

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/221826649>

Guidelines for accurate EC₅₀/IC₅₀ estimation

Article in *Pharmaceutical Statistics* · March 2011

DOI: 10.1002/pst.426 · Source: PubMed

CITATIONS

324

READS

22,291

1 author:



Jeanne L Sebaugh

31 PUBLICATIONS 868 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Retired, no project [View project](#)

Guidelines for accurate EC50/IC50 estimation

J. L. Sebaugh*

This article provides minimum requirements for having confidence in the accuracy of EC50/IC50 estimates. Two definitions of EC50/IC50s are considered: relative and absolute. The relative EC50/IC50 is the parameter c in the 4-parameter logistic model and is the concentration corresponding to a response midway between the estimates of the lower and upper plateaus. The absolute EC50/IC50 is the response corresponding to the 50% control (the mean of the 0% and 100% assay controls). The guidelines first describe how to decide whether to use the relative EC50/IC50 or the absolute EC50/IC50. Assays for which there is no stable 100% control must use the relative EC50/IC50. Assays having a stable 100% control but for which there may be more than 5% error in the estimate of the 50% control mean should use the relative EC50/IC50. Assays that can be demonstrated to produce an accurate and stable 100% control and less than 5% error in the estimate of the 50% control mean may gain efficiency as well as accuracy by using the absolute EC50/IC50. Next, the guidelines provide rules for deciding when the EC50/IC50 estimates are reportable. The relative EC50/IC50 should only be used if there are at least two assay concentrations beyond the lower and upper bend points. The absolute EC50/IC50 should only be used if there are at least two assay concentrations whose predicted response is less than 50% and two whose predicted response is greater than 50%. A wide range of typical assay conditions are considered in the development of the guidelines. Copyright © 2010 John Wiley & Sons, Ltd.

Keywords: EC50; IC50; 4-parameter logistic model; bioassay; bend points

1. INTRODUCTION

EC50s (effective concentrations) and IC50s (inhibitory concentrations) are used early in the discovery process to evaluate the suitability and the performance of drugs. Assay protocols are developed to determine whether the molecular structures of the tested compounds have specific, desired properties. There may be millions of compounds to test in an assay. The goal is to discover those most likely to have the desired qualities.

Pharmaceutical companies are for-profit institutions. The companies that discover the best lead compounds in the shortest amount of time and at the least cost are at an advantage. Discovering compounds with the greatest potential must be balanced against the costs involved in running the discovery assays. The successful company has discovery researchers who screen compounds efficiently and accurately, identify and refine lead chemistry, and deliver new pharmaceuticals to clinical trials and to market before their competitors.

Computational software routinely produces numeric values for the EC50/IC50 estimates. But, just because a value is obtained does not mean it is accurate. Theoretically, because there are four parameters in the logistic model, four observations are sufficient to characterize a curve. However, this article shows that the position of the observations on the curve has a great impact on the accuracy of the EC50/IC50 estimate.

The purpose of this article was to develop efficient guidelines for producing accurate EC50/IC50 estimates and to maximize the number compounds for which accurate estimates may be obtained. These guidelines are described in a step-by-step fashion that can be easily included with the computational software that produces the EC50/IC50 estimate. Assay costs may be reduced and the time frame for obtaining results may be compressed.

Specific definitions of the EC50/IC50 are given along with the associated theoretical considerations in their estimation. The methodology used to address these issues is described, and a simple process is developed to determine minimum requirements for reporting an EC50/IC50 estimate. Finally, the article shows how, for some types of assays, the number of compounds for which accurate estimates can be made may be increased by following the guidelines.

2. TWO DEFINITIONS OF EC50/IC50S

Definitions of EC50/IC50s contain some built-in assumptions: that there is a monotonic relationship between the dose of the compound and the response in the assay and that there is a consistent definition of a 50% response. The NIH Chemical Genomics Center (in the Third Section of their Assay Guidance Manual) [[1]] proposed the terminology shown in Figure 1 that defines EC50/IC50s in two ways.

