



Contents lists available at ScienceDirect

## Journal of Pharmacological and Toxicological Methods

journal homepage: [www.elsevier.com/locate/jpharmtox](http://www.elsevier.com/locate/jpharmtox)

Original article

## Computational tools for fitting the Hill equation to dose–response curves

Sudhindra R. Gadagkar<sup>a,\*</sup>, Gerald B. Call<sup>b</sup><sup>a</sup> Department of Biomedical Sciences, College of Health Sciences, Midwestern University, Glendale, AZ 85308, USA<sup>b</sup> Department of Pharmacology, Arizona College of Osteopathic Medicine, Midwestern University, Glendale, AZ 85308, USA

## ARTICLE INFO

## Article history:

Received 15 August 2014

Accepted 15 August 2014

Available online 23 August 2014

## Keywords:

Methods

Sigmoidal curve

Dose–response relationship

Four-parameter logistic model

EC<sub>50</sub>

Hill slope

Iteration

Computer program

Prediction band

Simulation

## ABSTRACT

**Introduction:** Many biological response curves commonly assume a sigmoidal shape that can be approximated well by means of the 4-parameter nonlinear logistic equation, also called the Hill equation. However, estimation of the Hill equation parameters requires access to commercial software or the ability to write computer code. Here we present two user-friendly and freely available computer programs to fit the Hill equation – a Solver-based Microsoft Excel template and a stand-alone GUI-based “point and click” program, called HEPB. **Methods:** Both computer programs use the iterative method to estimate two of the Hill equation parameters (EC<sub>50</sub> and the Hill slope), while constraining the values of the other two parameters (the minimum and maximum asymptotes of the response variable) to fit the Hill equation to the data. In addition, HEPB draws the prediction band at a user-defined confidence level, and determines the EC<sub>50</sub> value for each of the limits of this band to give boundary values that help objectively delineate sensitive, normal and resistant responses to the drug being tested. **Results:** Both programs were tested by analyzing twelve datasets that varied widely in data values, sample size and slope, and were found to yield estimates of the Hill equation parameters that were essentially identical to those provided by commercial software such as GraphPad Prism and *nls*, the statistical package in the programming language R. **Discussion:** The Excel template provides a means to estimate the parameters of the Hill equation and plot the regression line in a familiar Microsoft Office environment. HEPB, in addition to providing the above results, also computes the prediction band for the data at a user-defined level of confidence, and determines objective cut-off values to distinguish among response types (sensitive, normal and resistant). Both programs are found to yield estimated values that are essentially the same as those from standard software such as GraphPad Prism and the R-based *nls*. Furthermore, HEPB also has the option to simulate 500 response values based on the range of values of the dose variable in the original data and the fit of the Hill equation to that data.

© 2014 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

## 1. Introduction

Dose–response studies typically produce data that manifest as a sigmoid curve when a response is plotted against dosage (Fig. 1). A common inference done from such a curve is the estimation of the dose at which 50% of the subjects show the desired response. This is usually done by means of the four-parameter logistic nonlinear regression

model (Eq. 1), modified from the original equation developed by A.V. Hill to quantify the binding of oxygen to hemoglobin (Hill, 1910)

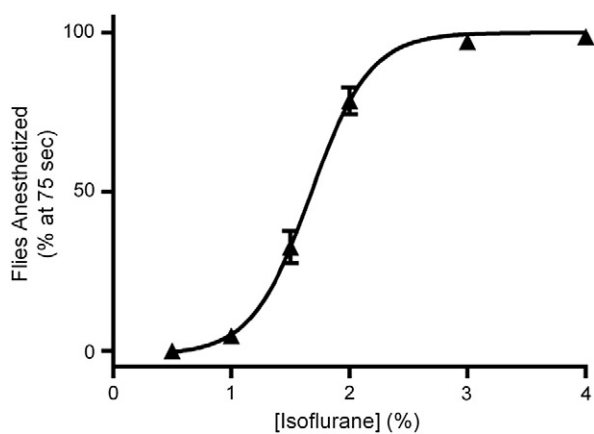
$$\hat{Y} = a + \frac{(b-a)}{\left[1 + \left(\frac{c}{X}\right)^d\right]} \quad (1)$$

where  $\hat{Y}$  is the expected response at dosage  $X$ ,  $a$  is the minimum asymptote or the response when dosage = 0,  $b$  = the maximum asymptote or the stabilized response for an infinite dosage,  $c$  is the dosage at which 50% of the subjects are expected to show the desired response (that is, the response halfway between the minimum response asymptote  $a$  and the maximum response asymptote  $b$ ); it also denotes the point of inflection in the dosage–response curve, and is referred to by various terms (e.g., EC<sub>50</sub>, ED<sub>50</sub>, LD<sub>50</sub>, IC<sub>50</sub>), and  $d$  is the slope at the steepest part of the curve, also known as the Hill slope. The model may be written to represent an ascending sigmoid curve of the type in Fig. 1 or a descending curve, depending on the sign of  $d$ . Specifically, positive  $d$

*Abbreviations:* GUI, graphic user interface; HEPB, Hill equation with prediction band; ET<sub>50</sub>, EC<sub>50</sub>, elution time for 50% of flies, or effective concentration for 50% response, the point on the X-axis corresponding to the point of inflection in the sigmoidal dose–response curve, this also corresponds to the dose at which 50% of the subjects being tested show the response being studied to the drug; IA, inhaled anesthetic; RSS, the residual sum of squares; VBA, visual basic for applications; R<sup>2</sup>, the coefficient of determination, which reflects the proportion of variation in the response variable that is accounted for by the independent variable.

\* Corresponding author. Tel.: +1 623 572 3855; fax: +1 623 572 3647.

E-mail address: [sgadagkar@midwestern.edu](mailto:sgadagkar@midwestern.edu) (S.R. Gadagkar).



**Fig. 1.** The typical sigmoidal nature of a dose–response curve. This example shows the percentage of flies that are anesthetized at 75 s (mean  $\pm$  SEM; triangles) plotted versus the dose of isoflurane anesthesia given, and the Hill equation fit to the data using Graphpad Prism software (solid line).

values yield ascending curves while negative values yield descending curves.

Eq. (1) represents one of a family of Hill equations that have been used to describe specific non-linear relationships under diverse scenarios, including, but not limited to, quantitative pharmacology (Gesztelyi et al., 2012), ligand binding (Poitevin & Edelstein, 2013; Siman et al., 2012), plant growth modeling (Zub, Rambaud, Bethencourt, & Brancourt-Hulmel, 2012), and modeling patterns of urban electricity usage (To, Lai, Lo, Lam, & Chung, 2012).

Computer programs have been available since the early 1970s to estimate the parameters of different versions of the Hill equation, most of which are specific to fitting kinetic data (Atkins, 1973; Knack & Rohm, 1977; Leone, Baranauskas, Furiel, & Borin, 2005; Wieker, Johannes, & Hess, 1970). None of these uses Eq. (1) specifically, although commercial software exists that can be made to fit the four-parameter logistic curve in Eq. (1) (e.g., GraphPad Prism, [www.graphpad.com](http://www.graphpad.com); The MiraiBio Group of Hitachi Solutions at [www.miraibio.com](http://www.miraibio.com)). Eq. (1) can also be fit to data using a computer program written using the open-access language, R, or the Solver Add-in in Microsoft Excel. In addition, some of these also permit the computation of confidence and prediction bands around the curve. However, the existing tools either require an investment in commercial software, which are also typically opaque to the user as to the code and algorithms used to generate the results, or require the ability of the user to write computer code in order to accomplish these tasks.

