Look up or don't look up?

A deep dive into bioassay data analysis

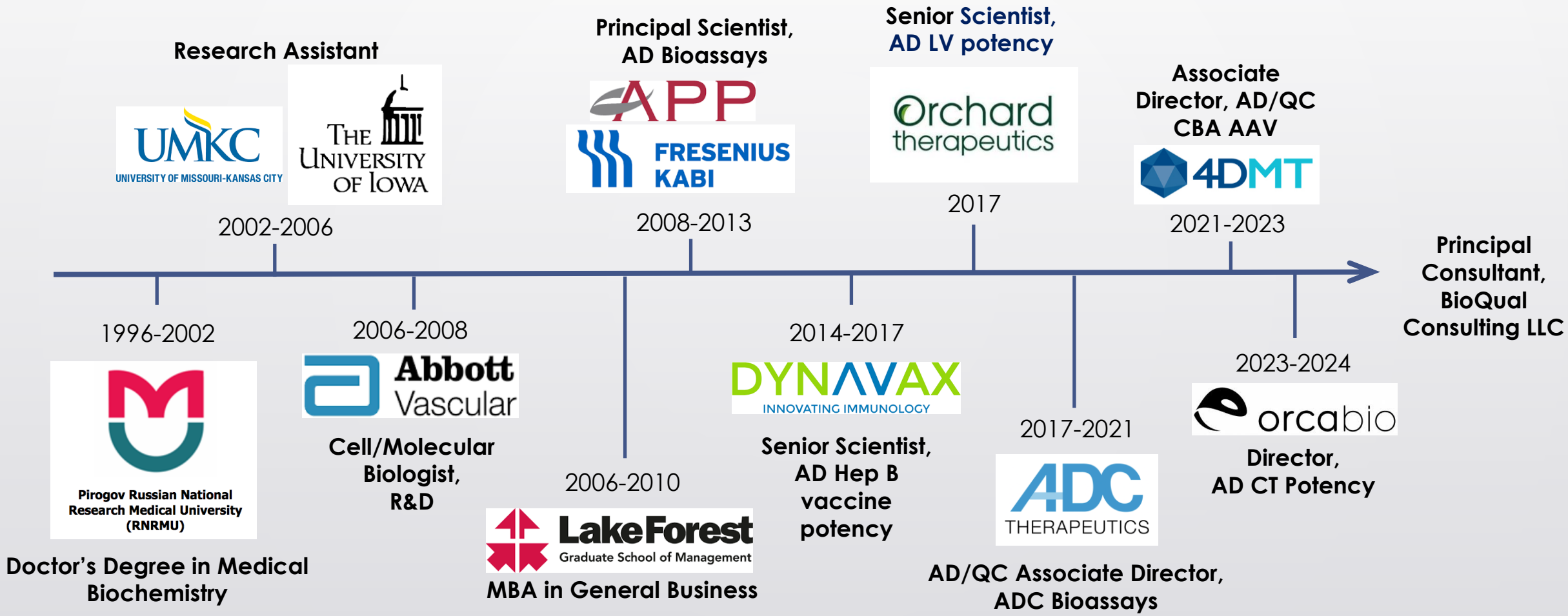
Anton Stetsenko,
Principal Consultant
BioQual Consulting LLC

anton@bioqualconsulting.com

Legal Disclaimer

The information contained in this presentation has been compiled from various sources and is believed to be reliable and to represent the best current opinion relative to this topic. BioQual Consulting LLC offers no warranty, guarantee or representation as to its absolute correctness or sufficiency. BioQual Consulting LLC assumes no responsibility in connection therewith; nor should it be assumed that all acceptable safety and regulatory measures are contained herein, or that other or additional information may be required under particular or exceptional conditions or circumstances.

Who am I?



Fundamental quality question

Will you give your child the medicine released by this potency assay?

Quantitative dose-response relation

David John Finney (January 3, 1917 – November 12, 2018)

- Relative Potency from "Statistical Method in Biological Assay" by David J. Finney (Third Edition-1978), page 41:

$$F_T(\mathbf{z}) = F_S(\rho\mathbf{z})$$

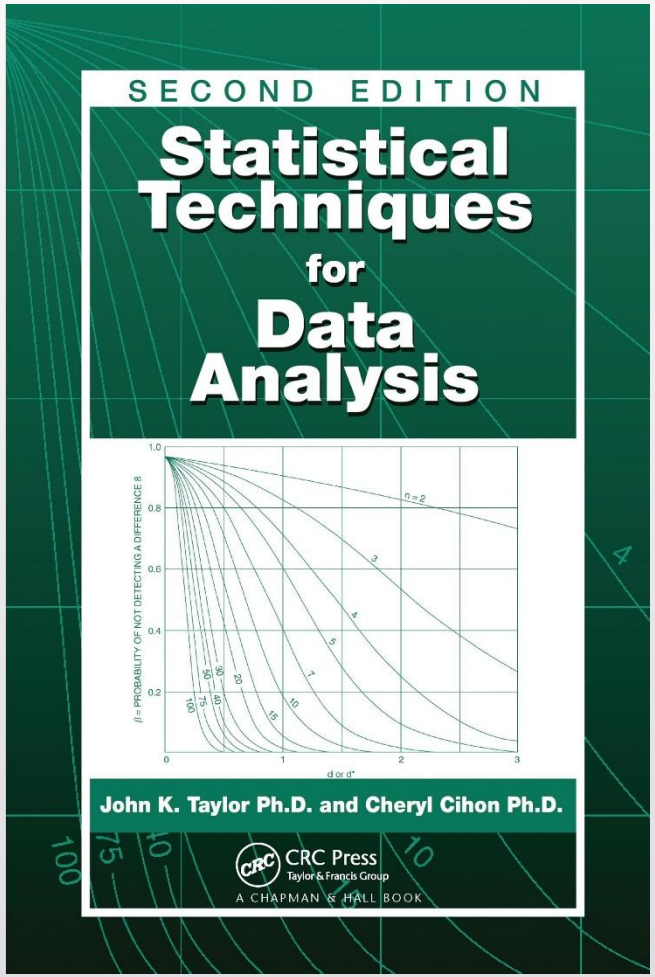
, where
z is a dose of a particular stimulus,
F_T is a dose response regression function for the Test(T),
F_S is a dose response regression function for the Standard(S),
ρ is a constant, the potency of the Test(T) relative to the Standard(S)

- The condition of "*similarity*" applies as a sample validity criterion for all types of relative potency bioassay:
  - F_T and F_S functions have the same properties (i.e., share common functional parameters) except for the relative potency parameter (i.e., EC_{50})
  - "*If the responses cannot be adequately be represented by the same form of regression function for both preparations, either the conditions of testing have differed for the two preparations or the basic assumption of similarity is false*."
- We vary the dose of a particular stimulus to generate a dose-dependent activity response, as we are unable to directly vary the stimulus's actual activity. In other words, we cannot perform a controlled degradation of the product for the x-axis while maintaining a constant concentration.

John K. Taylor, Statistical Techniques for Data Analysis

- "There are three sources of error in measurements and they can be classed as systematic, random, and just plain blunders.… The third kind of uncertainty [happens] when mistakes are made, knowingly or unknowingly. Errors resulting from [blunders] are not statistically manageable and, in fact, they can invalidate an otherwise good set of data. A measurement system that is unstable and fraught with blunders is not in statistical control and cannot be relied upon to produce useful data."

