On the Number of Trials Required to Estimate *Specificity* to a Given *Confidence* Level

The purpose of this white paper is to describe how to compute the number of "null" case trials needed in order to estimate *Specificity* to a given *Confidence* level. For example, suppose 200 null measurements are made with no false alarms. What is the relationship between estimated *Specificity* and *Confidence*? Or, suppose it is required to demonstrate that a particular detection algorithm achieves a *Specificity* of at least 99%, with an associated *Confidence* of 95%. How many trials must be run to claim that such a requirement has been met? 100? 300? 1000?

Specificity and the Null Trial

Consider a Rapid Diagnostic Test, such as a Lateral Flow Assay, which for our current purpose is used solely as a positive/negative diagnostic. While there are only two possible outcomes, there are actually 4 possible conditions which produce them: (1) true positive; (2) false positive; (3) true negative; and (4) false negative. Given these conditions, *Sensitivity* is the measure of the true positive rate, while *Specificity* is the measure of the true negative rate. Here we concern ourselves with *Specificity*, though the identical set of calculations shown below can be applied to *Sensitivity*.

In order to determine *Specificity* for a given algorithm/assay combination, a series of trials must be performed in which the underlying condition is known to be negative i.e. a "Null Trial". For each trial, the detection algorithm assigns a positive or negative result (of course, any positive result is a false positive, while any negative result is a true negative). Given the outcome of these trials, what is the estimated *Specificity* and its associated *Confidence* level?

The Bernoulli Experiment – Probability Distribution

For convenience, instead of discussing *Specificity* directly, it is conceptually easier to frame the discussion in terms of the false positive rate, or 1 – *Specificity*. A set of trials as described constitutes a Bernoulli experiment and offhand, the underlying statistics are described by the quite familiar Binomial probability distribution:

$$B_n(n_{fp} \mid P_{fp}) = \frac{n!}{n_{fp}!(n-n_{fp})!} P_{fp}{}^{n_{fp}} (1-P_{fp})^{n-n_{fp}} \tag{1}$$

where n is the total number of trials, n_{fp} is the number of false positives and P_{fp} is the probability of a false positive (i.e. 1 – *Specificity*)

Unfortunately, in words, this expression tells us the probability of obtaining n_{fp} false positives out of n trials ***given*** some specified probability P_{fp} of obtaining a false positive. However, in our case, we have an experimentally determined number of false positives, but an *unknown* false positive probability. Essentially, we need the inverse of what the Binomial distribution is telling us. That is, we would like to determine the probability P_{fp} of obtaining a false positive ***given*** our measured n_{fp} false positives out of n trials. To determine this inverse relationship, we invoke Bayes' theorem (here assuming a uniform "prior") and find

$$iB_n(P_{fp} \mid n_{fp}) = \frac{(n+1)!}{n_{fp}!(n-n_{fp})!} P_{fp}{}^{n_{fp}} (1-P_{fp})^{n-n_{fp}} \tag{2}$$

The inverse Binomial expression looks almost identical to the "forward" Binomial expression with only a slight change in the Binomial coefficient (i.e. $(n+1)!$ instead of $n!$). However, it is important to note that the Binomial distribution is a *discrete* function of the unknown variable n_{fp}, while the inverse Binomial distribution is a *continuous* function of the unknown variable P_{fp}. This is made clear by their respective normalization conditions:

$$\sum_{n_{fp}=0}^{n} B_n(n_{fp} \mid P_{fp}) = 1 \quad \text{and} \quad \int_0^1 iB_n(P_{fp} \mid n_{fp}) dP_{fp} = 1 \tag{3}$$

Estimating *Confidence*

Working with the inverse Binomial distribution, the *Confidence* α in *any* estimate P_α of the false alarm probability is given by the following integral relationship:

$$Confidence \equiv \alpha = \int_0^{P_\alpha} iB(P_{fa} \mid n_{fa}) dP_{fa} \tag{4}$$

Substituting the expression from Eqn (2) for the inverse Binomial distribution into Eqn (4) yields a known integral function called the *Incomplete Beta Function* β_{P_α} (with arguments as shown):

$$\alpha = \beta_{P_\alpha}(n_{fp}+1, n-n_{fp}+1) \tag{5}$$

For certain conditions of interest, the *Incomplete Beta Function* reduces to quite simple expressions. For example, suppose in our Bernoulli experiment that there are no false positives i.e. $n_{fp} = 0$. In this case, Eqn (5) simplifies to

$$\alpha = \beta_{P_\alpha}(1, n+1) = 1-(1-P_\alpha)^{n+1} \tag{6}$$

This then shows a very simple relationship between *Confidence* α, the associated *Specificity* $1-P_\alpha$ and the number of trials n. Now suppose that there is one false positive in the experiment. In this case, Eqn (5) reduces to

$$\alpha = B_{P_\alpha}(2, n) = 1-(1-P_\alpha)^n (nP_\alpha + 1) \tag{7}$$

This is only slightly more complicated, but still quite tractable.

