Some Statistics

In this kind of testing, it is very important to use enough test cases. Doing the test with only 20 or 50 cases may deliver quite random and statistically invalid results. We think that a minimum of 100 cases is a must. If an insufficient number of samples are used in comparative tests, differences in results may not indicate actual differences among the tested products. Some statistics below illustrates why.

Suppose that when 10,000 samples were tested and product A scored 80% while product B scored 70%. Then if the products were tested with a random selection of only 10 of those samples, these are the probabilities of the results:

Score (10 samples)	Product A (80% detection)	Product B (70% detection)
10 / 10	11%	3%
9 / 10	27%	12%
8 / 10	30%	23%
7 / 10	20%	27%
6 / 10	9%	20%
5 / 10	3%	10%
4 / 10	1%	4%
3 / 10	0.1%	0.9%
2 / 10	0.01%	0.14%
1 / 10	0.0004%	0.0138%
0 / 10	0.00001%	0.00059%
total	100%	100%

Even though the most likely score for Product A (with detection rate 80%) is 8 out of 10, there is only a 30% chance that Product A's score will be 8. Product A has a 37% chance of scoring higher than 8 out of 10 and a 33% chance of scoring lower than 8/10. When these probabilities are combined with those for Product B, the chances that Product A (80% detection) scores higher than Product B (70% detection) are about 60%. In a 10 sample test there is an 18% chance the products will have the same score and 22% chance Product B will outscore the superior Product A. This difficulty is overcome as the number of samples is increased. Figure 1 shows the effect on the 95% confidence interval of increasing the sample size from 10 samples to 100 samples to 1000 samples. The 95% confidence interval indicates the range over which there is 95% chance that the true detection rate falls within that range.