

Curve Fitting

Qinlu (Claire) Wang
BCBB/OCICB/NIAID

Outline

1. How to perform curve fitting?
2. Comparing models/curves
3. How to customize equation of models?

1. How to perform curve fitting

1) Simple Linear Regression

Let's take simple linear regression with curve fitting as an example. Linear regression fits a straight line through your data to find the best-fit value of the slope and intercept.

Choose the **sample data** with XY data table: **Simple Linear Regression**

The screenshot shows the GraphPad Prism 8.2.0 (272) interface. On the left is a dark purple sidebar with navigation options. The main window is titled 'Welcome to GraphPad Prism' and displays 'XY tables: Each point is defined by an X and Y coordinate'. It includes a table structure, a small scatter plot, and a 'Data table:' section with two radio buttons. The 'Start with sample data to follow a tutorial' option is selected. Below this, a 'Select a tutorial data set:' list shows various categories, with 'Simple linear regression' highlighted under the 'Correlation & regression (linear and nonlinear)' section. At the bottom right, there are 'Cancel' and 'Create' buttons.

GraphPad Prism Version 8.2.0 (272)

NEW TABLE & GRAPH

XY

Column

Grouped

Contingency

Survival

Parts of Whole

Multiple variables

Nested

EXISTING FILE

Open a File

LabArchives

Clone a Graph

Graph Portfolio

Prism Tips

XY tables: Each point is defined by an X and Y coordinate

	X	A			B		
	Minutes	Control			Treated		
	X	A:Y1	A:Y2	A:Y3	B:Y1	B:Y2	B:Y3
1	Title						
2	Title						
3	Title						

Minutes

Control

Treated

Learn more

Data table:

☐ Enter or import data into a new table

☒ Start with sample data to follow a tutorial

Select a tutorial data set:

- ▼ Error bars in XY tables
 - Entering replicate data
 - Entering mean (or median) and error values
- ▼ Correlation & regression (linear and nonlinear)
 - Correlation
 - Simple linear regression**
 - Linear regression - Compare slopes
 - Interpolate unknowns from a linear standard curve
 - Nonlinear regression - One phase exponential decay
 - Eliminating outliers during nonlinear regression
 - RIA or ELISA - Interpolate unknowns from sigmoidal curve
- ▼ Pharmacology
 - Dose-response - X is log(dose)
 - Dose-response - X is dose
 - Dose-Response - Ambiguous until constrained

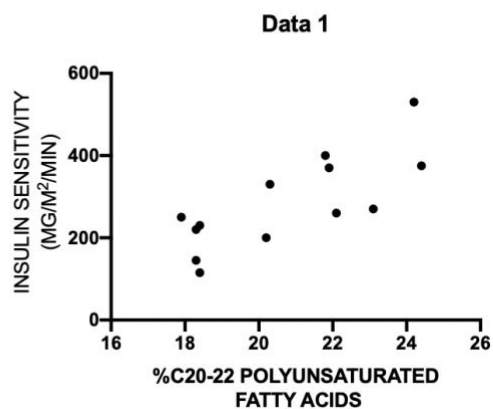
Cancel

Create

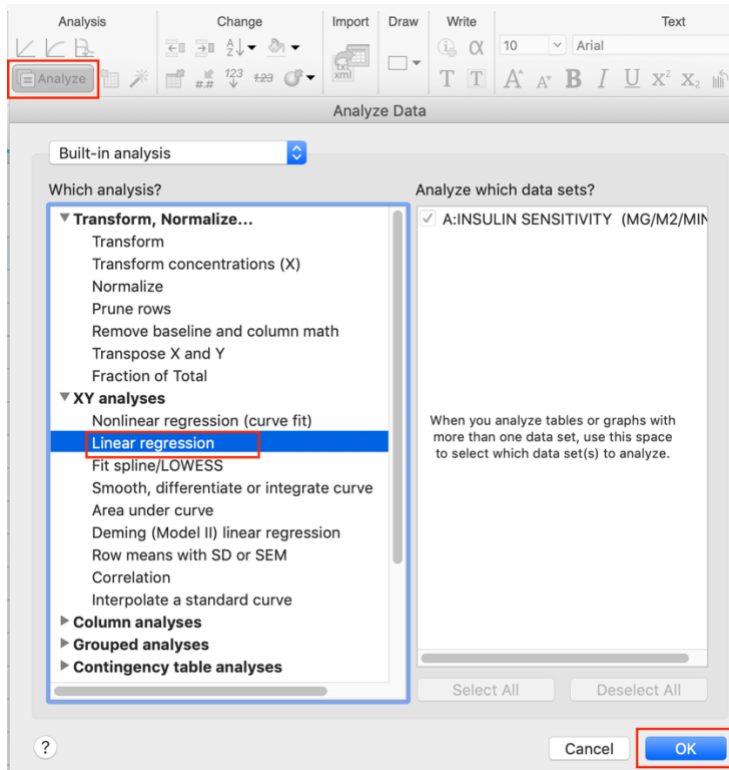
In this example, the X values are the percentage of a certain kind of fatty acid in muscle (from a biopsy) and the Y values are insulin sensitivity.