To interpret Figure 1, consider an assay plate that has wells for plate controls in addition to wells that contain compound concentrations. The estimate of the 0% (or no drug) effect is a straightforward measurement in any assay. A 0% control well has the same contents as a compound concentration well, except it contains no compound. There are more options for the contents of a 100% (or maximal response) control well. For example, it may have no stimulation, no enzyme, no cells, or a

Sebaugh's Information Services, Columbia, MO, USA

*Correspondence to: J. L. Sebaugh, Sebaugh's Information Services, 3609 Holly Hills Ct. Columbia, MO 65203, USA.
E-mail: sis.stat@centurytel.net

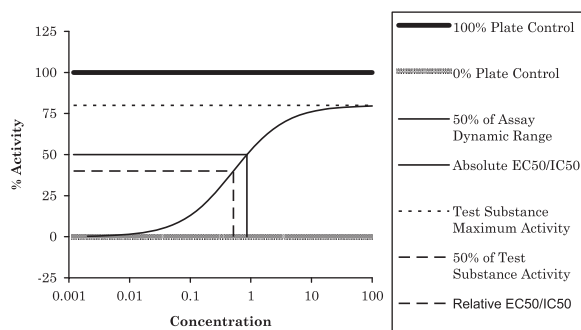


Figure 1. Illustration of two ways of defining EC50/IC50s.

high concentration of a compound known to produce the maximum possible effect.

As illustrated in Figure 1, the 100% plate control mean may not coincide with the maximum activity of the test substance. And the maximal activity may not be the same for all compounds in the assay. Some or all of the compounds being screened may be partial agonists or antagonists. The assay may be screening for compounds that produce a maximal response. The library of compounds may have compounds with very different or even unknown properties. Regardless, it is useful to have a consistent measure of high activity that may be labeled the 100% plate control.

In many cases, the maximal possible activity (or inhibition) of the compounds is anticipated to be the same as the 100% plate control. Of the compounds being screened, the molecular structure considered to be of interest may be limited to those having this characteristic. In such an assay, compounds whose EC50/IC50s exhibit potency of interest, but whose maximal response do not correspond to the 100% plate control, can be identified for further study.

A common way of defining a 50% response is to use a mathematical model that includes a parameter that directly defines it. The 4-parameter logistic model (4PL) is an example of this type of model. The 4PL model describes the sigmoid-shaped response pattern [2,3] that can be used in the majority of screening assays. For ease of discussion, it is assumed that the response can be expressed so that the slope increases as the concentration increases. The formula for the 4PL may be expressed as:

$$Y = \frac{a - d}{1 + (X/c)^b} + d \quad (1)$$

where Y is the response and X is the concentration. The lower asymptote is a , the bottom of the curve or lower plateau (commonly referred to as the min) and the upper asymptote is d , the top of the curve or upper plateau (commonly referred to as the max). The steepness of the linear portion of the curve is described by the slope factor, b . The parameter c is the concentration corresponding to the response midway between a and d . In Figure 1, c is labeled the 'relative EC50/IC50' as it is relative to the maximal response achieved by a compound having a full dose-response curve.

Another way to define a 50% response is the mean of the 0% plate control and the 100% plate control. It is referred to as the 50% control. Nonlinear regression is used to fit a 4PL to the data obtaining estimates of a , b , c , and d . Inverse regression is then used to solve Equation (1) for X setting Y equal to a 50%

response (the 50% control mean).

$$EC50/IC50 = c \left(\frac{a - 50\% \text{ response}}{50\% \text{ response} - d} \right)^{1/b} \quad (2)$$

In Figure 1, this definition of an EC50/IC50 is labeled as the 'absolute EC50/IC50' because the 50% response is defined by the control wells for all compounds on the same plate, not the estimates of the min and max for each compound. If the average of the estimates of the min and max, a and d , equals the 50% response, the relative EC50/IC50 is the same as the absolute EC50/IC50.

3. ISSUES IN EC50/IC50 ESTIMATION

Optimal design theory is useful for increasing the efficiency of EC50/IC50 estimation for individual compounds. But, the theory has not yet addressed the situation studied in this article. Discovery assays may include hundreds of compounds and the assays are usually conducted using automated systems. All compounds are tested at the same set of concentrations determined by a highest concentration, a serial dilution factor, and the total number of concentrations. However, the range of EC50/IC50 estimates is very large. A D-optimal design for the 4PL requires initial estimates of c and b [4]. For a given assay, the slope factor (b) can be estimated, but the expected value for c (the relative EC50/IC50) is unknown and covers a very wide range.

