y} = 0.596 + 1.289x - 0.006330x^2 + 0.000009x^3$	49.52	
	1	98.2	$\hat{y} = 9.625 + 0.1822x - 0.000095x^2$	255.68	
12	2	94.2	$\hat{y} = 1.537 + 0.3247x - 0.000435x^2 + 0.0000002051x^3$	195.94	229.07
	3	99.1	$\hat{y} = 7.370 + 0.2104x - 0.000125x^2$	235.59	
	1	96.7	$\hat{y} = 3.058 + 0.4787x + 0.002700x^2 - 0.000017x^3$	80.12	
13	2	89.0	$\hat{y} = 0.85 + 0.1192x + 0.00932x^2 - 0.000041x^3$	81.16	87.32
	3	77.6	$\hat{y} = -5.58 + 0.7470x - 0.001936x^2$	100.67	
	1	88.6	$\hat{y} = -14.22 + 0.6860x - 0.001371x^2 + 0.000001x^3$	119.79	
14	2	92.9	$\hat{y} = -15.93 + 0.6889x - 0.001296x^2 + 0.000001x^3$	120.47	115.73
	3	95.6	$\hat{y} = -0.41 + 0.5763x - 0.001088x^2 + 0.000001x^3$	106.94	

Table 4. RMSE of 4-parameter logistic and polynomial regression of compound 1

4-parameter logistic regression						
*i*th Experiment run, *n*	Coefficients				*RMSE*	IC_{50}
	$\hat{a}$	$\hat{b}$	$\hat{c}$	$\hat{d}$		
1	17.94	1.15	658.35	147.53	8.19	249.12
2	21.00	1.07	398.84	105.77	10.72	216.74
3	31.56	1.65	672.69	131.29	8.54	273.00
Polynomial Regressions Model						
*i*th Experiment run, *n*	Coefficients				*RMSE*	IC_{50}
1	$\hat{y} = 17.01 + 0.1545x - 0.000082x^2$				6.97	245.52
2	$\hat{y} = 20.89 + 0.1593x - 0.000115x^2$				8.77	216.61
3	$\hat{y} = 30.10 + 0.07475x$				7.11	266.22

The comparison between the polynomial regression model and fitted 4-PLR non-linear models was listed as the final fitted model of Table 5. Although the fitted model 4-PLR is recommended to fit the dose-response curve, it is not always the best-fitted model in all cases. For instance, the 4-PLR was not the best-fitted model for compound 1. However, for compound 12, the 4-PLR was the best-fitted model for all replications *n*.

Table 5. Final fitted model, IC_{50} and the mean of IC_{50} of each synthetic compound