		X	Group A
		%C20-22 POLYUNSATURATED FATTY ACIDS	INSULIN SENSITIVITY (MG/M ² /MIN)
	✕	X	Y
1	TiH	17.9	250
2	TiH	18.3	220
3	TiH	18.3	145
4	TiH	18.4	115
5	TiH	18.4	230
6	TiH	20.2	200
7	TiH	20.3	330
8	TiH	21.8	400
9	TiH	21.9	370
10	TiH	22.1	260
11	TiH	23.1	270
12	TiH	24.2	530
13	TiH	24.4	375

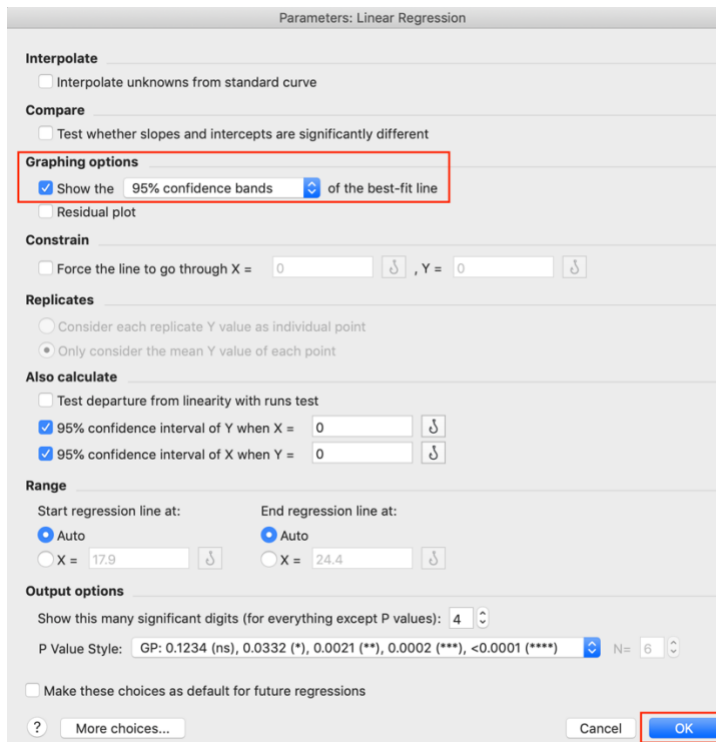
The corresponding graph is like:



Click “Analyze”. In the Analyze Data dialog, select **“Linear regression”** under the “XY analyses”. Then **click “OK”**.



In the “Parameters: Linear Regression” dialog, check the “Show the 95% confidence bands of the best-fit line” then click “OK”.



Then Prism presents a full set of linear regression results.

Linear reg. Tabular results		A
Part 1		Y
1	Best-fit values	
2	Slope	37.21
3	Y-intercept	-486.5
4	X-intercept	13.08
5	1/slope	0.02688
6		
7	Std. Error	
8	Slope	9.296
9	Y-intercept	193.7
10		
11	95% Confidence Intervals	
12	Slope	16.75 to 57.67
13	Y-intercept	-912.9 to -60.18
14	X-intercept	3.562 to 15.97
15	Part 2	
16	Goodness of Fit	
17	R square	0.5929
18	Sy.x	75.90
19		
20	Is slope significantly non-zero?	
21	F	16.02
22	DFn, DFd	1, 11
23	P value	0.0021
24	Deviation from zero?	Significant
25	Part 1	
26	Equation	$Y = 37.21 * X - 486.5$

Results Explanation:

a) Part 1: Slope and Intercept

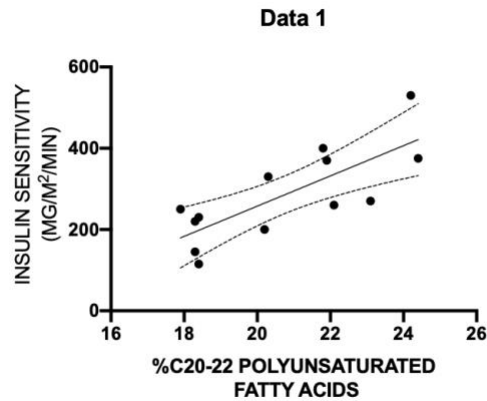
Prism first reports the best-fit values of the slope and intercept. It also reports the X intercept and the reciprocal of the slope.

At the bottom of the results page, the slope and intercept are reported again in the form of the equation that defines the best-fit line. You can copy this equation and paste onto a graph, or into a manuscript. In this case, the best-fit line is $Y = 37.21 * X - 486.5$

b) Part 2: Goodness-of-fit of linear regression

Prism provides R-square for goodness-of-fit. You can think of R-square as the fraction of the total variance of Y that is “explained” by variation in X.