The following quote [5] explains other limitations to D-optimal designs for discovery assays:

D-optimal designs, pure or modified, are practical and useful when the true underlying model is known, a good prior knowledge of parameters is available, and experimental units are relatively dear. A practical limitation of D-optimal designs, which we have encountered, centers at the relative ease in making serial dilutions of drugs in 96-well plate assays. Interestingly, when each experimental unit is very inexpensive, the extra time needed to make the appropriate drug dilutions to make a frugal D-optimal design may result in more trouble and expense than using less frugal serial dilution-based logarithmically spread designs.

These guidelines for accurate EC50/IC50 estimation are developed in the context of the automated systems commonly used. There are some standard curve-fitting attributes (such as a significant dose-response and acceptable variability of the data points around the curve) that should be achieved before accepting any EC50/IC50 estimate. That is, the case regardless of whether it is an estimate of a relative or an absolute EC50/IC50 because both are model-based. But in addition, both types of estimates require special considerations before the estimates are considered accurate enough to be reported.

The relative EC50/IC50, c , directly corresponds to the response midway between the estimates of the top and bottom of the curve, a and d . An accurate estimate of c is therefore dependent on accurate estimates of a and d . The estimates of a and d cannot be accurate unless there are sufficient assay concentrations having responses on the lower and upper plateaus ([6], p.173). The plateaus are those portions of the 4PL curve that extend beyond the linear portion in the middle of the curve. Formulas for

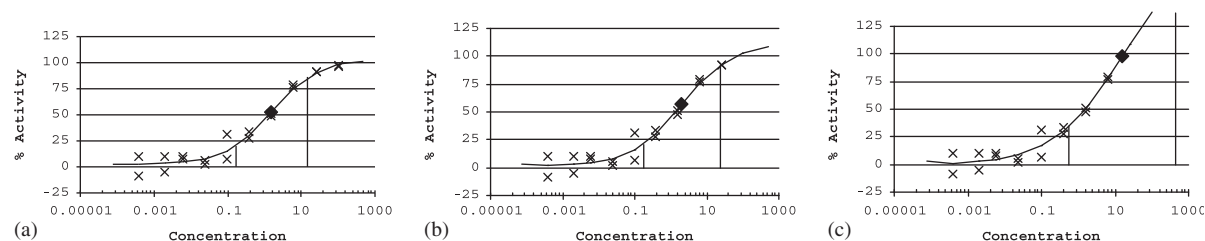


Figure 2. (a) Shows the estimated curve with 10 concentrations, (b) omits the highest concentration, and (c) omits the two highest concentrations. Vertical lines show the bend points and the diamond symbol shows the estimate of c .

Table I. Relative EC50/IC50 estimates for Figure 2(a–c).

Figure	Parameter estimates				Bend points (X, Y)		No. concentrations on plateaus	
	a	d	b	c	Lower	Upper	Bottom	Top
a	2.3	103.7	0.69	1.57	(0.17, 20.2)	(14.6, 85.8)	5	2
b	1.6	111.7	0.63	2.04	(0.17, 21.0)	(24.0, 92.4)	5	1
c	−0.3	196.1	0.46	15.33	(0.55, 34.3)	(425.5, 161.5)	6	0

Table II. Absolute EC50/IC50 estimates for Figure 2(a–c).

Figure	No. concentrations beyond 50%		Absolute EC50/IC50
	Lower	Upper	
a	5	3	1.32
b	5	2	1.38
c	6	1	1.54

the coordinates of the beginning and end of the linear portion of the curve (bend points) are given in Sebaugh and McCray [6]. The bend point concentrations may be calculated as:

$$c * (k)^{(1/b)} \quad \text{and} \quad c * (1/k)^{(1/b)} \quad (3)$$

where c and b are parameters of the 4PL curve (1) and $k=4.6805$. Once these bend point concentrations are calculated, the number of more extreme assay concentrations beyond them on the lower and upper plateaus may be counted.

