



# Bench Marks!

## Normalization of Data for Viability and Relative Cell Function Curves

Alice Krebs<sup>1,2</sup>, Johanna Nyffeler<sup>1,3</sup>, Jörg Rahnenführer<sup>4</sup> and Marcel Leist<sup>1,2,5</sup>

<sup>1</sup>In vitro Toxicology and Biomedicine, Dept inaugurated by the Doerenkamp-Zbinden Foundation, University of Konstanz, Konstanz, Germany; <sup>2</sup>Konstanz Research School Chemical Biology (KoRS-CB), University of Konstanz, Konstanz, Germany; <sup>3</sup>present address: National Centre for Computational Toxicology, US EPA, Research Triangle Park, NC, USA; <sup>4</sup>Department of Statistics, TU Dortmund University, Dortmund, Germany; <sup>5</sup>CAAT-Europe, University of Konstanz, Konstanz, Germany

### Summary

Many types of assays in cell biology, pharmacology and toxicology generate data in which a parameter is measured in a reference system (negative control) and then also under conditions of increasing stress or drug exposure. To make such data easily comparable, they are normalized, i.e., the initial value of the system (e.g., viability or transport function) is set to 100%, and all data are indicated relative to this value. Then, curves are fitted through the data points and summary data of the system behavior are determined. For this, a benchmark response (BMR) is given (e.g., a curve drop by 15 or 50%), and the corresponding benchmark concentration (BMC15 or BMC50) is determined. Especially for low BMRs, this procedure is not very robust and often results in incorrect summary data. It is often neglected that a second normalization (re-normalization) is necessary to make the data suitable for curve fitting. It is also frequently overlooked that this requires knowledge of the system behavior at very low stress conditions. Here, good *in vitro* practice guidance for the re-normalization procedure is provided so that data of higher fidelity can be generated and presented.

Keywords: BenchMarks series

### Initial situation

Assume you had a good week and performed three experiments testing the effect of a drug on cell viability or, perhaps, on transporter activity (as examples of any of thousands of cell functions). You found that increasing drug concentrations decrease your readout. To make experiments easier to compare, and to visualize the data in a standardized way, you normalized all data so that untreated controls are set to 100%. In your case, data at high drug concentrations are far below 100%, possibly even tending towards 0%. Now you want to determine certain summary data. These indicate, for instance, which drug concentration leads to a readout decrease by 10% or 15% (or 50%), compared to controls (Fig. 1). This is an everyday question in pharmacology and toxicology labs, and it looks as if answering it should be a matter of routine.

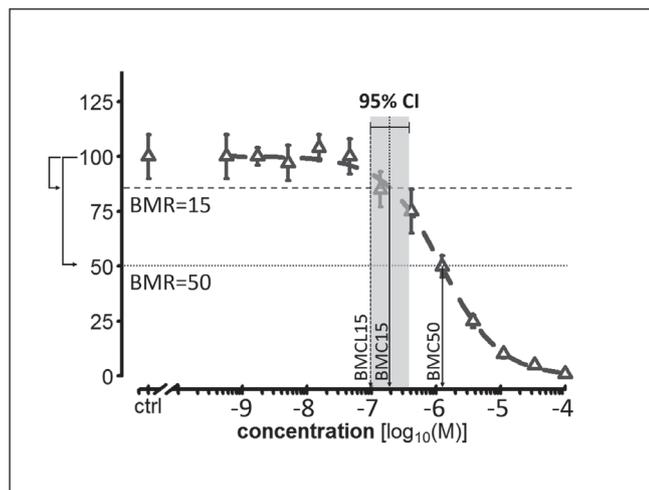
### Field practice

At a recent toxicology meeting we checked 100 posters displaying concentration-response curves. There was good agreement that a curve should be fitted to the data so as to minimize the distances between the data and the curve. Moreover, the definition of the desired summary data was unanimously accepted to be the concentration at which the curve had dropped by a pre-defined percentage, e.g. 15% (or 50%). Concerning the curve fitting, a number of different approaches were used. They ranged from using a fixed mathematical model (e.g., linear, logistic or Weibull) to modelling large numbers of curve functions and optimizing for or selecting the best fit. The most commonly used curves were typical sigmoidal curves generated by using a 4-parameter log-logistic function.

Received: March 23, 2018;  
doi:10.14573/altex.1803231



This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International license (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is appropriately cited.