In the corresponding graph, two confidence bands surrounding the best-fit line define the confidence interval of the best-fit line. The dashed confidence bands are curved and they are the boundaries of all possible straight lines.



2) Nonlinear Regression: Dose-response – X is dose

Then let's try a nonlinear regression example. Choose the **sample data** with XY data table: **Dose-response – X is dose**.

New Data Table and Graph

NEW TABLE & GRAPH

XY

Column
Grouped
Contingency
Survival
Parts of Whole
Multiple variables
Nested

EXISTING FILE

Clone a Graph

XY tables: Each point is defined by an X and Y coordinate

	X	A		B	
	Minutes	Control		Treated	
	X	A:Y1	A:Y2	B:Y1	B:Y2
1. Title					
2. Title					
3. Title					

[? Learn more](#)

Data table:

☐ Enter or import data into a new table

☒ Start with sample data to follow a tutorial

Select a tutorial data set:

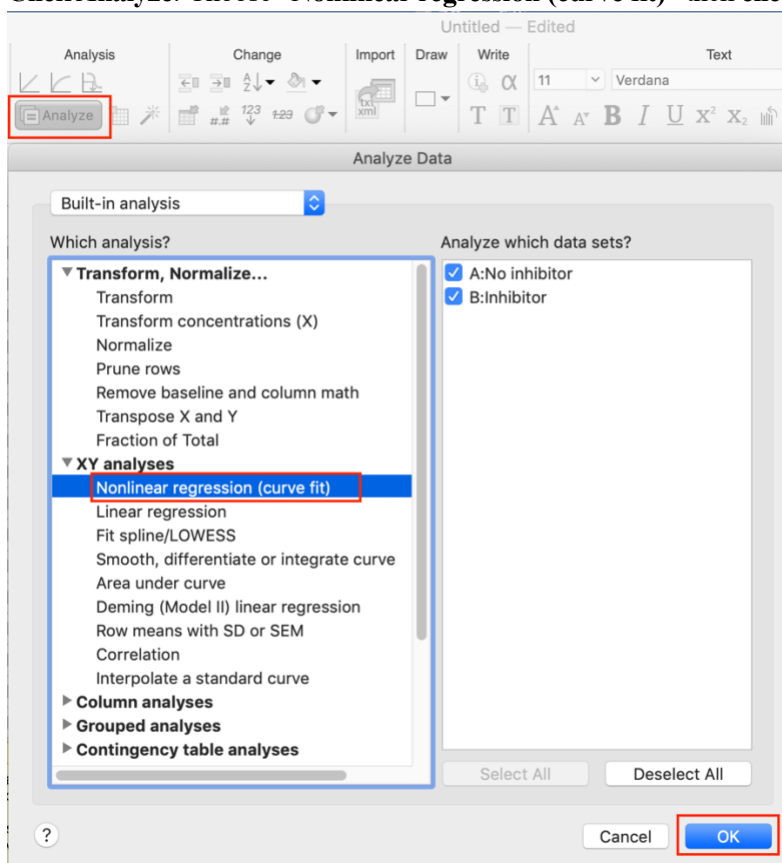
- Simple linear regression
- Linear regression - Compare slopes
- Interpolate unknowns from a linear standard curve
- Nonlinear regression - One phase exponential decay
- Eliminating outliers during nonlinear regression
- RIA or ELISA - Interpolate unknowns from sigmoidal curve
- ▼ Pharmacology**
 - Dose-response - X is log(dose)
 - Dose-response - X is dose**
 - Dose-Response - Ambiguous until constrained
 - Dose-response - EC50 shift by global fitting
 - Binding - Saturation binding to total and nonspecific
 - Binding - Saturation binding, specific binding only
 - Binding - Compare two models - One vs two site competition
 - Schild competitive antagonist - Global nonlinear regression