A motivating example is demonstrated in Figure 2. The 0% and 100% plate controls were used to calculate percent activity values for the observed data shown in Figure 2. Using nonlinear regression, the data were used to obtain the Maximum Likelihood Estimation (MLE) parameter estimates shown in Table I. The relative EC50/IC50 estimate changed from 1.57 to 2.04 to 15.33 as the minimum number of concentrations on the upper plateau decreased from 2 to 1 to 0. After two of the highest concentrations have been deleted, the top of the curve is much higher than 100% (196.1%) because there are no data to define the upper plateau. The midpoint between the top and bottom is 98%, not 50%.

Similarly, the number of concentrations greater than 50% decreased from 3 to 2 to 1. Table II shows the absolute EC50/IC50s changed from 1.32 to 1.38 to 1.54. These estimates are more consistent but still illustrate a potential issue in the accuracy for the absolute EC50/IC50.

The general purpose of this investigation is to develop the guidelines required for accurate EC50/IC50 estimation. Two specific questions are answered. How many concentrations are required beyond the lower and upper bend points for accurate relative EC50/IC50 estimates? How many concentrations on either side of a

50% response are required for accurate absolute EC50/IC50 estimates?

4. METHODOLOGY

To answer the questions posed above, experimental results were generated for a variety of assay scenarios. Factors were considered that result from the design of an assay and that have an impact on the accuracy of EC50/IC50 estimates. As indicated in Figures 3 and 4 of Sebaugh and McCray [6], two important factors to consider are the slope parameter of the 4PL and the dilution factor. This is because these two factors have an impact on the number of assay concentrations that are in the linear portion of the curve. The dilution factor is the factor that defines the spacing between adjacent test concentrations. Using a dilution factor of 2, each concentration equals the next highest concentration divided by 2.

Other factors that may influence the accuracy of the estimates are the total number of concentrations and the variability of the responses at the same concentration. The total number of concentrations is a function of the size and layout of the assay plate. It is often determined by throughput requirements for the assay. The variability of the responses is a function of the type of assay. For example, cell-based assays generally have greater biological variability than other types of assays. To consider a range of variability, two replicate observations were generated at each concentration.

The following scenarios were considered: all combinations having slopes of 0.5, 1, and 2; dilution factors of 2, 5, and 10; 6, 10, and 22 concentrations per compound; and root mean square errors (RMSEs) of 5%, 10%, and 20% (assuming the dependent variable is a percentage). As there were three options for each of four conditions, there are $3*3*3*3=81$ scenarios. The options considered cover the most common range of possible assay designs.

Five hundred 4PL curves were generated for each scenario (40 500 curves). The curves had fixed values for three of the four parameters. Without loss of generality, the bottom of the curve, a , was set to 0% and the top of the curve, d , was set to 100%. Fixing the top and bottom of the curve at these values makes

the assumption that if, in theory, a sufficient number of concentrations were used, a complete dose–response curve would be generated. The slope (b) was fixed (at 0.5, 1, or 2). To generate the test concentrations for each combination, the only parameter remaining to be specified was the EC50/IC50 (c). This parameter was generated in such a way as to simulate real-life assays that consist of compounds having varying potency (described below). Two responses were generated for each test concentration using the scenario’s RMSE value to represent the error in measurement between wells having the same compound concentration.

Thus, two seeds were used to generate the normally distributed, randomly generated values in the simulations: one for parameter c and one for the dependent Y values (in duplicate) at each concentration. The initial seeds were the same for each of the 81 scenarios to simulate the same compounds being tested under each set of conditions. The highest concentration was arbitrarily set to 100 μM . The mean for the randomly generated log of the EC50 was calculated as the log of the highest concentration minus one-third of the difference between the log of the highest and lowest concentration for that simulation (67th log percentile). These concentrations were chosen to simulate discovery-screening assays in which there is a preponderance of less potent compounds that have higher EC50/IC50s. The log standard deviation used for the randomly generated model log EC50 was generated with the intention that 15% of the log EC50s would be greater than the highest concentration. Thirteen percent of the curves generated had actual EC50s greater than the highest concentration.