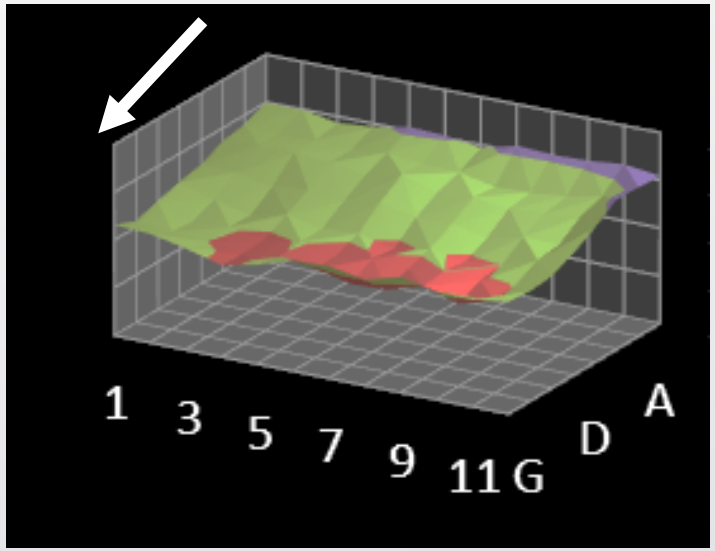
- Statistics deals only with Taylor's second source of error, random. It is unreliable to rely solely on statistics to discover or correct **systematic errors** and blunders. Profound knowledge and experience of the system under observation are usually required to recognize the surprise caused by a blunder.



Taylor, J.K., & Cihon, C. (2004). Statistical Techniques for Data Analysis

Mitigation of bias: blocking & randomization

- Bias (systematic error) in bioassay measurement can arise due to operational variables, including location-dependent effects within a plate.
- During development, test of uniformity should be conducted to assess potential biases within the assay or plate, for example:
  - cage effects in *in vivo* bioassays (filtration effect or temperature differences for cages at the ceiling vs cages located close to the floor);
  - plate effects in *in vitro* bioassays (left-to-right, top-to-bottom, "edge effect");
  - time-based variations across a series of sample tests (from beginning to end of measurement);
  - equipment-dependent effect (luminometer with 2 detectors that measure rows 1 – 6 and 7 – 12 respectively)
  - sample specific effect due to viscosity caused by different concentrations in RS and Test sample (same matrix, but different concentrations)