A long-term goal of the Call laboratory is to determine the mechanism of action of inhaled anesthetics (IAs), for which *Drosophila melanogaster* is used as the model system for providing in vivo responses to IAs in the presence of various genetic manipulations. *Drosophila* represents a good model for working with anesthetics as fruit flies follow the Meyer–Overton rule of anesthetics and display physiological responses to IAs similar to those in humans (Allada & Nash, 1993; Tinklenberg, Segal, Tianzhi, & Maze, 1991). Additionally, flies provide an inexpensive, yet robust model with access to a variety of genetic tools available to answer many scientific questions in vivo.

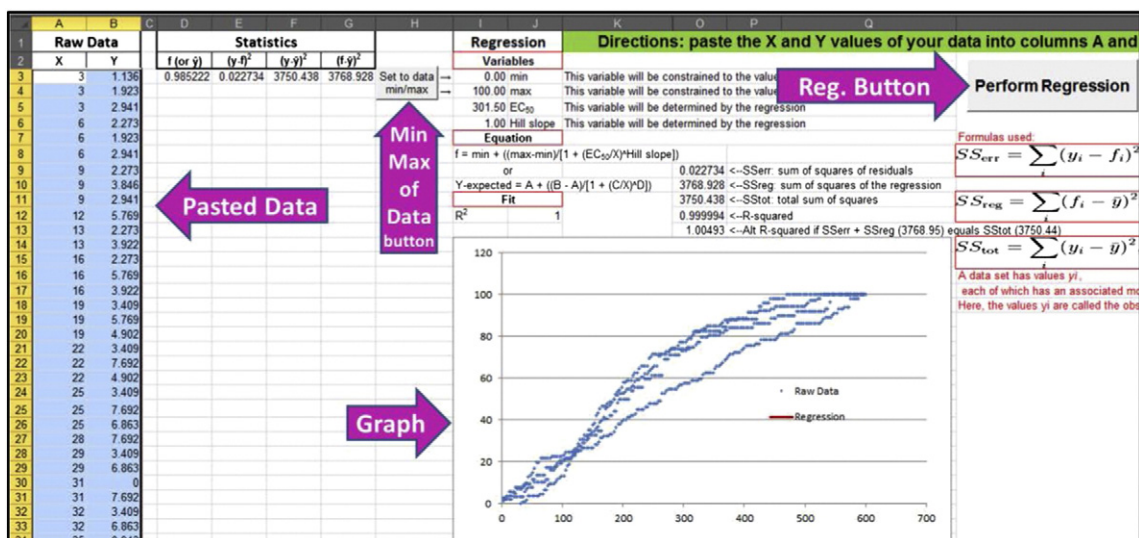
The Call laboratory has recently adapted an apparatus for the quantification of the *Drosophila* response to IAs (Dawson, Heidari, Gadagkar, Murray, & Call, 2013). Known as the inebriometer, it was originally designed to quantitatively measure the flies' response to ethanol vapors (Weber, 1988). This adaptation and modification of the inebriometer has enabled the rapid and efficient collection of simple quantitative data from large genetic screens. The inebriometer consists of a large column that is flooded with the IA. As the flies succumb to the IA, they elute out the bottom of the column and are counted. The Mean Elution Time

(MET) of the flies from the inebriometer column can then be computed, followed by standard statistical analysis (e.g., *t*-test).

In order to verify consistent inebriometer function, control flies are simultaneously assayed each day an experimental fly line is tested. In a genetic screen consisting of hundreds of experimental fly lines, this practice produces a large control dataset that presents a statistical problem: the Mean Elution Time when used with standard statistical tests is almost guaranteed to show a statistically significant difference between the experimental fly line being assayed and the control, simply due to the large numbers of flies used. Furthermore, the median test is also almost guaranteed to have low power due to the large sample sizes used; ~150 flies per assay. Therefore another approach was needed for the analysis of the genetic screen data. Since the raw fly elution data from the inebriometer was sigmoidal in nature, Eq. (1) was fit to the data, followed by the estimation of what we term the  $ET_{50}$ , which is analogous to  $EC_{50}$ , but represents the time, rather than the concentration, at which 50% of the flies elute from the inebriometer column. The  $ET_{50}$  value was then used as a measure of the flies' response to the IA. This is done by estimating the parameter  $c$  in Eq. (1), where  $X$  is the time it takes for  $Y$  percent of flies to elute through the inebriometer,  $a$  and  $b$  are the minimum and maximum asymptotes of the percentage of flies eluting through the system (0 and 100, respectively), and  $d$  is the Hill slope. Repeated assessments of the  $ET_{50}$  have shown it to be an efficient, direct and reliable indicator of the flies' response to various IAs.

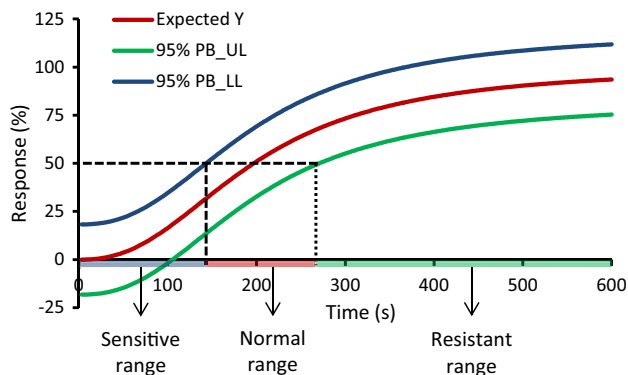
Here we present two computer programs: 1) a macros-enabled, Solver-based Excel template developed in the Call laboratory, and 2) a stand-alone Windows based computer program, HEPB (Hill Equation with Prediction Band), designed and developed in the Gadagkar lab. The Microsoft Excel template with Visual Basic for Applications (VBA) macros uses the above formula and estimates the  $ET_{50}$  and the Hill slope (variables  $c$  and  $d$  in Eq. (1)) for the inebriometer data. This template utilizes the Solver tool that comes with Excel. Solver is an optimization tool that uses techniques from Operations Research and has wide applicability including regression analysis and curve fitting. However, neither the availability nor the operation of Solver is straightforward to the average researcher more familiar with the graphic user interface (GUI) of most statistical software typically used to perform this type of analysis. For this reason, and due to the large amount of data routinely collected and the large number of student users in the Call laboratory, a macro was written to automate this analysis within Excel. This made the task very easy and straightforward even for the novice user as the analysis was done simply by the press of a button after data entry (Fig. 2). Furthermore, the macro ensured consistency in the output for easy and accurate export of the data and results to the relational database (Microsoft Access) being maintained in the laboratory. The Excel macros proved to be very useful and convenient, and have become a staple in the Call laboratory.