Cancel Create

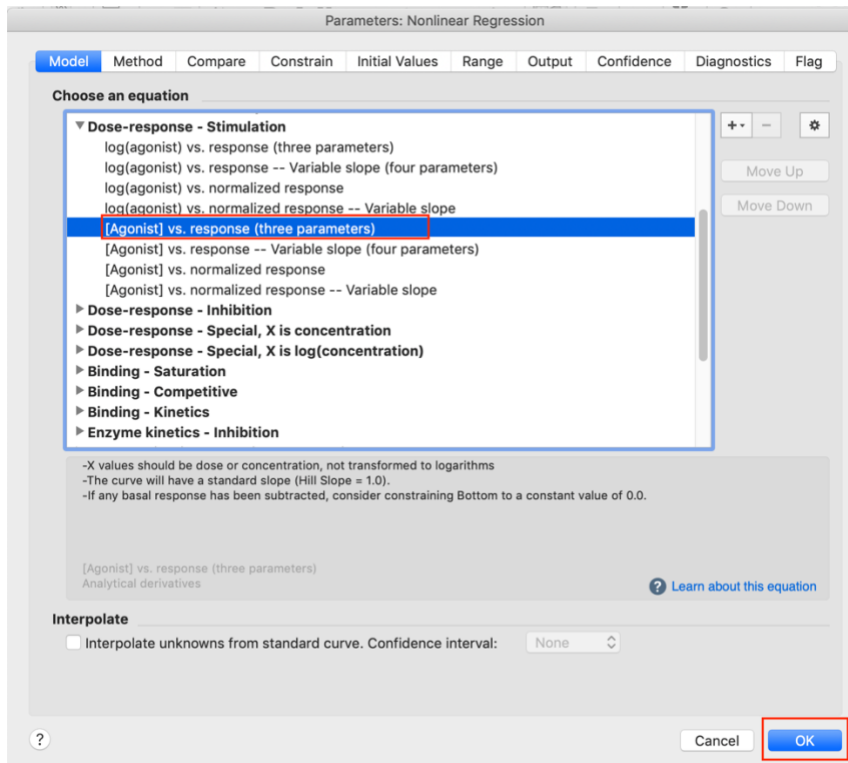
In this example, the X values are molar concentration. The Y values are responses (in triplicate) in two experimental conditions. We would like to fit a dose-response curve to determine the EC50 of the drug, and its Hill Slope, in both conditions.

		X	Group A			Group B		
		[Agonist], M	No inhibitor			Inhibitor		
		X	A:Y1	A:Y2	A:Y3	B:Y1	B:Y2	B:Y3
1	Title	0.000000	0	3	2	3	5	4
2	Title	1.000000e-008	11	33	25			
3	Title	3.000000e-008	125	141	160	11	25	28
4	Title	1.000000e-007	190	218	196	52	55	61
5	Title	3.000000e-007	258	289	345	80	77	44
6	Title	0.000001	322	353	328	171	195	246
7	Title	0.000003	354	359	369	289	230	243
8	Title	0.000010	348	298	372	272	333	310
9	Title	0.000030				359	306	297
10	Title	0.000100	412	378	399	352	320	365
11	Title	0.000300				389	338	

Click **Analyze**. Choose “**Nonlinear regression (curve fit)**” then click “**OK**”.



In the “Parameters: Nonlinear Regression” dialog, choose “**[Agonist] vs. response (three parameters)**”. Click “**OK**”.



The difference between stimulation or inhibition?

Prism offers one set of dose-response equations for stimulation and another set for inhibition. The inhibitory equations are set up to run downhill. The only difference is that the inhibitory equations fit the IC₅₀ ("I" for inhibition) while the stimulation equations fit the EC₅₀ ("E" for effective).

If the curve goes up hill, choose from the set of stimulation equations. If the curve goes down hill, choose from the set of inhibition equations.

Then the results will be like:

Nonlin fit Table of results		A	B
		No inhibitor	Inhibitor
		Y	Y
1	[Agonist] vs. response (three parameters)		
2	Best-fit values		
3	Bottom	-2.228	9.125
4	Top	365.8	343.0
5	EC50	6.814e-008	9.369e-007
6	logEC50	-7.167	-6.028
7	Span	368.0	333.8
8	95% CI (profile likelihood)		
9	Bottom	-31.50 to 25.94	-11.72 to 29.34
10	Top	347.9 to 384.1	324.9 to 361.5
11	EC50	4.705e-008 to 9.987e-008	6.537e-007 to 1.349e-006
12	logEC50	-7.327 to -7.001	-6.185 to -5.870
13	Goodness of Fit		
14	Degrees of Freedom	24	26
15	R squared	0.9640	0.9660
16	Sum of Squares	19278	18248
17	Sy.x	28.34	26.49
18	Constraints		
19	EC50	EC50 > 0	EC50 > 0
20			
21	Number of points		
22	# of X values	30	33
23	# Y values analyzed	27	29

More detailed interpretation of these results could be found at: [Interpreting results: Nonlinear regression](#).

2. Comparing models

Previously the dose-response model was generated WITHOUT a variable slope parameter. Now let's compare it with another dose-response model WITH a variable slope parameter.

Double click the header of the previous results.

Nonlin fit Table of results		A	B
		No inhibitor	Inhibitor
		Y	Y
1	[Agonist] vs. response (three parameters)		
2	Best-fit values		
3	Bottom	-2.228	9.125
4	Top	365.8	343.0
5	EC50	6.814e-008	9.369e-007
6	logEC50	-7.167	-6.028
7	Span	368.0	333.8
8	95% CI (profile likelihood)		
9	Bottom	-31.50 to 25.94	-11.72 to 29.34
10	Top	347.9 to 384.1	324.9 to 361.5
11	EC50	4.705e-008 to 9.987e-008	6.537e-007 to 1.349e-006
12	logEC50	-7.327 to -7.001	-6.185 to -5.870

Go to the 3rd tab “Compare”. Choose “For each data set, which of two equations (models) fits best?” as the question we would like to answer; Choose “Extra sum-of-squares F Test” OR “Akaike’s information Criterion (AICc)” as the comparison method; Choose “[Agonist] vs. response – Variable slope (four parameters)” as the second equation. Click “OK”.