Top-to-bottom plate effect

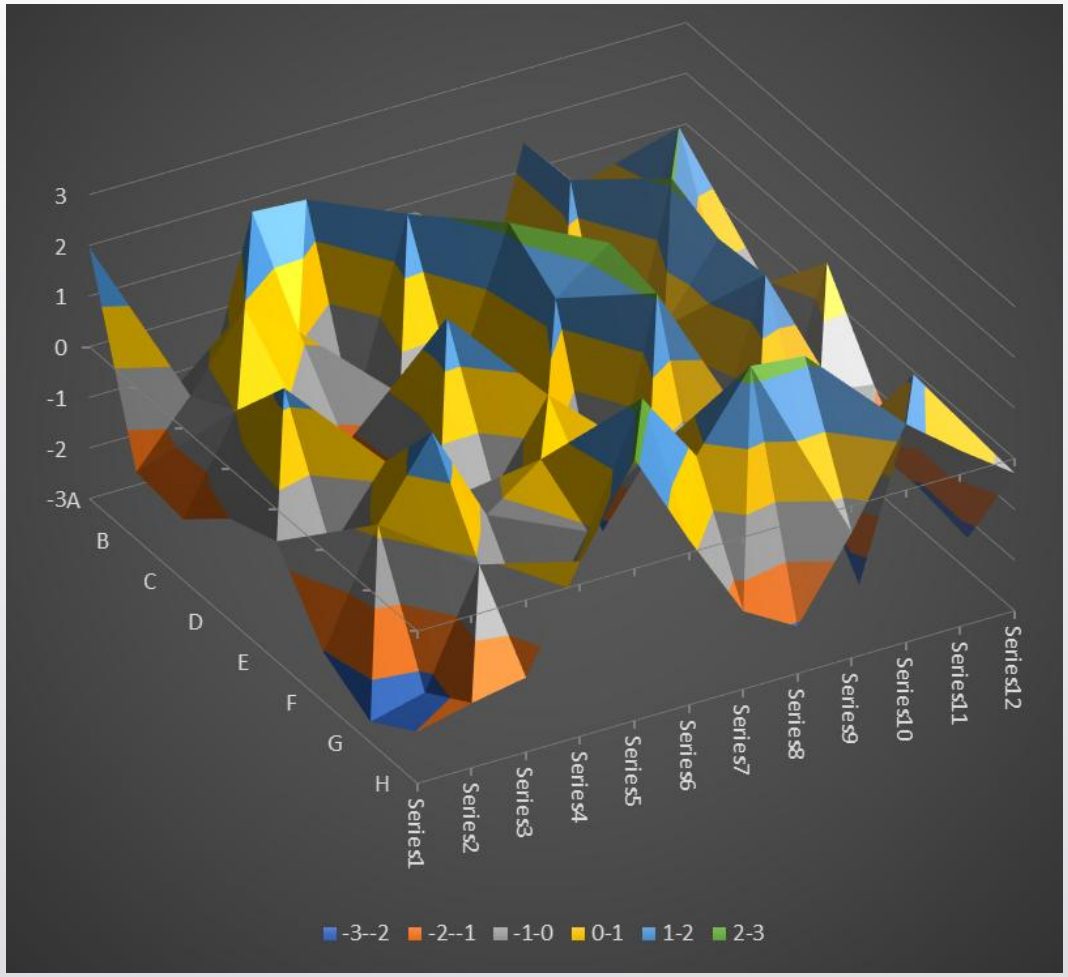
Z score: uniform ideal plate

$$Z = \frac{x - \mu}{\sigma}$$

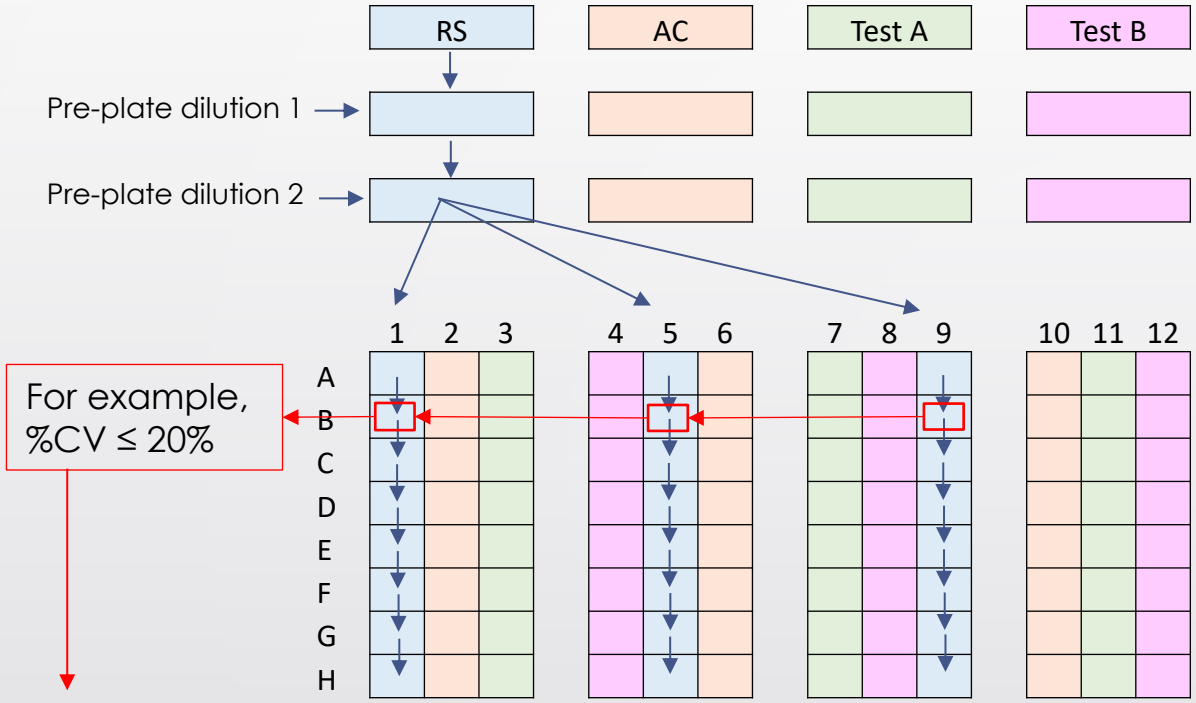
Z = standard score
X = observed value
μ = mean of the sample (n=96)
σ = standard deviation of the sample (n=96)

Simulated data set for
Z min -3
Z max 3

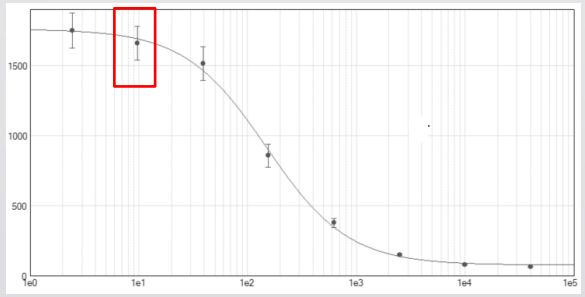
Row	1	2	3	4	5	6	7	8	9	10	11	12
A	-1.8416	-2.1217	1.7365	1.6675	-2.0962	0.7976	1.9586	-2.5553	1.5547	0.4105	-0.1068	-0.9205
B	-0.4724	0.9465	0.1624	-1.795	1.8777	1.2856	-1.6718	-2.4524	1.6213	-0.7225	2.0406	-0.7794
C	0.0415	-0.0394	-1.2589	-0.5653	-0.7333	2.1969	-1.803	-0.865	-2.2198	2.1481	-0.501	-1.0397
D	1.2781	-0.0168	-1.6176	1.7377	-1.5565	1.5014	-0.988	2.2901	0.3321	1.7564	-2.653	0.6621
E	-1.0794	0.3253	-0.7986	-0.8668	0.7061	-2.2975	-0.6343	2.0972	-1.3342	1.8379	-0.7836	-0.9373
F	0.1905	1.7062	-0.235	0.5495	-0.6576	0.0018	-2.0254	-1.7498	-0.3403	-1.176	-0.9663	-2.209
G	-2.3196	-0.1057	-1.8632	-0.0732	2.2472	0.8017	-2.5878	2.267	2.1519	-2.6687	1.1686	-2.3699
H	-1.7251	-1.5313	0.4529	2.1492	0.2578	-1.461	-1.9808	-2.0726	-0.5252	1.3017	0.4675	-0.2854



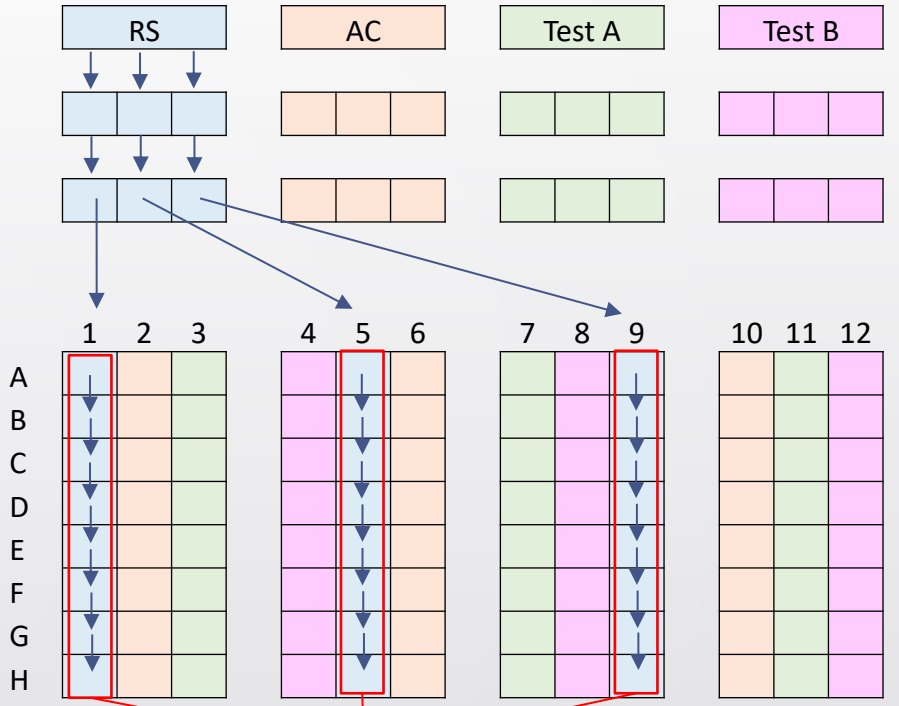
Dilutional statistical unit: what is it?

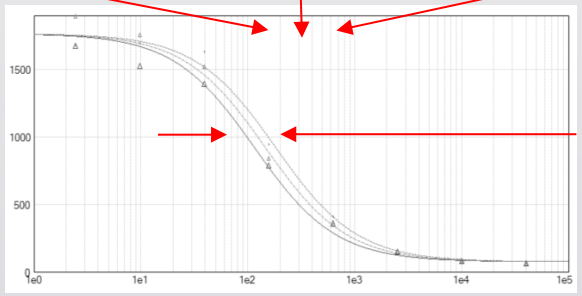


For example, %CV ≤ 20%



If readouts values from row B are 786, 983, 1156, then the mean=975, SD=185, %CV=19%, but 95%CI for RS mean readout in row E is 515-1435 (almost 3-fold difference!)