However, while the Hill equation was easily fit to the data and the  $ET_{50}$  and Hill slope were determined quickly by the macros, the problem of meaningfully comparing an experimental line with the control still remained. In addition, an important goal of these assays was also to classify a given fly line as having a sensitive, normal or resistant phenotype to the IA. To help resolve both problems, that is, comparing an experimental line to the control and classifying the experimental line as one of the above three types, the stand-alone computer program, HEPB, was developed. HEPB has an easy-to-use GUI that, in addition to estimating the parameters  $c$  and  $d$  in Eq. (1), also computes the prediction band (at a given level of confidence) for the control fly data and solves for the  $X$  value when  $Y = 50\%$  for each of the upper and lower limits of the prediction band. These form the cut-off values to objectively discriminate among sensitive, normal and resistant responses to a given anesthetic. These two limits each give the boundary value between sensitive and normal responses, and normal and resistant responses, respectively (Fig. 3). This is similar to standard statistical practice for a two-tailed test where the distribution under the null hypothesis is constructed, the critical regions delineated on either side of the curve, and



**Fig. 2.** Screenshot of the Regression worksheet with sample data pasted. Entering the data automatically populates the graph. The Hill equation is fit to the data when the Perform Regression button is pressed. The user can manually enter the values of min and max (set at 0 and 100, respectively, by default), or push the button to automatically change them to the minimum and maximum values of the data.

the experimental value simply compared to the critical values on this curve to accept or reject the hypothesis. Our critical values are the  $ET_{50}$  values for the upper and lower limits of the prediction band for the null distribution (the control). If the  $ET_{50}$  value for the experimental run falls within these two limits, it is determined to be no different from that of the control (null hypothesis accepted), and if it falls outside the limits, the null hypothesis is rejected and we conclude that the experimental run is statistically different from the control. Specifically, the experimental fly line is determined to be sensitive or resistant if the corresponding  $ET_{50}$  falls outside the lower limit, or outside the upper limit, respectively. Furthermore, HEPB has the option of generating 500 values of the response variable based on simulation, for equally spaced values of the dose variable within the range specified in the original data file, based on the fit of the Hill equation to the original data. This can be useful if the sample size in the original data is small and one is interested in understanding better the behavior of the dose-response relationship in question. Both programs are freely available, and can be obtained by contacting the authors.



**Fig. 3.** The Hill equation fit to the data in Fig. 2 (red line) and the 95% prediction band bounded by the blue and green lines. The  $ET_{50}$  values for each of prediction band limits (shown by the dashed and dotted lines to the X axis) delineate the boundaries between the zones of sensitive, normal and resistant flies. Based on this, any future dataset with estimated  $ET_{50}$  values in the blue, pink or green zones will be classified as sensitive, normal or resistant, respectively.

## 2. Methods

### 2.1. Regression and estimation of $c$ and $d$ in the macros-enabled Excel template

The principle of least-squares in the context of regression states that the line with the best fit to the data is that for which the sum of squared residuals,  $RSS = \sum_i^n (Y_i - \hat{Y})^2$ , is the smallest (where  $Y_i$  and  $\hat{Y}$  are the observed and expected values, respectively, of the response variable for the  $i^{\text{th}}$  value of the dose (or explanatory) variable, and  $i$  is the number of pairs of values in the data). The Excel template presented here contains VBA macros that utilize the built-in Solver tool to perform iterations to determine the best-fit curve by minimizing RSS (cell O9 in Fig. 2). The Excel 2010 + version of Solver uses Generalized Reduced Gradient (GRG), a robust algorithm for non-linear regression programming (Lasdon, Waren, Jain, & Ratner, 1978). The initial value for  $c$  in Eq. (1) is the calculated midpoint of the range of the data (explanatory variable), and  $d$  is set to equal 1. Solver is adequate for this purpose and generally determines the values of  $c$  and  $d$  quite accurately. However, accuracy is achieved only when the initial values used for these parameters are close approximations of their final values. The formulae used in the spreadsheet provide those approximations automatically and the VBA macro has been programmed to check the  $R^2$  value (coefficient of determination) that reflects the goodness of fit of the model to the data. For the first run, the starting value for  $c$  is the median of the  $X$  variable and for  $d$ , it is 1. If the first run yields a  $R^2 \geq 0.99$ , the regression results are accepted, as it is likely that Solver will not fit the data any better if run again. If not, Solver is run automatically again with the values of  $c$  and  $d$  determined from the initial fit, to yield better results. For this second run, the stringency is reduced, such that the results are accepted if  $R^2 \geq 0.95$ . If an  $R^2$  of 0.95 or higher is not achieved in the second run, Solver is run one last time with the third set of starting values for  $c$  and  $d$  determined in the same manner as for the second run, and the  $R^2$  value is reported. If  $R^2 \leq 0.50$  or the analysis with Solver does not converge (perhaps because the starting values are too far from the final values), producing an error, the macro has been programmed to recognize this and repeat the estimation with different starting values. These starting values are determined for  $c$  by systematically selecting values from the range of the dose variable, and  $d$  by choosing among



the empirically determined Hill slope values in the Call laboratory for sensitive and resistant relationships. This exercise is done in order to reach or exceed the threshold of  $R^2 \geq 0.95$ . This process has yielded excellent results with  $R^2$  values typically  $> 0.95$  in the Call laboratory. If  $R^2$  is still short of 0.95, however, and all the alternative sets of starting values have been used, the analysis stops, the default values of  $c$  and  $d$  are restored, and a message informing the user that the curve fit failed appears. In this situation, the user can manually change all four values in Eq. (1) in the template, as for instance, would be necessitated for  $a$  and  $b$  if the min and max values in a given dataset are not the default values of 0 and 100, respectively. To this end, a button next to the variables,  $a$  and  $b$ , allows the user to change automatically the min and max values to the minimum and maximum values of the entered dataset.

## 2.2. Regression and estimation of $c$ and $d$ in HEPB

HEPB also uses the least-squares criterion to determine the best fit to the data, while approaching the problem somewhat differently from Solver, namely by serial iteration. Each of three tandem iterations is done by looping through 200 equally spaced values within the range provided for the parameter  $d$ , nested within 200 equally spaced values within the range provided for the parameter  $c$ . The set of three tandem iterations with increasingly smaller ranges to iterate over ensures finer estimates of the parameters  $c$  and  $d$ . The minimum and maximum asymptotes ( $a$  and  $b$ , respectively) may be provided by the user or alternatively, can simply be the minimum and maximum values of the response variable in the data.

No starting values are required for the estimation of  $c$  and  $d$ . Instead, an all-inclusive range of  $-50$  to  $+50$  for the estimation of  $d$ , and the range defined by the min and max values of the dose ( $X$ ) variable for the estimation of  $c$ , are used in the first pass, and the iterations loop over 200 equally spaced values between the corresponding limits for both parameters in a nested fashion (explained below). Since parameters  $a$  and  $b$  are fixed for a given dataset, it is a straightforward procedure to estimate the values of  $c$  and  $d$ . The process begins by regressing iteratively the response variable against the dose variable, beginning with the value of  $a$  and progressing to the value of  $b$ , while saving the estimated values of  $c$  and  $d$  from each iteration along with the sum of the squared residuals ( $RSS$ ). When the program runs through all the iterations in the first pass over the ranges of both  $c$  and  $d$  (in increments of 200 equally spaced values between the corresponding limits for each), the values of these parameters are then estimated in this round of iteration as those associated with the smallest  $RSS$ , based on the least squares principle. The second pass or iteration is identical to the first, the only difference being that the iteration range for the estimation of each of  $c$  and  $d$  is now delimited by values 10% below and above each of the values of  $c$  and  $d$  obtained from the first-pass iteration. The final iteration is identical to the second iteration, except that the new iteration ranges are set as  $\pm 1\%$  around the values of  $c$  and  $d$  obtained from the second iteration. The number of steps between the two limits of each range is always maintained at 200 for both parameters. Thus, the increasingly small size of the increment and range to iterate over across the three sets of iterations (or passes) provide efficiency and accuracy in the estimation of the parameters. There is no quality control embedded in the program (as in the case of the Excel template). However, the  $R^2$  value has typically been above 95% for most datasets; when lower, it has been due to variation in the data and not a poor fit. HEPB also includes the residuals from the regression in the output. The speed of the program was determined by running it on a dataset with 5000 pairs of values (dataset XII, Table 1) on a Dell Optiplex 980 computer with Intel Core™ i7 CPU 860 @ 2.80 GHz processor, 8.00 GB of RAM, running on 64-bit, Microsoft Windows 7 Professional operating system, and the analysis was completed in 58 s. On a less powerful machine (Intel Core2 Duo E7500 @2.93GHz, 4 GB RAM, 32 bit Windows 7), it took 3 min and 56 s.