**Fig. 1: Illustration of the concepts of benchmark response and benchmark concentrations**

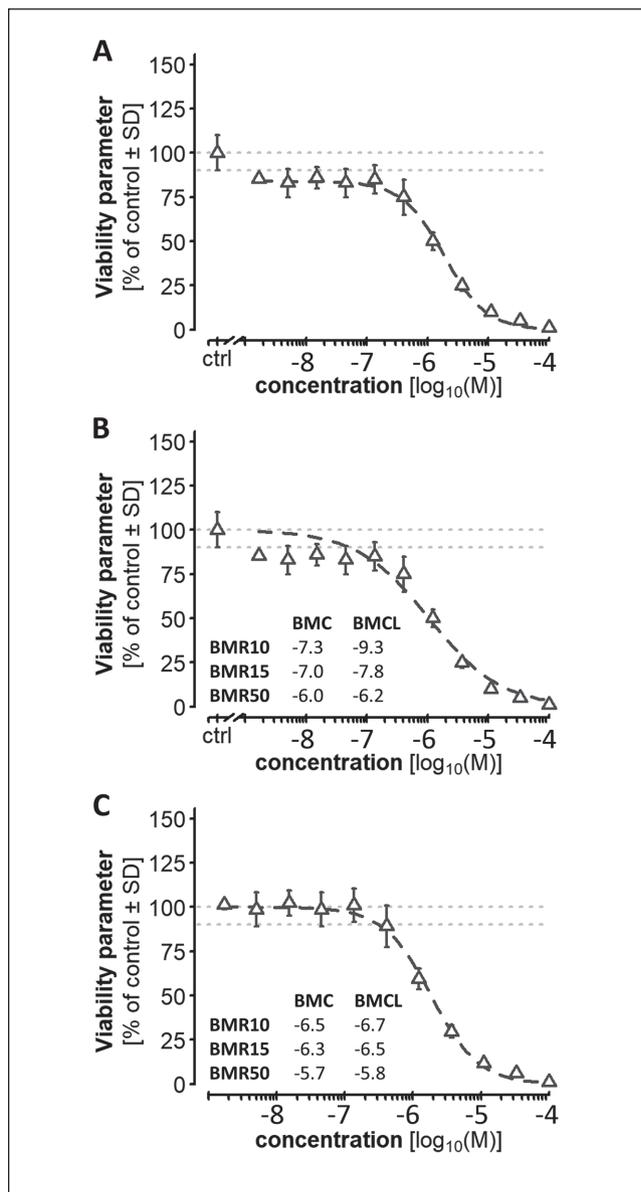
An exemplary normalized data set is shown, with a curve fit that has the upper asymptote at 100% (= negative control value). Two exemplary benchmark responses (BMR) are shown at 85% (BMR15, dashed line) and at 50% (BMR50, dotted line). The corresponding benchmark concentrations (BMC) are the concentrations at which the curve reaches the BMR. In particular contexts, the BMC50 can be named an effective concentration (EC50), an inhibitory concentration (IC50) or an active concentration (AC50). The BMC15 can be used in some contexts to define the highest non-active concentration (if a change from baseline of up to 15% is considered to be baseline noise). However, each of these summary data points has an uncertainty. The uncertainty of the BMC15 is shown as the 95% confidence interval (CI). The lower boundary of this CI (BMCL) is the BMCL15. Strictly speaking, the BMCL, and not the BMC, is the highest definitely non-active concentration.

### The stumbling block

However, many of the graphs looked like Figure 2A, i.e., the upper asymptote of the fitted curve did not run at 100%, but slightly above or below. Considering the example shown, and a BMR of 15% (indicated by the dotted line that cuts the y-axis at 85%), how would you determine the BMC15 (or IC15)?

### Definition of the problem

Typical sigmoidal curves, as shown in Figure 1, are obtained by a 4-parameter fit. These 4 parameters determine the lower and upper asymptote, the turning point of the curve and the steepness of the curve (at its turning point). Most programs allow these parameters to be adapted automatically (to best fit the data points) or to be predefined by the operator. If the parameter defining the upper asymptote is adapted automatically, it is unlikely to be exactly at 100%. Thus, if this program setting is used, there is a problem in defining the BMR. There is no easy solution, as evidenced by some bizarre situations it can create: (i) Assume that the starting point for the BMR definition is the 100% value. If the upper asymptote is, e.g., at 120%, then a 15%