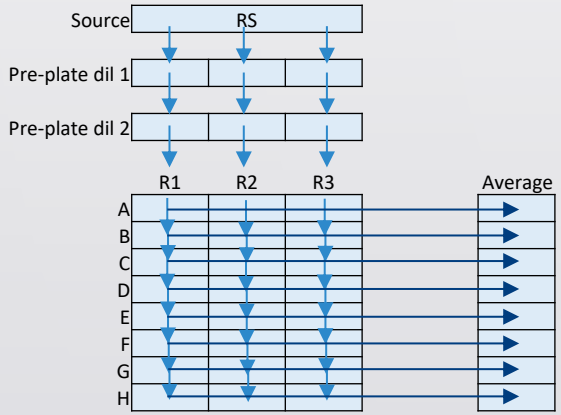


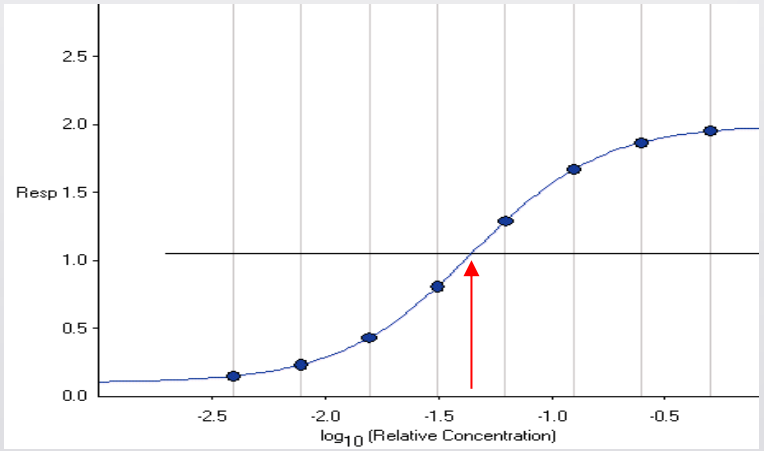


The closeness of RP values for each dilutional replicate against desirable CI

From 4PL model to …

- In relative format bioassay data analysis for potency estimation, common practice includes:
  - Averaging responses from all serially diluted dose replicates
  - Fitting separate four-parameter logistic (4PL) models to the reference standard and test sample
  - Assessing similarity by comparing asymptotes, slopes, and EC_{50} values
- However, noise in parameter estimation, especially when relying on extrapolation, can lead to false positives in non-parallelism assessments.



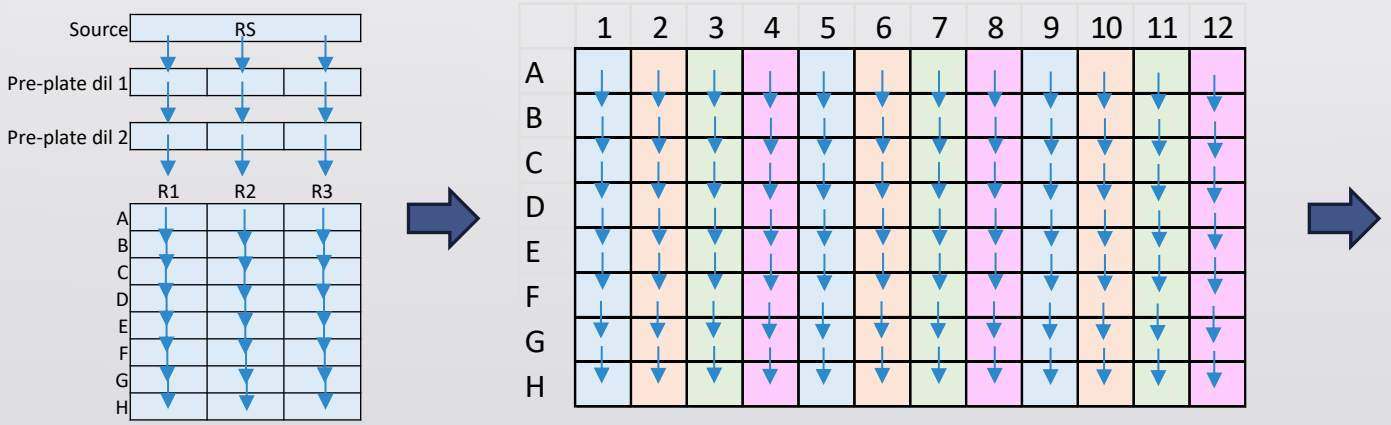


$$\text{Response} = \text{A} + \left(\frac{\text{D} - \text{A}}{1 + \left(\frac{10^{\text{C}}}{10^{\text{x}}} \right)^{\text{B}}} \right)$$

, where
A is the lower asymptote,
D is the upper asymptote,
B is the Hill factor (aka slope),
C = log10(fold over concentration),
x = log10(relative concentration)

… n+3PL model

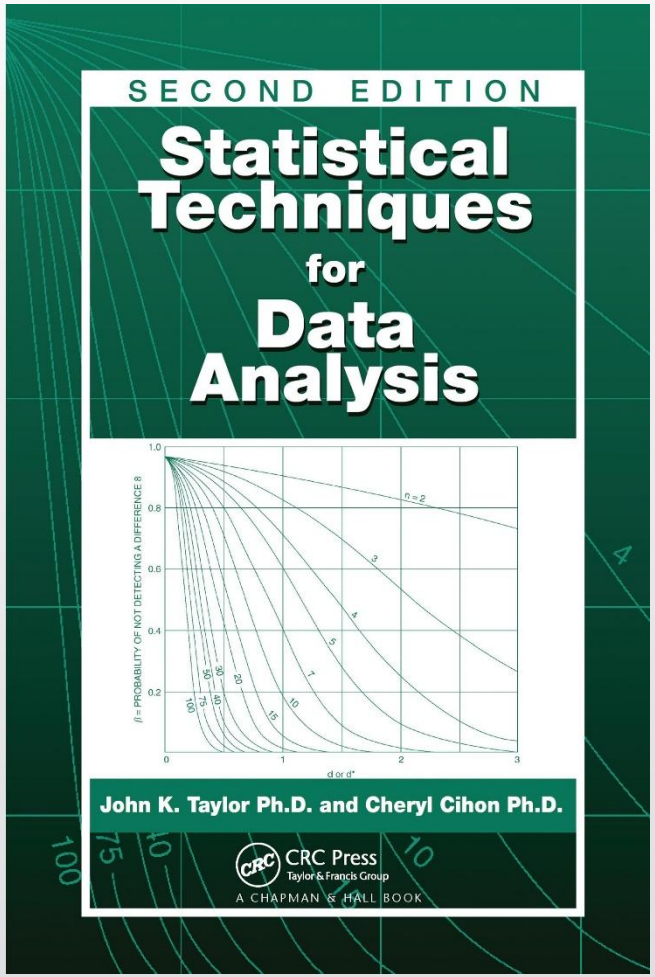
- Building on DeLean et al. (1978), who advocated for a more efficient common-parameters approach, a more robust method involves:
  - Treating each serially diluted replicate of the reference standard and test sample as a distinct dilutional unit
  - Sharing asymptotes and slopes across replicates for improved consistency
- Using an (n+3)PL model as an extension of the 4PL (where n represents the total number of all dilutional units) offers key advantages:
  - Enhanced accuracy and precision in relative potency estimation
  - Fewer false positives when assessing non-parallelism
  - Improved detection of errors (blunders) in pre-plate dilutions, first-well transfers, and serial dilutions



John K. Taylor, Statistical Techniques for Data Analysis

- "There are three sources of error in measurements and they can be classed as systematic, random, and just plain blunders.… The third kind of uncertainty [happens] when mistakes are made, knowingly or unknowingly. Errors resulting from [blunders] are not statistically manageable and, in fact, they can invalidate an otherwise good set of data. A measurement system that is unstable and fraught with blunders is not in statistical control and cannot be relied upon to produce useful data."