## 2.3. Prediction band

When the estimation involves a single value, it is customary to construct a confidence interval around the point estimate. This requires knowledge of the distribution that the estimate is expected to follow, and the width of a given confidence interval depends on the level of assurance required in ensuring that the unknown true value of the estimate resides within that interval. When the confidence interval is constructed for each  $\hat{Y}$  value in a regression, however, the two series of values at each end of the confidence interval then lie on either side of the  $\hat{Y}$  values (the regression line), thus forming a band along the length of the regression line.

When the goal is to predict a new individual value of  $Y$  for a given value of  $X$ ,  $s_p(\hat{Y})$ , the standard error of  $\hat{Y}$ , is given as the square-root of the following expression (Snedecor & Cochran, 1980):

$$s_p^2(\hat{Y}) = \frac{1}{n-2} \left( \sum_i^n y^2 - \frac{\left( \sum_i^n xy \right)^2}{\sum_i^n x^2} \right) \left( 1 + \frac{1}{n} + \frac{x^2}{\sum_i^n x^2} \right); y_i = (Y_i - \bar{Y}), x_i = (X_i - \bar{X}). \quad (2)$$

The lower and upper prediction band limits for a given  $\hat{Y}$  value are obtained using the following equation:

$$\hat{Y} \pm \left[ (t_{\alpha, n-2}) s_p(\hat{Y}) \right] \quad (3)$$

where  $\alpha$  is the level of significance and  $n$  is the sample size in terms of the number of pairs of values. If the predictions are being made for  $k$  new  $X$  values, it would be necessary to use the Bonferroni inequality and obtain the  $t$  value from the Student's  $t$  tables for  $\alpha/k$  and  $(n-2)$  degrees of freedom (Snedecor & Cochran, 1980). However, since the purpose of drawing the prediction band in the present case is to give cut-off values that allow us to distinguish among sensitive, normal and resistant responses to a given anesthetic being used in any given experiment for the  $X$  values already in the data (Fig. 3), Eq. (5) is used to obtain the lower and upper limits of the prediction band.

The  $c$  and  $d$  values for the upper and lower limits of the prediction band are estimated in the same manner of sequential sets of iterations as in the estimation of these parameters for the main regression equation, with the exception that the values of the corresponding prediction limits are used here instead of the observed values of the response variable. Once the values of  $c$  and  $d$  are obtained, the values for a given prediction limit are then plugged into Eq. (1) and the cut-off value for that limit obtained by solving for  $X$  when  $Y = 50\%$ . Thus, the cut-off values obtained from the upper prediction limit help distinguish between fly lines with sensitive and normal responses, and those from the lower prediction limit are used to distinguish between flies with normal and resistant response.

In addition, we have incorporated in HEPB the option of generating 500 values of the response variable, using simulation, within the observed range of the explanatory variable, based on the regression parameters estimated for the original data. The implementation of this project was done using the Embarcadero® Delphi® XE language (Embarcadero® RAD Studio XE Version 15.0.3953.35171).

## 3. Results

For the purposes of demonstration of our programs, a dataset from the Call laboratory is used where 809 flies from 6 separate experiments were assayed for their response to 1% isoflurane using the inebriometer (Dawson et al., 2013). The data needs to be formatted in two columns, the first ( $X$ ) is the independent variable or the dose associated with a desired response (e.g., time taken for a given fly to be fully anesthetized,

**Table 1**  
Performance of the Excel template and HEPB in estimating  $c$  and  $d$  in Eq. (1)<sup>a</sup> when compared to two standard computer programs, GraphPad Prism 6 and the *nls* statistical package in the R programming language.

| Dataset | $n$  | Excel              |                    | HEPB               |                    | Prism              |                      | R ( <i>nls</i> ) |         |
|---------|------|--------------------|--------------------|--------------------|--------------------|--------------------|----------------------|------------------|---------|
|         |      | $c$                | $d$                | $c$                | $d$                | $c$                | $d$                  | $c$              | $d$     |
| I       | 6    | 4.372              | 4.042              | 4.373              | 4.042              | 4.373              | 4.042                | 4.373            | 4.042   |
| II      | 6    | 2.315              | 3.356              | 2.316              | 3.356              | 2.316              | 3.356                | 2.316            | 3.356   |
| III     | 8    | 136.564            | 2.488              | 136.559            | 2.488              | 136.600            | 2.487                | 136.565          | 2.487   |
| IV      | 10   | 7.328              | 10.954             | 7.328              | 10.956             | 7.328              | 10.950               | 7.328            | 10.954  |
| V       | 14   | 1.246 <sup>b</sup> | 0.661 <sup>b</sup> | 1.249 <sup>c</sup> | 0.660 <sup>c</sup> | 1.247              | 0.661                | 1.245            | 0.661   |
| VI      | 25   | 12.953             | 20.096             | 12.953             | 20.096             | 12.950             | 20.100               | 12.953           | 20.096  |
| VII     | 27   | −6.296             | −10.846            | −6.926             | −10.845            | 6.926 <sup>d</sup> | −10.850 <sup>d</sup> | −6.926           | −10.846 |
| VIII    | 29   | −5.783             | −8.924             | −5.783             | −8.924             | 5.783 <sup>d</sup> | −8.924 <sup>d</sup>  | −5.782           | −8.923  |
| IX      | 1055 | 342.701            | 4.859              | 342.706            | 4.859              | 342.700            | 4.859                | 342.701          | 4.859   |
| X       | 740  | 197.577            | 2.410              | 197.570            | 2.410              | 197.600            | 2.410                | 197.578          | 2.410   |
| XI      | 1494 | 16.043             | 1.820              | 16.069             | 1.826              | 16.070             | 1.826                | 16.070           | 1.826   |
| XII     | 5000 | 103.009            | 1.668              | 102.971            | 1.667              | 103.000            | 1.668                | 103.009          | 1.667   |

Source and features of the datasets:

I: Source: Finney, D. J. (1971) *Probit Analysis*, Cambridge: Cambridge University Press. The data ( $n = 6$ ) examines the effect of a series of doses of rotenone (mg/l) on the aphid, *Macrosiphoniella sanborni*.

II: Source: Hornick, R.B., Music, S.I., Wenzel, R., Cash, R.A., Libonati, J.P., Snyder, M.J., Woodward, T.E. (1971) *The Broad Street Pump Revisited: Response of Volunteers to Ingested Cholera Vibrios*, *Bulletin New York Academy of Medicine*, 47, 1181–1191. The data ( $n = 6$ ) examines the response to human ingestion of *V. cholerae* (Inaba 569B strain) with sodium bicarbonate.

III: Source: Jeske, D. R., Xu, H. K., Blessinger, T., Jensen, P. and Trumble, J. (2009) Testing for the Equality of  $EC_{50}$  Values in the Presence of Unequal Slopes With Application to Toxicity of Selenium Types, *Journal of Agricultural, Biological, and Environmental Statistics*, 14, 469–483 The data ( $n = 8$ ) shows the percentage of deaths in flies from exposure to different concentrations of selenomethionine, a form of the element, selenium.

IV: Source: A generic dataset downloaded from the internet.

V: Source: <http://www.ats.ucla.edu/stat/sas/faq/doseresponse.htm>. The webpage uses this data ( $n = 14$ ) to demonstrate the parameterization shown in *Pharmaceutical Statistics Using SAS: A Practical Guide* by Dmitrienko et al. for the expected response,  $Y$ , given a dose,  $X$ , for the four-parameter logistic model.

VI: Source: Milicer, H. and Szczotka, F., 1966, Age at Menarche in Warsaw girls in 1965, *Human Biology*, 38, 199–203. The data ( $n = 25$ ) gives the proportion of female children reaching menarche by a given age.