Figure 3 shows a sample of the curves for the simulated scenario with 10 concentrations, a dilution factor of 5, a slope of 1, and an RMSE of 10. The values of a , d , b , and c will be referred to as the actual curve values.

Taking the antilog of the log EC50 provides a value for c that along with the other three parameters allows a preliminary Y value to be calculated. Using the second seed, a normal value between 0 and 1 was generated. It was multiplied by the scenario RMSE value and added to the preliminary Y value to produce the final Y value. The mean RMSE values from the curve fits were 5.5%, 9.9%, and 18.9% for the curves whose theoretic RMSEs were 5%, 10%, and 20%, respectively. Thus, the results obtained using the randomly generated values were very close to the targeted values.

SAS PROC NLIN was then used to estimate the four parameters of the logistic model, the min (a), the max (d), the slope (b), and the relative EC50 (c). The Marquardt method was used with the convergence criterion set to $10\text{E-}7$ and the maximum number of iterations set to 999. The c parameter was constrained to be greater than 0.

Eight-nine percent of the curve fits converged in 1 to 996 iterations. Eighty percent of the curve fits generated a numeric standard error for the EC50 that ranged from zero to $4\text{E}+75$. It is standard practice for investigators not to report the results for an EC50/IC50 from a curve with a negative slope or with more than 20% of the variability in the responses unexplained by the curve fit. Therefore, in order to focus more on the central questions, curves were not considered that had negative slope estimates or R^2 values less than 80% (29% of the curves). The estimate of the c parameter provided the estimate for the relative EC50/IC50. The absolute EC50/IC50 could be calculated as long as the curve spanned the 50% response (estimates of a less than 50% and estimates of d greater than 50%). The absolute EC50/IC50 was not used unless there was also at least one response less than the 50% response and one response greater than the 50% response (68% of the curves). Normally distributed, randomly generated values were generated for the 0% and 100% controls using means of 0 and 100 and the same standard deviation (RMSE value) as was used for the simulated response data. The estimate of the absolute EC50/IC50 was calculated using the mean of the two controls as the 50% control response.

The absolute value of the difference in the log EC50/IC50s (base 10) between actual and estimated values was calculated for each curve. Greater accuracy is demonstrated by differences closer to zero. These difference values are not normally distributed. Therefore, statistics are presented for the median (50th percentile) and the interquartile range, Q1 (25th percentile) and third quartile values (Q3) (75th percentile).

5. NUMBER OF CONCENTRATIONS REQUIRED ON PLATEAUS FOR ACCURATE RELATIVE EC50/IC50 ESTIMATION

The first question addressed here is the number of concentrations required beyond the lower and upper bend points in order to accurately estimate the relative EC50/IC50. Table III shows the

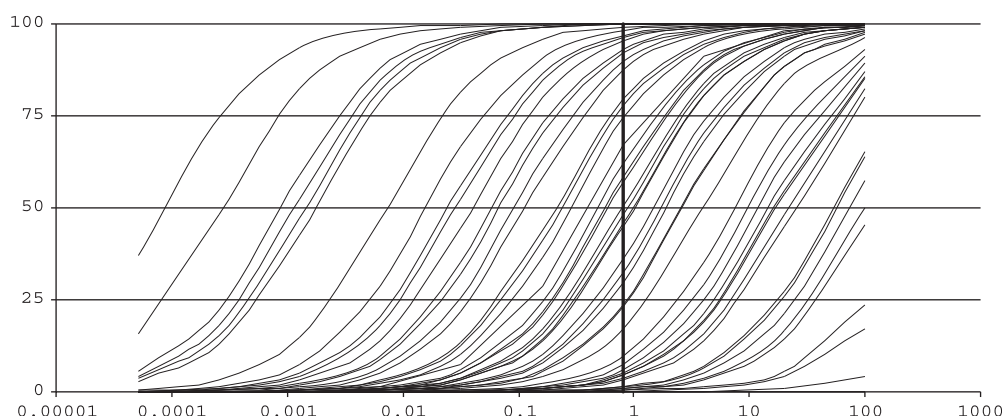


Figure 3. A sample of the simulated curves for the scenario with 10 concentrations, a dilution factor of 5, a slope of 1, and an RMSE of 10. The vertical line is the mean used for log c when generating the actual c values for the simulation.