Adamantane	*n*	*RMSE*	Final Model	IC_{50}	Mean of IC_{50}
1	1	6.97	$\hat{y} = 17.01 + 0.1545x - 0.000082x^2$	245.52	242.78
	2	8.77	$\hat{y} = 20.89 + 0.1593x - 0.000115x^2$	216.61	
	3	7.11	$\hat{y} = 30.10 + 0.07475x$	266.22	
2	1	2.97	$\hat{y} = 17.38 + 0.4085x - 0.001191x^2 + 0.000001x^3$	114.27	107.49
	2	6.39	$\hat{y} = 16.84 + 0.4279x - 0.001260x^2 + 0.000001x^3$	110.03	
	3	2.67	$\hat{y} = 27.99 + 0.2891x - 0.000759x^2 + 0.000001x^3$	98.16	
3	1	6.76	$\hat{y} = 13.85 + 0.4535x - 0.001141x^2 + 0.000001x^3$	104.82	98.19
	2	6.49	$\hat{y} = 139.86 + \frac{11.09 - 139.86}{1 + \left(\frac{x}{209.60}\right)^{1.10}}$	97.84	
	3	1.17	$\hat{y} = 226.16 + \frac{1.10 - 226.16}{1 + \left(\frac{x}{678.54}\right)^{0.64}}$	91.90	
4	1	4.40	$\hat{y} = 89.51 + \frac{3.88 - 89.51}{1 + \left(\frac{x}{93.17}\right)^{3.65}}$	97.19	119.29
	2	14.79	$\hat{y} = 26.86 + 0.1740x$	132.99	
	3	3.07	$\hat{y} = 93.41 + \frac{38.13 - 93.41}{1 + \left(\frac{x}{156.28}\right)^{6.41}}$	127.68	
5	1	9.38	$\hat{y} = -23.47 + 3.506x - 0.03957x^2 + 0.000123x^3$	30.40	31.84
	2	12.34	$\hat{y} = -4.07 + 2.738x - 0.02872x^2 + 0.000086x^3$	26.56	
	3	12.24	$\hat{y} = 19.26 + 0.9188x - 0.003155x^2$	38.56	
6	1	4.95	$\hat{y} = 5.238 + 1.171x - 0.004178x^2$	45.67	50.82
	2	11.55	$\hat{y} = 1.681 + 0.9317x - 0.002862x^2$	64.73	
	3	10.40	$\hat{y} = 17.68 + 0.9112 - 0.003392x^2$	42.05	
7	1	13.59	$\hat{y} = 29.06 + 0.2315x$	90.45	62.69
	2	9.68	$\hat{y} = 69.65 + \frac{4.36 - 69.65}{1 + \left(\frac{x}{29.45}\right)^{1.85}}$	46.39	
	3	7.68	$\hat{y} = 25.64 + 0.5771x - 0.001984x^2$	51.24	
10	1	1.33	$\hat{y} = 211.69 + \frac{4.66 - 211.69}{1 + \left(\frac{x}{608.12}\right)^{0.50}}$	48.27	56.86
	2	2.14	$\hat{y} = 21.86 + 0.4421x - 0.000659x^2$	71.21	
	3	3.02	$\hat{y} = 80.10 + \frac{21.11 - 80.10}{1 + \left(\frac{x}{51.55}\right)^{4.59}}$	51.09	
12	1	4.21	$\hat{y} = 110.94 + \frac{13.73 - 110.94}{1 + \left(\frac{x}{331.94}\right)^{1.67}}$	243.48	215.93
	2	8.02	$\hat{y} = 97.68 + \frac{8.76 - 97.68}{1 + \left(\frac{x}{205.87}\right)^{1.79}}$	189.94	
	3	2.42	$\hat{y} = 106.72 + \frac{8.96 - 106.72}{1 + \left(\frac{x}{265.53}\right)^{1.51}}$	214.38	
13	1	1.63	$\hat{y} = 67.19 + \frac{8.10 - 67.19}{1 + \left(\frac{x}{53.20}\right)^{3.29}}$	69.76	83.86
	2	10.19	$\hat{y} = 0.85 + 0.1192x + 0.00932x^2 - 0.000041x^3$	81.16	
	3	13.70	$\hat{y} = -5.58 + 0.7470x - 0.001936x^2$	100.67	
14	1	8.30	$\hat{y} = -14.22 + 0.6860x - 0.001371x^2 + 0.000001x^3$	119.79	115.73
	2	6.67	$\hat{y} = -15.93 + 0.6889x - 0.001296x^2 + 0.000001x^3$	120.47	
	3	4.20	$\hat{y} = -0.41 + 0.5763x - 0.001088x^2 + 0.000001x^3$	106.94	

5. CONCLUSION

Simple linear regression is the simplest model to fit the dose-response curve to look for the IC_{50}. When it comes to situations where other polynomial regression fits better than SLR, the model of SLR is often discarded. As recommended, non-linear regression such as four-parameter logistic regressions (4-PLR) often fits the dose-response data adequately than any other type of regression. However, referring to RMSE in this research, it was found that sometimes 4-PLR did not fit closely to the data. The linear regression or polynomial regressions fitted better than 4-PLR. Based on the RMSE value, it was found that that the better model to determine the IC_{50} of a series of synthetic compounds was polynomial regression of degree 2. The estimated regression model is $\hat{y} = 17.01 + 0.1545x - 0.000082x^2$. This model has an RMSE value of 6.97. Based on the model, the estimated IC_{50} value was 245.52. Further inspection of the percentage

of inhibition (PI) for each compound collected must be carried out. More analysis has to be done to have better-fitted regression models with a more reliable half-maximal inhibitory concentration (IC_{50}).

REFERENCES

[1] Alzheimers Dement, "2021 Alzheimer's disease facts and figures," *Alzheimer's & Dementia*, vol. 17, no. 3, pp. 327–406, Mar. 2021, doi: 10.1002/alz.12328.

[2] V. V. Vanessa and S. H. Mah, "Xanthone: potential acetylcholinesterase inhibitor for Alzheimer's disease treatment," *Mini-Reviews in Medicinal Chemistry*, vol. 21, no. 17, pp. 2507–2529, Nov. 2021, doi: 10.2174/1389557521666210212152514.