VII and VIII: Source: GraphPad Prism 6, 2014. The data ( $n = 27, n = 29$ ) is sample dose response data provided by the program as an example for curve fitting analysis.

IX–XII: Source: Unpublished data from the Call laboratory, 2014. The data ( $n = 194, 740, 1494, 5000$ , respectively) comprises flies that have a resistant, normal, sensitive and normal response to 1% isoflurane as measured in the inebriometer, respectively.

Note: All datasets are available in Table S1.

Note: All datasets are available in Table S1.

<sup>a</sup> The values of  $c$  and  $d$  are estimated in all four programs while constraining the values of  $a$  and  $b$  to the minimum and maximum values of the  $X$  variable in the data.

<sup>b</sup> Convergence achieved after starting values for  $c$  and  $d$  that were close to their respective final values were manually entered into the template.

<sup>c</sup> These values were achieved when the iteration ranges for  $c$  and  $d$  were changed in HEPB to reflect those suggested by the  $XY$  plot of the data.

<sup>d</sup> Convergence achieved after starting value for  $c$  that was close to its final value was manually entered into the program.

as manifested by falling through the entire inebriometer column), and the second ( $Y$ ) is the response variable (e.g., the percentage of flies that were anesthetized in a given time).

### 3.1. Macros-enabled Excel template with Solver

The analysis to estimate the parameters  $c$  and  $d$  and compute the regression was done using the Excel template (available from the authors). The instructions to enable the use of macros and Solver are given in the Initial Instructions worksheet. The  $X$  and  $Y$  variables need to be entered into the corresponding columns in the Regression worksheet, following which, the graph will auto-populate with the raw data (blue dots; Fig. 2). In this process, the user has the option to change any or all of the four parameter values (that is, set the range limits for  $a$  and  $b$  and starting values for  $c$  and  $d$ ). A warning message alerts the user if the range limits for  $a$  and  $b$  are set to be within the corresponding limits in the observed data. A button then allows the user to assign  $a$  and  $b$  to the minimum and maximum values of the current dataset.

The data are analyzed by pressing the Perform Regression button. This runs Solver, which begins the optimization process by means of iteration. When this process is complete, the Excel spreadsheet displays the final Hill equation fit to the data and the values of  $c$  and  $d$  (called  $EC_{50}$  and Hill slope in the template), along with the  $R^2$  value. The regression line is plotted in red in the graph with the original data (Fig. 4). The analysis on the example dataset yielded a  $c$  value ( $EC_{50}$ ) of 342.701 and a  $d$  value (Hill slope) of 4.859, with a  $R^2$  value of 0.970. If the initial fit is not deemed optimal by the macro (see methods section for details), the

red regression line may shift and change as the starting values for  $c$  and  $d$  are adjusted and the iterative process is repeated. This is normal.

### 3.2. HEPB

The data file ( $X$  and  $Y$  values) should be saved as a comma-delimited (.csv) file, and opened by clicking on the File menu in HEPB and selecting Open (Fig. 5). The two columns of data are displayed in the memo field of the HEPB main interface for verification that the correct file has been opened. In addition, the name of the file is displayed at the bottom of the GUI, and remains there until another file is opened. The user then clicks on the Analysis menu, and selects the Options sub-menu. This opens the Analysis Options window (Fig. 6) where the user can indicate to the program that the minimum and maximum values of the response variable in the data should be used as the fixed values of  $a$  and  $b$ , respectively (see Eq. (1)), or alternatively, the user can provide the values for the two constraints. The options for entering the values become visible upon choosing the “No” radio button. In a similar manner, the user can either accept the default options of iterating over the range of  $X$  values for estimating  $c$  and the range of  $−50$  to  $50$  for estimating  $d$ , or enter the desired range for either or both parameters. The user then chooses among five confidence levels for the prediction band (80%, 85%, 90%, 95% and 97.5%), which have been provided based on the algorithm by Shamma for the rapid approximation of the critical values of the Student's  $t$  distribution (Shamma, 2009). Finally, the user has the option of generating 500 values of the response variable within the observed range of the explanatory variable, based on the regression parameters estimated for the original data, by checking the Simulate data checkbox.

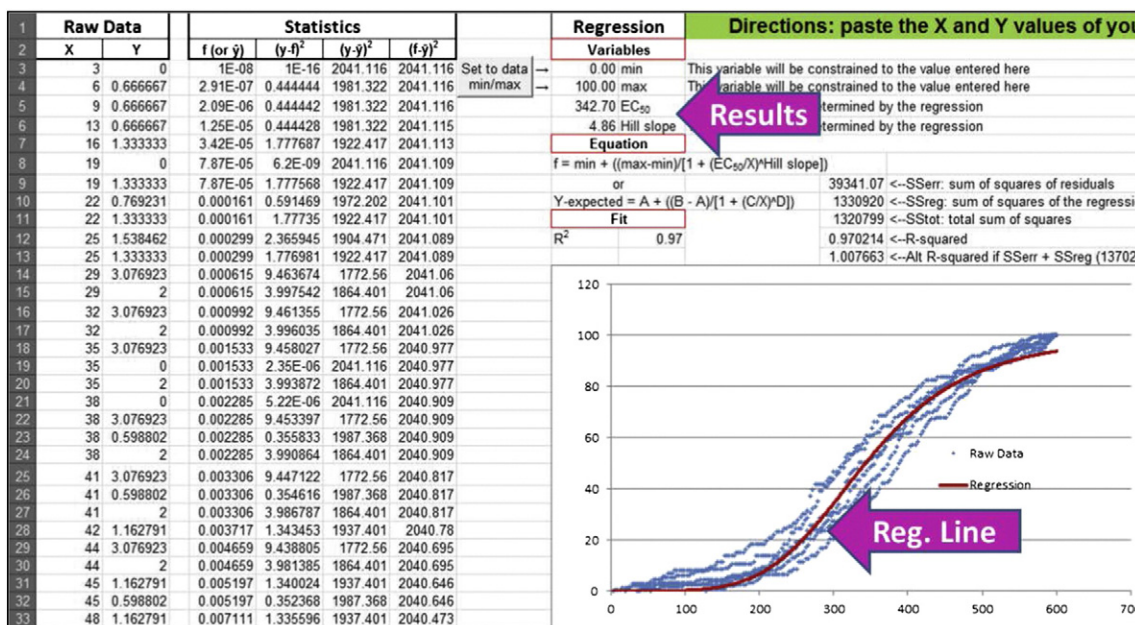


Fig. 4. Screenshot of the Regression worksheet after the Hill equation is fit to the data. The values of the final regression line (see text) are given in the Statistics columns and also plotted in red in the graph, and the estimated parameters ( $EC_{50}$  and Hill slope) along with  $R^2$  are shown.

After all the selections have been made (or default options accepted), the user then saves the options by pressing the Save Options button. While this button saves the options selected, it also alerts the user to any errors made on this page (e.g., invalid values) by means of messages at the bottom of the page (Fig. 7). After correcting all the errors, the user then presses the Save Options button again. This enables the Run sub-menu in the Analysis menu in the main HEPB form, which can now be selected. The analysis is then “Run.” The progress bar at the bottom of the HEPB main interface tracks the status of the analysis.

