Appendix A
Statistical Power

Statistical power is the probability that an experimental effect will be observed if the effect actually occurs. More technically, it is the probability that a statistical test will reject the null hypothesis (no difference in two sample means being compared) for a specific alternative hypothesis. Statistical power and four other related terms are defined as follows. The first two parameters are properties of the system that we are testing. Significance level is a standard feature of a statistical test which is fixed, usually at 0.05.

> M**agnitude of the effect** = often, $(\mu_1-\mu_0)$, i.e., difference in means across two groups, measured in units of the response variable. Effect size and the ability to detect it are indirectly related; the smaller the effect, the more difficult it will be to find.
>
> **Inherent variability in the response variable**. The standard deviation (σ) is often used as a measure of the variability in the response variable of interest.
>
> **Significance level** = (α) is the probability of a Type I error. An arbitrary but common choice is 0.05.
>
> **Power** to detect an effect = $(1-\beta)$ where β is the probability of a Type II error. .
>
> **Sample size** = the number of experimental treatment units (individuals, liters, breeding pairs, tanks, etc). A larger sample size generally leads to parameter estimates with smaller variances, giving greater ability to detect a significant difference.

These five terms are not independent: any four of them determines the fifth. The usual objective of a power analysis is to calculate the sample size for given values of the other four terms. Studies can be planned based on estimates of σ based on pilot or literature data. In validating assays, EPA will pick the minimum effect (as represented by one or more benchmark chemicals) that an assay should be able to detect in order to fulfill its purpose. In studies with limited resources, such as the EDSP studies, the maximum practical sample size is often chosen. In these cases one could calculate the power corresponding to a given study size to determine if the study will have adequate power to meet its objectives. Often times, a power greater than 0.80 is deemed acceptable. There are a number of different software packages available on-line for calculating the power of a study.

The results of our test is shown to the right: **REALITY as indicated by Reference Values is shown below:**	**There is a statistical difference between controls and treated values** • We reject the null hypothesis • We accept the alternative hypothesis that there is an effect • We conclude that the **test result is POSITIVE**	**There is NO statistical difference between controls and treated values** • We accept the null hypothesis • We reject the alternative hypothesis that there is an effect • We conclude that the **test is NEGATIVE**
H_0 (null hypothesis) false • There **is** an effect • The test result should be **positive**	**POWER** **1-β** **The probability of finding an effect when, in fact, there is one. [Sensitivity]** A power of 0.80 or greater is often deemed acceptable.	**TYPE II ERROR** **β** **The probability of concluding there is no effect when, in fact, there is one.** With a power of 0.80, 20% of the time we will conclude there is an effect when, in fact, there is one.
H_0 (null hypothesis) is true • There is **no** effect • The test result should be **negative**	**TYPE I ERROR** **(Significance Level)** **α** **The probability of concluding there is an effect when, in fact, there is none.** A significance level of 0.05 is typically chosen.	**CONFIDENCE LEVEL** **1-α** **The probability of correctly concluding there is no effect. [Specificity]**