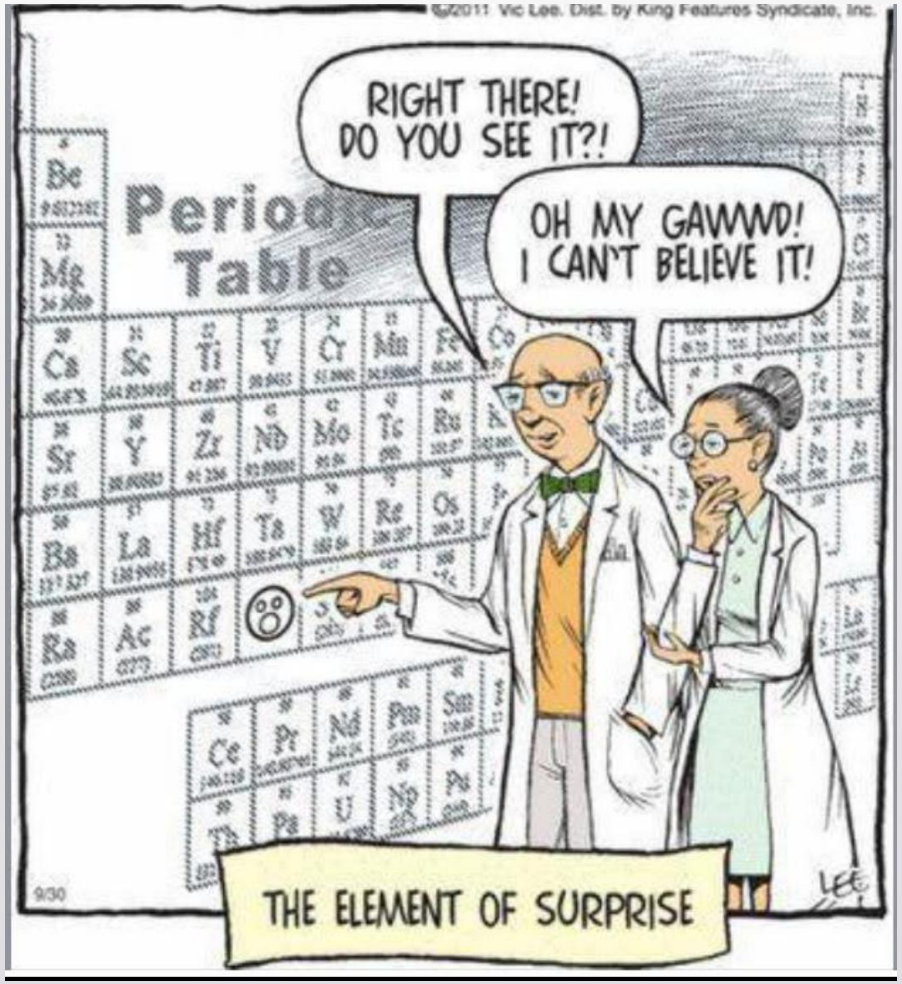
- Statistics deals only with Taylor's second source of error, random. It is unreliable to rely solely on statistics to discover or correct systematic errors and **blunders**. Profound knowledge and experience of the system under observation are usually required to recognize the surprise caused by a blunder.



Taylor, J.K., & Cihon, C. (2004). Statistical Techniques for Data Analysis

Underlying philosophy

- Controlling the number of outliers, whether due to blunders or statistically rare events, aligns with Taylor's observation: "*A measurement system that is unstable and fraught with blunders is* ***not in statistical control and cannot be relied upon to produce useful data***."
- If a dose-response curve contains more than two blunders, it likely indicates a lack of statistical control. Possible causes include operational conditions (such as plate well contamination) or insufficient analyst training, among other factors.
- This principle forms the foundation of the outlier detection approach outlined here.

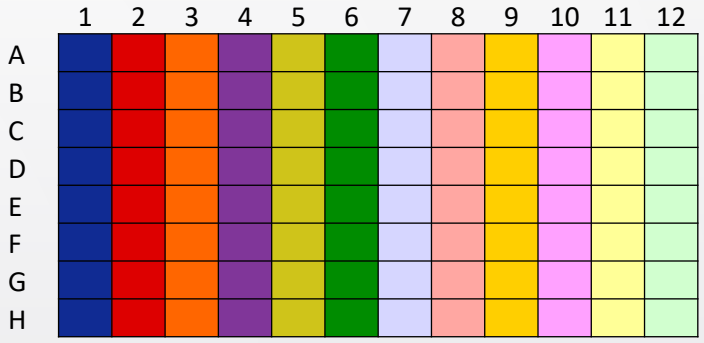


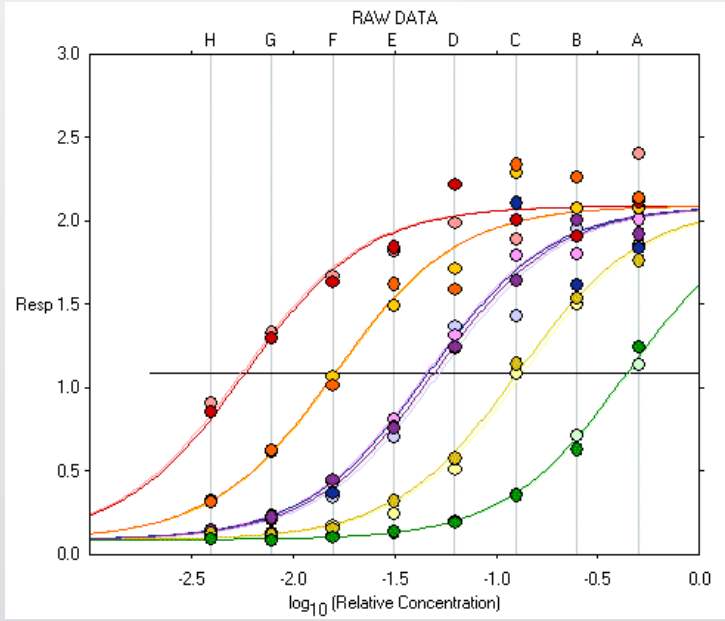
What am I talking about?

- If 96-well plate uses all 8 rows for n=8 serially diluted points and all 12 columns, where each column represents an independent sample's replicate of material, then
- All samples will share common upper and lower asymptotes as well as a common Hill factor (aka slope) in the analysis.
- Therefore, based on the methodology outlined by De Lean et al., the 4PL model can be expanded into a 15-parameter logistic (15PL) model by introducing 11 additional C parameters, one for each of the 12 columns.
- In the resulted 15PL model, each column maintains three common parameters while incorporating its own unique C parameter, effectively creating 12 distinct 4PL models within the plate.
- For n of samples' replicate the model becomes n+3PL, where n – the number of unique C parameters, and 3 is the number of three common parameters.
- C parameter in the constrained model becomes a composite of n C values, one C value per sample's replicate:

$$C = \sum_{col=1}^{n} d_{col} C_{col}$$

where dummy variable d_{Col} assumes a value of 1 if the data point belongs to that column and 0 otherwise, which allows the estimation of all C parameters simultaneously