The results (the estimated  $EC_{50}$  and Hill slope values for the regression, the cut-off values for the upper and lower limits of the prediction band, and the  $R^2$  value) are displayed in the memo field of the main form. These results are followed by the input values ( $X$  and  $Y$ ), the expected  $Y$  values based on the Hill equation regression ( $\hat{Y}$ ), the lower and upper limits of the prediction band for each  $X$  value at the confidence level chosen by the user, and the residual ( $Y - \hat{Y}$ , Fig. 8). If the option to simulate has been checked, the final output is supplemented with the simulated values of the response variable

for 500 values of the  $X$  variable, based on the Hill slope regression on the original data. A window with the message, “Done!” indicates the successful completion of the analysis. The output can be saved as a .csv file to a folder of the user’s choice. The default name of the file is “Results,” which can be changed by the user. The example dataset used above yielded values of 342.706 and 4.859 for  $c$  and  $d$ , respectively, with a  $R^2$  value of 0.970. The GUI also allows the instructions, data or results to be displayed and saved at any time.

As can be seen, the results from both the Excel template and the HEPB program for the  $c$  and  $d$  variables ( $EC_{50}$  and Hill slope, respectively) are essentially identical when using the example dataset from the Call laboratory. In order to test if our two programs consistently yielded similar results, we chose twelve different datasets (Supplementary Table 1) from the Call laboratory and elsewhere that varied widely in size (6–5000 pairs of values) and exhibited a variety of curve shapes and slopes (Fig. 9). The example dataset used in the analysis above is dataset IX. Furthermore, we also analyzed these datasets using the *nls* statistical package written by D.M. Bates and S. DebRoy in the *R*

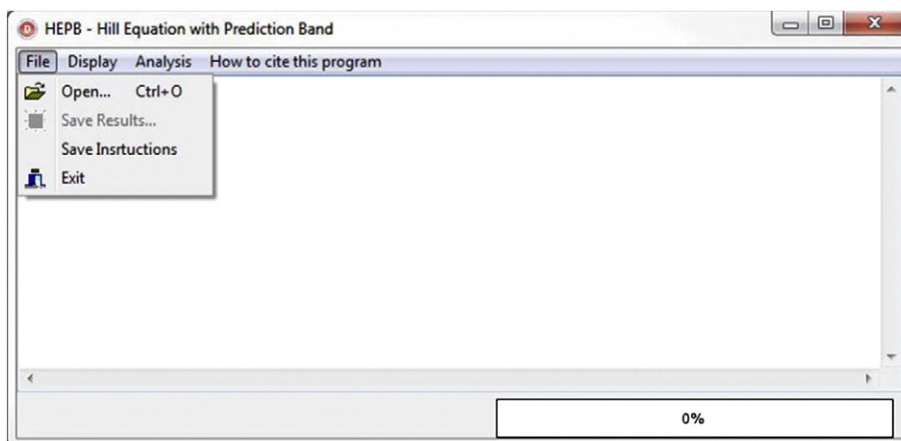


Fig. 5. The main window of HEPB with the File menu open.

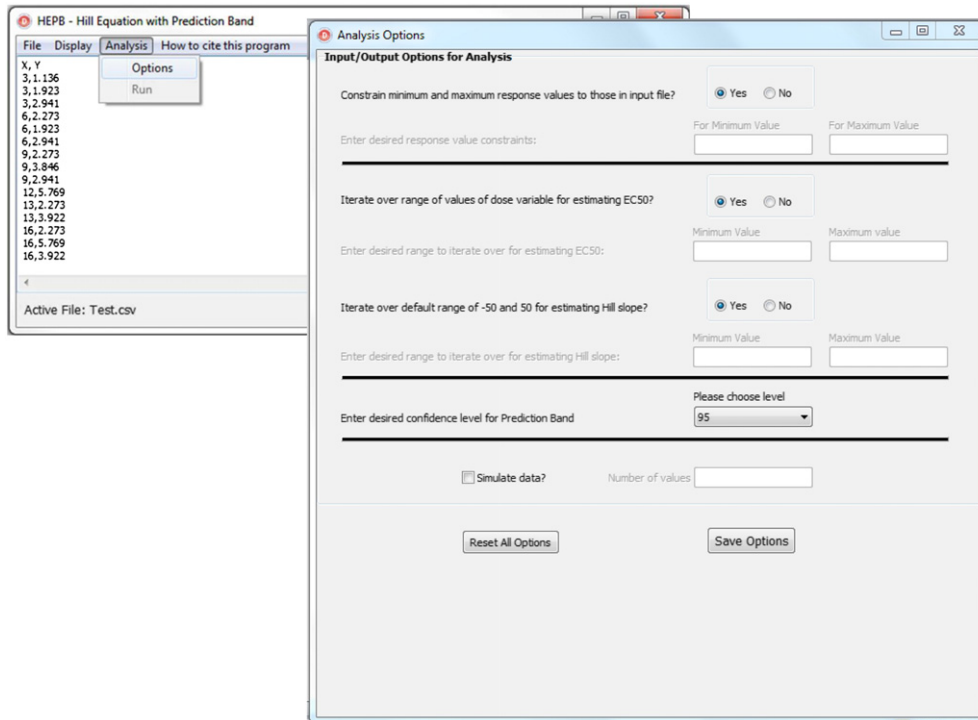


Fig. 6. The Analysis Options window is opened by clicking on the Options submenu in the Analysis menu on the main HEPB form.

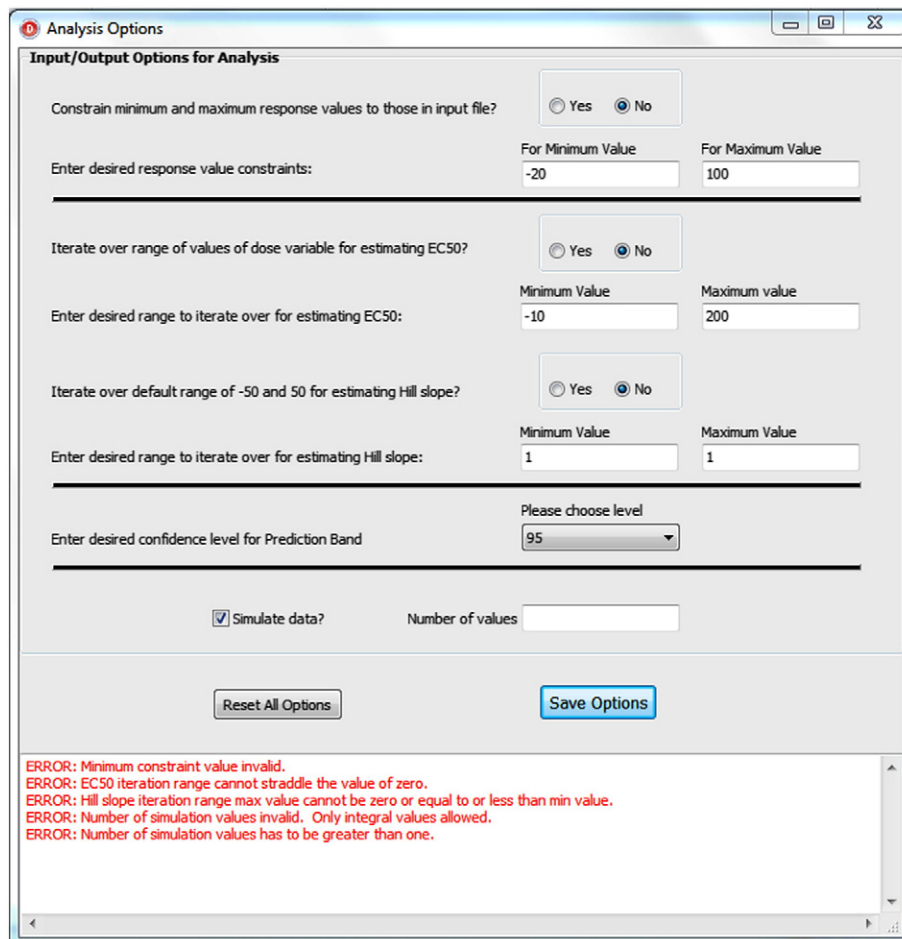
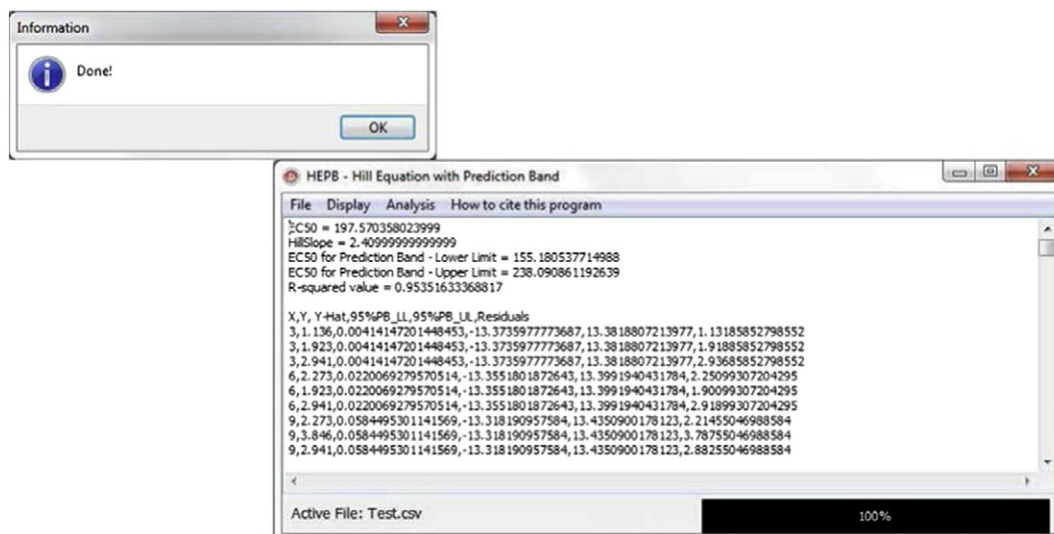