[3] W. Thies and L. Bleiler, "2012 Alzheimer's disease facts and figures," *Alzheimer's & Dementia*, vol. 8, no. 2, pp. 131–168, Mar. 2012, doi: 10.1016/j.jalz.2012.02.001.

[4] R. Akbari and H. A. Javar, "Efficacy of capecitabine and 5-Fluorouracil (5-FU) on the human breast cancer cell line (MCF7)–effect of concentration," *American Journal of Research Communication*, vol. 1, no. 6, pp. 75–91, 2013.

[5] J. L. Sebaugh, "Guidelines for accurate EC50/IC50 estimation," *Pharmaceutical Statistics*, vol. 10, no. 2, pp. 128–134, Mar. 2011, doi: 10.1002/pst.426.

[6] W. N. Cumberland, Y. Fong, X. Yu, O. Defawe, N. Frahm, and S. De Rosa, "Nonlinear calibration model choice between the four and five-parameter logistic models," *Journal of Biopharmaceutical Statistics*, vol. 25, no. 5, pp. 972–983, Jun. 2014, doi: 10.1080/10543406.2014.920345.

[7] M. A. O'Connell, B. A. Belanger, and P. D. Haaland, "Calibration and assay development using the four-parameter logistic model," *Chemometrics and Intelligent Laboratory Systems*, vol. 20, no. 2, pp. 97–114, Sep. 1993, doi: 10.1016/0169-7439(93)80008-6.

[8] J. W. Ashford, M. Shan, S. Butler, A. Rajasekar, and F. A. Schmitt, "Temporal quantification of Alzheimer's disease severity: 'time index' model," *Dementia and Geriatric Cognitive Disorders*, vol. 6, no. 5, pp. 269–280, 1995, doi: 10.1159/000106958.

[9] M. S. Mendiondo, J. W. Ashford, R. J. Kryscio, and F. A. Schmitt, "Modelling mini mental state examination changes in Alzheimer's disease," *Statistics in Medicine*, vol. 19, no. 11–12, pp. 1607–1616, Jun. 2000, doi: 10.1002/(sici)1097-0258(20000615/30)19:11/12<1607::aid-sim449>3.0.co;2-o.

[10] B. S *et al.*, "Early diagnosis of Alzheimer's disease using a grid implementation of statistical parametric mapping analysis," *Studies Health Technology Informatics*, pp. 69–81, 2006.

[11] C. Habeck and and Yaakov Stern, "Multivariate data analysis for neuroimaging data: overview and application to Alzheimer's Disease," *Cell Biochemistry and Biophysics*, vol. 58, no. 2, pp. 53–67, Jul. 2010, doi: 10.1007/s12013-010-9093-0.

[12] W. K. Al-Jibory and A. El-Zaart, "Edge detection for diagnosis early Alzheimer's disease by using Weibull distribution," *2013 25th International Conference on Microelectronics (ICM)*, Dec. 2013, doi: 10.1109/icm.2013.6735024.

[13] S. B. Jazayeri, A. Amanlou, N. Ghanadian, P. Pasalar, and M. Amanlou, "A preliminary investigation of anticholinesterase activity of some Iranian medicinal plants commonly used in traditional medicine," *DARU Journal of Pharmaceutical Sciences*, vol. 22, no. 1, Jan. 2014, doi: 10.1186/2008-2231-22-17.

[14] A. Terracciano *et al.*, "Personality and risk of Alzheimer's disease: New data and meta-analysis," *Alzheimer's & Dementia*, vol. 10, no. 2, pp. 179–186, May 2013, doi: 10.1016/j.jalz.2013.03.002.

[15] T. C. Durazzo, N. Mattsson, and M. W. Weiner, "Smoking and increased Alzheimer's disease risk: A review of potential mechanisms," *Alzheimer's & Dementia*, vol. 10, no. 3S, pp. S122--S145, Jun. 2014, doi: 10.1016/j.jalz.2014.04.009.

[16] W. H. Kim *et al.*, "Multi-resolution statistical analysis of brain connectivity graphs in preclinical Alzheimer's disease," *NeuroImage*, vol. 118, pp. 103–117, Sep. 2015, doi: 10.1016/j.neuroimage.2015.05.050.