Table III. Median and interquartile range (Q1 and Q3) of the absolute value of the log difference (base 10) between actual and estimated relative EC50/IC50 values.

Smallest number of concentrations beyond bend points	Absolute log difference				
	<i>n</i>	%	Q1	Median	Q3
(a) 6 concentrations					
0	3623	35.9	0.50	1.80	16.32
1	3763	37.3	0.12	0.31	0.71
2	2592	25.7	0.09	0.21	0.45
3	100	1.0	0.13	0.22	0.42
Total	10 078				
(b) 10 concentrations					
0	2011	20.0	0.50	1.66	15.32
1	1754	17.5	0.14	0.34	0.79
2	2377	23.7	0.11	0.25	0.56
3	2477	24.7	0.10	0.24	0.53
4	1347	13.4	0.09	0.21	0.46
5	81	0.8	0.06	0.17	0.35
Total	10 047				
(c) 22 concentrations					
0	702	8.1	0.44	1.11	3.75
1	486	5.6	0.17	0.37	0.80
2	534	6.2	0.08	0.22	0.46
3	663	7.7	0.08	0.19	0.43
4	683	7.9	0.08	0.21	0.47
5	920	10.7	0.09	0.22	0.47
6	1158	13.4	0.10	0.22	0.50
7 or more	3482	40.4	0.09	0.21	0.48
Total	8628				

results for the absolute value of the difference in the log EC50/IC50s between actual and estimated values. As can be seen in Table III, regardless of the number of concentrations, results for curves that do not have at least one concentration beyond both bend points should not be used. Thirty-six percent of these curves have relative EC50/IC50 estimates higher than the highest assay concentration (100), 13% have estimates lower than the lowest assay concentration, and 51% have estimates within the range of assay concentrations. The EC50/IC50 estimates for all of the curves with at least one assay concentration beyond both ends of the linear range are within the range of assay concentrations (55% of the curves).

For 6 or 10 concentration assays that have no concentrations beyond at least one of the bend points, the Q3 values show that 25% of the results would have more than a ten log difference from the actual value. For assays with 22 concentrations, 25% of the results would have almost a four log difference. In the case of one or more concentrations beyond both bend points, all the medians are less than a half log (0.5) and all the Q3 are less than one log (1.0).

In all cases, using a nonparametric analysis of variance shows that there is a statistically significant difference in the average log differences between having one and two concentrations beyond the bend points but not between two and three concentrations beyond the bend points. Thus, a guideline can be stated that says at least two concentrations are required

Table IV. Median and interquartile range (Q1 and Q3) of the absolute value of the log difference (base 10) between actual and estimated absolute EC50/IC50 values.

Smallest number of concentrations beyond 50%	Absolute log difference				
	<i>n</i>	%	Q1	Median	Q3
(a) 6 concentrations					
1	2192	23.7	0.08	0.21	0.45
2	4523	48.9	0.08	0.19	0.38
3	2542	27.5	0.08	0.19	0.38
Total	9257				
(b) 10 concentrations					
1	969	10.1	0.10	0.21	0.42
2	1684	17.5	0.07	0.18	0.38
3	2452	25.5	0.07	0.17	0.37
4	3011	31.4	0.08	0.19	0.40
5	1483	15.4	0.10	0.22	0.46
Total	9599				
(c) 22 concentrations					
1	345	4.0	0.07	0.16	0.35
2	265	3.1	0.07	0.14	0.29
3	604	7.1	0.08	0.16	0.30
4	579	6.8	0.05	0.12	0.24
5	586	6.9	0.06	0.16	0.35
6	874	10.2	0.08	0.19	0.37
7 or more	5278	61.9	0.08	0.19	0.45
Total	8531				

beyond the lower and upper bend points in order to accurately estimate the relative EC50/IC50.

6. NUMBER OF CONCENTRATIONS REQUIRED SPANNING 50% FOR ACCURATE ABSOLUTE EC50/IC50 ESTIMATION

The next question is whether accurate EC50/IC50 estimates can be made for incomplete curves and/or curves that do not have at least two concentrations on both plateaus. Table IV shows the absolute difference between the log of the actual EC50/IC50 and the log of the estimate of the absolute EC50/IC50. All the medians are less than a quarter log (0.25) and all the Q3 are less than a half log (0.5).