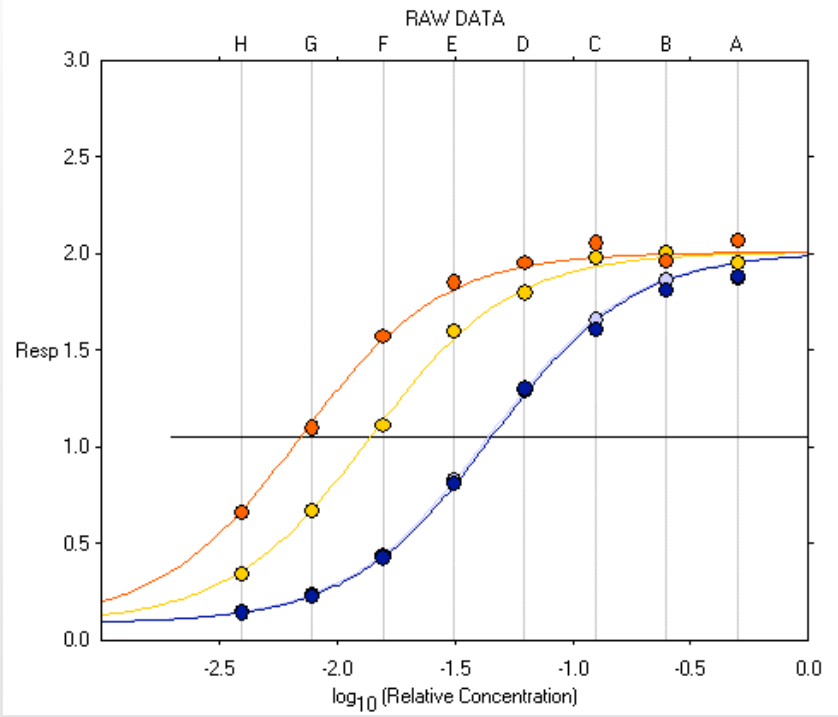




De Lean, P. J. Munson, and D. Rodbard, "Simultaneous analysis of families of sigmoidal curves: application to bioassay, radioligand assay, and physiological dose-response curves," American Journal of Physiology, 235, E97-E102 (1978).

1st reason: detection of a whole curve outlier (blunder)

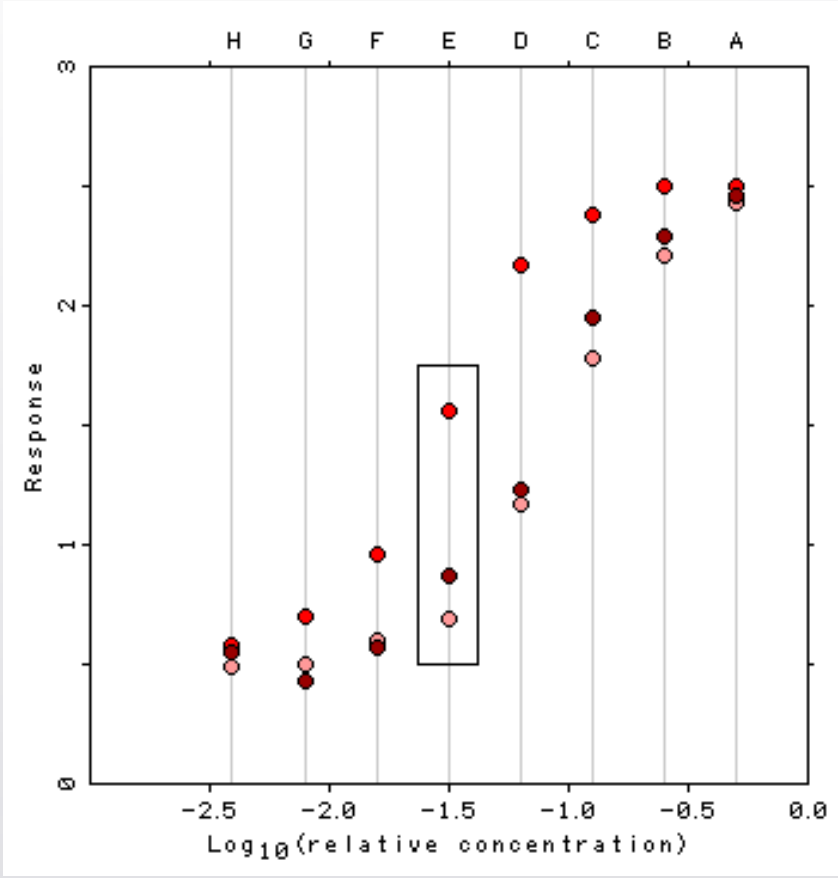
- Standard practice assumes that uncertainty in 4PL model fitting comes only from random errors (statistical noise).
- To reduce uncertainty, responses from replicated doses of the same sample are often averaged before analysis.
- However, determinate errors such as blunders in pre-plate and/or 1st plate well dilution and material transfers can shift the 4PL curve left or right.
- Analyzing columns separately helps identify these errors, while averaging conceals them.
- If replicate columns show too much disagreement, the sample should be excluded from further analysis.



Example: two reference standard replicates (dark and light blue) agree with each other, but one of the test sample replicates (orange) is shifted to the left because of such blunder.

2nd reason: preserve data and improve statistical power

- Statistical uncertainty decreases as the number of data points (n) increases, following a $1/\sqrt{n}$ relationship. A larger n leads to more reliable decisions.
- A 96-well plate provides n = 96 data points.
- Averaging replicates before analysis reduces n, losing valuable information.
- For example, combining triplicates measurements across four plate items decreases n from 96 to 32 and increases uncertainty ($1/\sqrt{32} = 0.177$, worse than $1/\sqrt{96} = 0.102$).
- Analyzing replicates separately preserves data and improves statistical reliability.

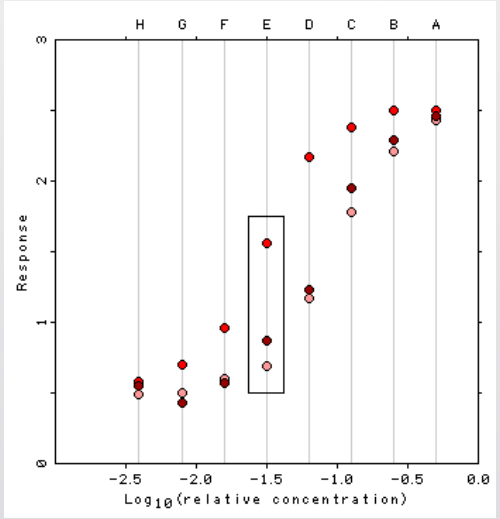


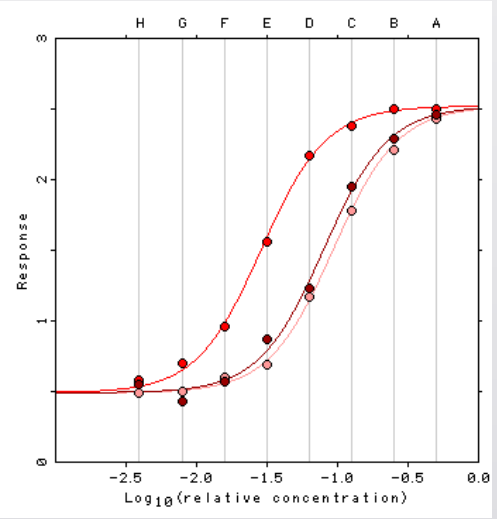
3rd reason: detection of a whole curve outlier

- When averaging replicates, a common practice for outlier detection is used to remove dose-specific data points with high %CV (e.g., NLT 15%).
- In the given example, plate rows C-G show high %CV, but this isn't due to excessive random error.
- A blunder (determinate error) has shifted one of the 4PL curves, inflating these %CV values.
- Instead of simply rejecting the entire dose as outliers, it may be better to remove the only problematic replicate OR reject the entire sample replicate for more accurate analysis OR the entire sample preparation.