Parameters: Nonlinear Regression

Model Method **Compare** Constrain Initial Values Range Output Confidence Diagnostics Flag

What question are you asking?

- ☐ No comparison
- ☒ For each data set, which of two equations (models) fits best?
- ☐ Do the best-fit values of selected unshared parameters differ between data sets?
- ☐ For each data set, does the best-fit value of a parameter differ from a hypothetical value?
- ☐ Does one curve adequately fit all the data sets?

Comparison method

- ☐ Akaike's Information Criterion (AICc)
Select the model that is most likely to have generated the data.
- ☒ Extra sum-of-squares F Test
Select the simpler model unless the P value less than 0.05

☒ If one fit is ambiguous or flagged, choose the other without formal comparison

Choose the second equation

- ▼ Dose-response - Stimulation
 - log(agonist) vs. response (three parameters)
 - log(agonist) vs. response -- Variable slope (four parameters)
 - log(agonist) vs. normalized response
 - log(agonist) vs. normalized response -- Variable slope
 - [Agonist] vs. response (three parameters)
 - [Agonist] vs. response -- Variable slope (four parameters)**
 - [Agonist] vs. normalized response

For each data set, compare the fit of "[Agonist] vs. response (three parameters)" (chosen on the Model tab) with the fit of a second model (which you choose above).
Prism will fit both models to your data and compare them. Note that if you choose to compare with the extra sum-of-squares F test, the models must be nested - one model must be a special case of the other. If your models are not nested, choose the AICc comparison.

? Cancel **OK**

Prism offers two approaches to comparing models with different numbers of parameters: Extra sum-of squares F test and Information theory approach Akaike’s criterion (AIC):

a) Extra sum-of squares F test

The F test compares the difference in sum-of-squares with the difference you would expect by chance. If the p-value is small, conclude that the simple model is wrong, and accept the more complicated model.

The results of comparing these two models are generated accordingly:

Nonlin fit Table of results		A	B
		No inhibitor	Inhibitor
		Y	Y
1	Comparison of Fits		
2	Null hypothesis	[Agonist] vs. response (three parameters)	[Agonist] vs. response (three parameters)
3	Alternative hypothesis	[Agonist] vs. response -- Variable slope (four parameters)	[Agonist] vs. response -- Variable slope (four parameters)
4	P value	0.4582	0.5651
5	Conclusion (alpha = 0.05)	Do not reject null hypothesis	Do not reject null hypothesis
6	Preferred model	[Agonist] vs. response (three parameters)	[Agonist] vs. response (three parameters)
7	F (DFn, DFd)	0.5692 (1, 23)	0.3400 (1, 25)

In this case, the p-values are ~0.5 which is larger than the traditional value of 0.05, we cannot reject our null hypothesis that the simpler model is correct.

b) Information theory approach Akaike's criterion (AICc)

This alternative approach is based on information theory, and does not use the traditional “hypothesis testing” statistical paradigm. It determines how well the data supports each model, taking into account both the goodness-of-fit (sum-of-squares) and the number of parameters in the model. The results are expressed as the probability that each model is correct, with the probabilities summing to 100%. If one model is much more likely to be correct than the other (say, 1% vs. 99%), you will want to choose it. If the difference in likelihood is not very big (say, 40% vs. 60%), you will know that either model might be correct, so will want to collect more data.

The AICc results of this model comparison are:

Nonlin fit Table of results		A	B
		No inhibitor	Inhibitor
		Y	Y
1	Comparison of Fits		
2	Simpler model	[Agonist] vs. response (three parameters)	[Agonist] vs. response (three parameters)
3	Probability it is correct	76.66%	78.16%
4	Alternative model	[Agonist] vs. response -- Variable slope (four parameters)	[Agonist] vs. response -- Variable slope (four parameters)
5	Probability it is correct	23.34%	21.84%
6	Ratio of probabilities	3.285	3.579
7	Preferred model	[Agonist] vs. response (three parameters)	[Agonist] vs. response (three parameters)
8	Difference in AICc	-2.379	-2.550

In this case, for the group “No inhibitor”, the probability that the model WITHOUT variable slope (76.66%) is correct is higher than the probability that the model WITH variable slope (23.34%) is correct. Therefore, the model WITHOUT variable slope is preferred.

Which approach to choose? F test or AICc?

- If the two models are nested (one model is a simpler case of the other), you may use either the F test or the AIC approach.
- If the models are not nested, then the F test is not valid, so you should choose the information theory approach.