Using a nonparametric analysis of variance shows that there is a statistically significant difference in the average log differences between having one and two concentrations beyond 50%, but not between two and three concentrations beyond 50%. Thus, it can be stated that two concentrations are required beyond 50% in order to accurately estimate the absolute EC50/IC50.

7. COMPARISON OF ESTIMATES FOR THE RELATIVE AND ABSOLUTE EC50/IC50

The purpose of the previous analyses was to determine guidelines or rules for obtaining accurate EC50/IC50 estimates, relative and absolute. For relative EC50/IC50s, we found that greater accuracy is not obtained by requiring more than two assay concentrations beyond the linear region, regardless of the number of assay concentrations. Thus, the rule for accurate relative EC50/IC50 estimates is that at least two assay

Table V. Median and interquartile range (Q1 and Q3) of the absolute value of the log difference (base 10) between known and estimated relative and absolute EC50/IC50 values for curves meeting recommended guidelines.

Type of EC50/IC50	No. assay concentrations	<i>n</i>	Absolute log difference		
			Q1	Median	Q3
Relative	6	2692	0.09	0.21	0.45
	10	6282	0.10	0.23	0.53
	22	7440	0.09	0.21	0.48
Absolute	6	7687	0.08	0.18	0.39
	10	8938	0.08	0.18	0.38
	22	8254	0.07	0.17	0.38

concentrations should occur on both the lower plateau and the upper plateau. For absolute EC50/IC50 estimates at least two assay concentrations should occur both below the 50% response concentration and above the 50% response concentration.

Table V shows the results using these guidelines. There are fewer relative than absolute EC50/IC50 estimates (65% fewer at six concentrations, 30% fewer at ten concentrations, and 10% fewer at twenty-two concentrations). The absolute differences from the actual EC50/IC50 values are smaller for the absolute EC50/IC50 values.

It must be acknowledged that this precision is obtained by an assumption that the control mean responses at 0% and 100% provide an accurate estimate of a 50% response. However, errors in the measurement of the control means (and thus errors in the 50% control mean) may occur because the control wells have not been sufficiently tested to ensure that the 0% mean corresponds to the lower plateau and the 100% plateau corresponds to the upper plateau. The following analysis examines the effect of possible errors in the control means that result in the 50% response calculated from the controls being in error by up to 10%. An error of 10% means that the 50% mean calculated from the 0% and 100% controls is biased, actually corresponding to 40% or 60%.

Table VI shows the median error in the absolute EC50/IC50 estimates. When the error in the response calculated from the 0% and 100% controls is 2.5% or less, the error in the absolute EC50/IC50 estimates is smaller than the error in the relative EC50/IC50 estimates. As the magnitude of the error in the 50% control response increases, the accuracy of the absolute EC50/IC50 drops, although the magnitude of the error is still less than half a log.

8. GUIDELINES

Guidelines have been developed for both relative and absolute EC50/IC50 estimation. The first set of rules or guidelines determines which type of EC50/IC50 may be considered. Assays for which there is no stable 100% control must use the relative EC50/IC50. Assays having a stable 100% control but for which there may be more than 5% error in the estimate of the 50% control mean should use the relative EC50/IC50. Assays that can be demonstrated to produce an accurate and stable 100% control and less than 5% error in the estimate of the 50% control mean may gain efficiency as well as accuracy by using the absolute EC50/IC50 because under these conditions accurate estimates are obtained for more compounds.

Table VI. Median and interquartile range (Q1 and Q3) of the absolute value of the log difference (base 10) between known and estimated absolute EC50/IC50 values for curves with different amounts of error in the 50% control.