Source	Test sample		
Pre-plate dil 1	T1_1	T1_2	T1_3
Pre-plate dil 2	T2_1	T2_2	T2_3

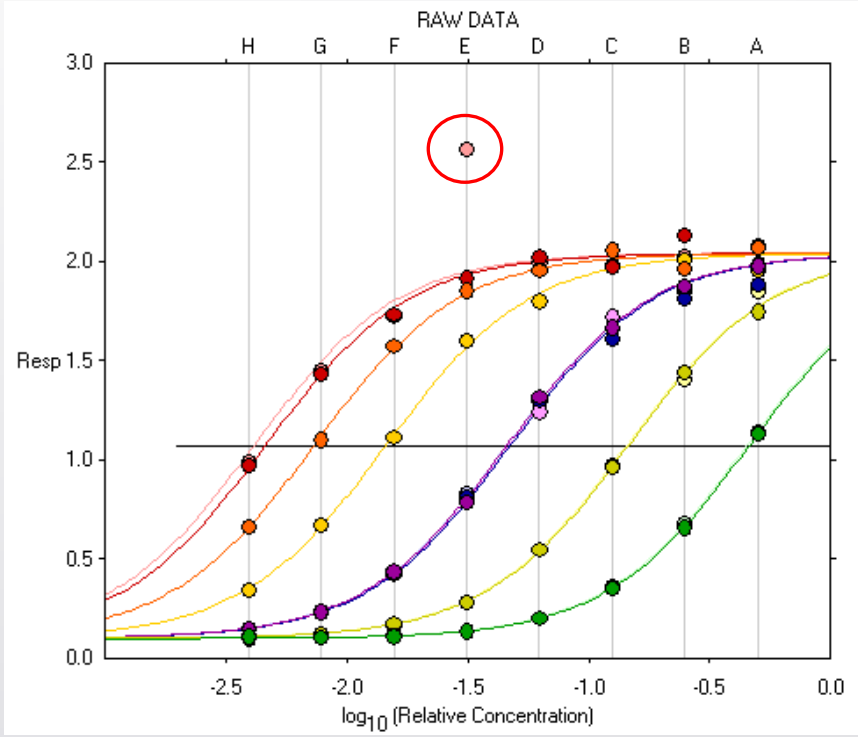
	T3_1	T3_2	T3_3	Average	CV
A	2.4	2.5	2.5	2.5	2.3
B	2.2	2.5	2.3	2.3	6.5
C	1.8	2.4	1.9	2.0	15.8
D	1.2	2.2	1.2	1.5	37.7
E	0.7	1.6	0.9	1.1	44.3
F	0.6	1.0	0.6	0.7	31.5
G	0.5	0.7	0.4	0.5	28.6
H	0.5	0.6	0.5	0.5	10.8





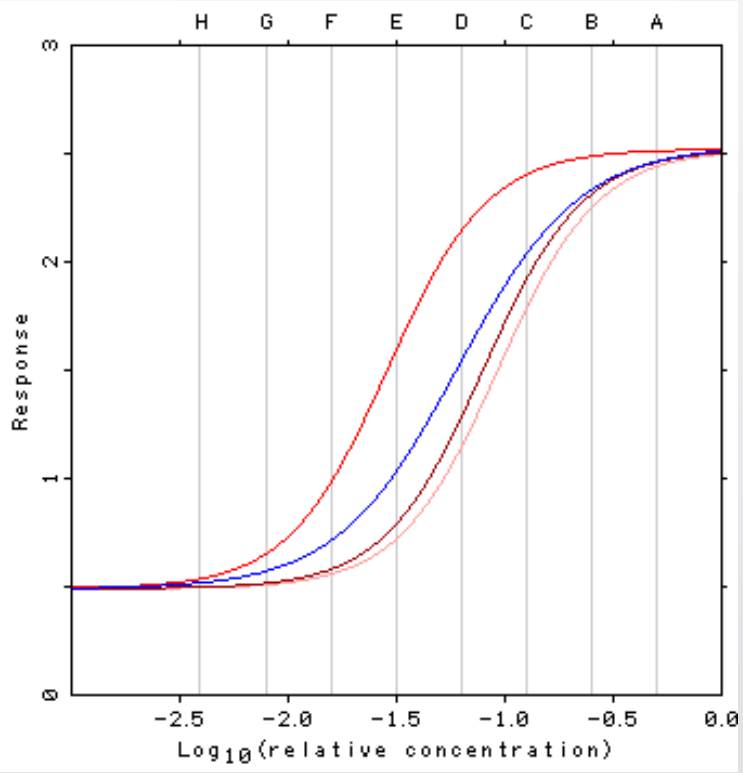
4th reason: detection of model-based outliers

- Practical and statistical model-based outlier detection examines how each data point fits the full model, rather than analyzing subgroups first.
- With the n+3PL model, it focuses on individual data points, avoiding pre-processed averages.
- It is more sensitive because outliers aren't masked by averaging with unaffected points.
- If an outlier is found, only the problematic data point is removed and guiltless points stay in the analysis.



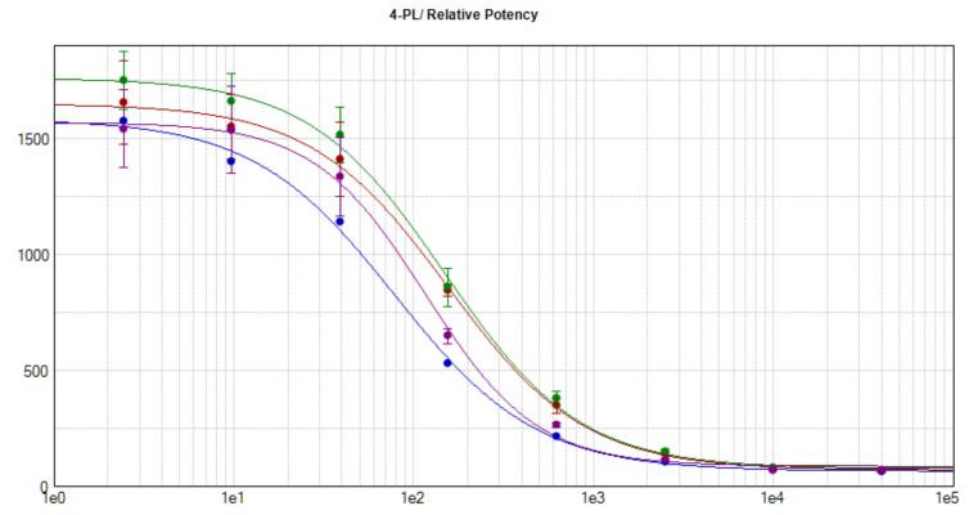
5th reason: distortion of Hill factor (aka slope)

- Averaging across rows can distort the Hill slope when a blunder is present.
- In the example, three original data sets (red) share the same Hill slope (B parameter in 4PL).
- After averaging, the new data set (blue) always has a shallower slope, which can be mistaken for non-parallelism.
- A better approach may be to remove the faulty replicate and analyze the remaining two, or reject the entire sample if necessary.