[17] Z. Wang *et al.*, "Chronic exposure to aluminum and risk of Alzheimer's disease: A meta-analysis," *Neuroscience Letters*, vol. 610, pp. 200–206, Jan. 2016, doi: 10.1016/j.neulet.2015.11.014.

[18] H. Liu-Seifert *et al.*, "Delayed-start analysis: Mild Alzheimer's disease patients in solanezumab trials, 3.5 years," *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, vol. 1, no. 2, pp. 111–121, Jul. 2015, doi: 10.1016/j.trci.2015.06.006.

[19] A. Satlin *et al.*, "Design of a Bayesian adaptive phase 2 proof-of-concept trial for BAN2401, a putative disease-modifying monoclonal antibody for the treatment of Alzheimer's disease," *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, vol. 2, no. 1, pp. 1–12, Jan. 2016, doi: 10.1016/j.trci.2016.01.001.

[20] K. M. de Almondes, M. V. Costa, L. F. Malloy-Diniz, and B. S. Diniz, "Insomnia and risk of dementia in older adults: Systematic review and meta-analysis," *Journal of Psychiatric Research*, vol. 77, pp. 109–115, Jun. 2016, doi: 10.1016/j.jpsychires.2016.02.021.

[21] J. B. Francisco, R. Guitian, J. Moreno, F. J. De Toro, and F. Galdo, "Effect of anti-inflammatory drugs on COX-1 and COX− 2 activity in human articular chondrocytes," *The Journal of Rheumatology*, vol. 26, no. 6, pp. 1366–1373, 1999.

[22] I. Pardoe, *Applied Regression Modeling*. Wiley, 2020.

[23] J. W. Tyner *et al.*, "Kinase pathway dependence in primary human leukemias determined by rapid inhibitor screening," *Cancer Research*, vol. 73, no. 1, pp. 285–296, Oct. 2012, doi: 10.1158/0008-5472.can-12-1906.

[24] H. Zhang, J. Holden-Wiltse, J. Wang, and H. Liang, "A strategy to model nonmonotonic dose-response curve and estimate IC50," *PLoS ONE*, vol. 8, no. 8, p. e69301, Aug. 2013, doi: 10.1371/journal.pone.0069301.

[25] D. A. Pierce and R. J. Gray, "Testing normality of errors in regression models," *Biometrika*, vol. 69, no. 1, pp. 233–236, 1982, doi: 10.1093/biomet/69.1.233.

[26] R. J. Carroll and D. Ruppert, *Transformation and Weighting in Regression*. Chapman and Hall/CRC, 2017.

[27] G. A. F. Seber and A. J. Lee, *Linear regression analysis*. John Wiley & Sons, 2012.

[28] G. Jones *et al.*, "Extension of the four-parameter logistic model for ELISA to multianalyte analysis," *Journal of Immunological Methods*, vol. 177, no. 1–2, pp. 1–7, Dec. 1994, doi: 10.1016/0022-1759(94)90136-8.

[29] C. Ritz, F. Baty, J. C. Streibig, and D. Gerhard, "Dose-response analysis using R," *Plos One*, vol. 10, no. 12, p. e0146021, Dec. 2015, doi: 10.1371/journal.pone.0146021.

[30] R. H. Lyles, C. Poindexter, A. Evans, M. Brown, and C. R. Cooper, "Nonlinear model-based estimates of IC50 for studies involving continuous therapeutic dose-response data," *Contemporary Clinical Trials*, vol. 29, no. 6, pp. 878–886, Nov. 2008, doi: 10.1016/j.cct.2008.05.009.

[31] G. E. Dinse, "An EM algorithm for fitting a four-parameter logistic model to binary dose-response data," *Journal of Agricultural, Biological, and Environmental Statistics*, vol. 16, no. 2, pp. 221–232, Oct. 2010, doi: 10.1007/s13253-010-0045-3.