No. assay concentrations	Error in 50% control mean	<i>n</i>	Absolute log difference		
			Q1	Median	Q3
6	0	7687	0.08	0.18	0.39
	2.5	7639	0.09	0.20	0.41
	5	7576	0.12	0.26	0.50
	7.5	7529	0.16	0.33	0.60
	10	7451	0.21	0.41	0.73
10	0	8938	0.08	0.18	0.38
	2.5	8897	0.09	0.20	0.41
	5	8845	0.12	0.26	0.50
	7.5	8815	0.16	0.33	0.61
	10	8785	0.21	0.42	0.75
22	0	8254	0.07	0.17	0.38
	2.5	8262	0.09	0.20	0.42
	5	8271	0.12	0.27	0.51
	7.5	8277	0.17	0.35	0.64
	10	8291	0.23	0.44	0.78

Next, guidelines were developed for when each type of EC50/IC50 should be reported or used. For relative EC50/IC50 estimates, obtain the parameter estimates for the 4PL model and use Equation (3) to calculate the lower and upper bend point concentrations. This allows the number of assay concentrations that are smaller than the lower bend point and larger than the upper bend point to be counted and the guideline to be applied. The EC50/IC50 estimate should only be used if there are at least two concentrations beyond both the lower and upper bend points, i.e. beyond the linear portion of the curve.

For absolute EC50/IC50 estimates, obtain the parameter estimates for the 4PL model. If the estimate for the lower plateau is less than 50% and the estimate for the upper plateau is greater than 50%, the absolute EC50/IC50 can be calculated using Equation (2). If the absolute EC50/IC50 can be estimated, calculate the predicted *Y* values for each concentration. Then count the number of concentrations whose response is less than 50% and the number whose response is greater than 50%. The absolute EC50/IC50 estimate should only be used if there are at least two concentrations whose predicted response is less than 50% and two whose predicted response is greater than 50%.

9. DISCUSSION

The purpose of this investigation was not to compare the results among the 81 scenarios. Rather by using these diverse scenarios, the guidelines that were developed can be considered to be broadly applicable. However, there are a few comments that can be made about the scenarios. As might be expected, the greatest precision was found for assays having the smallest RMSE, an RMSE of 5%. In addition, within each RMSE level, the greatest precision and the least precision were found for assays having the highest slope, 2.0 and the smallest slope, 0.5, respectively.

The guidelines suggested here are minimum requirements for having confidence in the EC50/IC50 estimate. These rules should

be supplemented with others to ensure accurate and informative results. Examples of other possible rules could include agreement of the 0% and 100% control means with the estimates of the lower and upper plateaus, consistent concentration–response across the concentration spectrum, closeness of predicted and observed values, and reasonable values for the coefficient of variation of the EC50/IC50 estimate.

Hopefully, these findings will stimulate more informed theoretical development. In particular, research that will aide in the design of assays that routinely include test compounds having a wide range of EC50/IC50 values. This article makes it clear how critical it is to establish a range of concentrations such that the curves for the targeted compounds have sufficient concentrations on the plateaus or around the 50% response.

Acknowledgement

The author thanks the reviewers for their many helpful comments that substantially improved the quality of this paper.

REFERENCES

- [1] Assay Guidance Manual Version 5.0. Eli Lilly and Company and NIH Chemical Genomics Center. 2008. Available at: http://www.ncgc.nih.gov/guidance/manual_toc.html (accessed 13.02.2010).
- [2] De Lean A, Munson PJ, Rodbard D. Simultaneous analysis of families of sigmoidal curves: application to bioassay, radioligand assay, and physiological dose-response curves. *American Journal of Physiology* 1978; **235**(2):E97–E102.
- [3] Finney DJ. *Statistical method in biological assay* (3rd edn). Charles Griffin & Co: London, 1978.
- [4] Li G, Dibyen M. D-optimal designs for logistic models with three and four parameters. *Journal of Statistical Planning and Inference* 2008; **138**(7):1950–1959. DOI: 10.1016/j.jspi.2007.07.010.
- [5] Khinkis LA, Levasseur L, Faessel H, Greco WR. Optimal design for estimating parameters of the 4-parameter Hill model. *Nonlinearity in Biology, Toxicology, and Medicine* 2003; **1**(3):363–377. DOI: 10.1080/15401420390249925.
- [6] Sebaugh JL, McCray PD. Defining the linear portion of a sigmoid-shaped curve: bend points. *Pharmaceutical Statistics* 2003; **2**:167–174. DOI: 10.1002/pst.062.