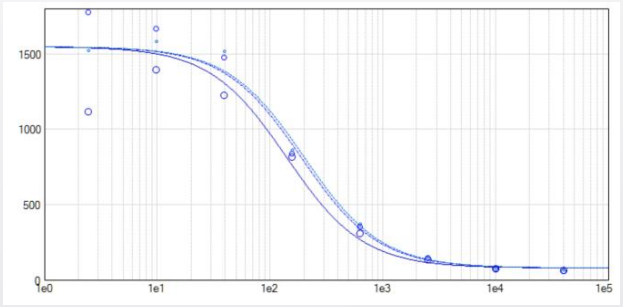


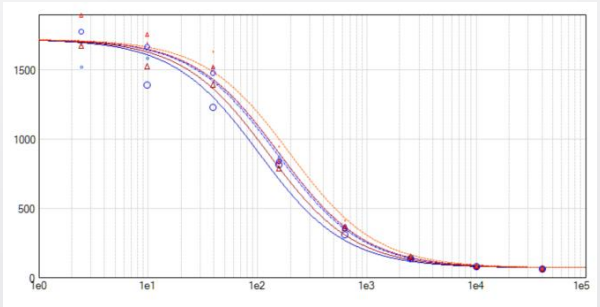
Example of n+3PL model in SoftMax Pro

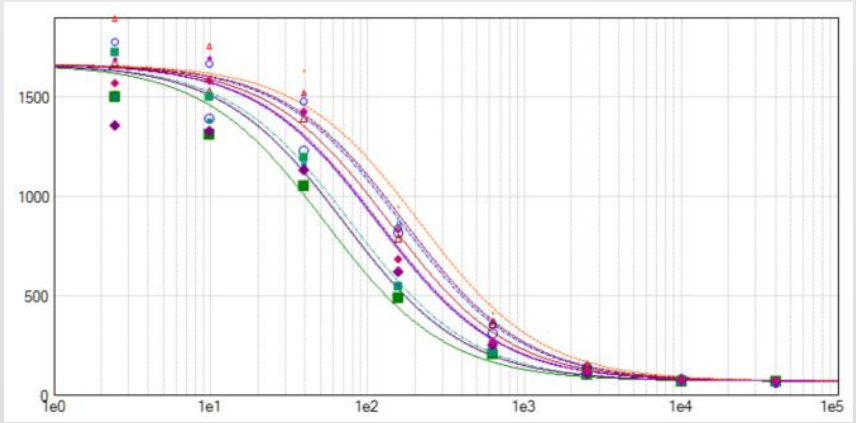
Instead of averaging responses for each dilution's response based on %CV criteria for dilution replicate's wells and calculation of RP using these mean values…



… we can remove outliers in each dilutional unit, obtain potency value specific to each dilutional unit and then calculate an average as a reportable value





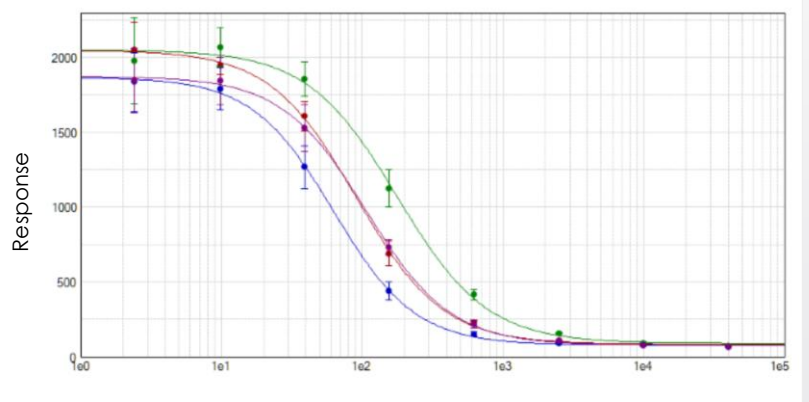


The "closeness" criteria between dilutional unit's replicates (such as simple max-min difference between C values or %RSD between plate column's C values) is needed to ensure that individual results are in agreement before averaging for reportable value

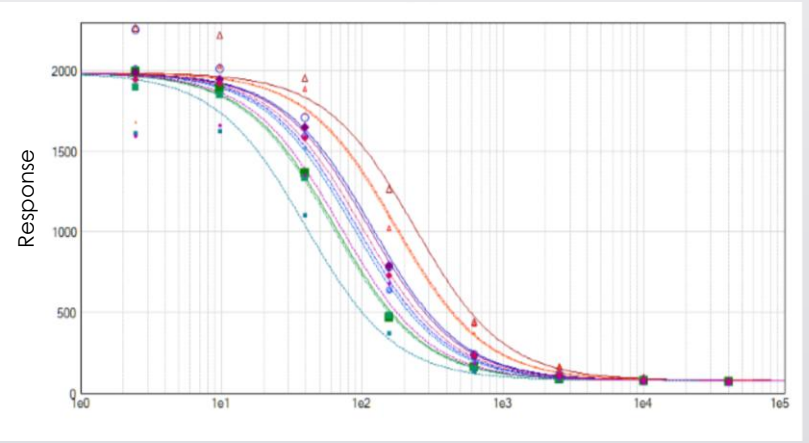
I can see blunders!

- When fitting 4PL models to bioassay data, it is commonly assumed that response uncertainty stems solely from random error (statistical noise).
- To reduce this variability, responses from replicate columns/rows are often averaged across columns/rows before data processing begins.
- However, systematic and random errors can occur due to mistakes in pre-plate dilutions or transfers to the first well or serial dilution's preparation, resulting in incorrect material concentrations.
- These errors lead to shifts in the fitted 4PL curve along the x-axis, causing inaccurate potency assessments.
- Analyzing individual dilutional units separately helps identify these errors, preserving critical information.
- Averaging responses from replicates of dilutions obscures evidence of these errors.
- If replicate dilutional units show excessive disagreement, the affected dilutional unit replicate or whole sample should be removed from further analysis to maintain data integrity.

Common practice with pre-averaging



Recommended practice for individual dilutions



Benefits of n+3PL model ("… and now I can't unsee it")

- By fitting statistical dilutional units individually, we can observe the effects of blunders in pre-plate dilutions and transfers to the first well (differences in C values of individual replicate columns or rows).
- By fitting statistical dilutional units individually, we avoid loss of information from pre-averaging (df = 96 for 96-well plate).
- By fitting statistical dilutional units individually, we avoid loss of data during pre-averaging when %CV of replicates is used as a criterion for discarding suspected outliers.
- By fitting statistical dilutional units individually, we can use "model-based outlier detection" to find absolute and statistical outliers.
- By fitting statistical dilutional units individually, we can avoid distortion of curves if we pre-averaging in the presence of outliers (i.e., the Hill slope isn't distorted).
- By fitting statistical dilutional units individually and averaging the C values for replicates, more accurate relative potencies are obtained.

Q&A

Thank you!