Statistician's Summary for Alternative QC Frequency Testing Proposal

The statisticians developed two related alternative proposals which exhibit similar statistical performance but take fewer days to complete. Their observations and conclusions are summarized below.

1. The current protocol calls for one QC test per day for 20 consecutive test days. If there are 0 or 1 failures, the protocol is completed and the laboratory may continue to weekly QC testing. Otherwise, the laboratory must perform one test per day for 10 additional days. At this point, if among the 30 total test results there are 4 or more failures, the system fails, otherwise it passes and the laboratory may continue to weekly QC testing.

2. The team identified the protocol as a two-stage sampling plan, of the sort used to test manufactured product for release. While a lot-release plan determines whether to release a batch of product for sale, the protocol above determines whether to "release" a system for ongoing use. However, the key similarity is that a go/no-go decision is made, and the mathematical considerations are the same.

3. The probability of failing the 20 to 30 day protocol increases with the true long-run (but never fully revealed) rate of QC errors. Key statistical properties of any plan are the error rate at which the probability of passing the protocol is high (95%) and the error rate at which the probability of passing is low (10%). For the current protocol, these rates are 4.8% and 21.7%, respectively. This determination is made from fundamental probability calculations assuming error rates rather than on observed data.

4. The team proposed a two-stage plan, with three tests per day for 5 days in the first stage. Depending on the number of QC errors, the protocol may be completed with the first stage, or it may proceed to a second stage. The second stage would also consist of three tests per day for 5 days. The three tests on each day must be conducted as if they were on separate days, i.e., three completely separate inoculum preparations rather than three inoculations of one prepared organism suspension.

5. The team proposed two similar criteria corresponding to this design, differing only in the criteria for passing the protocol in the first stage. In the **first proposal**, the protocol is successfully completed upon 0 or 1 errors in the first stage. If 4 or more errors occur in the first stage, the protocol is completed in the first stage with a failed outcome. Upon 2 or 3 errors in the first stage, continue to the second stage. If there are 3 or fewer errors among all 30 test results, the system passes; if 4 or more errors are observed, the system fails. The **second proposal** is like the first, except that the criterion for passing the protocol is 0 errors in the first stage. 1, 2, or 3 errors in the first stage would lead to second-stage testing. As with the first proposal, 4 or more errors lead to failure, either in the first stage or the second stage (at which all 30 results are considered).

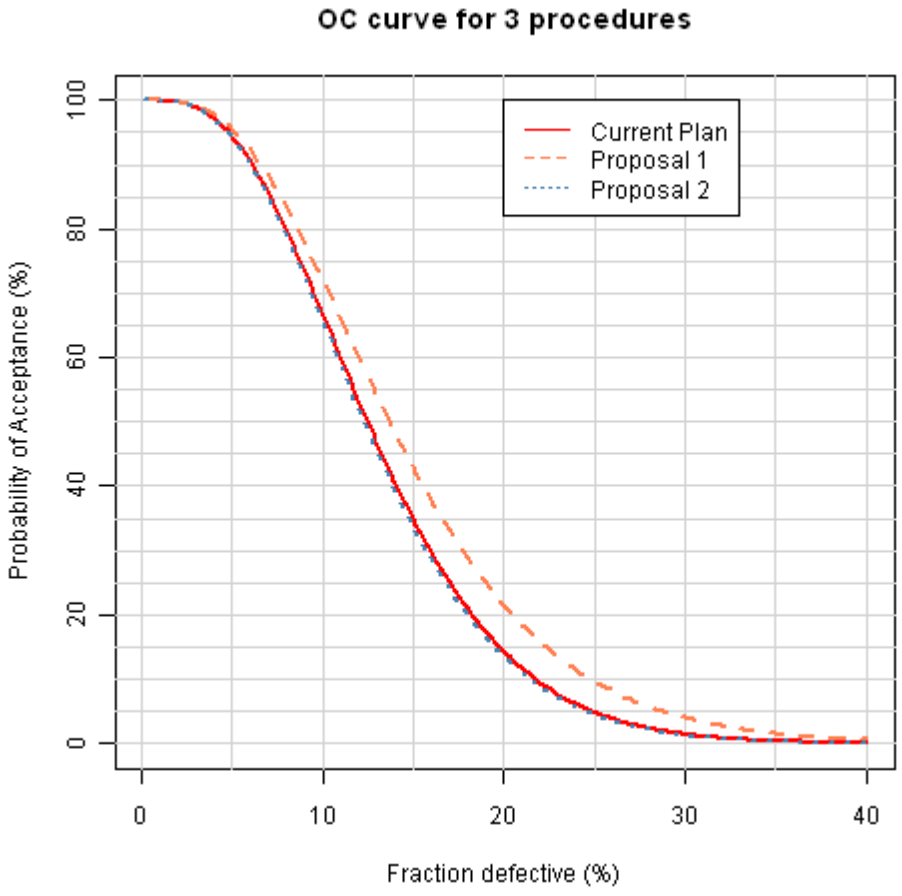
6. Based on their statistical properties, the first proposal is modestly more forgiving than the current protocol, while the second proposal matches the current CLSI protocol very closely.

a. The first proposal is highly likely (95%) to pass when the true long-run QC error rate is 5.2%, while the corresponding true long-run QC error rate for the second proposal is 4.8%. Both of these are very close to the current plan's 4.8%.

b. The first proposal is unlikely to pass (10%) when the true long-run QC error rate is 24.7%. This is modestly higher than the current protocol's 21.7%. The second proposal is unlikely to pass when the true long-run QC error rate is 21.4%.

An out-of-control result could be due to either systematic or random errors: systematic errors are likely to demonstrate more than 1 outlier of 15 results, which should be recognized with either proposal. Assuming the allowable random error rate generally accepted is 5% (95% confidence limits), the probability of getting one outlier for every 15 results due to random error is very high. Thus proposal #2 would likely be problematic and unlikely to improve quality of results.

7. The results above may be clearer with a plot. The "Operating Characteristic" ("OC") curve below plots the probability of passing criteria as a function of true long-run QC error rate:



8. The team was concerned that laboratories are allowed to report clinical results while QC testing is being conducted (by whatever method). This suggests an implicit assumption that QC testing can detect problems instantly, on a run-for-run basis. This is not generally true; rather, QC testing can, over many QC test results, determine that a system that was in control has deviated out of control. The team recommended that emphasis should be placed on other mechanisms to identify errors and ensure quality in addition to testing QC strains. Such mechanisms might include periodic user competency testing and checking unlikely results on patients' isolates.