3. Customizing your equation of models

Let's take Boltzmann sigmoidal equation, a classical equation in Prism as an example. Import the **sample data** with XY data table: **Does-response – EC50 shift by global fitting**.

NEW TABLE & GRAPH

XY

Column
Grouped
Contingency
Survival
Parts of Whole
Multiple variables
Nested

EXISTING FILE

Clone a Graph

XY tables: Each point is defined by an X and Y coordinate

	X	A		B	
	Minutes	Control		Treated	
	X	A:Y1	A:Y2	B:Y1	B:Y2
1	Title				
2	Title				
3	Title				

Data table:

☐ Enter or import data into a new table
☒ Start with sample data to follow a tutorial

Select a tutorial data set:

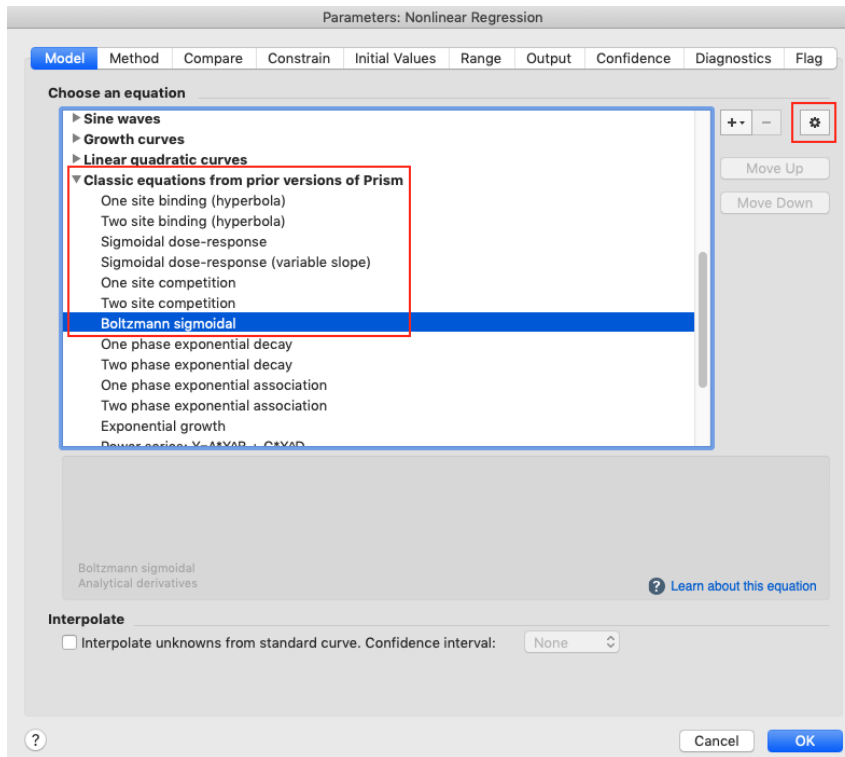
- ▼ **Pharmacology**
 - Dose-response - X is log(dose)
 - Dose-response - X is dose
 - Dose-Response - Ambiguous until constrained
 - Dose-response - EC50 shift by global fitting**
 - Binding - Saturation binding to total and nonspecific
 - Binding - Saturation binding, specific binding only
 - Binding - Compare two models - One vs two site competition
 - Schild competitive antagonist - Global nonlinear regression
 - Operational model of agonist action - Global nonlinear regression
- ▼ **Enzyme kinetics**
 - Enzyme kinetics - Michaelis-Menten
 - Enzyme kinetics - Competitive inhibition
- ▼ **Special uses of XY tables**
 - Entering dates into the X column

Cancel Create

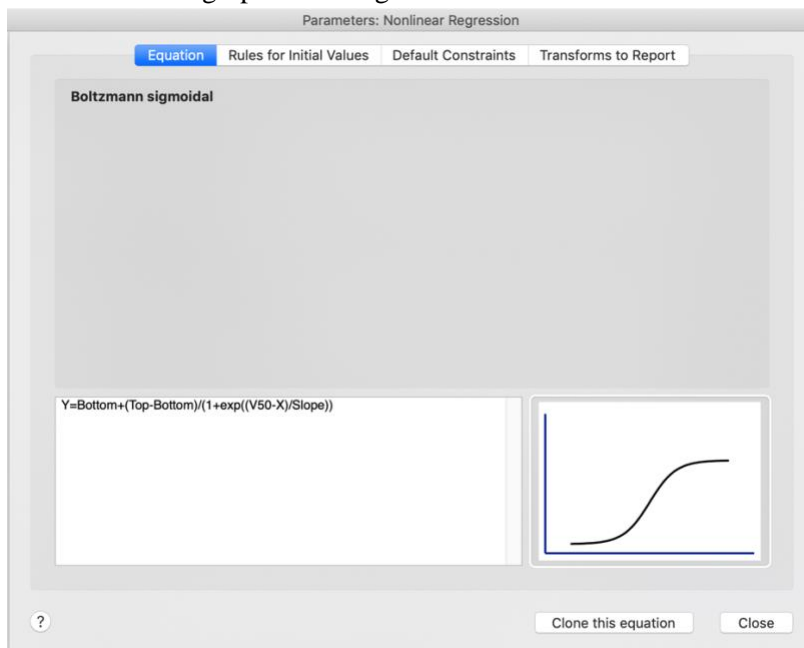
In this case, The X values are the logarithm of the concentration of agonist. The Y values are responses, in duplicate, in two conditions.