Fig. 7. Screenshot of the Analysis Options window with error messages that are displayed when the Save Options button is pressed with the user having entered invalid values.





**Fig. 8.** Screenshot of the window that pops up upon completion of the analysis informing the user of the successful completion of the analysis by means of the message “Done!” and the simultaneous display of the results in the menu field of the main form of HEPB.

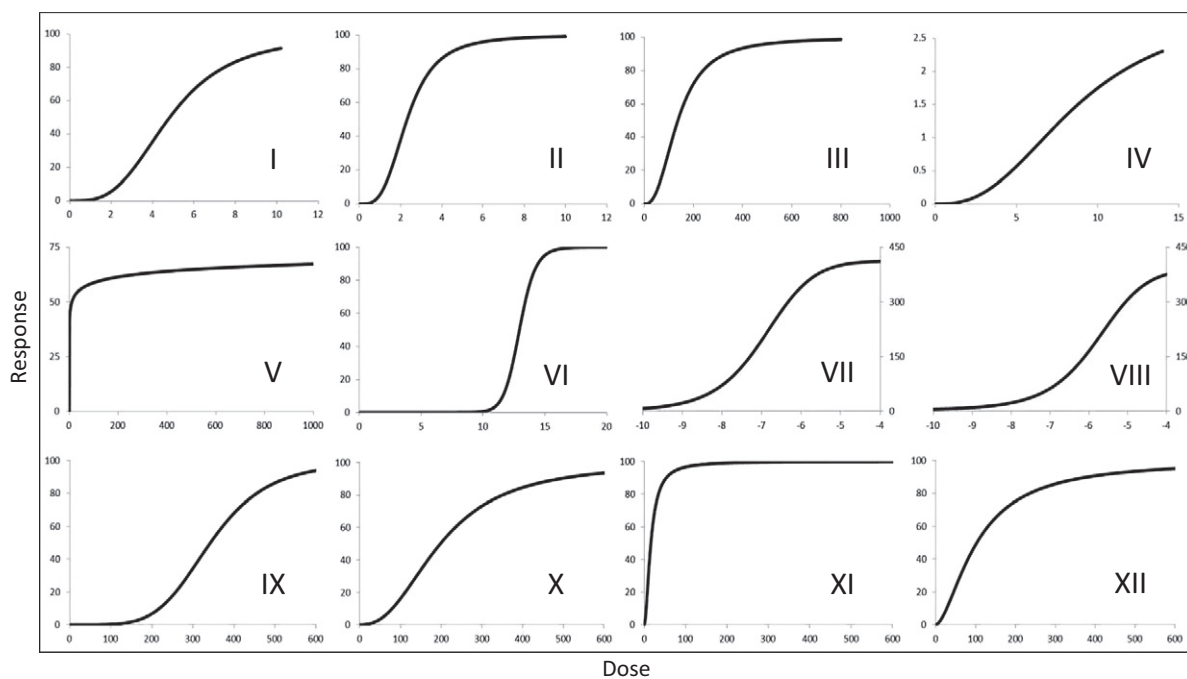
programming language (R\_Core\_Team, 2013) and the commercial software, GraphPad Prism 6.04 for Windows (GraphPad Software, La Jolla California USA, [www.graphpad.com](http://www.graphpad.com)), to ensure that the results of our programs were consistent with those from commonly used, standard software. In order to ensure appropriate comparisons among the different programs, the values of  $a$  and  $b$  were constrained to the min and max values in any given dataset. Table 1 shows the regression results in terms of the values of  $c$  and  $d$ . As can be seen, the values between the different programs are very similar, validating the use of the programs presented in this paper.

#### 4. Discussion

The four-parameter logistic equation, also known as the Hill equation (Eq. (1)) is commonly used to model the non-linear relationship

typically seen in the association between dose and response. This involves the estimation of four parameters ( $a$ – $d$ ) in the equation. Here we provide two user-friendly computational methods that perform the analysis by constraining the values of  $a$  and  $b$  and estimating the values of  $c$  and  $d$  by means of iteration, using the criterion of least squares.

The macros-enabled Excel template uses Solver to estimate the parameters  $c$  and  $d$  of Eq. (1) and plots the regression line based on this equation. Manipulation of Solver is done using VBA programming to automatically repeat the analysis using a different set of starting values each time for the estimation of  $c$  and  $d$  if the regression yields an error or if the criterion of  $R^2 \geq 0.5$  is not met, thus ensuring quality control without any input required from the user. This template was created for a specific need in the Call laboratory and is being routinely used there to assay different genetic lines of *D. melanogaster* for their



**Fig. 9.** Hill equation fit to the various datasets of Table 1 used to compare the performance of the Excel template and HEPB with other programs (GraphPad Prism and an R-based program). These curves were fit with HEPB with the min and max variables ( $a$  and  $b$  in Eq. (1)) constrained to the minimum and maximum values in the data. The roman numerals correspond to those used in Table 1.



response to various IAs, as it is very easy to use even by novice users. Nevertheless, it is being presented in this paper as it is applicable to analyzing any similar sigmoidal curve relationship in Excel, which is almost universally used. Furthermore, although the template provided here will work satisfactorily in the majority of cases, savvy users may modify the formulas and VBA code to suit their particular circumstances more precisely.

However, the results provided by the Excel template are restricted to the regression line and the estimates of  $c$  and  $d$  of Eq. (1), and do not permit the response of the flies to the anesthetics to be classified into sensitive, normal or resistant types — one of the major goals of the laboratory. The stand-alone GUI-based Windows program HEPB does the same analyses as above, but in addition it constructs a prediction band at a user-defined confidence level and then determines the cut-off values from those prediction band limits that help to objectively distinguish among sensitive, normal and resistant phenotypes. These values also enable researchers to determine rapidly and objectively if experimental values are statistically different from their control ranges in their assays. As far as we are aware, HEPB is the only program that does the four-parameter logistic regression, constructs the prediction band for the data, and provides objective, empirically determined cut-off values to distinguish among response phenotypes. Furthermore, it optionally generates 500 simulated values of the response variable within the range of the observed dose variable. This can be useful particularly when the sample size is limited and the user is unable to visualize the dose–response behavior in the data.