**Fig. 2: Normalization and curve fitting through a set of example data**

A set of example data was chosen for a typical cytotoxicity effect of a toxicant active in the  $\mu\text{M}$  range. Data were normalized to the control value. (A) A 4-parameter log-logistic regression curve choosing the upper asymptote automatically was fitted through the data. The values of about 84% for the lowest data points suggest that viability may be reduced by about 15% in the nM range. Note the relatively large error bar of the control, suggesting that there may be a problem with this value. (B) The same data as in A were used for curve fitting forcing the upper asymptote through 100%. The BMC and BMCL values (in log(M)) are indicated in the insert for three different benchmark responses. (C) The control value was taken out of the data set. All data were re-normalized to the upper asymptote in A (84 was set to 100%), and a new regression curve was fitted to the data. The insert shows the BMC and BMCL values for this curve fit after re-normalization.



drop would be to 105%, meaning that the beginning of cytotoxicity or functional failure would be predicted for fully viable and functional cells; (ii) Assume again that the starting point for the BMR definition is the 100% value. If the upper asymptote is, e.g., at 80%, then a BMR of 15 would be above the curve, meaning that cells would need to increase viability in order to die. (iii) Assume that the starting point for the BMR definition is the upper asymptote of the curve, i.e., 84. A BMR of 50 would then be at 42%. This means that the half-maximal effect concentration is found where only 42% of the cells are viable/functioning. Although mathematically correct, this is biologically counter-intuitive. These examples illustrate that many problems arise if the upper asymptote is not forced through 100%.

### There is also a reverse problem

An apparently simple solution to the above problem is to force the upper asymptote through 100% (Fig. 2B). Here, the issue is that then the curve may not really follow the data points, i.e., the curve fit would not correspond to the biological response it is intended to model and thus summary data derived from the curve would not be correct.

### Extent of the problem for various BMR

An important question is how relevant the problem is in practice. The extent of the problem differs greatly depending on the chosen BMR. If the BMR is 50 (classical EC50 values), a small shift of the asymptote above or below 100 plays only a minor role, especially if the slope of the curve is high. However, if the BMR is 10, i.e., if the beginning of the curve is considered, then an offset of the asymptote can play a large role or even lead to unsolvable situations. As the IC50 has been used more commonly in publications than IC10, there is still little awareness of the problem for the latter cases.

### Why is there a problem with the asymptote?

Since the data are normalized to (untreated) controls, and the controls are set to 100%, one should think that the upper asymptote should run approximately through 100%. To understand deviations, the conditions determining the asymptote need closer examination. It is important to realize that each data set that is used for such curve fitting must contain at least 2-3 data points from concentrations at which there is no effect. Without such data points, the acceptable conditions for curve fitting are not fulfilled.

In simple terms, the asymptote runs along the average of these (no effect) data points. For instance, there may be the control plus 2 no-effect concentration data points (one data point being considered the mean of its technical replicates). The exemplary 3 points (control plus two very low, no effect test concentrations) will have an average and a standard error. Assume that the standard error

(i.e., the noise level of no effect data points) is 10% (relative to the average of the data). If one assumes that the data are normally distributed, the likelihood of the negative control means being outside this noise band is large, i.e., there will be many cases in which the negative control mean value clearly differs from the asymptote modelled through the no effect data points. In practice, the number of data sets with controls largely displaced from the upper asymptote may be high also for other, non-statistical reasons: the controls are often placed at the edge of assay plates (the plate edge often shows different behavior from the center), and they may be pipetted/diluted differently from the other samples. A qualitative review of the literature indeed suggests a disproportionately high number of cases in which the negative control is clearly different from the no-effect drug concentrations<sup>1</sup>.

### Solutions to the problem

If it is clear that something is wrong with the control value, then the solution is relatively straightforward. One can assume the lowest test concentration (in the no effect range) to behave like a negative control and re-normalize all data to this value. A more robust approach would be to take the lowest 2-3 data points (assuming that they are in the no-effect range) and to renormalize to their average. In such cases, the original controls are typically eliminated from the display. A more generalized extension of the re-normalization approach is the following sequence of steps:

- (1) Decide (by visual inspection), whether or not controls are to be removed from the data set<sup>2</sup>.
- (2) Fit a curve to the data, with the upper asymptote setting to “automatic best fit” (i.e., not forced to 100%).
- (3) Use the value of the upper asymptote (e.g., 84 in Fig. 2A) to re-normalize all data points.
- (4) Now fit a curve through the new data set, with asymptote forced through 100. An exemplary result is shown in Figure 2C.