		X	Group A		Group B	
		log(agonist)	Control		Treated	
		X	A:Y1	A:Y2	B:Y1	B:Y2
1	Title	-7.5	341	298	295	395
2	Title	-7.0	671	752	616	481
3	Title	-6.5	874	721	362	412
4	Title	-6.0	1000	951	444	700
5	Title	-5.5	1305	1265	882	652
6	Title	-5.0	1254	1351	1354	1089
7	Title	-4.5	1265	1411	1452	1354

Click **“Analyze”**; Choose **“Boltzmann sigmoidal”** under the “Classic equations from prior versions of Prism”. Then **click the settings button** on the right.



It shows a dialog which presents the Boltzmann sigmoidal equation as well as how this equation looks like in the graph on the right.



Click the tab “**Rules for Initial Values**” then you could see the initial values defined in the equation.

Parameters: Nonlinear Regression

Equation **Rules for Initial Values** Default Constraints Transforms to Report

Define a set of rules to compute initial values to use as a default every time you curve fit this equation.

Initial values

Parameter	Value	Rule
Bottom	1	*YMIN
Top	1	*YMAX
V50	1	*(Value of X at YMID)
Slope	1	*SIGN(YATXMAX-YATXMIN)

Default range

Start graphing the curve at: ☒ The smallest X value ☐ X = 0

? Clone this equation Close

Why should we choose initial values? Nonlinear regression is an iterative procedure. The program must start with estimated initial values for each parameter. It then adjusts these initial values to improve the fit.

In this case, we define the initial **Bottom** as $1 * Y_{min}$, **Top** as $1 * Y_{max}$, **V50** as $1 * (\text{Value of X at } Y_{mid})$, and **Slope** as $1 * \text{SIGN}(Y \text{ at } X_{max} - Y \text{ at } X_{min})$.

These initial numbers of the equation could be customized by ourselves. Please **click the “Clone the equation”** on the bottom right. Then it pops up a new window for a new customized equation. You could name it and change the formula by your own.

Parameters: Nonlinear Regression

Equation Rules for Initial Values Default Constraints Transforms to Report

Equation type
Explicit Equation: Y = a function of X and parameters.

Name
Boltzmann sigmoidal [2]

Definition
Y=Bottom+(Top-Bottom)/(1+exp((V50-X)/Slope)) Available functions

Tip

Description

Cancel OK

Then click the tab “Rules for Initial Values” on the top. You could customize the initial values for each parameter.

Equation Rules for Initial Values Default Constraints Transforms to Report

Define a set of rules to compute initial values to use as a default every time you curve fit this equation.

Initial values

Parameter	Value
Bottom	1
Top	1
V50	1
Slope	1

Default range
Start graphing the curve at: ☒ The smallest X value ☐ X

☒ (Initial value, to be fit)
☒ *YMIN
☐ *YMAX
☐ *YMID
☐ *XMIN
☐ *XMAX
☐ *XMID
☐ /YMIN
☐ /YMAX
☐ /YMID
☐ /XMIN
☐ /XMAX
☐ /XMID
☐ *(XMID/YMID)
☐ *(YMID/XMID)
☐ *(YMAX-YMIN)
☐ *(XMAX-XMIN)
☐ /(YMAX-YMIN)
☐ /(XMAX-XMIN)
☐ *(YMAX-YMIN)*(XMAX-XMIN)
☐ *(YMAX-YMIN)/(XMAX-XMIN)
☐ XMIN + (Value)*(XMAX-XMIN)
☐ YMAX/(Value)
☐ *(Value of X at YMIN)
☐ *(Value of X at YMID)
☐ *(Value of X at YMAX)
☐ *Log(Value of X at YMID)
☐ /(Value of X at YMIN)
☐ /(Value of X at YMID)
☐ /(Value of X at YMAX)
☐ /Log(Value of X at YMID)
☐ *(Value of Y at XMIN)
☐ *(Value of Y at XMID)
☐ *(Value of Y at XMAX)
☐ /(Value of Y at XMIN)
☐ /(Value of Y at XMID)
☐ /(Value of Y at XMAX)
☐ *SIGN(YATXMAX-YATXMIN)