While it might seem redundant to provide these two different avenues for performing this analysis, we believe that each program fills a niche within the laboratory. Most users will find the Excel template straightforward and will be comfortable with its interface. Additionally, it can interface with other Microsoft Office software, like Access, to store data in a laboratory database, if needed. There are other sources that also involve the use of Solver to fit non-linear equations (Harris, 1998). In addition, there are instructions available in several websites on the internet. However, none of these sources provide a template such as the one presented here that not only makes it easy for the uninformed user (who merely needs to enter the data in the template) but more importantly, has been programmed to auto-check for the goodness of fit and redo the analysis with sets of alternative starting values for  $c$  and  $d$  in Eq. (1) until the goodness of fit criterion is met. It has been tested with a number of datasets that span a wide range of relationships between the dose and response and sample size (Fig. 9), and has performed remarkably well (Table 1). A further advantage of our template is that since it is being provided with permission to modify the code to accommodate particularly intractable datasets. In such cases, the non-savvy user would simply need to redo the regression after manually adjusting the four variables. However, after extensive testing done with a variety of datasets, we are confident that the need for manual intervention or code-modification will be rare; such an intervention was necessary in only one case (dataset V) among the datasets used in Table S1, and several of these datasets were chosen to be out of the ordinary.

As mentioned before, the Excel file, while giving the user a very easy to use and useful template, does not provide the user with a means to objectively screen new experimental strains to classify them as sensitive, normal or resistant with respect to the response to the drug used. Therefore, HEPB is being presented as a stand-alone program that, in addition to performing this analysis on any set of data, provides the prediction band based on a user-defined level of confidence and the boundary values that help distinguish among sensitive, normal and resistant phenotypes. It also has the option to simulate data.

In order to evaluate the robustness and consistency of the two programs, we analyzed diverse datasets from the Call laboratory and elsewhere with very different dose–response relationships (Fig. 9)

using both programs. In addition, we evaluated the accuracy of the two programs by comparing the output to that from Prism and an R-based program. The results, presented in Table 1, show that the output from the macros-enabled Excel template and HEPB are robust and consistent with each other and with other software commonly used for this purpose. These easy to use programs are freely available by contacting the authors.

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.vascn.2014.08.006>.

## Acknowledgments

We would like to thank Jorge Hasbun and Kim Cooper for discussions and testing the programs for bugs and errors. SRG would also like to acknowledge the start-up funds provided by the College of Health Sciences, and GBC would like to acknowledge intramural funds from Midwestern University and a generous donation from the Charity Fidelity Gift Fund, which supported this work.

## References

- Allada, R., & Nash, H. A. (1993). *Drosophila melanogaster* as a model for study of general anesthesia: The quantitative response to clinical anesthetics and alkanes. *Anesthesia and Analgesia*, 77(1), 19–26.
- Atkins, G. L. (1973). A simple digital-computer program for estimating the parameters of the Hill equation. *European Journal of Biochemistry*, 33(1), 175–180, <http://dx.doi.org/10.1111/j.1432-1033.1973.tb02667.x>.
- Dawson, A. G., Heidari, P., Gadagkar, S. R., Murray, M. J., & Call, G. B. (2013). An airtight approach to the inebriometer: From construction to application with volatile anesthetics. *Fly*, 7(2), 112–117, <http://dx.doi.org/10.4161/fly.24142>.
- Gesztesy, R., Zsuga, J., Kemeny-Beke, A., Varga, B., Juhasz, B., & Tosaki, A. (2012). The Hill equation and the origin of quantitative pharmacology. *Archive for History of Exact Sciences*, 66(4), 427–438, <http://dx.doi.org/10.1007/s00407-012-0098-5>.
- Harris, D. C. (1998). Nonlinear least-squares curve fitting with Microsoft excel solver. *Journal of Chemical Education*, 75(1), 119–121.
- Hill, A. V. (1910). The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *Journal of Physiology*, 40, iv–vii.
- Knack, I., & Rohm, K. H. (1977). Analysis of cooperative ligand-binding and steady-state kinetic data—Computer-program for calculating parameters of the Hill equation with a variable Hill coefficient. *Hoppe-Seyler's Zeitschrift Fur Physiologische Chemie*, 358(3), 262–262.
- Lasdon, L. S., Waren, A.D., Jain, A., & Ratner, M. (1978). Design and testing of a generalized reduced gradient code for nonlinear programming. *ACM Transactions on Mathematical Software (TOMS)*, 4(1), 34–50.
- Leone, F. A., Baranaukas, J. A., Furriel, R. P.M., & Borin, I. A. (2005). SigraFW: An easy-to-use program for fitting enzyme kinetic data. *Biochemistry and Molecular Biology Education*, 33(6), 399–403, <http://dx.doi.org/10.1002/bmb.2005.49403306399>.
- Poitevin, F., & Edelstein, S. J. (2013). Derivation of the Crick–Wyman equation for allosteric proteins defining the difference between the number of binding sites and the Hill coefficient. *Journal of Molecular Biology*, 425(9), 1497–1499, <http://dx.doi.org/10.1016/j.jmb.2013.03.014>.
- R\_Core\_Team (2013). R: A language and environment for statistical computing. from <http://www.R-project.org/>
- Shammas, N. C. (2009). Inverse probability distribution functions. from <http://namirshammas.com/NEW/InvPDF.pdf>
- Siman, L., Carrasco, I. S. S., da Silva, J. K. L., de Oliveira, M. C., Rocha, M. S., & Mesquita, O. N. (2012). Quantitative assessment of the interplay between DNA elasticity and cooperative binding of ligands. *Physical Review Letters*, 109(24), <http://dx.doi.org/10.1103/PhysRevLett.109.248103>.
- Snedecor, G. W., & Cochran, W. G. (1980). *Statistical methods* (7th ed.). Ames, Iowa: The Iowa State University Press.
- Tinklenberg, J. A., Segal, I. S., Tianzhi, G., & Maze, M. (1991). Analysis of anesthetic action on the potassium channels of the shaker mutant of *Drosophila*. *Annals of the New York Academy of Sciences*, 625, 532–539, <http://dx.doi.org/10.1111/j.1749-6632.1991.tb33884.x>.
- To, W. M., Lai, T. M., Lo, W. C., Lam, K. H., & Chung, W. L. (2012). The growth pattern and fuel life cycle analysis of the electricity consumption of Hong Kong. *Environmental Pollution*, 165, 1–10, <http://dx.doi.org/10.1016/j.envpol.2012.02.007>.
- Weber, K. E. (1988). An apparatus for measurement of resistance to gas-phase reagents. *Drosophila information services*, 67, 91–93.
- Wieker, H. J., Johannes, K. J., & Hess, B. (1970). A computer program for the determination of kinetic parameters obtained from sigmoidal steady-state kinetics. *Febs Letters*, 8(4), 178–185, [http://dx.doi.org/10.1016/0014-5793\(70\)80258-7](http://dx.doi.org/10.1016/0014-5793(70)80258-7).
- Zub, H. W., Rambaud, C., Bethencourt, L., & Brancourt-Hulmel, M. (2012). Late emergence and rapid growth maximize the plant development of *Miscanthus* clones. *Bioenergy Research*, 5(4), 841–854, <http://dx.doi.org/10.1007/s12155-012-9194-2>.