### Reduction of data uncertainty by re-normalization

The data uncertainty can be quantified by giving the lower 95% confidence interval of, e.g., the BMC10 (BMCL10) or the BMC50 (BMCL50). In the example data set, we assume the correct BMC10 to be  $10^{-6.5}$  and the BMC50 to be  $10^{-5.7}$ . If a curve is forced to 100% without data normalization, the BMC10 is off by a factor of 6.3, while the BMC50 is only off by a factor of 2 (showing that the problem is more pronounced for low BMRs). The uncertainty can be quantified by calculating the ratio of BMC and BMCL. This value is 100 for a BMR of 10 and non-normalized data, but it is dramatically reduced to 1.6 for normalized data! For a BMR of 50, the issue is much less pronounced, and the values are less than 2-fold in both cases (Fig. 2).

<sup>1</sup> For future work, it would be interesting to simulate this situation and its impact, and also to mine the literature for a quantitative overview

<sup>2</sup> A purely mathematical approach to this issue is difficult. However, simulation studies may provide a basis for a decision algorithm that provides an unbiased basis for semi-automatic data handling or user-guide immersive analytics.

### Alternative explanations

One easy way to avoid the data normalization problem is to ignore it or to deny statistical variation as its cause. The most common approach to this is the assumption that the difference between the negative control and the asymptote of the low concentration data is due to a real biological effect. This practice is encountered relatively frequently although it requires researchers to postulate low concentration effects of low plausibility. Moreover, a discontinuous (step-wise) and sometimes even non-monotonic concentration-response behavior has to be assumed. There are examples for which this is indeed the case. Nevertheless, these postulates lack scientific rigor for two reasons: (1) the postulation of a complex and little plausible system response without good reasons and evidence violates the principle of Occam's razor, according to which the answer that makes the fewest assumptions is most likely to be correct; (2) the assumption of concentration-response relationships without support by data sets in which the response is fully diluted out violates the good principles of curve fitting in pharmacology and toxicology.

### Caveats and trouble shooting

Good science means that we can use data to support hypotheses instead of relying on beliefs. There is an easy and straightforward method to determine, whether an "off" control is a technical artefact or an indication of a strange low-concentration behavior: The experiment needs to be repeated with inclusion of lower concentrations until a clear no-effect concentration is reached. The data obtained for these concentrations should be identical to negative controls; then the curve provided by the whole data set will indicate at which concentration a real effect starts. There is no exception to the rule that any effect, even the strangest low-concentration response, has to dilute out at some point and approach negative control values.

The same approach can also be used to increase the robustness of re-normalizations. The weakness and the danger of re-normalization is that the data assumed to be no-effect data are not robust or there are too few data points to yield reliable estimates. Inclusion of more low-concentration data points makes the asymptote more reliable, and therefore the whole re-normalization procedure becomes more exact.

### Outlook and next levels of complexity

Data re-normalization of one data set is a straightforward procedure, given that the underlying data set is suited for this. In practice, one usually does not deal with one single data set but rather with multiple data sets, corresponding to biological (independent) replicates of a given experiment. These may have been produced on different days and they therefore have their own controls. Thus, the question arises, whether data should be re-normalized independently and then averaged, or the other way around. The theoretically more appealing approach is to normalize each experiment first. In our experience, the more robust approach is to first average the normalization anchor (i.e., the no-effect data used for re-normalization, or the upper asymptotes of the different curves), then to normalize all data to this common anchor point, and then to average the data points of the different biological replicates. Simply put: "First average the anchor and then normalize." This approach provides a better buffer for errors and random variation in the anchor data.

Another feature that can increase complexity is non-monotonic curve behavior close to the highest non-cytotoxic concentrations. This is often manifest as an upward bump in the curve, possibly a last-resort stress response counter-regulation of cells. There are no universally accepted approaches to deal with this phenomenon, but it is highly recommended to control (by repeating the experiment, possibly using an alternative readout) whether the effect is biologically real.

### Acknowledgement

This work was supported by EU-ToxRisk, BMBF and DFG (KoRS-CB) grants.

### Correspondence to

Marcel Leist, PhD  
In vitro Toxicology and Biomedicine,  
Dept inaugurated by the Doerenkamp-Zbinden Foundation  
at the University of Konstanz, Konstanz, Germany  
University of Konstanz  
Universitätsstr. 10  
78464 Konstanz, Germany  
e-mail: marcel.leist@uni-konstanz.de