Specificity and *Confidence* – Graphical Representation

The following figures illustrate the utility of Eqn (5) and the relationship between *Confidence*, *Specificity*, the number of false positives and the total number of trials. In Figure 1, we assume there are no false positives in our Bernoulli experiment and show *Specificity* as a function of the number of trials for 3 different *Confidence* levels. Consider an experiment with 20 trials. In this case, we see from the plot that with a *Confidence* of 99%, the *Specificity* is at least 80%. As we decrease our desired *Confidence*, the limiting *Specificity* increases. This makes sense. That is, we can claim a higher *Specificity*, but only at the expense of a reduced *Confidence* in that value.

The plot can be utilized in another way. Suppose we require a *Specificity* of 95%. In this case, we see that if we also require a *Confidence* of 90%, this can be achieved with about 45 trials, provided there are zero false positives. On the other hand, if we require a *Confidence* of 99% for the same *Specificity*, then we need to double the number of trials to ~90.

In Figure 2, we again show *Specificity* versus the number of trials. In this case, the *Confidence* is fixed at 99%, and the 3 curves show the effects of varying the number of false positives within any given set of trials. Note that the blue curves in each plot represent the same conditions with 99% *Confidence* and zero false positives, but that the range of the vertical axes are slightly different. Again, in this plot, the behavior is easily understood. For a fixed number of trials, the *Specificity* decreases as the number of false positives increase. For a fixed *Specificity*, the number of trials required increases as the number of false positives increase i.e. an increased number of false positives needs a larger set of trials to preserve the false positive *rate*.

The final plot, shown in Figure 3, displays the relationship between *Specificity* and *Confidence* for a fixed number of trials. The “inverse” relationship between Specificity and Confidence is clear, with claims of increasing *Specificity* only being able to be made at the expense of decreasing *Confidence*.

Summary

This white paper provides the definitions and formulas required to assess the results of any study attempting to assess *Specificity*. It has been made clear that any estimated value of *Specificity* must be accompanied by an associated *Confidence* level, that these values are essentially "inversely" related, and that they in turn are dependent upon the total number of trials performed and the number of false positives obtained.

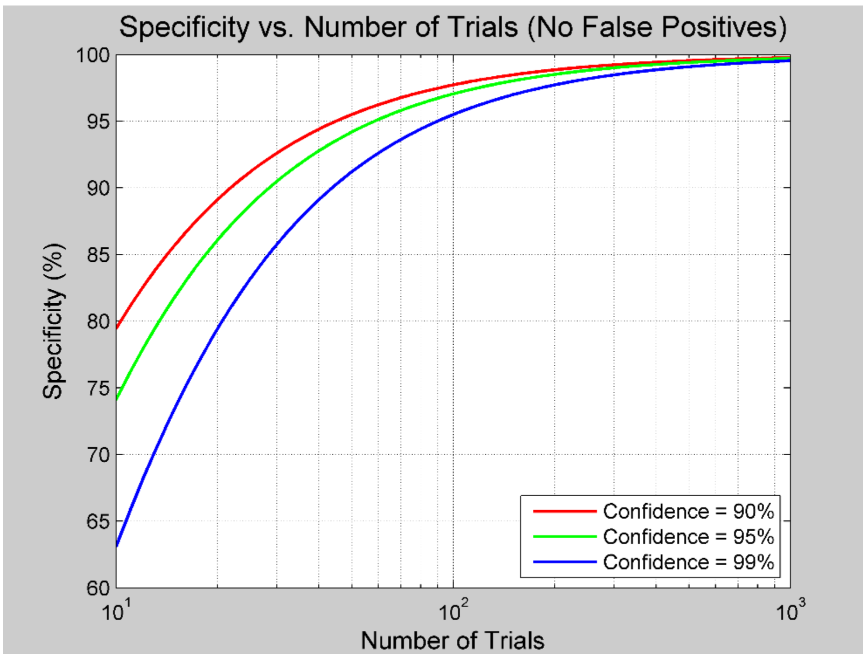

FIGURE 1: *Specificity* vs. number of trials with no false positives, and with curves shown for *Confidence* levels of 90%, 95% and 99%.