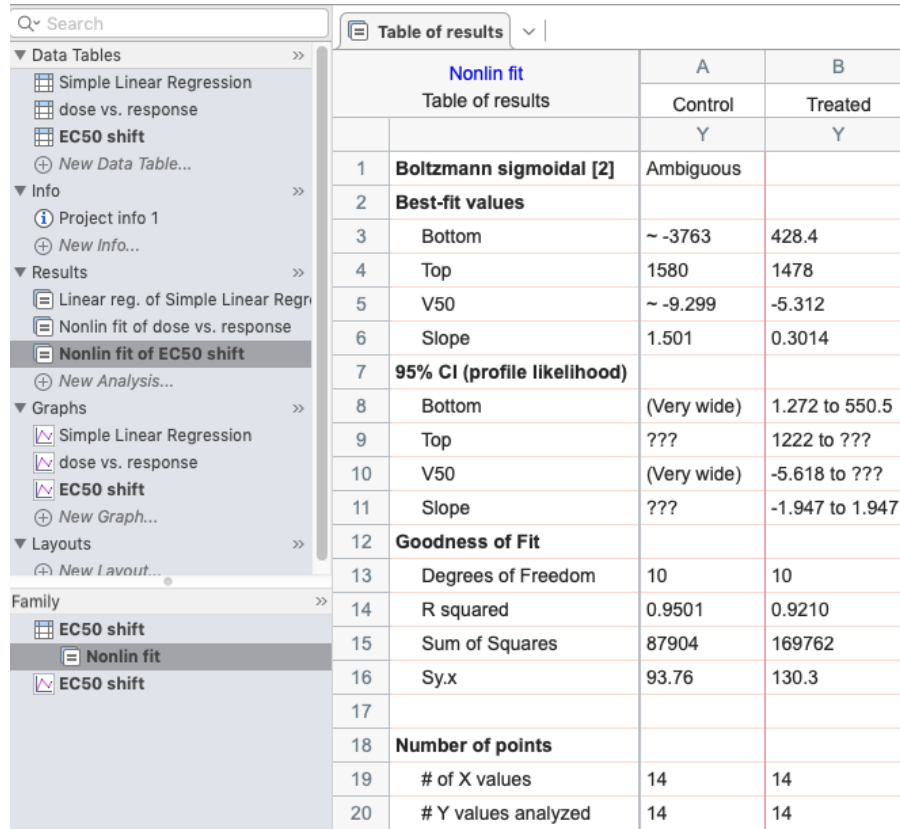
OK

OK

41911	46153	31747	43148
43317	44033	30446	42798
37894	42001	30694	43006

Data 1

Once you click “OK”, Prism 8 generates a result table “Nonlin fit of EC50 shift”.



The screenshot shows the Prism 8 software interface. On the left, there is a sidebar with a search bar and several expandable sections: 'Data Tables', 'Info', 'Results', 'Graphs', and 'Layouts'. The 'Results' section is expanded, showing a list of analyses. The 'Nonlin fit of EC50 shift' analysis is selected and highlighted. Below this, the 'Family' section shows the 'EC50 shift' model, with 'Nonlin fit' and 'EC50 shift' sub-items. On the right, the 'Table of results' is displayed. The table has three columns: 'Table of results', 'A', and 'B'. The 'A' column is labeled 'Control' and the 'B' column is labeled 'Treated'. The table contains 20 rows of data, including parameter estimates, 95% confidence intervals, and goodness of fit statistics.

Nonlin fit		A	B
Table of results		Control	Treated
		Y	Y
1	Boltzmann sigmoidal [2]	Ambiguous	
2	Best-fit values		
3	Bottom	~ -3763	428.4
4	Top	1580	1478
5	V50	~ -9.299	-5.312
6	Slope	1.501	0.3014
7	95% CI (profile likelihood)		
8	Bottom	(Very wide)	1.272 to 550.5
9	Top	???	1222 to ???
10	V50	(Very wide)	-5.618 to ???
11	Slope	???	-1.947 to 1.947
12	Goodness of Fit		
13	Degrees of Freedom	10	10
14	R squared	0.9501	0.9210
15	Sum of Squares	87904	169762
16	Sy.x	93.76	130.3
17			
18	Number of points		
19	# of X values	14	14
20	# Y values analyzed	14	14

How to interpret the results?

- **Best-fit values:** the estimate of each parameter. Please note that these estimates are not necessarily equal to those initial values we defined in the equation.
- **95% CI:** Nonlinear regression is an iterative procedure. The program must start with estimated initial values for each parameter. It then adjusts these initial values to improve the fit.
- **Goodness of Fit:**
 - o R squared: quantifies goodness of fit. Higher values indicate that the model fits the data better.
 - o Sum of Squares: This is useful if you want to compare Prism with another program, or compare two fits manually. Otherwise, the value is not very helpful.

Conclusions

1. BCBB keeps updating this file for more examples and explanations. Please feel free to watch/star our Github repository [BCBB Prism Lab: Curve Fitting](#) to be notified for update.
2. If you have any further question, please let me know qinlu.wang@nih.gov.

Reference:

[Prism 8 Curve Fitting guide](#)