[32] H. Gubler, U. Schopfer, and E. Jacoby, "Theoretical and experimental relationships between percent inhibition and IC50 data observed in high-throughput screening," *Journal of Biomolecular Screening*, vol. 18, no. 1, pp. 1–13, Aug. 2012, doi:

10.1177/1087057112455219.
[33] D. A. Volpe, S. S. Hamed, and L. K. Zhang, "Use of different parameters and equations for calculation of IC50 values in efflux assays: potential sources of variability in IC50 determination," *The AAPS Journal*, vol. 16, no. 1, pp. 172–180, Dec. 2013, doi: 10.1208/s12248-013-9554-7.
[34] S. R. Gadagkar and G. B. Call, "Computational tools for fitting the Hill equation to dose-response curves," *Journal of Pharmacological and Toxicological Methods*, vol. 71, pp. 68–76, Jan. 2015, doi: 10.1016/j.vascn.2014.08.006.
[35] G. Y. Di Veroli *et al.*, "An automated fitting procedure and software for dose-response curves with multiphasic features," *Scientific Reports*, vol. 5, no. 1, pp. 1-11, Oct. 2015, doi: 10.1038/srep14701.
[36] G. L. Ellman, K. D. Courtney, V. Andres, and R. M. Featherstone, "A new and rapid colorimetric determination of acetylcholinesterase activity," *Biochemical Pharmacology*, vol. 7, no. 2, pp. 88–95, Jul. 1961, doi: 10.1016/0006-2952(61)90145-9.
[37] S. Weisberg, *Applied Linear Regression*. John Wiley & Sons, Inc., 2005.
[38] H. J. Motulsky and L. A. Ransnas, "Fitting curves to data using nonlinear regression: a practical and nonmathematical review," *The FASEB Journal*, vol. 1, no. 5, pp. 365–374, Nov. 1987, doi: 10.1096/fasebj.1.5.3315805.
[39] E. L. Konrath, C. dos Santos Passos, L. C. Klein-Júnior, and A. T. Henriques, "Alkaloids as a source of potential anticholinesterase inhibitors for the treatment of Alzheimer's disease," *Journal of Pharmacy and Pharmacology*, vol. 65, no. 12, pp. 1701–1725, Jun. 2013, doi: 10.1111/jphp.12090.
[40] J. Marco-Contelles *et al.*, "Multipotent cholinesterase/monoamine oxidase inhibitors for the treatment of Alzheimer's disease: design, synthesis, biochemical evaluation, ADMET, molecular modeling, and QSAR analysis of novel donepezil-pyridyl hybrids," *Drug Design, Development and Therapy*, p. 1893, Oct. 2014, doi: 10.2147/dddt.s69258.
[41] B. R. Pinho, F. Ferreres, P. Valentão, and P. B. Andrade, "Nature as a source of metabolites with cholinesterase-inhibitory activity: an approach to Alzheimer's disease treatment," *Journal of Pharmacy and Pharmacology*, vol. 65, no. 12, pp. 1681–1700, May 2013, doi: 10.1111/jphp.12081.
[42] Z. H. Loh *et al.*, "New 3-O-substituted xanthone derivatives as promising acetylcholinesterase inhibitors," *Journal of Enzyme Inhibition and Medicinal Chemistry*, vol. 36, no. 1, pp. 627–639, Jan. 2021, doi: 10.1080/14756366.2021.1882452.
[43] A. Chakraborty and D. Goswami, "Prediction of slope stability using multiple linear regression (MLR) and artificial neural network (ANN)," *Arabian Journal of Geosciences*, vol. 10, no. 17, Sep. 2017, doi: 10.1007/s12517-017-3167-x.

BIOGRAPHIES OF AUTHORS

Anwar Fitrianto is currently active as a lecturer at IPB University. He obtained his Bachelor Degree in Statistics in 1999 from the Department of Statistics, IPB University, Indonesia. Both, Master of Science and PhD in Statistics were obtained in 2005 and 2010, respectively, from Universiti Putra Malaysia. He has strong expertise in Robust Statistics, Statistics Modeling, Experimental Design, and Statistical Process Control. He can be contacted at email: anwarstat@gmail.com.

Siau Man Mah completed her Master of Applied Statistics and Bachelor of Science Major in Statistics from Universiti Putra Malaysia in 2016 and 2014. She is currently a mathematics educator in Pin Hwa High School, Malaysia. Her expertise are in statistics modeling and cluster analysis. She can be contacted at email: euphynn@yahoo.com.

Siau Hui Mah is a Senior Lecturer at Taylor's University and currently holding positions of Head of the Centre for Drug Discovery and Molecular Pharmacology in Faculty of Health and Medical Sciences and Programme Director (Postgraduate Programmes) in the School of Biosciences. Her research interests focus on Natural Products and Medicinal Chemistry, especially on the discovery of lead compounds for anti-cancer, anti-inflammatory and anti-Alzheimer's disease drugs from the phytochemicals and their synthesized derivatives. She can be contacted at email: siauhui.mah@taylors.edu.my.