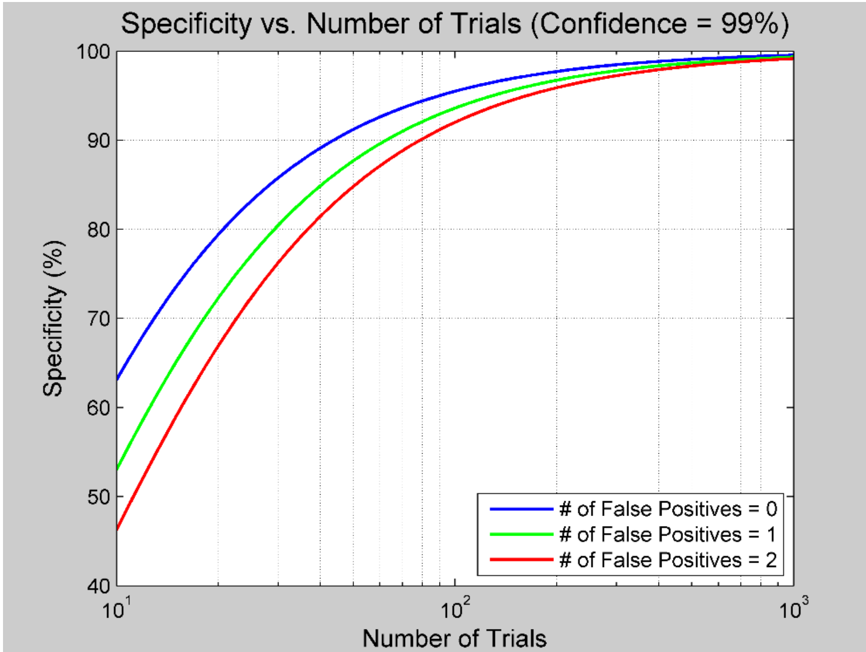

FIGURE 2: *Specificity* vs. number of trials with *Confidence* = 99%, and with curves shown for 0, 1 and 2 False Positives.

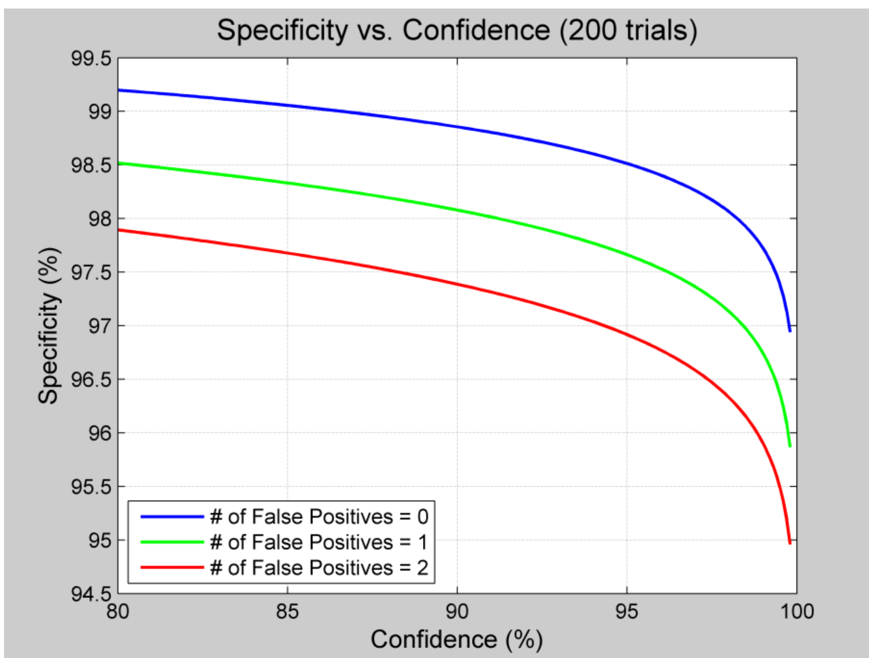

FIGURE 3: *Specificity* vs. *Confidence* with number of trials fixed at 200, and with curves shown for 0, 1 and 2 False Positives.